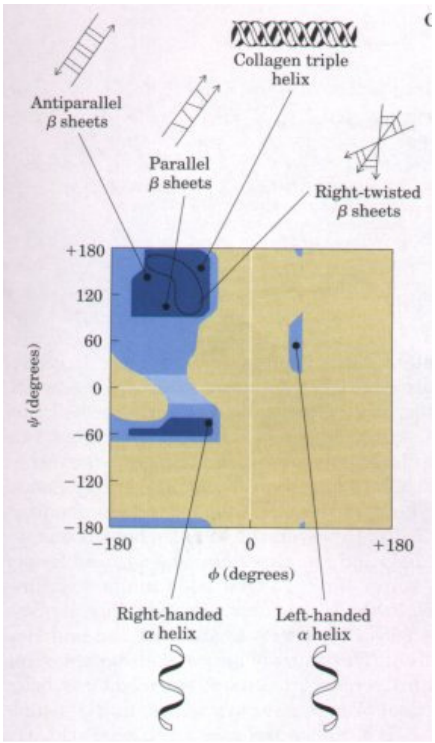
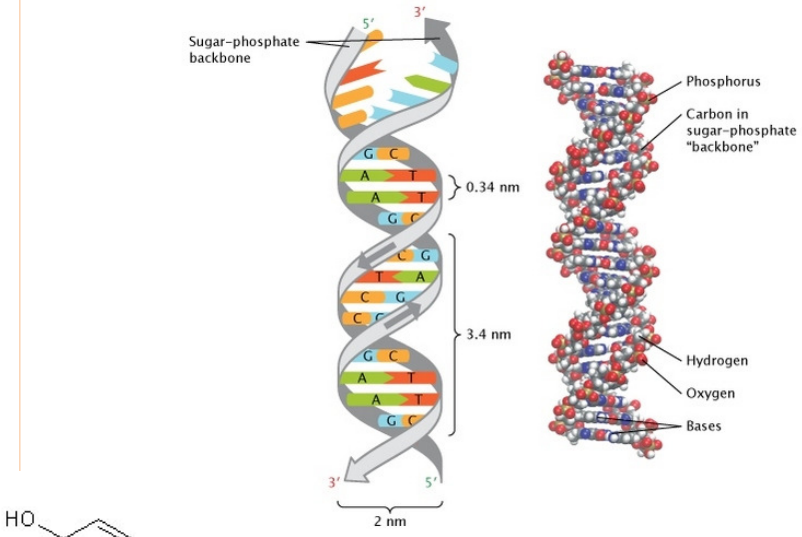
KARNATAKA STATE OPEN UNIVERSITY

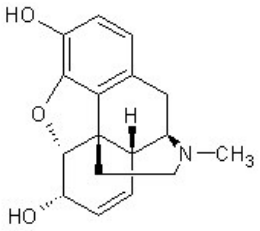
Mukthagangotri, Mysore – 570 006

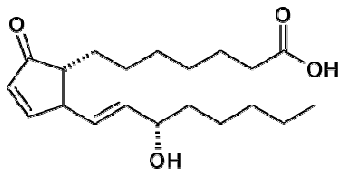
M.Sc. CHEMISTRY

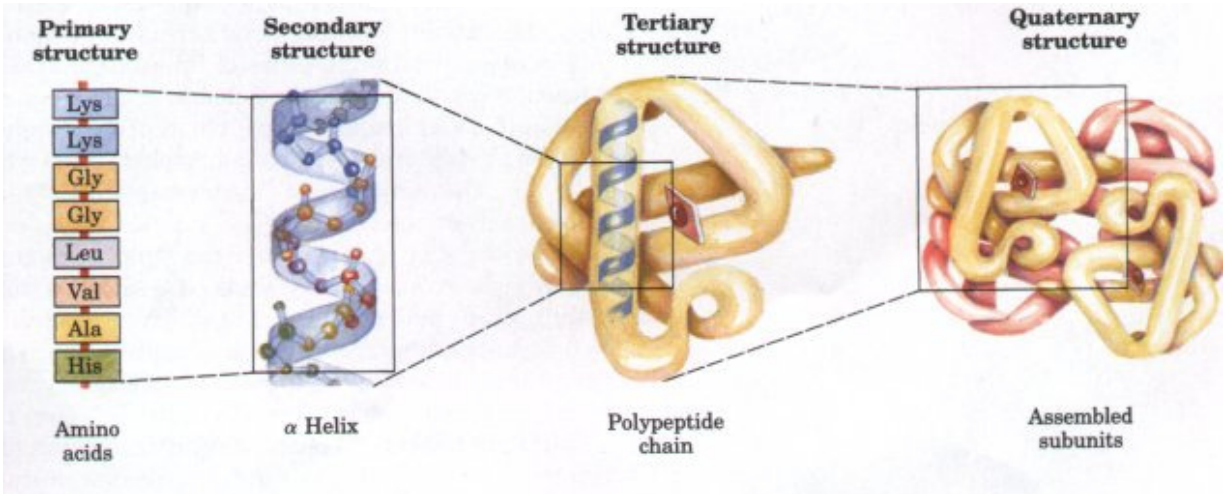
(FOURTH SEMESTER)











Course: MCH T 4.4 **Block 1,2,3 and 4**

BIO-ORGANIC CHEMISTRY

M.Sc. CHEMISTRY

FOURTH SEMESTER

COURSE: MCHT 4.4

BIO-ORGANIC CHEMISTRY

Course Design Committee

Vice-Chancellor & Chairperson
Karnataka State Open University
Mukthagangothri, Mysore - 570 006

Dean (Academic) & Convener
Karnataka State Open University
Mukthagangothri, Mysore - 570 006

COURSE WRITER

Dr. Suresh. P. R
Senior scientist
Chemical Engineering division
National Chemical Laboratory
Pune.

Block 4.4.1, 4.4.2, 4.4.3 and 4.4.4

Course Coordinator

Dr. M. Umashankara
Assistant Professor and Chairman
Department of Studies in Chemistry
Karnataka State Open University
Mukthagangothri, Mysore - 570 006

Course editor

Dr. Raman vyasabhattar
Senior Scientist and Production manager
GVK Bioscience Ltd
Hyderabad

PUBLISHER

The Registrar
Karnataka State Open University
Mukthagangothri, Mysore - 570 006
Developed by Academic Section, KSOU, Mysore

Printed and Published on behalf of Karnataka State Open University. Mysore – 6 by **Registrar (Administration)**

TABLE OF CONTENTS

COURSE INTRODUCTION

Bioorganic chemistry represents a merger of organic chemistry and biochemistry. In certain investigations, such as those involving metalloenzymes and cofactors, the contiguous areas of bioorganic and bioinorganic chemistry were also merge. Even the composite term, biophysical organic chemistry, has been used as a detailed descriptor in molecular recognition. Bioorganic chemistry had multiple origins. Nutritional research identified factors essential in the human diet, and their structures and syntheses led to the recognition of the modes of action of the so-called vitamins and related cofactors, or coenzymes. Secreted factors that exert a stimulatory effect on cellular activity, the hormones, could be better understood at the molecular level once their structure determinations and syntheses made them available in reasonable amounts. Concepts of the biogenesis of natural products played and continues to play a major role in the development

of bioorganic chemistry. The establishment of the complete stereochemical pathways of biosynthesis adds exquisiteness to this Endeavour.

A simple chemical applied according to a well recognized concept can be responsible for a great advance in biological chemistry. Thus, through the reaction of cyanogens bromide, Bernhard Witkop translated neighbouring group participation into selective, limited, non-enzymatic cleavage at methionine in a peptide chain. Application of the reagent has aided not only the correct sequencing of peptide segments of many proteins but also the production through genetic engineering of human insulin by means of a methionyl-containing precursor.

Organic chemistry approaching 100% conversion at each step provides the basis of modern biotechnology: the automated synthesis of polypeptide and polynucleotide chains and the sequencing of DNA and RNA.

UNIT-1

Structure

2.0 Objectives of the unit

After studying this unit you are able to

- Identify the different component present in nucleic acids
- Differentiate the Nucleosides and Nucleotides
- Recognize the types of sugar molecules present in DNA and RNA
- Write the structure of different heterocyclic bases present in DNA/RNA
- Explain the different methods available for synthesis of pyrimidines and purines bases
- Explain the different steps involved in synthesis of nucleosides
- Identify the different base protecting groups in oligonucleotide synthesis
- Explain the solid phase oligonucleotide synthesis

2.1 Introduction

Nucleic acids are the biopolymers like polysaccharides and proteins. They are called nucleic acids because scientists first found them in the nucleus of cells. There are two major classes of nucleic acids present in living organisms. They are deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). DNA is just one type of nucleic acid. Some other types are RNA, mRNA, and tRNA. ('m' stands for messenger and 't' stands for transfer). There is another type of RNA called rRNA in which 'r' stands for ribosomal RNA. All of these nucleic acids are work together to help cells replicate and build proteins. DNAs are carrier of genetic information and RNA are an intermediate in the expression of genetic information and other diverse roles

DNA $\rightarrow$ mRNA $\rightarrow$ Protein

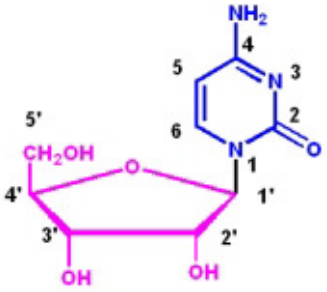
(genome) (transcriptome) (proteome)

1.2 Nucleic acids

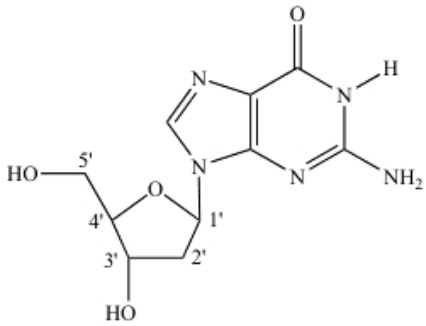
The nucleic acids are very large molecules that have three main parts, sugar, phosphate and heterocyclic base. The back bone of a nucleic acid is made of alternating sugar and phosphate molecules bonded together in a long chain. Each of the sugar groups in the backbone is attached to a molecule called a nucleotide base. Only four different nucleotide bases are occur in a nucleic acid, each nucleic acid contains millions of bases bonded to it. The order in which these nucleotide bases appear in the nucleic acid is the coding for the information carried in the molecule. In other words, the nucleotide bases serve as a sort of genetic alphabet on which the structure of each protein in our bodies is encoded.

1.3 Nucleosides.

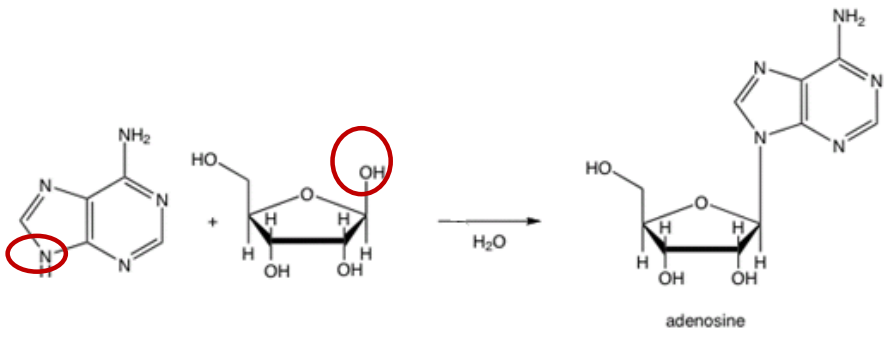
A nucleoside is single unit in which base is covalently attached to the C_1' position of a sugar. Or nucleosides are N-Glycosides of a purine or pyrimidine heterocyclic base and a carbohydrate. As we know the sugar in deoxynucleosides is 2'-deoxyribose. The sugar in ribonucleosides is ribose. Nucleosides differ from nucleotides in that they lack phosphate groups. The four different nucleosides of DNA are deoxyadenosine (dA), deoxyguanosine (dG), deoxycytosine (dC), and deoxythymidine (dT).



A nucleoside:
β-glycosidic bond between D-ribose and cytosine nitrogenouos base



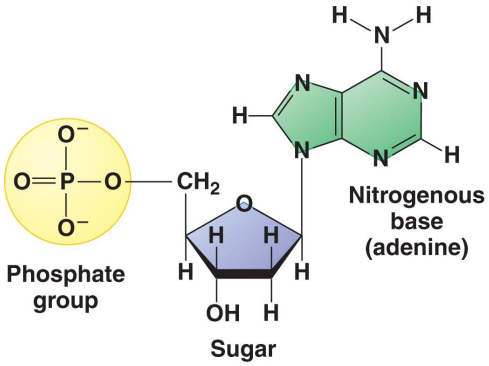
Nucleosides are the basic building blocks of ribonucleic acid (RNA) and deoxyriboneculeic acid (DNA). They are formed by the loss of water from a sugar plus a purine or pyrimidine base. The Water is formed by -OH group from the anomeric position of the sugar, and H from nitrogen of the base.



adenine ribose adenosine

1.4 Nucleotides

Nucleotides are phosphate esters of nucleosides. Most commonly, the phosphoryl group is attached to the oxygen of the 5'-hydroxyl group of the sugar and nucleotides are typically assumed to be 5'- unless otherwise stated. Thus nucleotides are monomeric units of nucleic acids and each nucleotide consists of a 5-carbon sugar, a nitrogen containing base attached to the sugar, and a phosphate group.



1.4.1 Sugar

Two five member ring sugars are found in nucleic acids. They are ribose sugar and deoxyribose sugar. Deoxyribose lacks an hydroxyl group at the 2'-position (Figure 3). The nucleic acids having deoxyribose sugar are called DNA (deoxyribonucleic acid) and the nucleic acids having ribose sugar are called RNA (ribonucleic acid). The carbon atoms of sugars in RNA and DNA are numbered as 1', 2', 3', 4', and 5' (reading as one prime, two prime etc….) to distinguish from the numbering of the atoms of the purine and pyrmidine rings.

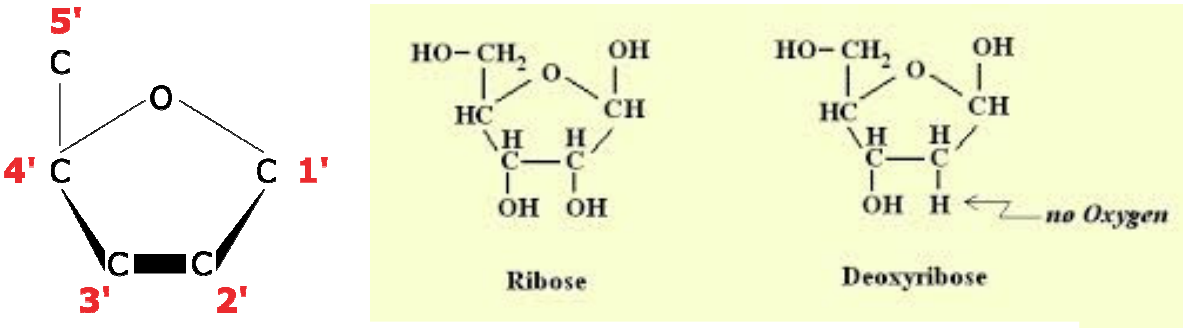


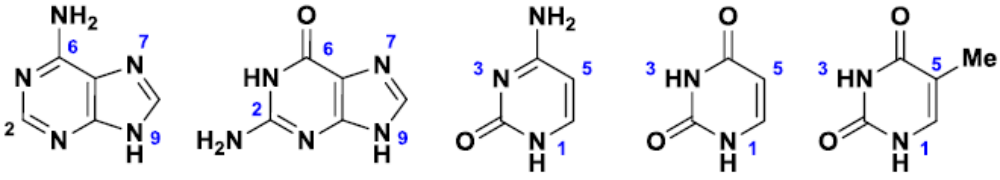
Figure 3: Structure of sugar unit present in DNA and RNA

The hydroxyl groups on the 5'- and 3'- carbons link to the phosphate groups to form the DNA/RNA backbone.

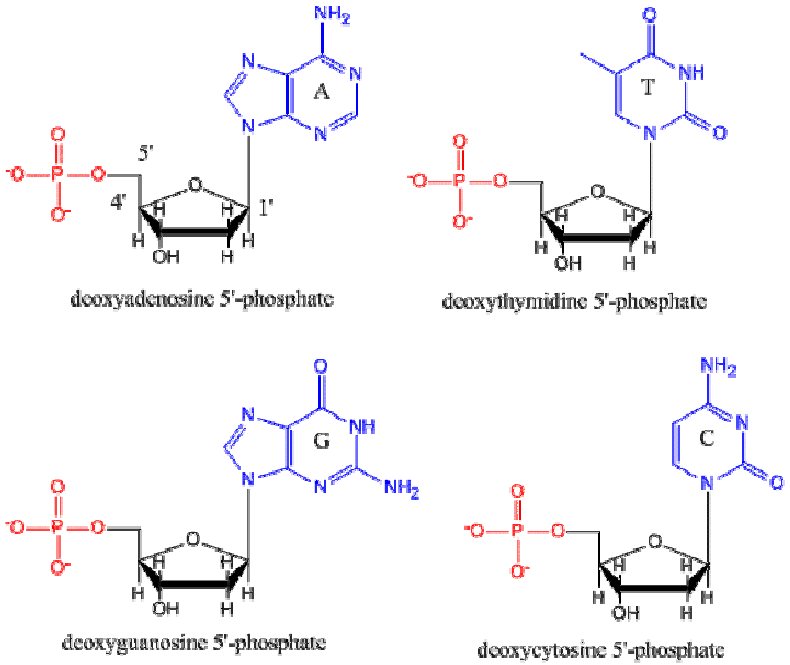
1.4.2 Bases

There are five bases are present in nucleic acids and they are **A**denine, **G**uanine, **C**ytosine **T**hymine and **U**racil. The bases are represented in single letter (First letter of the base) while writing the DNA/RNA sequence. These bases are divided in to two groups. They are pyrimidine bases and purine bases. Cytocine, Thymine and Uracil are single 6 membered heterocyclic ring bases and are termed as pyrimidines. Adenine and Guanine are fused ring heterocyles, in which 6 membered ring is fused with five membered ring and are termed as purines

Adenine, **G**uanine, **C**ytosine bases are present in both DNA and RNA. **T**hymine is present in DNA only and Uracil is present in RNA only.

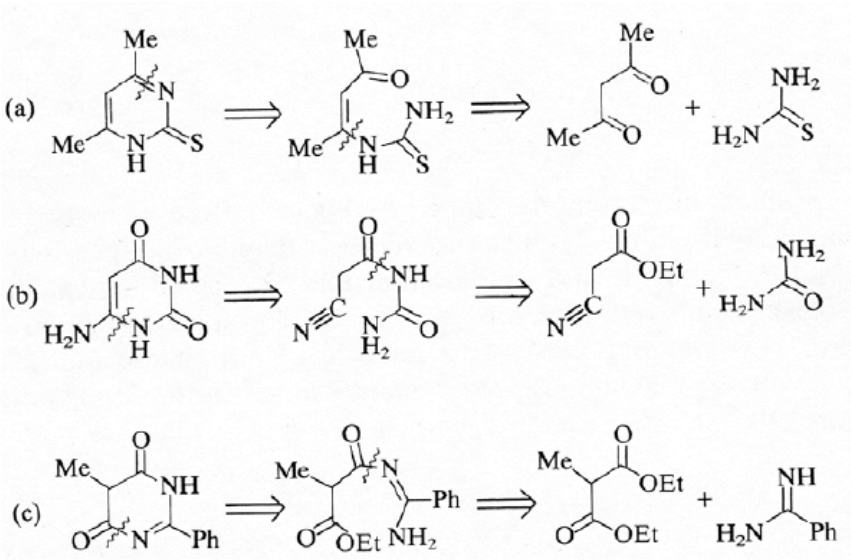


Each sugar molecule is covalently linked to one of 4 bases mentioned above. linking one of the sugars with a purine or pyrimidine base through an N-glycosidic linkage. Purines bond to the C1' carbon of the sugar at their N9 atoms. Pyrimidines bond to the C1' carbon of the sugar at their N1 atoms.

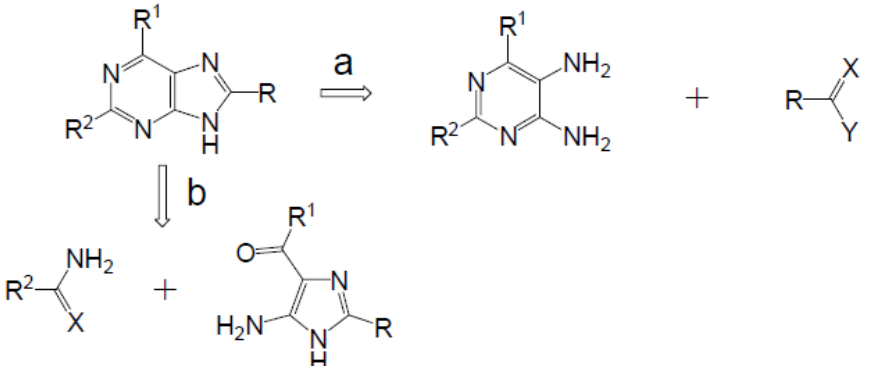


1.5 Synthesis of pyrimidines and purines base

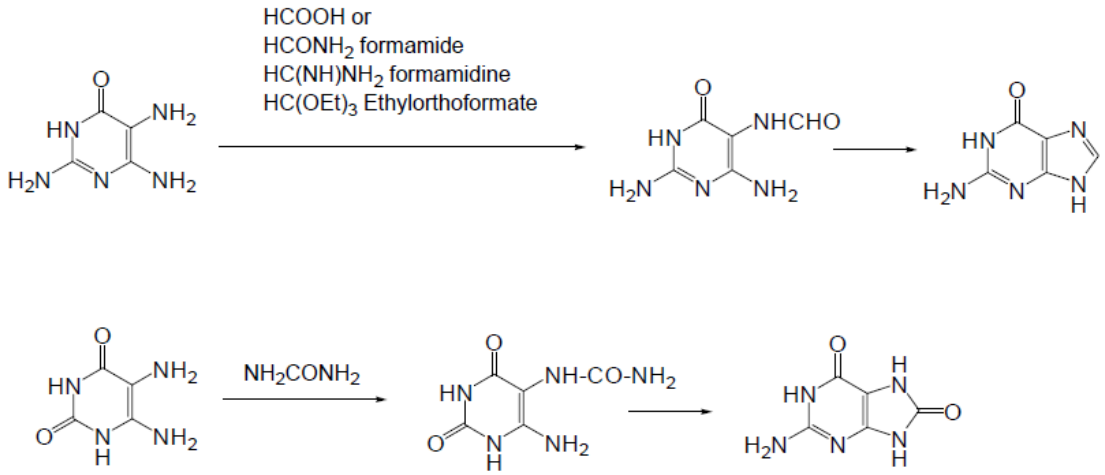
The synthesis of pyrimidines frequently includes the reaction of a C-C-C biselectrophile with an N-C-N bis-nucleophile. Bis-electrophiles can be 1,3-diketones (a), 1,3-esternitriles (b) or even 1,3-diesters. As bis-nucleophiles function predominately thiourea (a), urea (b) or guanidine (c). The reaction conditions are in most cases rather harsh (a) conc. HCl, reflux or (b), and (c) NaOEt, EtOH, reflux.



The chemical synthesis of purines often proceeds along two different routes (a) or (b). Along route a, the imidazole ring is constructed at the pyrimidine ring. Route b forms the pyrimidine ring at the imidazole ring system.

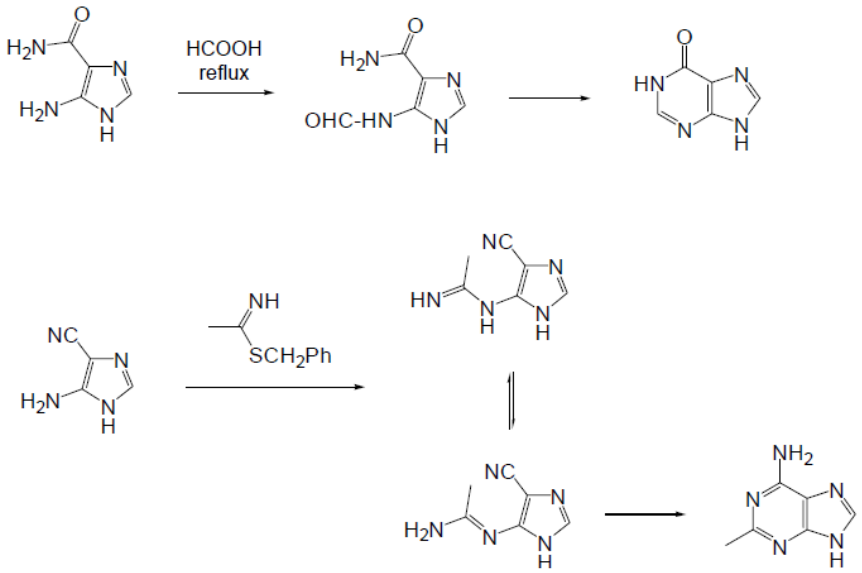


Route a is called Traube synthesis. Below are examples for this way to construct purine bases.



Purine synthesis via the Traube way (a)

Below are examples for the synthesis via route b.



These methods allow the synthesis of a variety of different pyrimidine and purine derivatives.

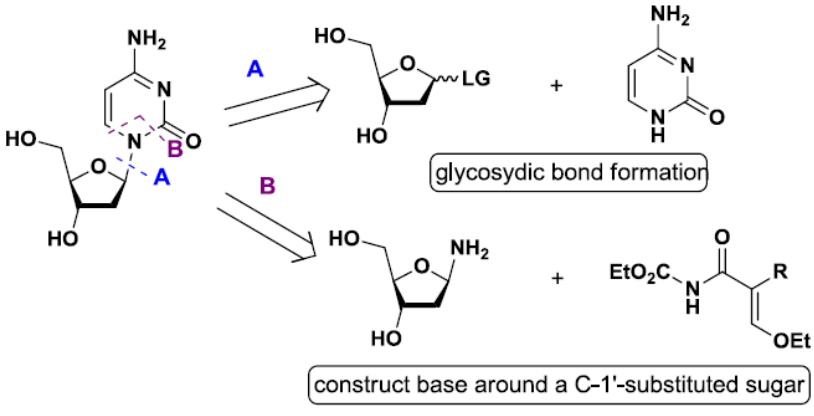
1.6 Synthesis of nucleosides

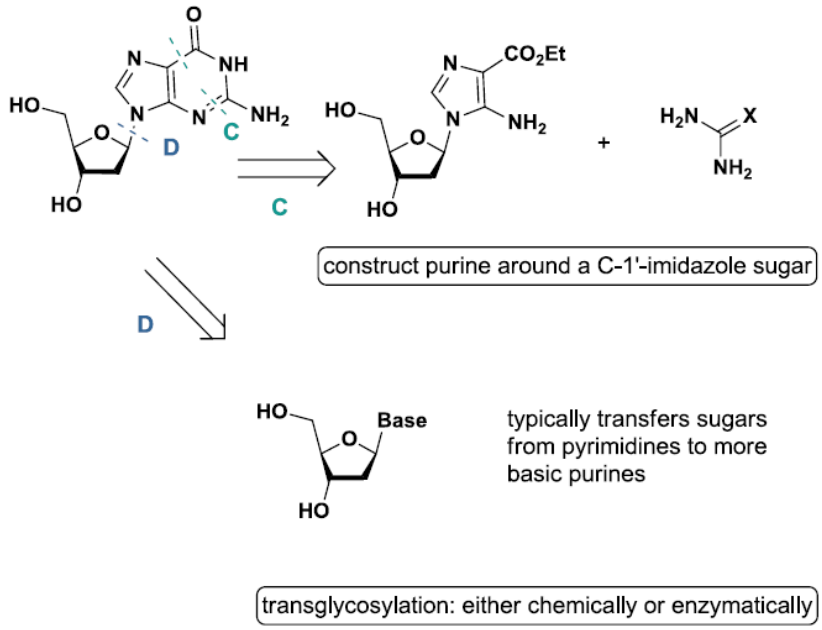
As already mention in the above discussion, nucleoside consists of five member sugar ring attached to heterocyclic base through N-glycosidic likage. For the preparation of nucleosides, the heterocycles have to be attached to the C1' position of the sugar. Generally two common approaches are followed for the synthesis of nucleosides. They are

1. Formation of N-glycosidic bond by reaction of sugar and heterocyclic base
2. Construction of heterocyclic base around a C1-substituted sugar.

The second approach is elegant but also difficult because the sugars rarely withstand the harsh condition often employed in heterocyclic chemistry.

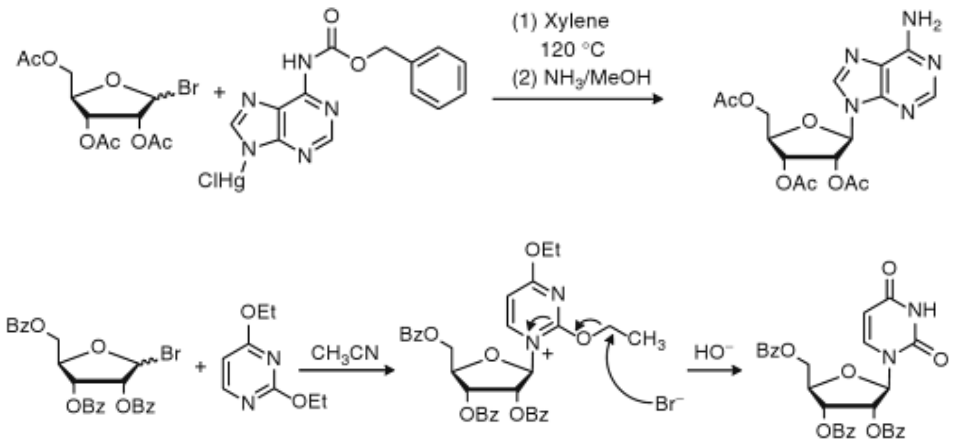
The schematic representation of both approaches for construction of purines and pyrimidines is as follows

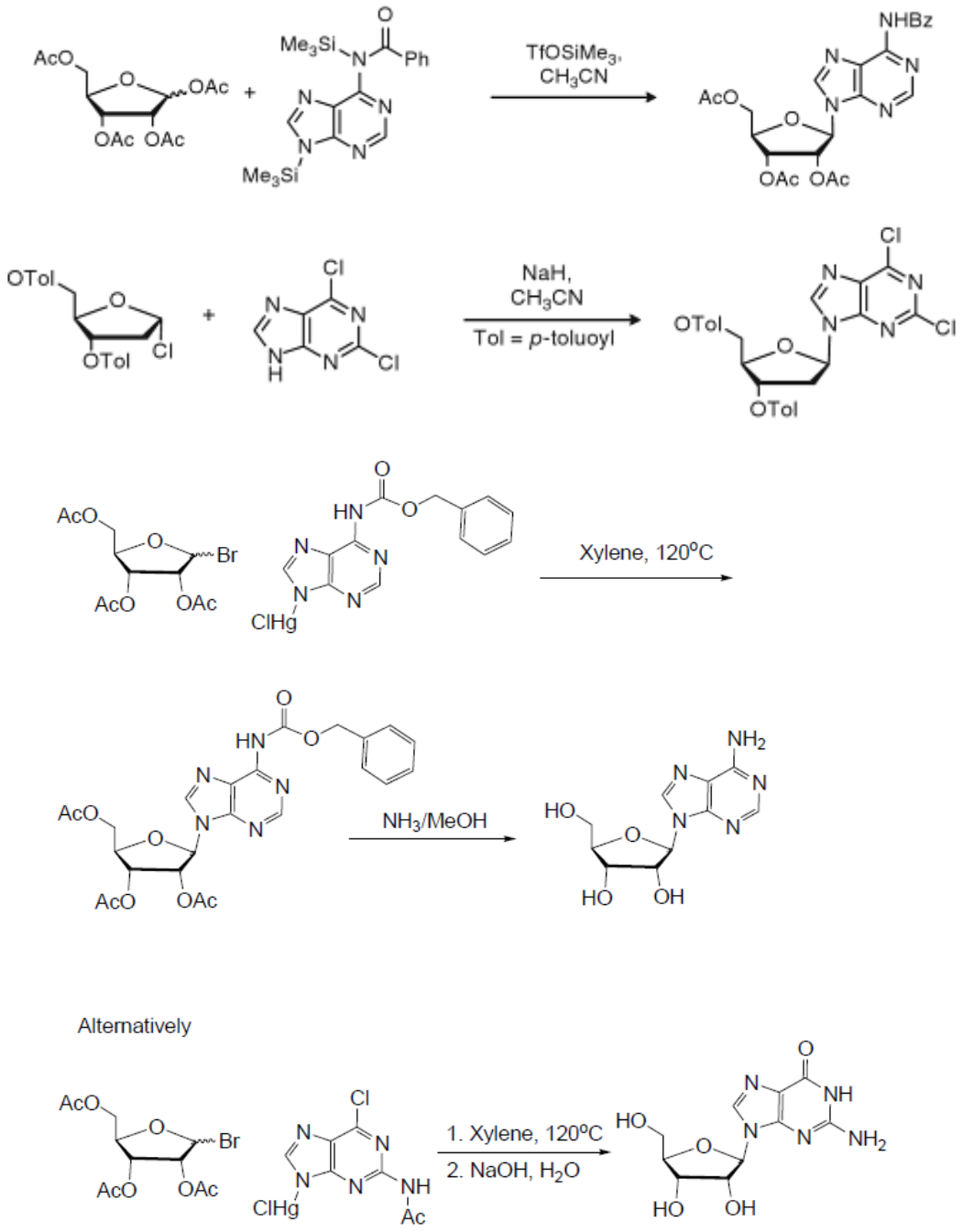




1.7 Synthesis of nucleosides using route A

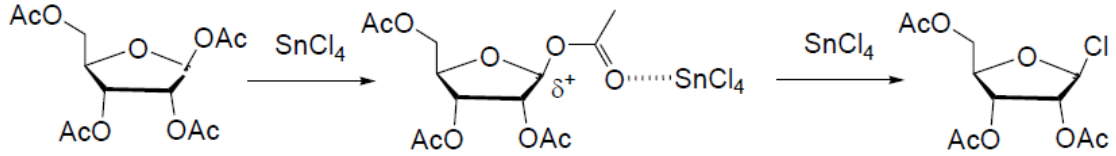
This is old methods for the construction of nucleosides. Which use heavy metal salts of the bases, predominantly Ag^{+} or Hg^{II+}salts. These heavy metal salts are reacted with sugars bearing in C1' position either a chloro- or bromo-substituent. These methods, which are called Fischer-Helferich or Koenigs-Knorr are still in use primarily for the synthesis of guanosine derivatives. Other potentially nucleophilic positions in the molecules require protection. Problematic with the method is the low solubility of the Hg^{II+} and Ag^{+} salts and the fact that the halogen sugars are rapidly hydrolyzed. The mechanisms of the reactions are very complex. One obtains often the kinetically determined O-alkylated products first, which then rearrange under the harsh condition.



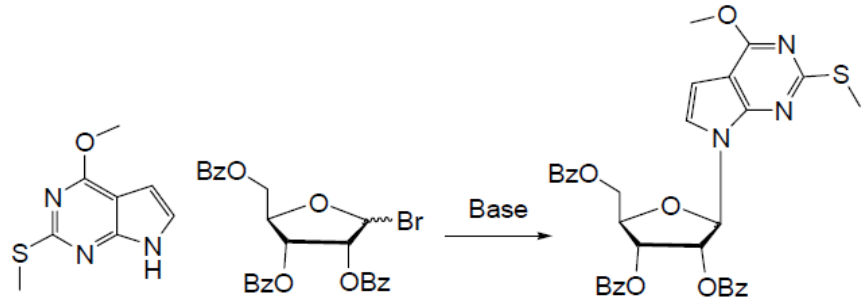


Fischer-Helferich or Koenigs-Knorr method

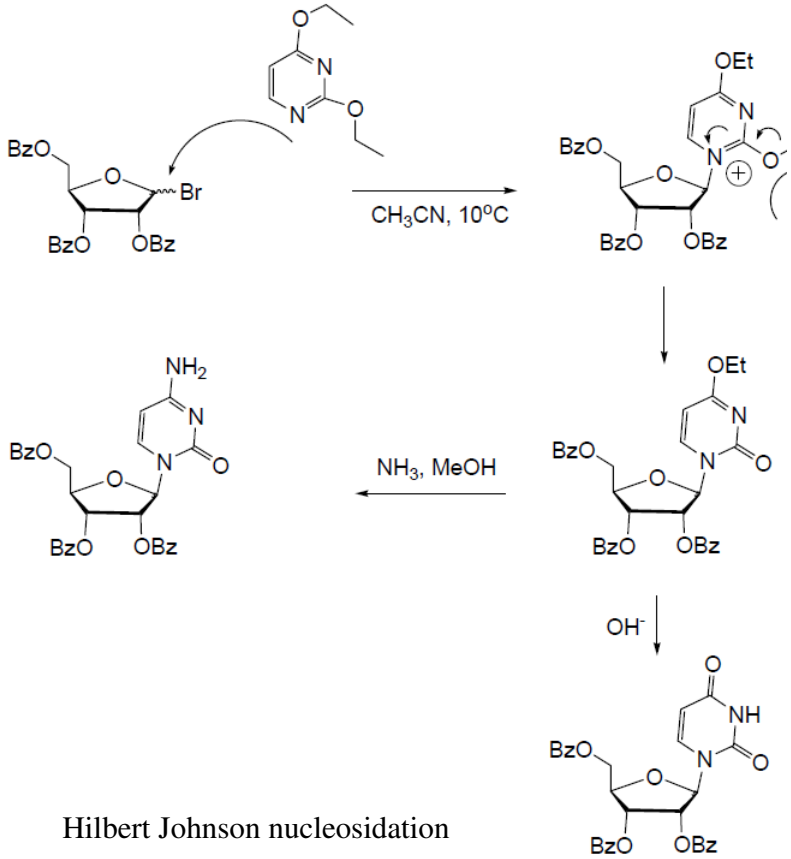
In general the procedure gives the correct regiochemistry since N1 of pyrimidines and N9 of the purines are reacting. A variation of the reaction uses instead of the halogen sugar the fully acetylated sugars, which are activated upon addition of a Lewis acid such as $SnCl_4$ or $TiCl_4$ generating the halogen sugar in situ.



In recent developments instead of the heavy metal salts alkylated bases are used. In these species the nucleophilicity of the sp-lone pairs is used for the reaction. This method is called Hilbert-Johnson nucleosidation.



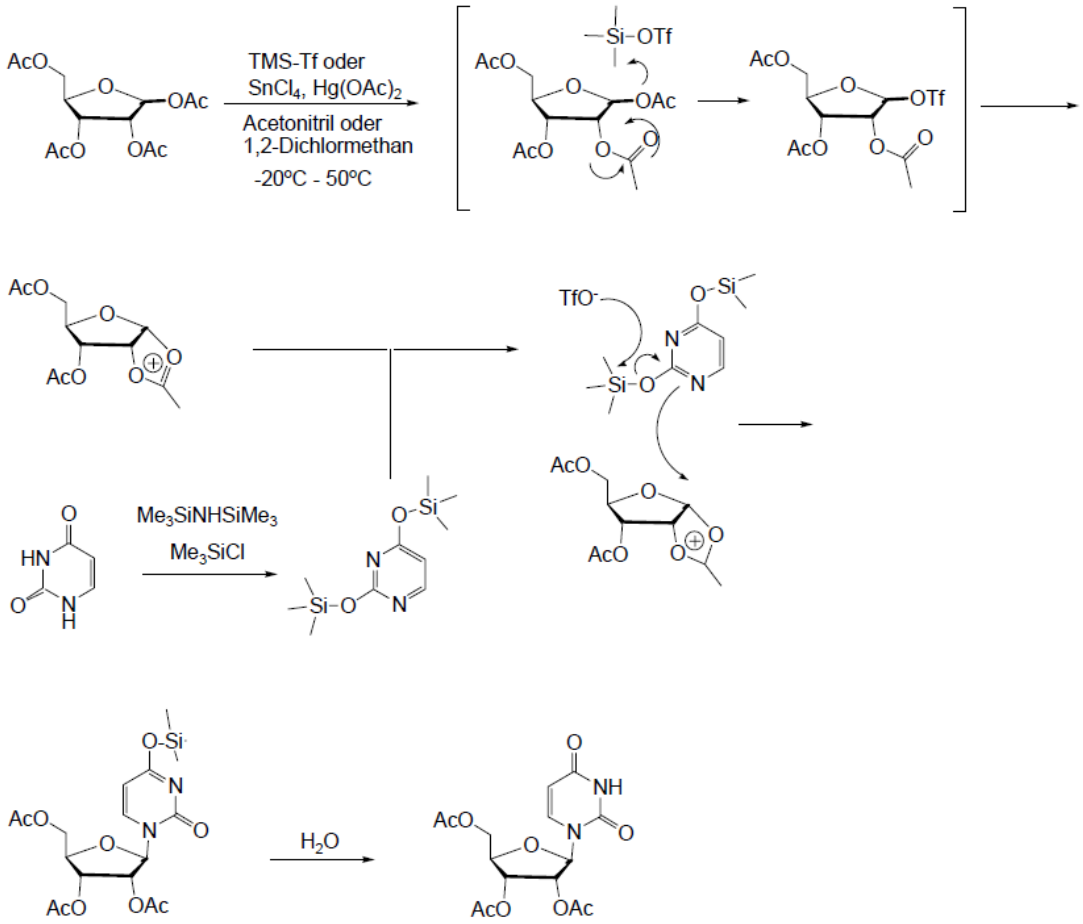
Direct nucleosidation gives first the quaternary salt, which than eliminates as shown below



Hilbert Johnson nucleosidation

A modern version of the Hilbert-Johnson method is the silyl-Hilbert-Johnson method. It is also called as Vorbrüggen nucleosidation. In this method the nucleobases are not alkylated but

silylated, which has the additional advantage that the bases become more soluble. The silylation is performed with hexamethyldisilazane (HMDS) in the presence of catalytic amount of Me_3SiCl. One can use bis(trimethylsilyl)acetamide [BSA] ($(Me_3Si)_2NCOCH_3$) as alternative reagent. The silylated base reacts subsequently with the peracetylated sugars, which are in situ converted into the halogen sugars or sugar triflates using a Lewis acid such as TMS-Tf or $SnCl_4$. The Tf-anion or the complex $SnCl_4$….-OAc are helping with the desilylation of the base during the reaction.



Vorbrüggen nucleosidation

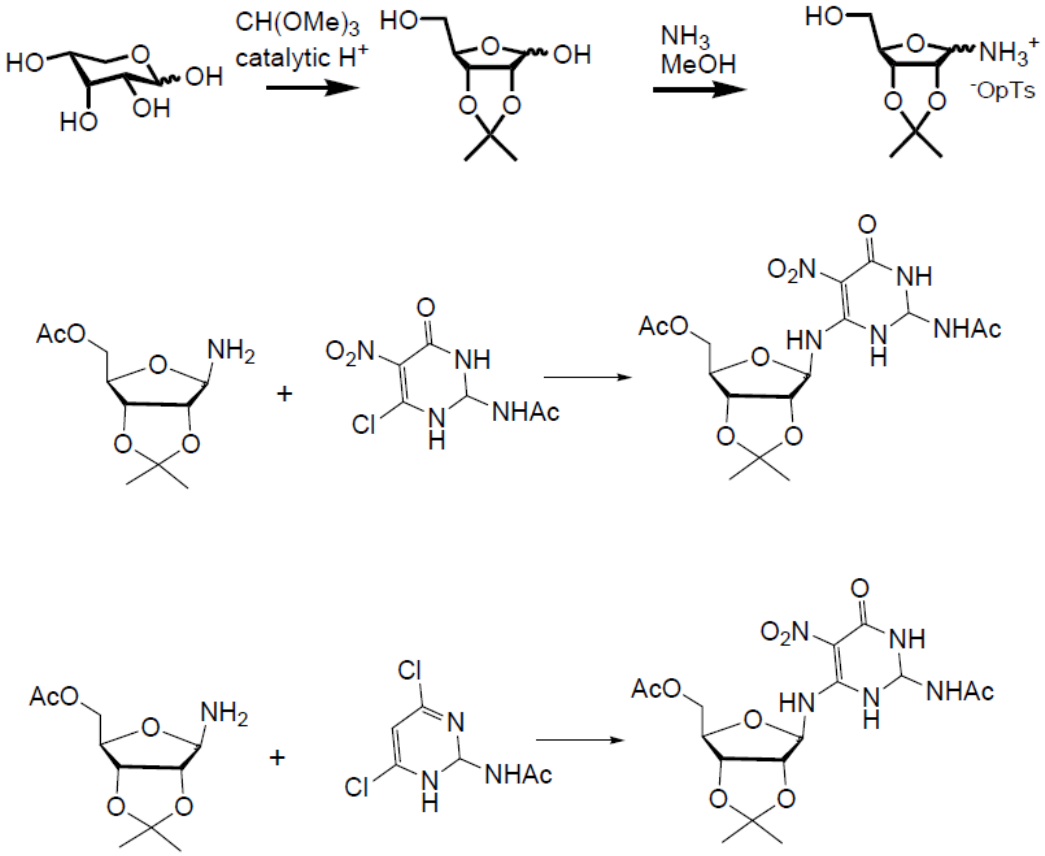
Depending on the conditions, the reaction proceeds in an S_N2 type manner with already strong S_N1 contribution. The transition state has consequently a strong oxycarbenium ion character. The glycosylation yields therefore a mixture of the α– and the β–products.

The reaction towards the α– to β–products depends strongly of the nature and the configuration of the group attached at C2‘. Acetate or a benzoate at C2’ are able to stabilize the oxycarbenium

ion (neighbouring group effect). The incoming base will in such case be always reacting trans relative to this group. This 1,2-trans rule is called Bakers rule.

1.8 Syntheses based on strategy B

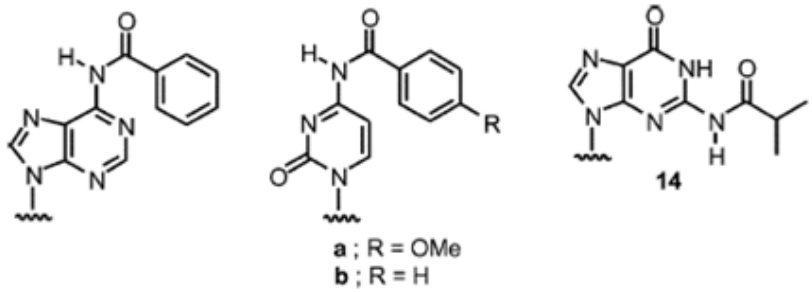
The nucleobases are very electron deficient aromatic systems. Therefore they are amenable to nucleophilic (instead or electrophilic) aromatic substitutions. Attacking bases with a nucleophile requires converting the sugar into the 2'-amino sugar. Starting from D-(-)-ribose, one changes the configuration of the sugar first by converting the ribose into the acetonite protected D-(-)-ribofuranose, which upon treatment with ammonia gives the 1'-aminosugar. This sugar can be reacted with nucleobases as shown below in the presence of weak bases (to free the amine). In the example below the heterocycle comprises quasi a vinylogous acid chloride. The reaction mechanism follows therefore an addition elimination process.



1.9 Base protecting groups in oligonucleotide synthesis

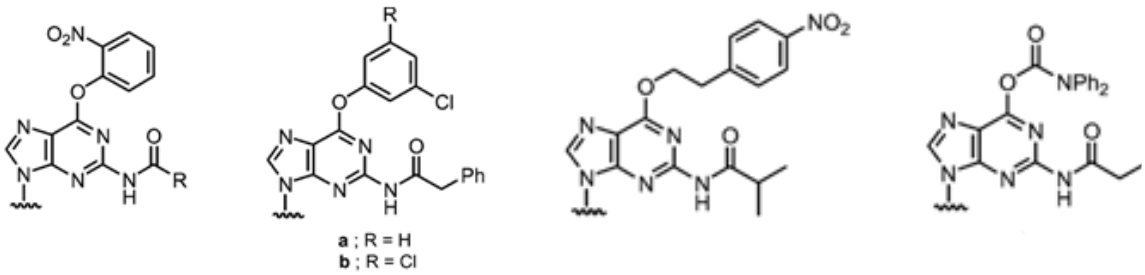
The heterocylic bases in nucleotide are reactive at many sites and cause many side reactions while synthesis of nucleotides. Thus protection of heterocylic base at suitable position with

proper protecting group is must. Adenine, cytosine and guanine residues are protected as their n-acyl derivatives. Thymine residues (and also uracil residues) were left unprotected in oligoribonucleotide synthesis.



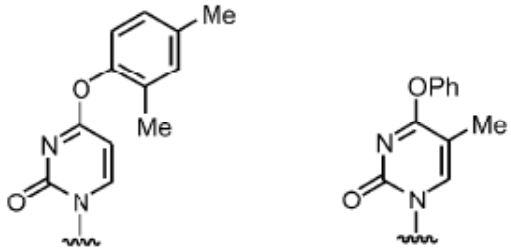
1.10 Protection of guanine base

The guanine residues were protected on N-2 and O-6 by (tert-butyl)phenylacetyl and 2-nitrophenyl groups, respectively. Protection on O-6 made the intermediates more lipophilic and consequently much easier to manipulate and also avoid the side-reactions. from several studies it is indicated that better yields were obtained when guanine residues were also doubly protected.



1.11 Protection of uracil and thymine residues

The uracil residues were particularly susceptible to modification in the solution phase approach to oligoribonucleotide. It is found that uracil could be effectively protected on O-4 with the 2,4-dimethylphenyl group, like 6-O-aryl protected guanine residues. The O-4 protected uracil residues may be rapidly deblocked with oximate ions under the conditions used to protected internucleotide linkages.



O-4 protected Uracil and thymine residues

Thymine residues may conveniently be protected with the 4-O-phenyl group which is both easy to introduce and then remove by oximate treatment. However the unprotect thymine is also used for the synthesis and it does not cause any problem. Hence thymine residue does not want any protection.

1.12 Solid phase oligonucleotide Synthesis

The chemical synthesis (solution phase) of oligonucleotides is labor intensive and inefficient till 1970's and in early 1980's. A breakthrough was achieved in 1983, for synthetic chemistry of nucleosides that made it possible to make longer oligonucleotides in more efficient way. The new synthesis process was based on the use of *phosphoramidite* monomers as building blocks and the use of tetrazole catalysis.

Oligonucleotide synthesis is carried out by a stepwise addition of nucleotide residues to the 5'-terminus of the growing chain until the desired sequence is assembled and consists of four chemical reactions

Step 1: De-blocking (detritylation)

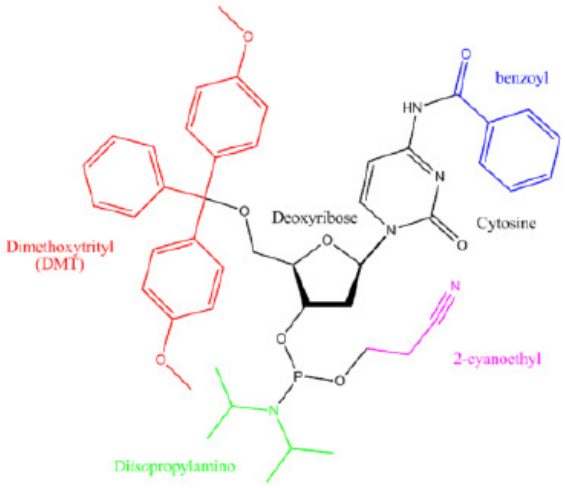
Step 2: Activation

Step 3: Coupling

Step 4: Capping

Step 5: Oxidation

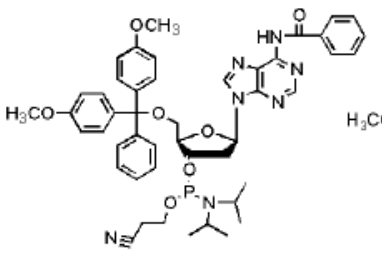
A phosphoramidite is used to protect the phosphate groups. Along with this protection group, a trityl group is used for protection of hydroxyl group and suitable protecting groups like beznoyl etcc…. were used to protect the reactive amine groups in the base. These protection groups prevent unwanted side reactions and force the formation of the desired product during synthesis.



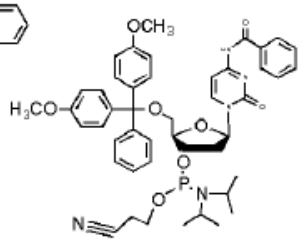
The protection groups are removed after the completion of the synthesis process. The link to the solid support is made through the 3' carbon and synthesis proceeds 3' to 5' rather than the 5' to 3'synthesis used previously. The solid support is a 5 micron controlled pore glass bead (CPG) with holes and channels where the protected nucleotide is attached.

The advances in oligonucleotide synthesis chemistries have resulted in substantial increases in quality and yield with the added advantage of decreasing cost.

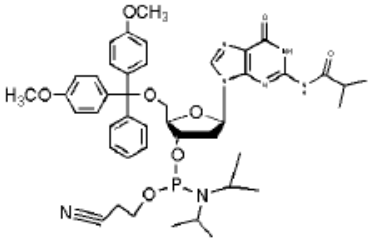
Some of the protected monomers used for solid phase synthesis of oligonucleicacids are given below



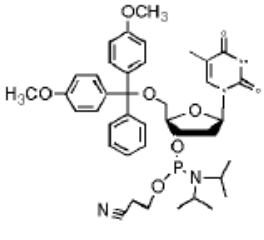
N-6-benzoyl-deoxyadenosine phosphoramidite



N-4-benzoyl-deoxycytidine phosphoramidite



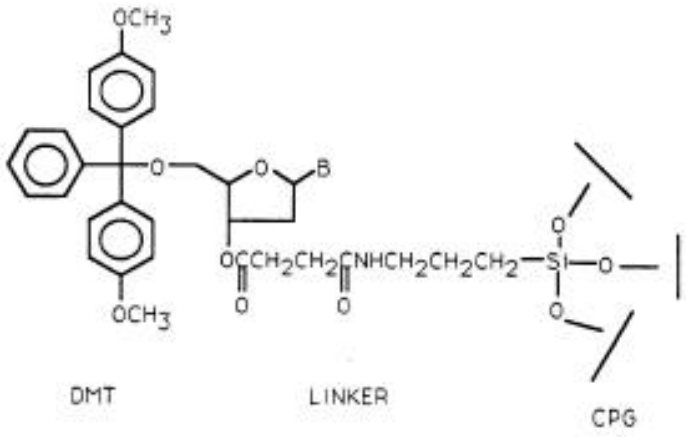
N-2-isobutryl deoxyguanosine phosphoramidite



deoxythymidine phophoramidite

1.12.1The Solid Support

The support used in the DNA synthesis is controlledpore glass (CPG) and polystyrene. The polystyrene has an aminomethyl linker attached to its surface. The support coupling efficiency is about 98%. The supports are covalently derivatized with one of the four nucleosides. The reactive group on these nucleosides are blocked or protected to prevent unwanted side reactions. They are all blocked at the 5'-hydroxyl with a dimethoxytrityl (DMT) group.



1.12.2Phosphoramidites nucleosides

Phosphoramidites are chemically modified nucleosides used as the building blocks for synthesis chemistry. Two types of phosphoramidites are available they are methyl or β-cyanoethyl. each nucleosides has the following modifications

1. A disisopropylamine on a 3' trivalent phosphorous moiety. The phosphoramidite is very stable and is made highly reactive by the activator tetrazole.
2. A methyl protecting group on the 3' phosphorous moiety of methyl phosphoramidites. A β-cyanoethyl protecting group on 3' phosphorous moiety of β-cyanoethyl phosphoramidites. these groups prevent side reactions and aid in the solubility of phosphoramidites and support-bound nucleotides in organic solvents. they are removed upon compleation of the synthesis.
3. dimethoxytrityl protecting group is used for the 5'-hydroxyl group.
4. Standard amidites use a benzoyl group to protect adenosine and cytidine. A isobutyryl group is used to protect guanosine. Thymidine is unreactive and does not need a protecting group because there are no exocyclic amines. These protecting groups prevent side reactions and are removed with ammonia after completing of synthesis.

1.12.3Deprotection

In the deprotection step, the trityl group, which is attached to the 5' carbon of the pentose sugar of the recipient nucleotide, is removed by trichloroacetic acid (TCA) or 3% dichloroacetic acid (DCA), leaving a reactive hydroxyl group to which the next base is added.

1.12.4 Coupling

In modern oligonucleoside synthesis, tetrazole activation method is used, which replaces the use of condensing agents like DCC, EDC etc.. The different tetrazoles used are 1H-tetrazole, 2-ethylthiotetrazole, 2-benzylthiotetrazole, or 4,5-dicyanoimidazole. Tetrazole is a weak acid and attacks the coupling phosphoramidite nucleoside forming a tetrazolyl phosphoramidite intermediate. The 5'-hydroxy group reacts with the activated phosphoramidite moiety of the incoming nucleoside phosphoramidite to form a phosphite triester linkage. The reaction is highly sensitive to water, for this reason oligonucleotide reaction is carried out coupling reactions in anhydrous acetonitrile. Upon the completion of the coupling reactions, any unbound reagents and by-products are then removed by washing. The use of tetrazole increased coupling efficiency to greater than 99% which allowed longer and longer oligonucleotides to be synthesized.

1.12.5 Capping

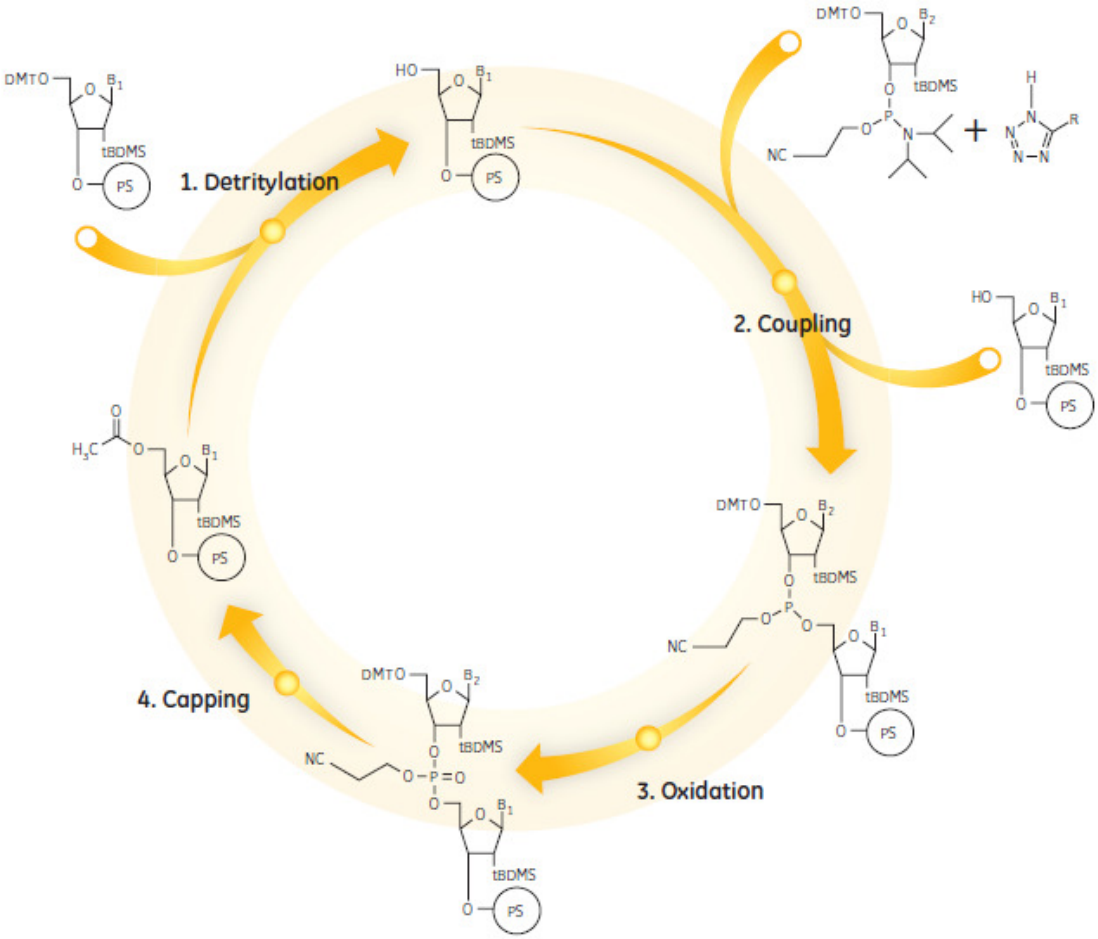
After the coupling reaction, a small amount of the solid support-bound 5'-OH groups (0.1 to 1%) remains unreacted and needs to be permanently blocked from further chain elongation to prevent the formation of (n-1) shortmers, which have an internal base deletion. These unreacted 5'-hydroxy groups are acetylated by the capping mixture. The capping step is performed by treating the solid support-bound material with a mixture of acetic anhydride and 1-methylimidazole (or, less often, DMAP)

1.12.6 Stabilization

The newly formed phosphite triester linkages are not natural and possess limited stability. Once the capping step is accomplished, the last step is oxidation, which stabilizes the phosphate linkage between the growing oligonucleotide chain and the most recently added base. The phosphate linkage between the first and second base must be stabilized by making the phosphate group pentavalent. This is achieved by adding iodine and water in the presence of a weak base (pyridine, lutidine, or collidine) which leads to the oxidation of the phosphite into phosphate leaving the phospho-triester bond stabilized.

1.12.7 Cleavage, Detritylation, Deprotection

This cycle is repeated for each nucleotide in the sequence. At the end of the synthesis, the oligonucleotide exists as, for example, a 25-mer with the 3’ end still attached to the solid support and the 5’ end protected with a trityl group. In addition, protecting groups remain on three of the four bases to maintain the integrity of the ring structures of the bases. The protecting groups are benzoyl on A and C and N-2-isobutyryl on G. The completed synthesis is detritylated and then cleaved off the solidsupport leaving a hydroxyl on both the 3’ and 5’ ends. At this point the oligo (base and phosphate) is deprotected by base hydrolysis using ammonium hydroxide at high temperature. The final product is a functional single-stranded DNA molecule.



1.13 Summary of the unit

Both DNA and RNA are known as nucleic acids. They have been given this name for the simple reason that they are made up of structures called nucleotides. Those nucleotides, themselves comprising a number of components. A nucleotide consists of three things: A nitrogenous base, which can be adenine, guanine, cytosine, or thymine (in the case of RNA, thymine is replaced by uracil). A five-carbon sugar called deoxyribose because it is lacking an oxygen group on one of its carbons. One or more phosphate groups.

The nitrogen bases are pyrimidine in structure and form a bond between their 1' nitrogen and the 1' -OH group of the deoxyribose. This type of bond is called a glycosidic bond. The phosphate group forms a bond with the deoxyribose sugar through an ester bond between one of its negatively charged oxygen groups and the 5' -OH of the sugar. Nucleotides join together through phosphodiester linkages between the 5' and 3' carbon atoms to form nucleic acids. The 3' -OH of

the sugar group forms a bond with one of the negatively charged oxygens of the phosphate group attached to the 5' carbon of another sugar.

1.14 Key Words

Nucleic acids; Nucleosides; Nucleotides; Sugar; Bases; Pyrimidines base; purines base; Phosphoramidites nucleosides; Coupling; Capping; Stabilization

1.15 References for further studies

1) Nucleic Acids: Structures, Properties, and Functions; Victor A. Bloomfield, Donald M. Crothers, Ignacio Tinoco; *University Science Books*, **2000**.

2) Nucleic Acids in Chemistry and Biology; G. Michael Blackburn; *Royal Society of Chemistry*, **2006**.

3) Nucleic Acid Structure: An Introduction; W. Guschlbauer; *Springer Science & Business Media*, **2012**.

4) Advanced Organic Chemistry of Nucleic Acids; Zoe A. Shabarova, Alexey A. Bogdanov; *John Wiley & Sons*, **2008**.

5) Basic Principles in Nucleic Acid Chemistry, Volume 1; Paul O.P. Ts'o; *Elsevier*, **2012**.

1.16 Questions for self understanding

What are nucleic acids?

Explain the difference between DNA and RNA.

What are nucleosides? Write the structure.

What are nucleotides? Write the structure.

What are the sugar molecule present in DNA and RNA?

What are the different types of heterocyclic bases present in nucleic acids? Write their structure.

Discuss the chemical synthesis of pyrimidines and purines base.

What are the major methods followed for synthesis of nucleoside?

Explain the synthesis of nucleosides using route A.

Explain the synthesis of nucleosides using strategy B.

Discuss the different base protecting groups in oligonucleotide synthesis and their significance.

What are the protecting groups used for guanine base? Which nitrogen of guanine base is need to be protected while oligonucleotide synthesis?

What are the protecting groups used for uracil and thymine residues

Discuss the following methods in solid phase oligonucleotide Synthesis

a) The Solid Support

b) Phosphoramidites nucleosides

c) Coupling

d) Capping

e) Stabilization

f) Cleavage, Detritylation, Deprotection

UNIT-2

Structure

2.0 Objectives of the unit

After studying this unit you are able to

- Write the linkage in nucleic acid back bone
- Explain the double helix structure of DNA
- Explain the Chargaff's rule and its significance in deducing DNA doubleheix strucutre
- Write the structure of Watson and Crick base paring of nucleic acid bases
- Explain the base stacking in double-helix
- Identify the different factors contributing to the stability of double-helix structure
- Identify the Major and Minor Grooves in DNA double helix structure
- Explain the Codon, Anti codon and genetic code and their biological significance
- Explain the biosynthesis of Proteins

2.1 Introduction

Many people believe that American biologist James Watson and English physicist Francis Crick discovered DNA in the 1950s. In reality, this is not the case. Rather, DNA was first identified in the late 1860s by Swiss chemist Friedrich Miescher. Then, in the decades following Miescher's discovery, other scientists--notably, Phoebus Levene and Erwin Chargaff--carried out a series of research efforts that revealed additional details about the DNA molecule, including its primary chemical components and the ways in which they joined with one another. Without the scientific foundation provided by these pioneers, Watson and Crick may never have reached their groundbreaking conclusion of 1953: that the DNA molecule exists in the form of a three-dimensional double helix.

Russian biochemist Phoebus Levene a physician turned chemist was the first to discover the order of the three major components of a single nucleotide (phosphate-sugar-base). Several scientists put forth suggestions for how these units are link together, but it was Levene's "polynucleotide" model that proved to be the correct one. Levene proposed that nucleic acids were composed of a series of nucleotides, and that each nucleotide was in turn composed of just one of four nitrogen-containing bases, a sugar molecule, and a phosphate group.

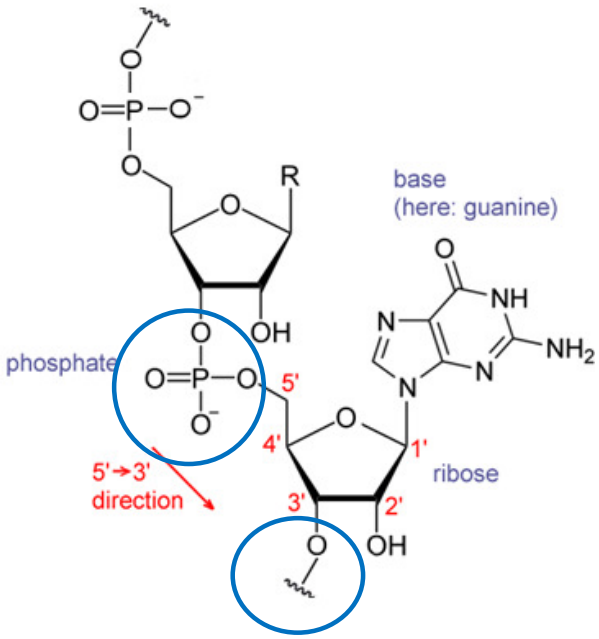
Erwin Chargaff an Austrian biochemist expanded the Levene's work and reached two major conclusions. First, he noted that the nucleotide composition of DNA varies among species, ie, the same nucleotides do not repeat in the same order. Second, Chargaff concluded that almost all DNA no matter what organism or tissue type it comes from, maintains certain properties, even as

its composition varies. In particular, the amount of adenine (A) is usually similar to the amount of thymine (T), and the amount of guanine (G) usually approximates the amount of cytosine (C). In other words, the total amount of purines (A + G) and the total amount of pyrimidines (C + T) are usually nearly equal. This second major conclusion is now known as "Chargaff's rule." Chargaff's research was vital to the later work of Watson and Crick.

2.2 Nucleic acid back bone

The sugar-phosphate linkage forms the structural framework of nucleic acids, including DNA and RNA and this is called nucleic acid backbone. This backbone is composed of alternating sugar and phosphate groups.

Both DNA and RNA are composed of nucleotides that are linked to one another in a chain by chemical bonds, called ester bonds. This bond is formed between the 3' end of one nucleotide sugar and the 5' end of phosphate group of the adjacent nucleotide. *Thus the phosphate group attached to the 5' carbon of the sugar on one nucleotide forms an ester bond with the free hydroxyl on the 3' carbon of the next nucleotide*. These bonds are called phosphodiester bonds, and the sugar-phosphate backbone is described as extending, or growing, in the 5' to 3' direction when the molecule is synthesized. The chemical linkage between monomer units in nucleic acids is a phosphodiester



At physiological pH, the phospates are ionized, as depicted in the picture. The ionization of the phospates means that RNA and DNA bear multiple negative charges, therefore they are

polyelectrolytes. The cations of various kinds, especially Mg^{++}, tend to cluster near the phosphates.

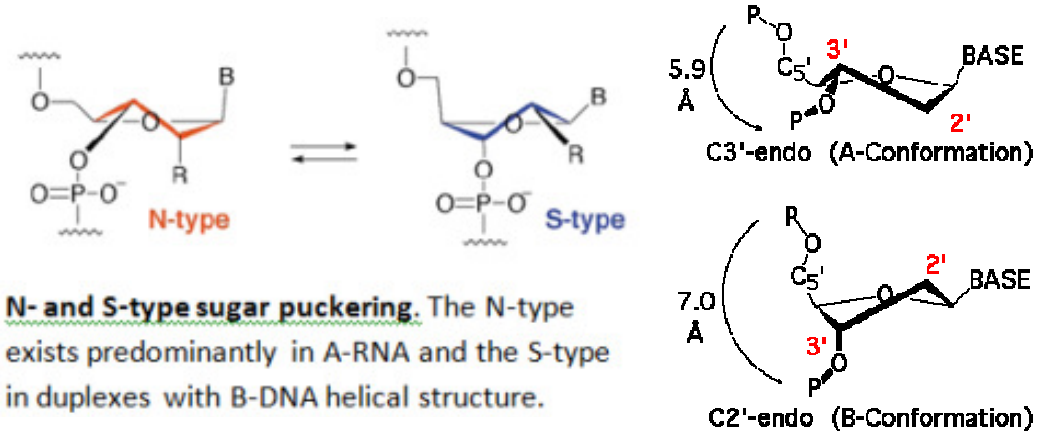
2.3 Structure of nucleosides

There are two conformational variations are possible for nucleosides, they are

i) Rotation of the base-to sugar bond, and

ii) Puckering of the sugar ring.

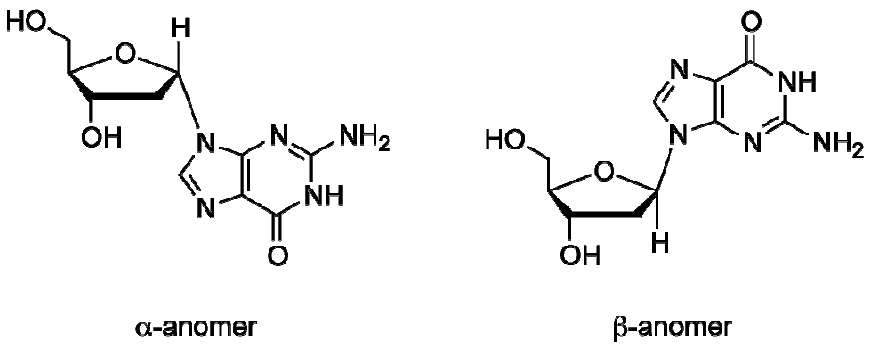
The puckering of the sugar ring usually involves having either C2' or C3' out of the plane formed by C1', O, and C4'.



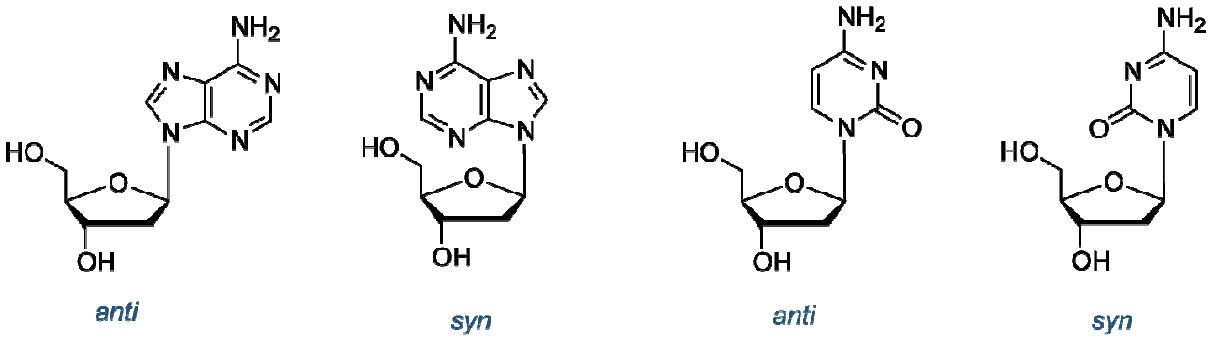
N- and S-type sugar puckering. The N-type exists predominantly in A-RNA and the S-type in duplexes with B-DNA helical structure.

C_2`-endo, conformation is also called south conformation (or B-conformation) and denoted by letter 'S'. Similarly C3`-endo conformation is also called a north conformation (or A-conformation) and denoted by N.

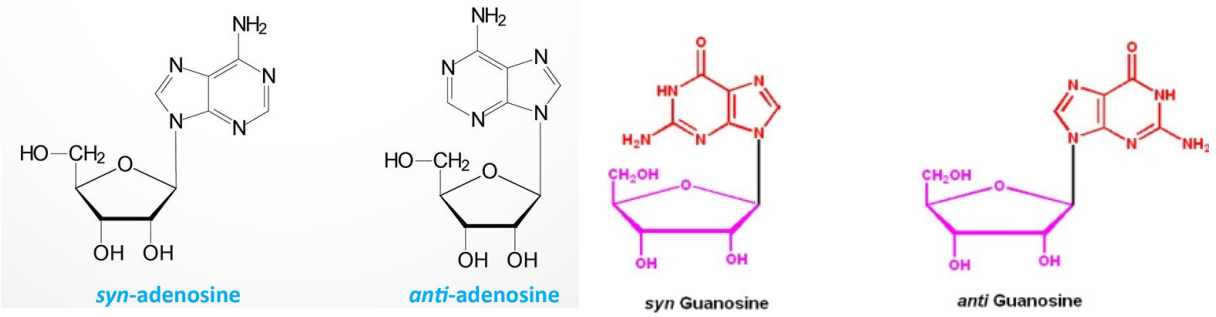
If C2' or C3' is on the same side of the ring as the glycosidic bond, the conformation is described as endo-; if on the other side, it is exo-.



α-anomer β-anomer



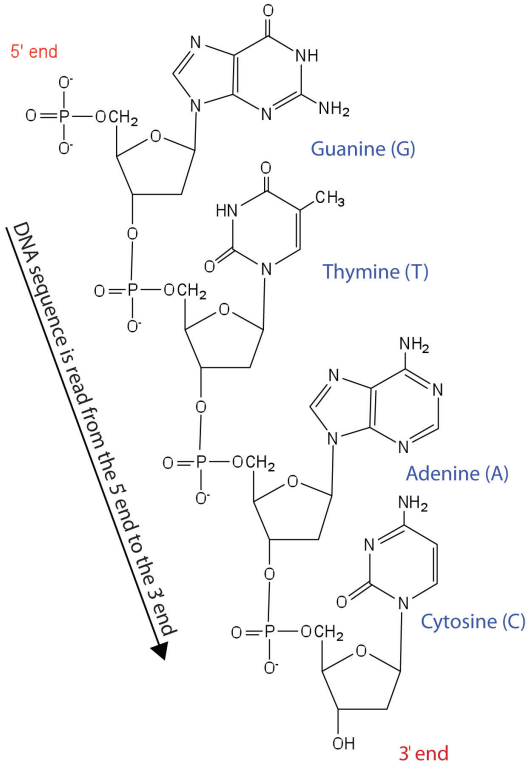
deoxyadenosine (a purine nucleoside) deoxycytidine (a pyrimidine nucleoside)



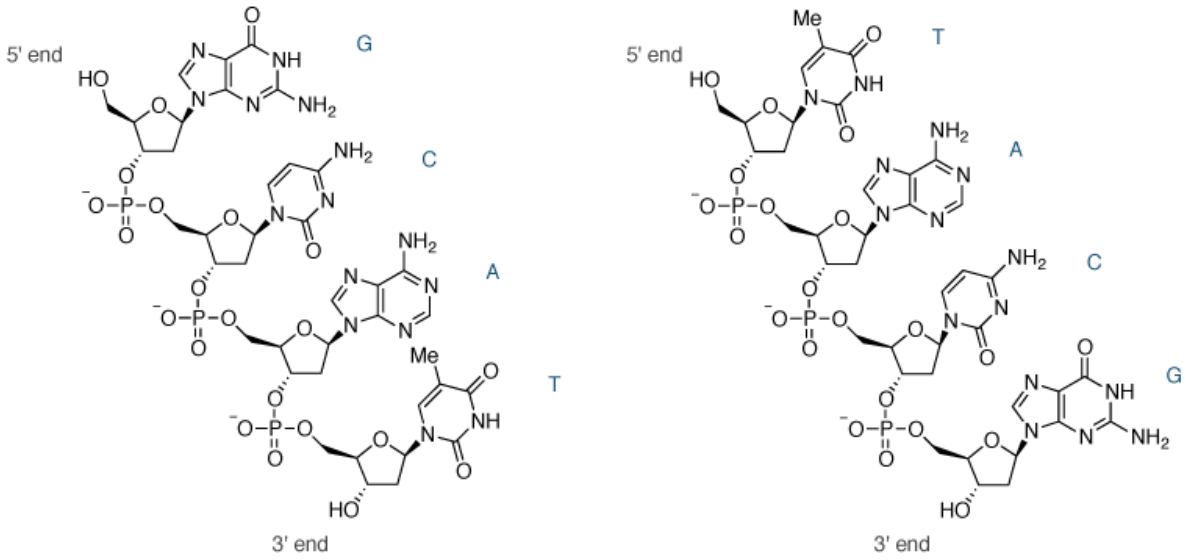
The anti- conformation of nucleotide is required for base-pairing to occur in a nucleic acid. The syn- and anti- conformations are in equilibrium for purine bases, but pyrimidines are exist entirely in the anti- conformation. In general, the two conformations are in equilibrium in solution.

2.4 Oligonucleotides

We known that dinucleotide (dimer) of DNA or RNA is formed by covalently linking the 5′-phosphate group of one nucleotide to the 3′-hydroxyl group of another to form a phosphodiester bond. An oligonucleotide (oligomer) is formed when several such bonds are made, At physiological pH (7.4) each phosphodiester group exists as an anion (hence the term nucleic acid), and nucleic acids are therefore highly charged polyanionic molecules.



One end of a nucleic acid strand has a 5′-hydroxyl group (primary hydroxyl) and the other end has a 3′-hydroxyl group (secondary hydroxyl). The nucleic acid chain therefore has directionality. The two tetranucleotides of the sequence 5′-GCAT-3' and 5′-TACG-3' respectively are shown below.



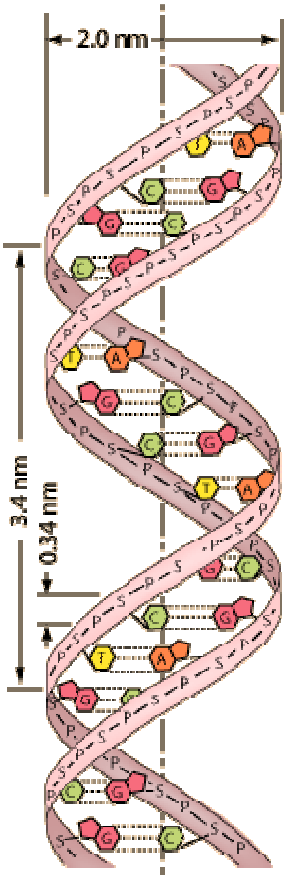
Chemical structures of the oligonucleotides dGCAT (left) and dTACG (right)

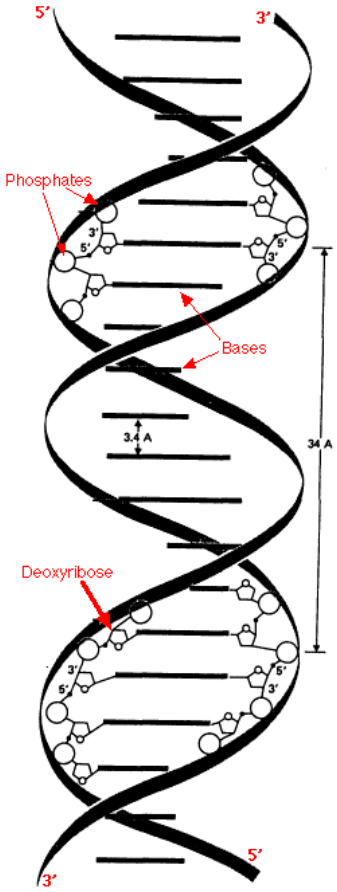
The nucleic acid sequences are always written in the 5' to 3' direction and the prefixes 5′- and 3′- are not written usually in the text books. The two oligonucleotides GCAT and TACG shown in above figure are distinct molecules with different chemical and biophysical properties.

2.5 Structure of DNA

The chemical structure of a single strand of DNA gives little insight into its biological function. James Watson and Francis Crick in 1953 discovered the DNA structure by X-Ray diffraction studies. They showed that DNA adopts a double-stranded structure called double-helix or duplex.

The sequence of one strand of DNA precisely defines the sequence of the other hence the two strands are said to be complementary and are sometimes called reverse complements of each other. The two strands are antiparallel, with the 5′-end of one strand adjacent to the 3′-end of the other. The two strands coil around each other to form a right-handed double helix, with the hydrophobic base pairs in the centre and the sugars and negatively charged phosphates forming the external hydrophilic backbone.





The term "right-handed" indicates that the backbone at the front of the molecule facing the observer slopes down from top right to bottom left. The planar heterocyclic bases stack one on another and the separation between successive base pairs along the helix axis is around 0.34 nm. One helical turn (a full 360° turn of the double helix) is repeated every 10 to 11 base pairs.

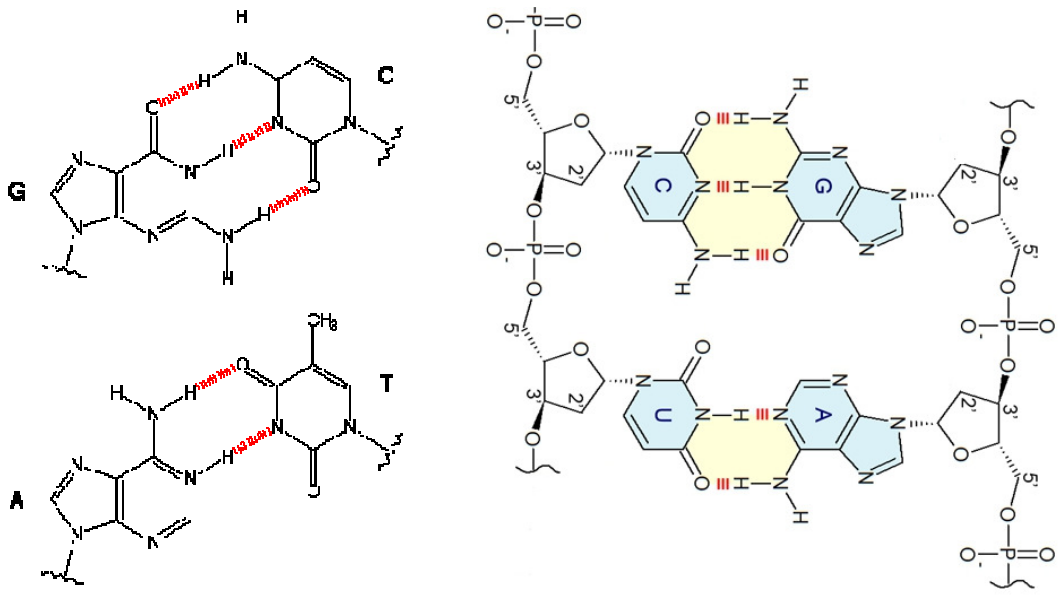
The double-helical structure was principally elucidated from X-ray fibre diffraction data (acquired by Rosalind Franklin and Maurice Wilkins) and Chargaff's rules.

2.6 Chargaff's Rule

Erwin Chargaff discovered that the molar amount of adenine in DNA was always equal to that of thymine and the same was true for guanine and cytosine (i.e. number of moles of G = number of moles of C). *Chargaff's rule states that the molar ratio of A to T and of G to C is almost always approximately equal in a DNA molecule*. Chargaff's Rule is true as a result of the strict hydrogen bond forming rules in base pairing. For every G in a double-strand of DNA, there must be an accompanying complementary C, similarly, for each A, there is a complementary paired T.

2.7 Watson and Crick base paring

The heterocyclic bases of single-stranded DNA have polar amido, amidino, guanidino and carbonyl groups that form a complex network of hydrogen bonds. Watson and Crick were able to explain Chargaff's Rule by building models to show that the two strands of DNA are held together by hydrogen bonds between individual bases on opposite strands. The purine base **A** always pairs with the pyrimidine **T** and the purine **G** always pairs with the pyrimidine **C** as shown in figure. This is called as rules of base pairing (or nucleotide pairing)



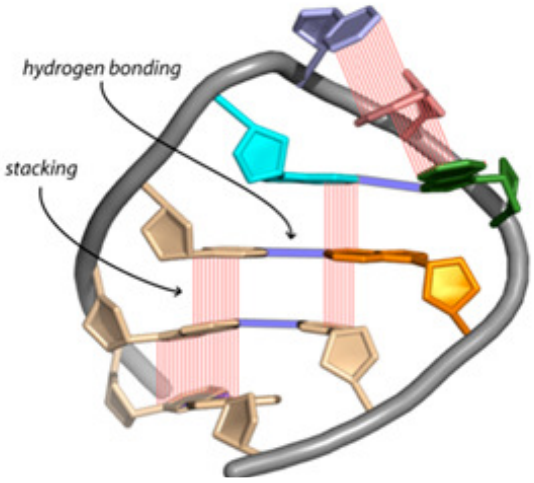
Thus the purine adenine (A) always pairs with the pyrimidine thymine (T) [A with T] and the pyrimidine cytosine (C) always pairs with the purine guanine (G) [C with G]

This is consistent with the X-ray structure of DNA. There is not enough space (20 Å) for two purines to fit within the helix and too much space for two pyrimidines to get close enough to each other to form hydrogen bonds between them. But one can ask the question why not A with C and G with T. The answer for this question is, there is an opportunities to establish hydrogen bonds only with A & T and C & G pairs. Between A & T there are two hydrogen bonds are possible and between C & G there are three hydrogen bonds are possible. This relationship between the bases is called the Watson-Crick base pairing.

The rules of base pairing tell us that if we can "read" the sequence of nucleotides on one strand of DNA, we can immediately deduce the complementary sequence on the other strand. The rules of base pairing also explain the phenomenon that whatever the amount of adenine (A) in the DNA of an organism, the amount of thymine (T) is the same. Similarly, whatever the amount of guanine (G), the amount of cytosine (C) is the same.

2.8 Base stacking in double-helix

The GC base pairs are stronger than AT base pairs because they form three hydrogen bonds rather than two. Inter-strand hydrogen bonding is clearly important in driving the formation of DNA duplexes, but it is not the only contributing factor. The flatness of the bases in the double-helix structure form conjugated pi-electron systems means that bases will be able to derive a lot of interaction energy from van der Waals forces. The individual bases form strong stacking interactions which are major contributors to duplex formation and stability.



Base-stacking interactions are hydrophobic and electrostatic in nature, and depend on the aromaticity of the bases and their dipole moments. Base-stacking interactions in nucleic acid duplexes are partly inter-strand and partly intra-strand in nature. However, it is probably more informative to consider base pairs rather than individual bases as discrete units in order to visualize the stabilizing effects of base stacking.

2.9 Stability of double-helix structure

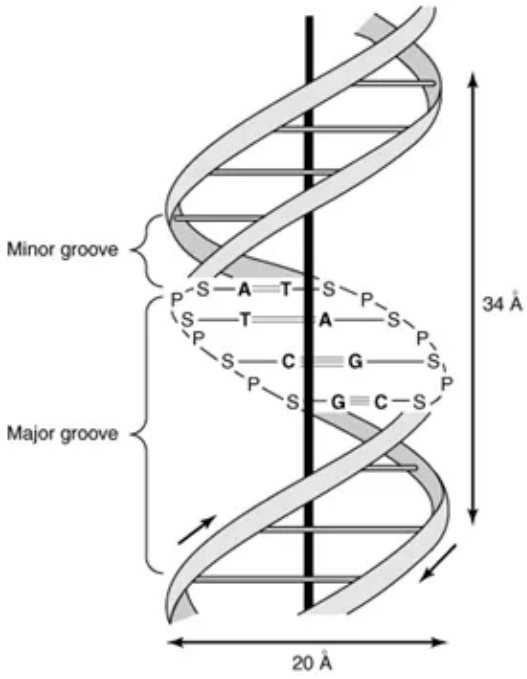
Two factors are mainly responsible for the stability of the DNA double helix: base pairing between complementary strands and stacking between adjacent bases. Thus the stability of the duplex is derived from both base stacking and hydrogen bonding. The two strands of DNA stay together by H bonds that occur between complementary nucleotide base pairs. As we known two hydrogen bonds occur between the adenosine and the thymine base pairs and between the cytosine and the guanine there are three. While each hydrogen bond is extremely weak, the millions of H-bonds together represent an extremely strong force that keeps the two DNA strands together. In addition, other groups of the base rings (polar groups) can form external hydrogen bonds with surrounding water that give the molecule extra stability.

The degree of stabilization afforded by base stacking depends on the DNA sequence. Some combinations of base pairs form more stable interactions than others, so nearest neighbour base-stacking interactions are important determinants of duplex stability.

Base-stacking interactions increase with increasing salt concentration, as high salt concentrations mask the destabilising charge repulsion between the two negatively charged phosphodiester backbones. DNA duplex stability therefore increases with increasing salt concentration. Divalent cations such as Mg^{2+} are more stabilising than Na+ ions, and some metal ions bind to specific location on the DNA duplex.

2.10 The Major and Minor Grooves

As a result of the double helical nature of DNA, the molecule has two asymmetric grooves. One groove is smaller than the other. This asymmetry is a result of the geometrical configuration of the bonds between the phosphate, sugar, and base groups that forces the base groups to attach at 120^0 angles instead of 180^0 degrees. The larger groove is called the major groove while the smaller one is called the minor groove.



2.11 Types of DNA Double-Helix

DNA will adopt two different forms of helices under different conditions. They are

1) A-forms and

2) B- forms.

These two forms differ in their helical twist, rise, pitch and number of base pairs per turn. *The twist of a helix refers to the number of degrees of angular rotation needed to get from one base unit to another*. In the B-form of helix, this is 36^0while in the A-form it is 33^0. *Rise refers to the height change from one base pair to the next* and is 3.4 Å in the B-form and 2.6 Å in the A-form. *The pitch is the height change to get one full rotation (360^0) of the helix*. This value is 34 Å in the B-form since there are ten base pairs per turn. In the A-form, this value is 28 Å since there are eleven base pairs per full turn.

Since the major and minor grooves expose the edges of the bases, the grooves can be used to tell the base sequence of a specific DNA molecule. The possibility for such recognition is critical, since proteins must be able to recognize specific DNA sequences on which to bind in order for the proper functions of the body and cell to be carried out. The major groove is more information rich than the minor groove. This fact makes the minor groove less ideal for protein binding.

Of the two forms, the B-form is far more common, existing under most physiological conditions. The A-form is only adopted by DNA under conditions of low humidity. RNA, however,

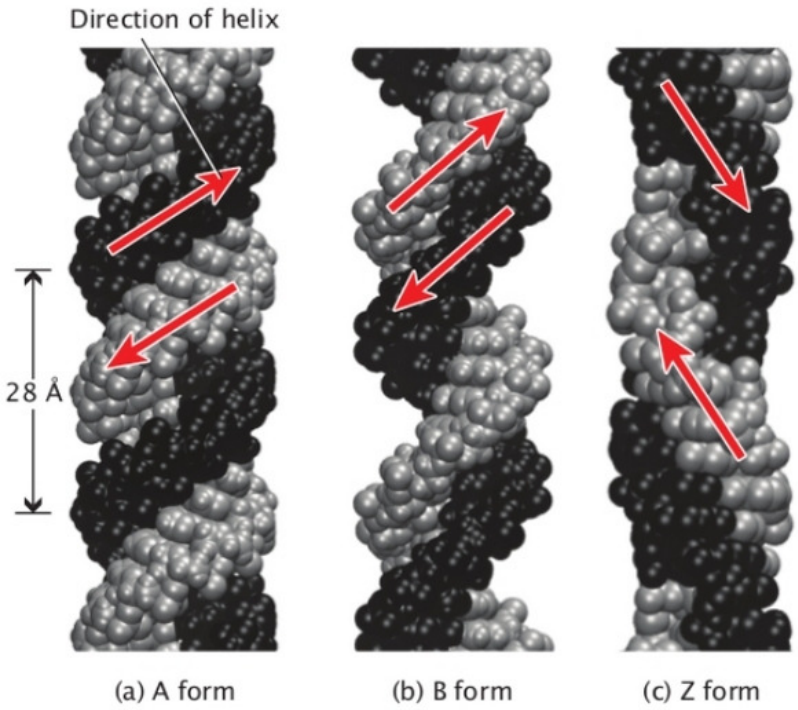
generally adopts the A-form in situations where the major and minor grooves are closer to the same size and the base pairs are a bit tilted.

In addition to these two forms DNA duplex structure also exits in third form called Z-form. Z-duplex DNA has a strikingly different, left-handed helical structure. This Z DNA is formed by stretches of alternating purines and pyrimidines, e.g. GCGCGC, especially in negatively supercoiled DNA. A small amount of the DNA in a cell exists in the Z form. It has been tantalizing to propose that this different structure is involved in some way in regulation of some cellular function, such as transcription or regulation.

Thus Z-DNA is a radically different duplex structure, with the two strands coiling in left-handed helices and a pronounced zig-zag (hence the name) pattern in the phosphodiester backbone. As previously mentioned, Z-DNA can form when the DNA is in an alternating purine-pyrimidine sequence such as GCGCGC, and indeed the G and C nucleotides are in different conformations, leading to the zig-zag pattern. The big difference is at the G nucleotide. It has the sugar in the C3' endoconformation and the guanine base is in the syn conformation. This places the guanine back over the sugar ring, in contrast to the usual anticonformation seen in A- and B-form nucleic acid. Note that having the base in the anticonformation places it in the position where it can readily form H-bonds with the complementary base on the opposite strand. The duplex in Z-DNA has to accomodate the distortion of this G nucleotide in the synconformation. The cytosine in the adjacent nucleotide of Z-DNA is in the "normal" C2' endo, anticonformation.

Comparisons of B-form, A-form and Z‑DNA

	B	A	Z
Helix sense	RH	RH	LH
bp per turn	10	11	12
Vertical rise per bp	3.4	2.56	3.7 Å
Rotation per bp	+36	+33	‑30 degrees
Helical diameter	19	23	18 Ås



The double helix of DNA has these features

1) It contains two polynucleotide strands wound around each other.
2) The backbone of each consists of alternating deoxyribose and phosphate groups.
3) The phosphate group bonded to the 5' carbon atom of one deoxyribose is covalently bonded to the 3' carbon of the next.
4) The two strands are "antiparallel"; that is, one strand runs 5′ to 3′ while the other runs 3′ to 5′.
5) The DNA strands are assembled in the 5′ to 3′ direction and by convention, we "read" them in the same way.
6) The purine or pyrimidine attached to each deoxyribose projects in toward the axis of the helix.
7) Each base forms hydrogen bonds with the one directly opposite to it, forming base pairs (also called nucleotide pairs). 3.4 Å separate the planes in which adjacent base pairs are located.
8) The double helix makes a complete turn in just over 10 nucleotide pairs, so each turn takes a little more (35.7 Å to be exact) than the 34 Å.
9) There is an average of 25 hydrogen bonds within each complete turn of the double helix providing a stability of binding about as strong as what a covalent bond would provide.
10) The diameter of the helix is 20 Å.

2.12 Helicases

Helicases are enzymes that bind and may even remodel nucleic acid or nucleic acid protein complexes. There are DNA and RNA helicases. DNA helicases are essential during DNA replication because they separate double-stranded DNA into single strands allowing each strand to be copied. During DNA replication, DNA helicases unwind DNA at positions called origins where synthesis will be initiated. DNA helicase continues to unwind the DNA forming a structure called the replication fork, which is named for the forked appearance of the two strands of DNA as they are unzipped apart. The process of breaking the hydrogen bonds between the nucleotide base pairs in double-stranded DNA requires energy. To break the bonds, helicases use the energy stored in a molecule called ATP, which serves as the energy currency of cells. DNA helicases also function in other cellular processes where double-stranded DNA must be separated, including DNA repair and transcription. RNA helicases are involved in shaping the form of RNA molecules, during all processes involving RNA, such as transcription, splicing, and translation.

2.13 Codon and genetic code

A codon is a sequence of three DNA or RNA nucleotides that corresponds with a specific amino acid or stop signal during protein synthesis. DNA and RNA molecules are written in a language of four nucleotides; meanwhile, the language of proteins includes 20 amino acids. Codons provide the key that allows these two languages to be translated into each other. *Each codon corresponds to a single amino acid (or stop signal), and the full set of codons is called the genetic code.* Thus DNA provides instructions for making proteins. The sequence of the bases, A, C, G and T, in DNA determines genetic code and provides the instructions for producing molecules in the body. The cell reads the DNA code in groups of three bases. Each triplet of bases, also called a codon, specifies which amino acid will be added next during protein synthesis. There are 20 different amino acids, which are the building blocks of proteins. The genetic code includes 64 possible permutations, or combinations, of three-letter nucleotide sequences that can be made from the four nucleotides. Only 61 of the 64 codons are used to specify which of the 20 amino acids is next to be added. There are three codons that don't code for an amino acid. These codons mark the end of the protein and stop the addition of amino acids to the end of the protein chain.

Leucine	L	CTT, CTC, CTA, CTG, TTA, TTG
Valine	V	GTT, GTC, GTA, GTG
Phenylalanine	F	TTT, TTC
Methionine	M	ATG
Cysteine	C	TGT, TGC
Alanine	A	GCT, GCC, GCA, GCG
Glycine	G	GGT, GGC, GGA, GGG
Proline	P	CCT, CCC, CCA, CCG
Threonine	T	ACT, ACC, ACA, ACG
Serine	S	TCT, TCC, TCA, TCG, AGT, AGC
Tyrosine	Y	TAT, TAC
Tryptophan	W	TGG
Glutamine	Q	CAA, CAG
Asparagine	N	AAT, AAC
Histidine	H	CAT, CAC
Glutamic acid	E	GAA, GAG
Aspartic acid	D	GAT, GAC
Lysine	K	AAA, AAG
Arginine	R	CGT, CGC, CGA, CGG, AGA, AGG
Stop codons	Stop	TAA, TAG, TGA

In the above table, the twenty amino acids found in proteins are listed, along with the single-letter code used to represent these amino acids in protein data bases. The DNA codons representing each amino acid are also listed.

All 64 possible 3-letter combinations of the DNA coding units T, C, A and G are used either to encode one of these amino acids or as one of the three stop codons that signals the end of a sequence. While DNA can be decoded unambiguously, it is not possible to predict a DNA sequence from its protein sequence. Because most amino acids have multiple codons, a number of possible DNA sequences might represent the same protein sequence.

2.14 Anti codon

Anti codon is a sequence of three nucleotides in a region of transfer RNA that recognizes a complementary coding triplet of nucleotides in messenger RNA during translation by the ribosomes in protein biosynthesis.

2.15 Differences between codon and anticodon

Some of the major differences between codon and anticodon are as follows

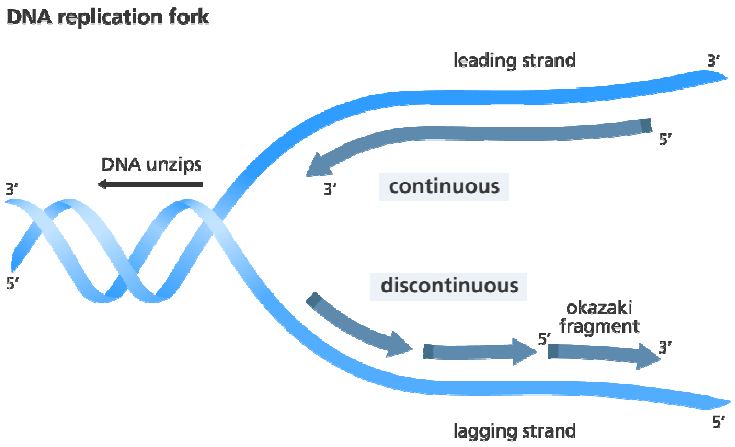
Codon	Anticodon
It is found in DNA and mRNA	It occurs in tRNA.

It is complementary to a triplet of template strand	It is complementary to a codon.
It determines the position of an amino acid in a polypeptide.	It helps in bringing a particular amino acid at its proper position during translation.

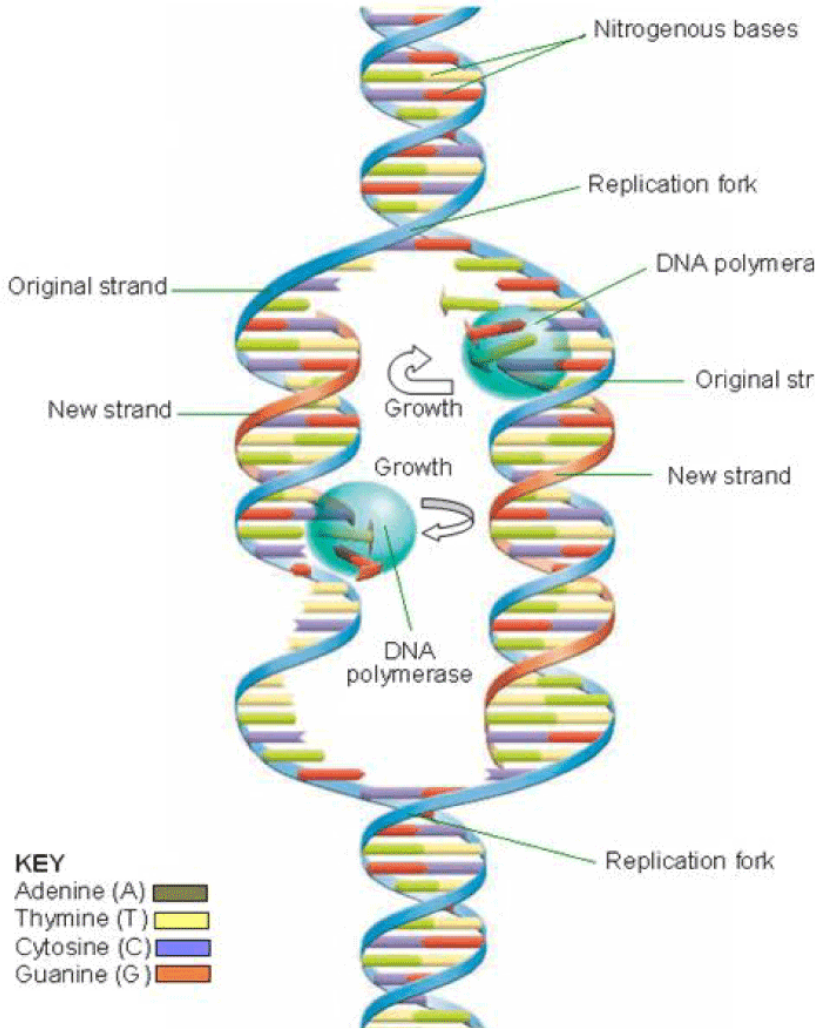
2.16 DNA replication

DNA replication is the process by which DNA makes a copy of itself during cell division. Each strand of the DNA double helix contains a sequence of nucleotides that is exactly complementary to the nucleotide sequence of its partner strand. Each strand can therefore act as a template for the synthesis of a new complementary strand. The process of DNA replication is begun by initiator proteins that bind to the DNA and pry the two strands apart, breaking the hydrogen bonds between the bases. *The positions at which the DNA is first opened are called replication origin.* They usually marked by a particular sequence of nucleotides. During the DNA replication it is possible to see Y-shaped junctions in the DNA, called *replication forks*. At these forks, the replication machine is moving along the DNA, opening up the two strands of double helix and using each strand as a template to make new daughter strand. Two replication forks are formed starting from each replication origin, and they move away from the origin in both directions, unzipping DNA as they go.

Elongation of new DNA at replication fork is catalyzed by enzymes called DNA polymerase. As nucleotides align with complementary bases along "old" template strand of DNA, they are added by polymerase, one by one, to the growing end of the new DNA strand. The rate of elongation is about 500 nucleotides per second in bacteria and 50 per second in human cells. As each monomer joins the growing end of DNA strand, it losses two phosphate groups. Hydrolysis of phosphate is the exergonic reaction that drives polymerization of nucleotides to from DNA.



The 5'-to-3' direction of the DNA polymerization mechanism poses a problem at replication fork One new DNA is being made on a template that runs in one direction (3'to 5'), whereas the other new strand is being made on a template that runs in the opposite direction (5' to 3').



The DNA replication has these features

i) The first step in DNA replication is to 'unzip' the double helix structure of the DNA molecule. This is carried out by an enzyme called helicase which breaks the hydrogen bonds holding the complementary bases of DNA together (A with T, C with G).

ii) The separation of the two single strands of DNA creates a 'Y' shape called a replication 'fork'. The two separated strands will act as templates for making the new strands of DNA.

One of the strands is oriented in the 3' to 5' direction (towards the replication fork), this is the leading strand. The other strand is oriented in the 5' to 3' direction (away from the replication

fork), this is the lagging strand. As a result of their different orientations, the two strands are replicated differently.

Leading Strand

iii) A short piece of RNA called a primer (produced by an enzyme called primase) comes along and binds to the end of the leading strand. The primer acts as the starting point for DNA synthesis.

iv) DNA polymerase binds to the leading strand and then 'walks' along it, adding new complementary nucleotide bases (A, C, G and T) to the strand of DNA in the 5' to 3' direction.

v) This sort of replication is called continuous.

Lagging strand

iiia) Numerous RNA primers are made by the primase enzyme and bind at various points along the lagging strand.

iva) Chunks of DNA, called Okazaki fragments, are then added to the lagging strand also in the 5' to 3' direction.

va) This type of replication is called discontinuous as the Okazaki fragments will need to be joined up later.

vi) Once all of the bases are matched up (A with T, C with G), an enzyme called exonuclease strips away the primer(s). The gaps where the primer(s) were are then filled by complementary nucleotides.

vii) The new strand is proofread to make sure there are no mistakes in the new DNA sequence.

viii) Finally, an enzyme called DNA ligase seals up the sequence of DNA into two continuous double strands.

The result of DNA replication is two DNA molecules consisting of one new and one old chain of nucleotides. This is why DNA replication is described as semi-conservative, half of the chain is part of the original DNA molecule, half is brand new.

2.17 Biosynthesis of Protiens

The genetic material is stored in the form of DNA in most organisms. In humans, the nucleus of each cell contains 3×10^9 base pairs of DNA distributed over 23 pairs of chromosomes, and each cell has two copies of the genetic material. This is collectively known as the human genome. The human genome contains around 30,000 genes, each of which codes for one protein.

Protein synthesis occurs in cellular structures called ribosomes found out-side the nucleus. The process by which genetic information is transferred from the nucleus to the ribosomes is called transcription. During transcription, a strand of ribonucleic acid (RNA) is synthesized. This messenger RNA (mRNA) is complementary to the portion of DNA that directed it. It has a complementary nucleotide at each point in the chain.

As with any of the polymerization reactions, protein synthesis can be divided into three phases

Initiation: In this phase a functionally competent ribosome is assembled in the correct place on an mRNA ready to commence protein synthesis.

Elongation: In this phased the correct amino acid is brought to the ribosome, is joined to the nascent polypeptide chain, and the entire assembly moves one position along the mRNA.

Termination: This happens when a stop codon is reached. There is no amino acid to be incorporated and the entire assembly dissociates to release the newly-synthesized polypeptide.

There are two rules about protein synthesis to keep in mind they are

i) mRNA is translated 5' -> 3' and

ii) Proteins are synthesized from the N-terminus to the C-terminus

2.17.1 Transcription

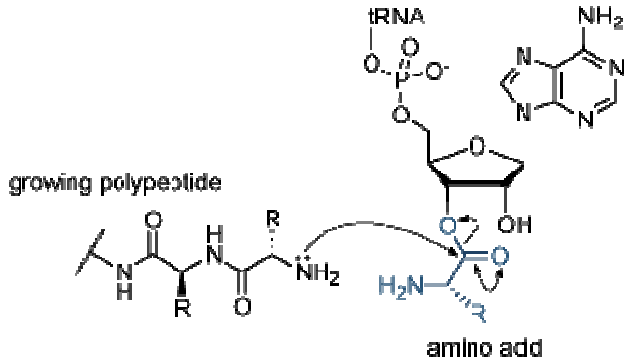
Transcription is the process by which DNA is copied (transcribed) to mRNA, which carries the information needed for protein synthesis.

2.17.2 Translation

The mRNA formed in transcription is transported out of the nucleus, into the cytoplasm, to the ribosome (the cell's protein synthesis factory). Here, it directs protein synthesis. Messenger RNA is not directly involved in protein synthesis – transfer RNA (tRNA) is required for this. *The process by which mRNA directs protein synthesis with the assistance of tRNA is called translation.*

Activation of aminoacids

Each amino acid has its own special tRNA (or set of tRNAs). For example, the tRNA for phenylalanine (tRNAPhe) is different from that for histidine (tRNAHis). Each amino acid is attached to its tRNA through the 3′-OH group to form an ester which reacts with the α-amino group of the terminal amino-acid of the growing protein chain to form a new amide bond (peptide bond) as shown in below figure during protein synthesis. The reaction of esters with amines is generally favourable but the rate of reaction is increased greatly in the ribosome.

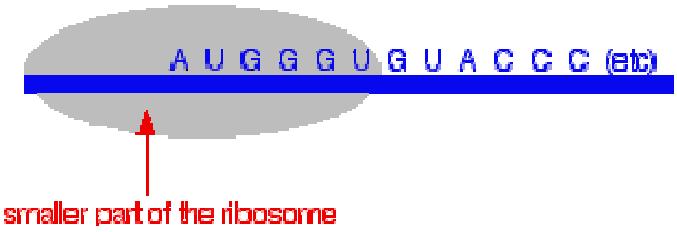


In the last step of the protein synthesis, the polypeptide chain is terminated by adequate termination signals (three special stop codons) in the mRNA. Eventually, the completed chain separates from the ribosome. The release of the polypeptide-tRNA from the ribosome is initiated, when stop codon is reached, by a specific protein factor (release factor) which is attached to the ribosome and hydrolytically cleaves the ester bond between polypeptide and tRNS.

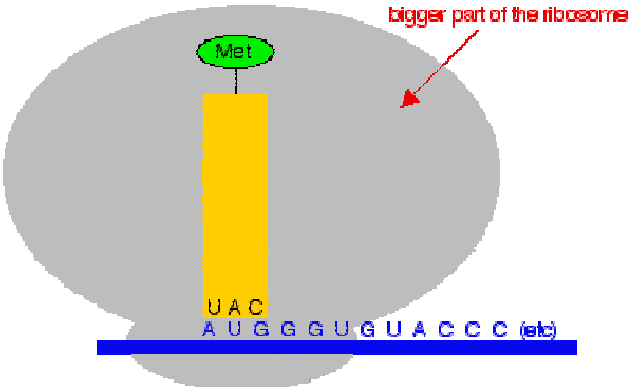
The below carton will explain the biosynthesis of proteins through step by step.

Step 1. Translation is the name given to the process of turning the coded message in the messenger RNA into the final protein chain.

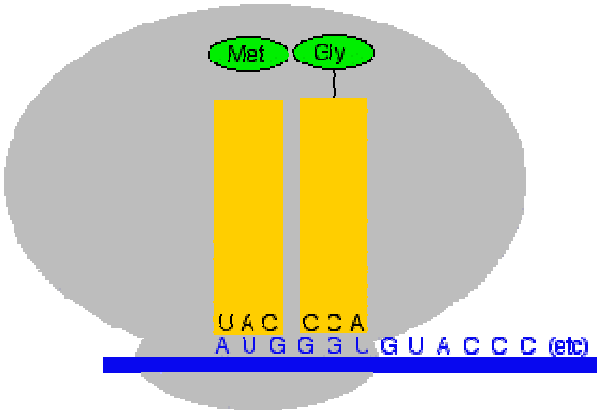
The messenger RNA has a part of ribosome attached to it at the AUG start codon. A small part of the RNA base sequence downstream of the start codon make a coded message of protein chain which is to be synthesized.



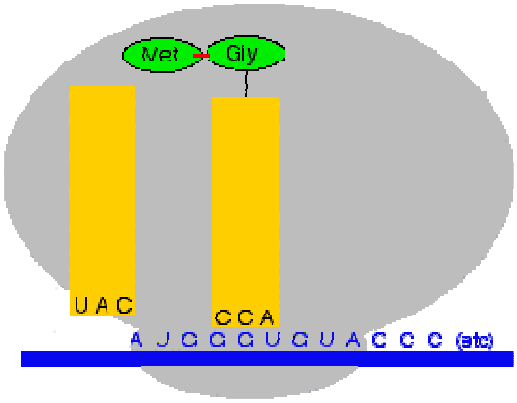
Step 2. Here two things are happen. The transfer RNA carrying a methionine (for example) attaches itself to the AUG codon by pairing its anti-codon bases with the complementary bases on the messenger RNA. And the second, bigger part of the ribosome attaches to the system as well.



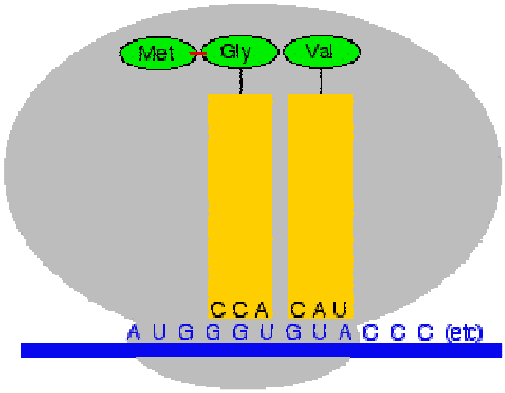
Then another transfer RNA molecule with its attached amino acid binds to the next codon along the chain. The next codon on the messenger RNA is GGU which codes for glycine (Gly). The anti-codon would therefore have to be CCA.



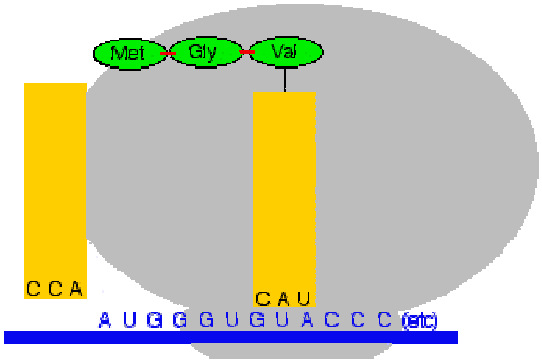
Next, the ribosome moves along the messenger RNA chain to the next codon. At the same time a peptide bond is made between the two amino acids, and the first one (the methionine) breaks away from its transfer RNA. That transfer RNA molecule leaves the ribosome and goes off to pick up another methionine.



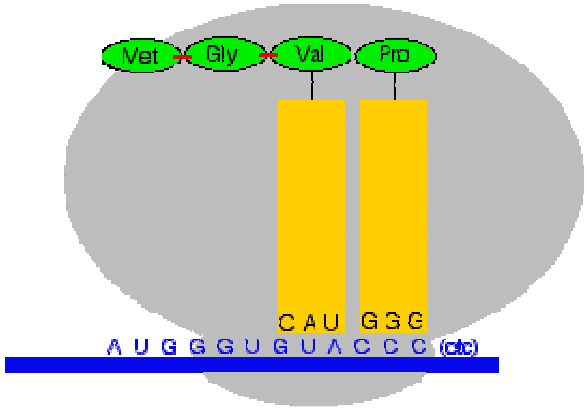
Like this the process repeats. The next codon is GUA which codes for valine (Val). The anti-codon must be CAU.



And again, the ribosome moves forward one codon, a new peptide bond is formed, and the transfer RNA on the left breaks away to be used again later.



And the next transfer RNA with its amino acid comes along . . . and so on, and on and on . . .



At the end the ribosome will come to a stop codon. The three stop codons don't code for any amino acids, and so the process comes to a halt. The protein chain produced up to that point is then released from the ribosome, and then folds itself up into its secondary and tertiary structures.

2.18 Summary of the unit

Chargaff's realization that A = T and C = G, combined with some crucially important X-ray crystallography work by English researchers Rosalind Franklin and Maurice Wilkins, contributed to Watson and Crick's derivation of the three-dimensional, double-helical model for the structure of DNA.

DNA is a double-stranded helix, with the two strands connected by hydrogen bonds. A bases are always paired with Ts, and Cs are always paired with Gs, which is consistent with and accounts for Chargaff's rule.

Most DNA double helices are right-handed; that is, if we were to hold your right hand out, with our thumb pointed up and our fingers curled around our thumb, our thumb would represent the axis of the helix and your fingers would represent the sugar-phosphate backbone. Only one type of DNA, called Z-DNA, is left-handed.

The DNA double helix is anti-parallel, which means that the 5' end of one strand is paired with the 3' end of its complementary strand (and vice versa), nucleotides are linked to each other by their phosphate groups, which bind the 3' end of one sugar to the 5' end of the next sugar.

Not only the DNA base pairs connected via hydrogen bonding, the outer edges of the nitrogen-containing bases are exposed and available for potential hydrogen bonding as well. These hydrogen bonds provide easy access to the DNA for other molecules, including the proteins that play vital roles in the replication and expression of DNA.

A sequence of three adjacent nucleotides constituting the genetic code that determines the insertion of a specific amino acid in a polypeptide chain during protein synthesis or the signal to stop protein synthesis.

Each polynucleotide chain serves as a template for the synthesis of a new DNA molecule. DNA replication follows the base pairing rules by which A pairs only with T, and G with C, and as a consequence of this, each daughter molecule is an exact replica of the parent molecule. One of the strands of each daughter DNA molecule is newly synthesized, whereas the other is derived from the parent DNA molecule. This distribution of parental atoms is called semiconservative.

Protein synthesis can be divided into three phases: Initiation, where a functionally competent ribosome is assembled in the correct place on an mRNA ready to commence protein synthesis.

Elongation, whereby the correct amino acid is brought to the ribosome, is joined to the nascent polypeptide chain, and the entire assembly moves one position along the mRNA.

Termination, which happens when a stop codon is reached, there is no amino acid to be incorporated and the entire assembly dissociates to release the newly-synthesized polypeptide.

There are two rules about protein synthesis to keep in mind: mRNA is translated 5' → 3' and Proteins are synthesized from the N-terminus to the C-terminus

2.19 Key words

Nucleic acid back bone; Structure of nucleosides; Oligonucleotides; Structure of DNA; Chargaff's Rule; Watson and Crick base paring; Base stacking in double-helix; Stability of double-helix structure; The Major and Minor Grooves; Types of DNA Double-Helix; Helicases;

Codon and genetic code; Anti codon; DNA replication; Biosynthesis of Protiens; Transcription; Translation

2.20 References for further studies

1) Nucleic Acids: Structures, Properties, and Functions; Victor A. Bloomfield, Donald M. Crothers, Ignacio Tinoco; *University Science Books*, **2000**.

2) Nucleic Acids in Chemistry and Biology; G. Michael Blackburn; *Royal Society of Chemistry*, **2006**.

3) Nucleic Acid Structure: An Introduction; W. Guschlbauer; *Springer Science & Business Media*, **2012**.

4) Advanced Organic Chemistry of Nucleic Acids; Zoe A. Shabarova, Alexey A. Bogdanov; *John Wiley & Sons*, **2008**.

5) Basic Principles in Nucleic Acid Chemistry, Volume 1; Paul O.P. Ts'o; *Elsevier*, **2012**.

2.21 Questions for self understanding

1) What are Oligonucleotides?

2) How nucleic acid back bone is formed?

3) Discuss the structure of nucleosides.

4) Explain the silent features of DNA structure.

5) Explain Chargaff's Rule of base composition in nucleic acid.

6) Deduce the Watson and Crick base paring.

7) Write a note on Base stacking in double-helix.

8) Explain the factors governing the stability of double-helix structure

9) What are The Major and Minor Grooves in DNA double helix? Explain their importance.

10) Discuss the different types of DNA Double-Helix structure.

11) What are Helicases?

12) What is Codon? Explain the term genetic code

13) What are anti codons?

14) Write the differences between codon and anticodon.

15) Discuss briefly about DNA replication.

16) Explain the following steps involved in biosynthesis of Proteins

a) Transcription

b) Translation

UNIT-3

Structure

3.0 Objectives of the unit

After studying this unit you are able to

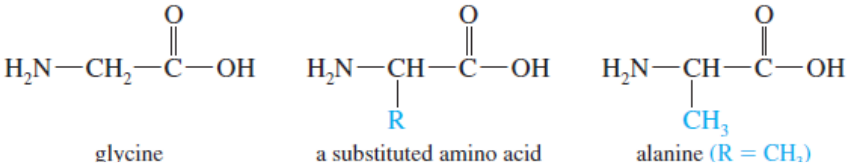
- Write the structure of 20 naturally occurring amino acids
- List the standard amino acids and Essential amino acids
- Explain the Isoelectric points
- Classify the amino acids based on nature of chains
- Write the Stereo chemistry of α-carbon atom in amino acid
- Write the different methods of synthesis of amino acids

3.1 Introduction

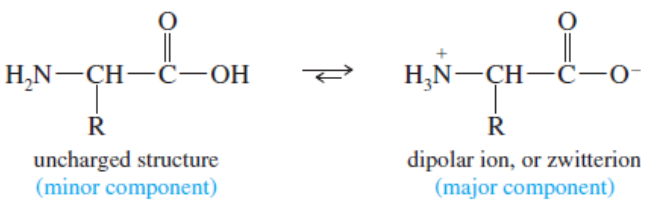
Proteins are the most abundant organic molecules in animals, playing important roles in all aspects of cell structure and function. Proteins are biopolymers of α- amino acids, so named because the amino group is bonded to the α-carbon atom, next to the carbonyl group. The physical and chemical properties of a protein are determined by its constituent amino acids. The individual amino acid subunits are joined by amide linkages called peptide bonds.

3.2 Amino acids

The term amino acid means any molecule containing both an amino group and any type of acid group, however, the term is almost always used to refer to an carboxylic acid. The simplest acid is amino acetic acid, called glycine. Other common amino acids have side chains (symbolized by R) substituted on the α-carbon atom. For example, alanine is the amino acid with a methyl side chain.



Commonly amino acids are written with an intact carboxyl (-COOH) group and amino ($-NH_2$) group, their actual structure is ionic and depends on the pH. The carboxyl group loses a proton, giving a carboxylate ion, and the amino group is protonated to an ammonium ion. This structure is called a dipolar ion or a zwitterion

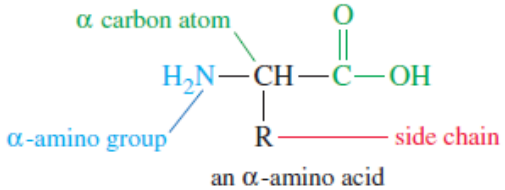


3.3 The standard amino acids

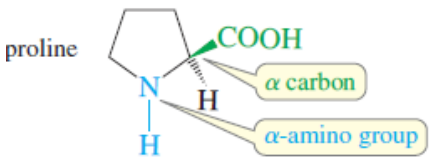
The hydrolysis of polypeptide yields a set of 20 different amino acids. These amino acids are referred as standard amino acids. The standard amino acids are 20 common -amino acids that are found in nearly all proteins. The standard amino acids differ from each other in the structure of the side chains bonded to their α-carbon atom. Common abbreviations for the standard amino acids are listed in below Table.

Amino Acid	Three-Letter Abbreviation	One-Letter Abbreviation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Out of 20 standard amino acids, 19 have the same general structure. These molecules contain a central carbon atom called α-carbon to which an primary amino group, a carboxylate group, a hydrogen atom, and an R group (generally called side chain) are attached.

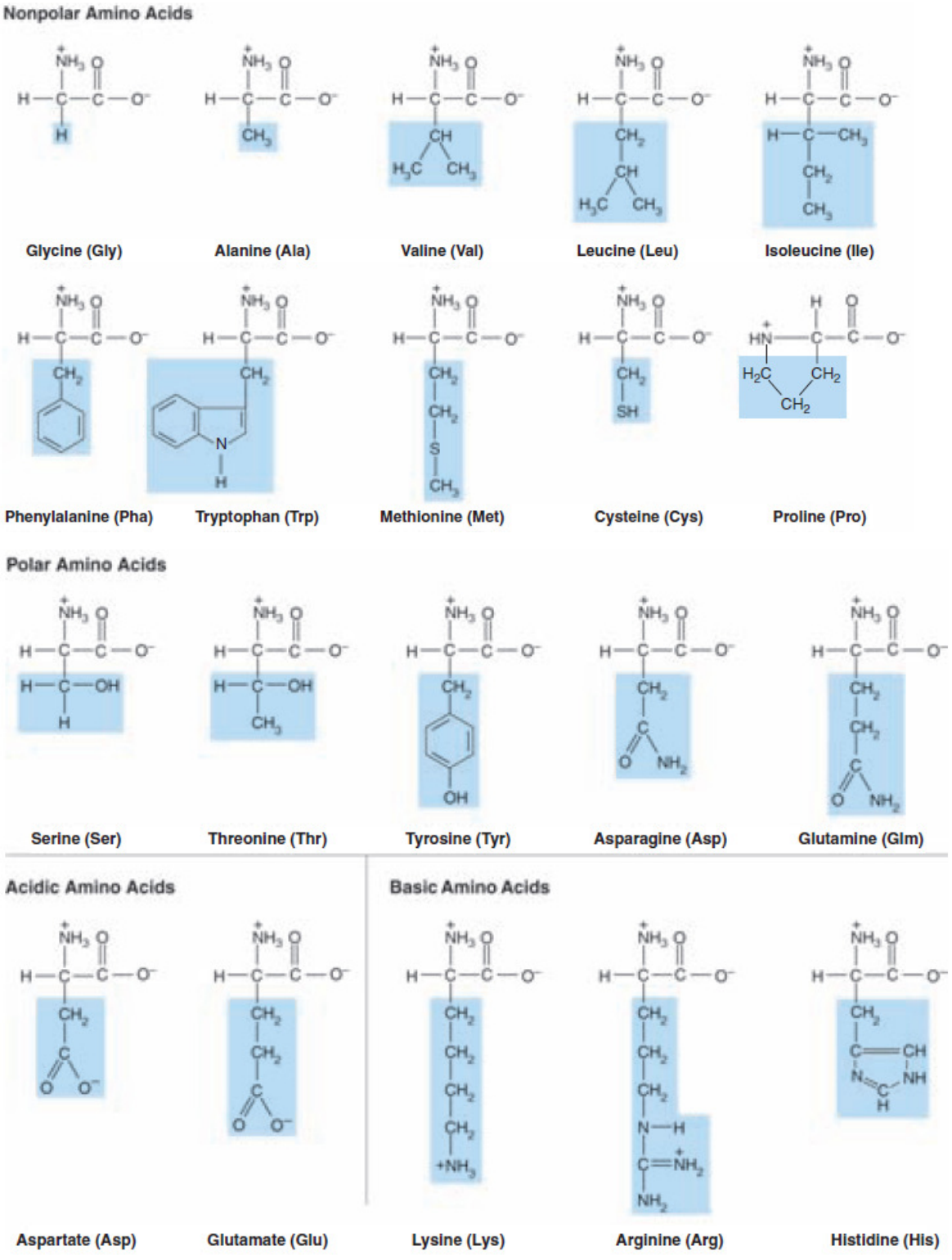


an α-amino acid



The Proline is an exception, which differs from the other standard amino acids in that its amino group is secondary, formed by ring closure between the R group and the amino nitrogen.

The structures of the 20 amino acids that are commonly found in naturally occurring polypeptides are shown in below figure



3.4 Essential amino acids

Human body synthesizes only about half of the amino acids needed to make proteins. Other amino acids are called the essential amino acids, they must be provided in the diet. The ten essential amino acids are the following:

Arginine (Arg); Valine (Val); Methionine (Met); Leucine (Leu); Threonine (Thr); Phenylalanine (Phe); Histidine (His); Isoleucine (Ile); Lysine (Lys); Tryptophan (Trp)

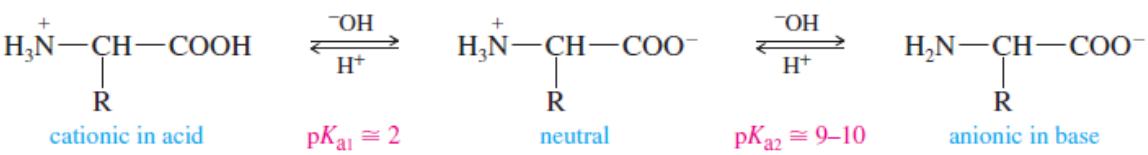
Proteins that provide all the essential amino acids in about the right proportions for human nutrition are called complete proteins. Proteins that are severely deficient in one or more of the essential amino acids are called incomplete proteins. If the protein in a person's diet comes mostly from one incomplete source, the amount of human protein that can be synthesized is limited by the amounts of the deficient amino acids. Plant proteins are generally incomplete.

Rice, corn, and wheat are all deficient in Lysine. Rice also lacks Threonine, and corn also lacks Tryptophan. Beans, peas, and other legumes have the most complete proteins among the common plants, but they are deficient in Methionine.

Vegetarians can achieve an adequate intake of the essential amino acids if they eat many different plant foods. Plant proteins can be chosen to be complementary, with some foods supplying amino acids that others lack. An alternative is to supplement the vegetarian diet with a rich source of complete protein such as milk or eggs.

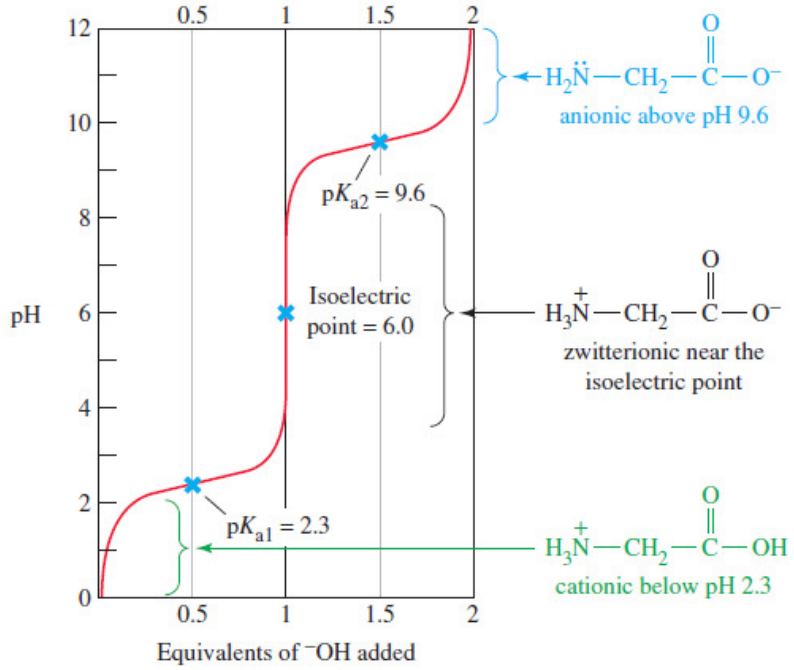
3.5 Isoelectric Points

Because amino acids contain both acidic ($-NH_3^+$) and basic ($-COO^-$) groups, they are amphoteric (having both acidic and basic properties). The predominant form of the amino acid depends on the pH of the solution. In an acidic solution, the $-COO^-$ group is protonated to a free –COOH group, and the molecule has an overall positive charge. As the pH is raised, the –COOH loses its proton at about pH 2. This point is called pK_{a1}, the first acid-dissociation constant. As the pH is raised further, the $-NH_3^+$ group loses its proton at about pH 9 or 10. This point is pK_{a2} called the second acid-dissociation constant. Above this pH, the molecule has an overall negative charge.



Thus an amino acid bears a positive charge in acidic solution (low pH) and a negative charge in basic solution (high pH). There must be an intermediate pH where the amino acid is evenly balanced between the two forms, as the dipolar zwitterion with a net charge of zero. This pH is called the isoelectric pH or the isoelectric point.

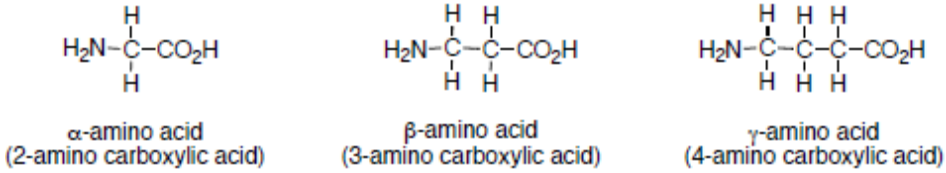
The isoelectric pH depends on the amino acid structure. Acidic amino acids have isoelectric point around 2.8 to3.2. Neutral amino acids have isoelectric point around 5.0 to 6.3 and basic amino acids have isoelectric point around 7.6 to 10.8.



3.6 Classification of amino acids

Amino acids are classified in two ways

The first way is based on the location of the amino group. Based on this, amino acids are classified in to α-amino acids, β-amino acids, γ-amino acids etc….



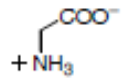
In second way amino acids are classified according to their capacity to interact with water. Using this criterion amino acids are classified in to following four classes: 1) nonpolar, 2) polar, 3) acidic, and 4) basic.

Nonpolar amino acids

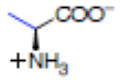
The nonpolar amino acids contain mostly hydrocarbon R groups that do not bear positive or negative charges. Nonpolar (i.e., hydrophobic) amino acids play an important role in maintaining the three-dimensional structures of proteins, because they interact poorly with water.

Two types of hydrocarbon side chains are found in this group, they are aromatic and aliphatic. Aromatic hydrocarbons contain cyclic structures that constitute a class of unsaturated hydrocarbons with planar conjugated π electron clouds.

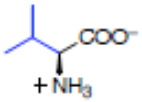
The term aliphatic refers to nonaromatic hydrocarbons. Glycine, alanine, valine, leucine, isoleucine, and proline have aliphatic R groups. A sulfur atom appears in the aliphatic side chains of methionine and cysteine. Methionine contains a thioether group (—S—CH_3) in its side chain.



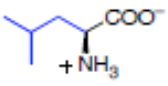
Glycine *(Gly, G)*



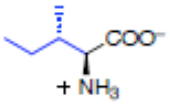
(S)-(+)-Alanine *(Ala, A)*



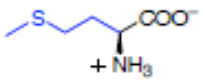
(S)-(+)-Valine *(Val, V)*



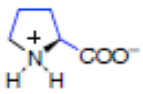
(S)-(–)-Leucine *(Leu, L)*



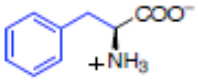
(2S,3S)-(+)-Isoleucine *(Ile, I)*



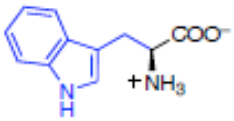
(S)-(–)-Methionine *(Met, M)*



(S)-(–)-Proline *(Pro, P)*



(S)-(–)-Phenylalanine *(Phe, F)*

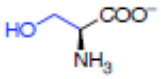


(S)-(–)-Tryptophan *(Trp, W)*

Polar but non ionizable amino acids

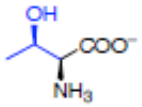
Polar amino acids have functional groups capable of hydrogen bonding, they easily interact with water. Polar amino acids are described as hydrophilic, or "water-loving. Serine, threonine, tyrosine, asparagine, and glutamine belong to this category. Serine, threonine, and tyrosine contain a polar hydroxyl group, which enables them to participate in hydrogen bonding.

Asparagine and glutamine are amide derivatives of the acidic amino acids aspartic acid and glutamic acid, respectively. Because the amide functional group is highly polar, the hydrogen-bonding capability of asparagine and glutamine has a significant effect on protein stability.



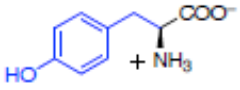
(S)-(–)-Serine *(Ser, S)*

pKa ~ 13



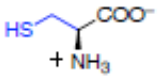
(2S,3R)-(–)-Threonine *(Thr, T)*

pKa ~ 13



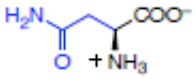
(S)-(–)-Tyrosine *(Tyr, Y)*

pKa ~ 10.1

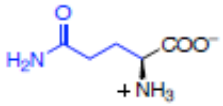


(R)-(–)-Cysteine *(Cys, C)*

pKa ~ 8.2



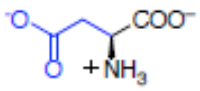
(S)-(–)-Asparagine *(Asn, N)*



(S)-(+)-Glutamine *(Gln, Q)*

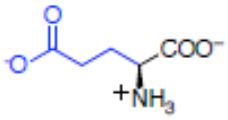
Acidic amino acids

Two standard amino acids have side chains with carboxylate group they are aspartic acid and glutamic acid. Because the side chains are negatively charged at physiological pH, they are often referred to as aspartate and glutamate.



(S)-(+)-Aspartic Acid *(Asp, D)*

pKa ~ 3.6



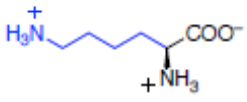
(S)-(+)-Glutamic Acid *(Glu, E)*

pKa ~ 4.2

Basic amino acids

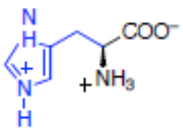
Basic amino acids bear a positive charge at physiological pH. They can therefore form ionic bonds with acidic amino acids Lysine, which has a side chain amino group, accepts a proton from water to form the conjugate acid (—NH_3^+).

Because the guanidino group of arginine has a pKa range of 11.5 to 12.5 in proteins, it is permanently protonated at physiological pH and, therefore, does not function in acid-base reactions. The imidazole side chain histidine, on the other hand, is a weak base because it is only partially ionized at pH 7 (pKa is approximately 6). Its capacity under physiological conditions to accept or donate protons in response to small changes in pH plays an important role in the catalytic activity of numerous enzymes.



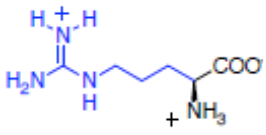
(S)-(+)-Lysine *(Lys, K)*

pKa ~ 10.5



(S)-(–)-Histidine *(His, H)*

pKa ~ 6.0

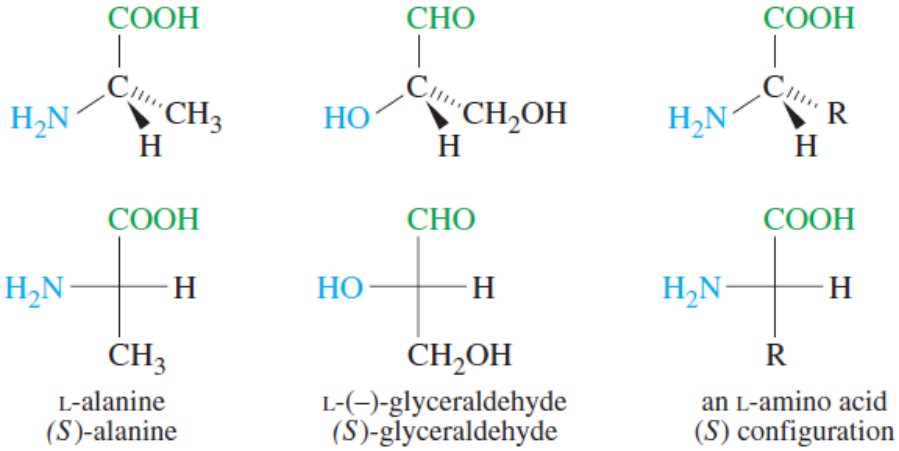


(S)-(+)-Arginine *(Arg, R)*

pKa ~ 12.5

3.7 Stereo chemistry of Amino Acid

The α-carbons of 19 of the 20 standard amino acids are attached to four different groups (i.e., a hydrogen, a carboxyl group, an amino group, and an R group). Hence the α-carbon in these amino acids is asymmetric or chiral. Glycine is an exception because its α-carbon is attached to two hydrogens, hence it is a non chiral amino acid. As we aware, molecules with chiral carbons can exist as stereoisomers. In all of the chiral amino acids, the chirality center is at α-carbon atom. Nearly all the naturally occurring amino acids are found to have the (S) configuration at this α-carbon atom. They are called L-amino acids because their stereochemistry resembles that of L(-)-glyceraldehyde. ie, the natural configuration of the α-carbon is L.



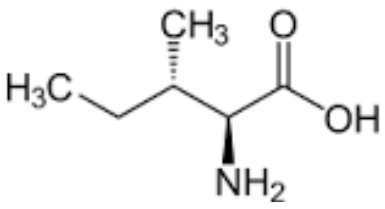
L-alanine
(*S*)-alanine

L-(–)-glyceraldehyde
(*S*)-glyceraldehyde

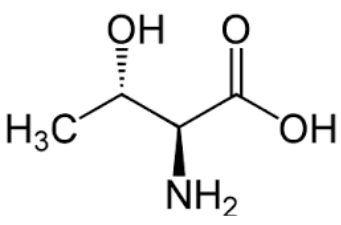
an L-amino acid
(*S*) configuration

D-Amino acids are found in the cell walls of bacteria. The D-amino acids are not genetically encoded, but derived from the epimerization of L-isomers

Two of the standard amino acids, isoleucine and threonine, have a second asymmetric carbon at β-carbons in addition to their α-carbons. The β-carbon stereochemistry in these two amino acids is 3*S* and 3*R* respectively. Thus naturally occurring Isoluecine has (2*S*, 3*S*) and Threonine has (2*S*, 3*R*) configuration.



L-isoleucine (2*S*,3*S*)



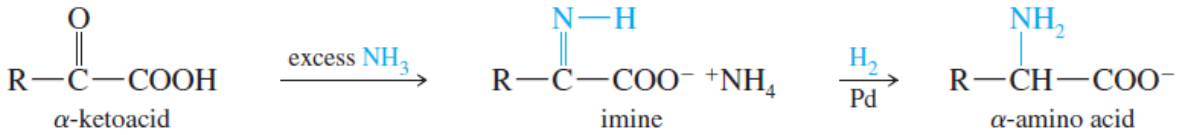
L-Threonine (2*S*,3*R*)

3.8 Synthesis of amino acids

Naturally occurring amino acids can be obtained by hydrolyzing proteins and separating the amino acid mixture. Many methods are available for synthesis of natural and non-natural amino acids.

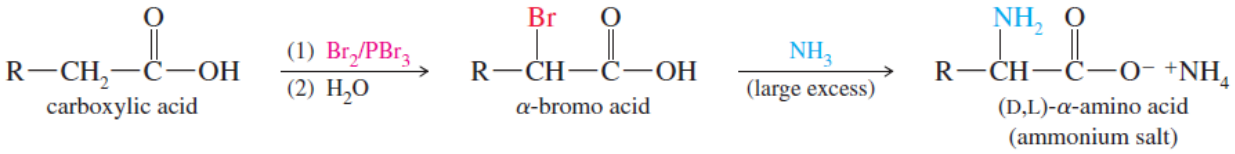
Reductive Amination

Reductive amination of ketones and aldehydes is one of the best methods for synthesizing amines. When an α-ketoacid is treated with ammonia, the ketone reacts with ammonia to form an imine. The imine is reduced to an amine by hydrogen in presence of palladium catalyst. Under these conditions, the carboxylic acid is not reduced. The product so obtained is a racemic α-amino acid.



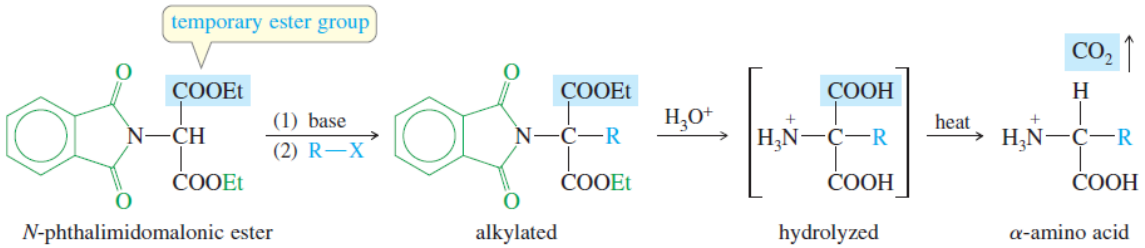
Amination of an α-Halo Acid

The Hell–Volhard–Zelinsky reaction is an effective method for introducing bromine at the α-position of a carboxylic acid. The racemic α-bromo acid is converted to a racemic α-amino acid by direct amination, using a large excess of ammonia.



The Gabriel–Malonic Ester Synthesis

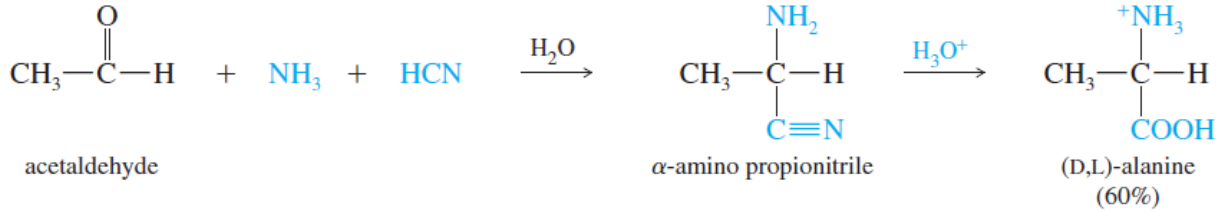
One of the best methods of amino acid synthesis is a combination of the Gabriel synthesis of amines with the malonic ester synthesis of carboxylic acids. The Gabriel–malonic ester synthesis begins with N-phthalimidomalonic ester. The amino group protected as an amide (a phthalimide in this case) to keep it from acting as a nucleophile. The acid is protected as an ethyl ester. N-Phthalimidomalonic ester is alkylated in the same way as malonic ester. When the alkylated N-phthalimidomalonic ester is hydrolyzed, the phthalimido group is hydrolyzed along with the ester groups. The product is an alkylated aminomalonic acid. Decarboxylation gives a racemic α -amino acid.



The Gabriel–malonic ester synthesis is used to make many amino acids that cannot be formed by direct amination of haloacids.

The Strecker Synthesis

Strecker added acetaldehyde to an aqueous solution of ammonia and HCN. The product was α-amino propionitrile, which Strecker hydrolyzed to racemic alanine. The Strecker synthesis can form a large number of amino acids from appropriate aldehydes.



3.9 Summary of the unit

Hydrolysis of proteins by boiling aqueous acid or base yields an assortment of small molecules identified as α-aminocarboxylic acids. More than twenty such components have been isolated.

The term amino acid might mean any molecule containing both an amino group and any type of acid group, however, the term is almost always used to refer to an α-aminocarboxylic acid. The simplest α-aminoacid is aminoacetic acid, called glycine. Other common amino acids have side chains (symbolized by R) substituted on the α-carbon atom. For example, alanine is the amino acid with a methyl side chain.

Some common features of amino acids are, with the exception of proline, they are all 1°-amines; and with the exception of glycine, they are all chiral. The configurations of the chiral amino acids are the same.

In free aminoacid at normal conditions the carboxyl group loses a proton, giving a carboxylate ion, and the amino group is protonated to an ammonium ion. This structure is called a dipolar ion or a zwitterions. The dipolar nature of amino acids gives them some unusual properties:

1. Amino acids have high melting points, generally over 200 °C.

2. Amino acids are more soluble in water than they are in ether, dichloromethane, and other common organic solvents.
3. Amino acids have much larger dipole moments (μ) than simple amines or simple acids.
4. Amino acids are less acidic than most carboxylic acids and less basic than most amines. In fact, the acidic part of the amino acid molecule is the $-NH_3^+$ group, not a –COOH group. The basic part is the $–COO^-$ group, and not a free $–NH_2$ group.

3.10 Key words

Amino acids; Standard amino acids; Essential amino acids; Isoelectric Points; Classification of amino acids; Stereo chemistry of Amino Acid

3.11 Reference for further studies

1) Amino Acids, Peptides and Proteins in Organic Chemistry, Analysis and Function of Amino Acids and Peptides; Andrew B. Hughes; *John Wiley & Sons*, **2013**.
2) Amino Acids and Peptides; G. C. Barrett, D. T. Elmore; *Cambridge University Press*, **1998**.
3) Amino Acids, Peptides And Proteins; Sp Bhutani; *Ane Books Pvt Ltd*, **2010**.
4) Lehninger Principles of Biochemistry; David L. Nelson, Michael M. Cox; *W. H. Freeman*, **2012**.
5) Introduction to Peptides and Proteins; Ulo Langel, Benjamin F. Cravatt, Astrid Graslund, N.G.H. von Heijne, Matjaz Zorko, Tiit Land, Sherry Niessen; *CRC Press*, **2009**.
6) Solid-Phase Synthesis: A Practical Guide; Fernando Albericio; *CRC Press*, **2000**.

3.12 Questions for self understanding

1) What are Amino acids?
2) What are standard amino acids?
3) What are essential amino acids?
4) Write the structure of 20 naturally occurring aminoacids.
5) What is Isoelectric Point?
6) Discuss the stereo chemistry of α-carbon atom of amino acids.
7) How aminoacids are classified?
8) Glycine is a non-sterogenic aminoacid explain why?
9) How proline is different from other aminoacids?
10) Name the two aminoacids having 2 stero centres? Write their structure.
11) Explain the different methods of synthesis of aminoacids.

UNIT-4

Structure

4.0 Objectives of the unit

After studying this unit you are able to

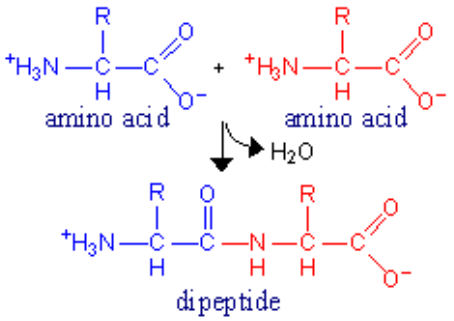
- Explain the structure of peptide
- Write the different methods available for peptide Synthesis
- Recognize the different N-protecting groups used in peptide synthesis
- Identify the Carboxy (-COOH) protection groups in peptide synthesis
- Choose the suitable side chain protection groups in peptide synthesis
- Explain the necessity of coupling reagents in peptide synthesis
- Write the mechanism of racemization of peptide bond
- Write the structure of different racemisation suppressants
- Name the different peptides

4.1 Introduction

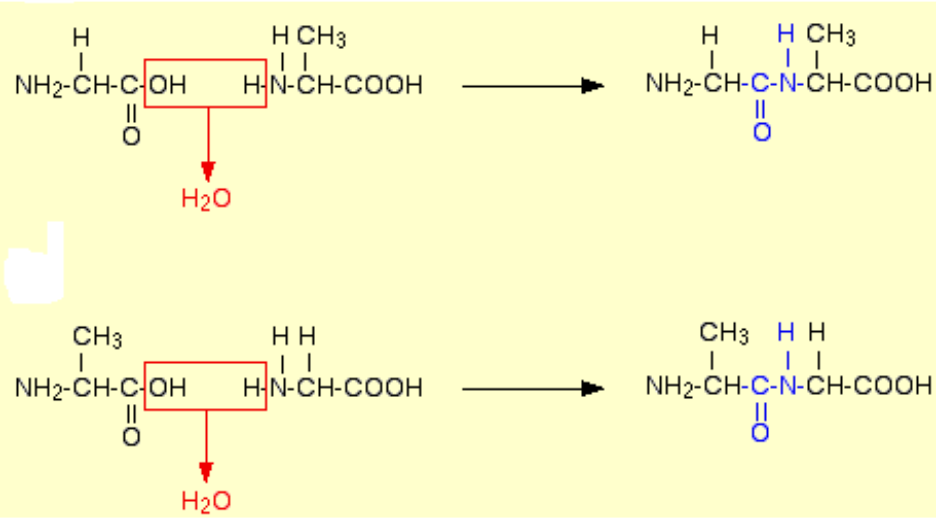
The proteins are made up of one or more polypeptides, unbranched polymers of 20 different amino acids. The diversity of proteins is directly related to the combinatorial possibilities of the 20 amino acid monomers. Amino acids are linked to form protein molecules in any imaginable size or sequence, for example, a hypothetical protein composed of 100 amino acids. The total possible number of combinations for such a molecule is an astronomical 20^{100}. However, of the trillions of possible protein sequences, only a small fraction (possibly no more than 2 million) is actually produced in all living organisms. An important reason for this remarkable discrepancy is demonstrated by the complex set of structural and functional properties of naturally occurring proteins that have evolved over billions of years in response to selection pressure.

4.2 peptides

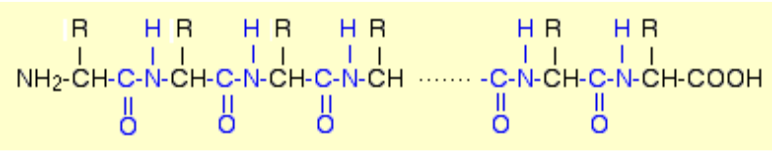
Peptides are polymers of amino acids made using anything from two to hundreds of amino acids. They are all based on the α-amino acid structure. *If the α-amine functional group of one amino acid joins together with carboxylic acid functional group another aminoacid to form amide bonds, it is called a peptide*.



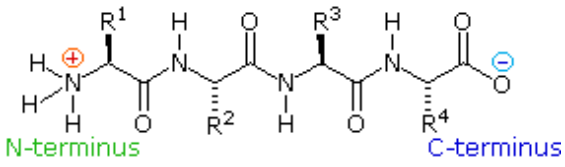
For example; the glycine and alanine combine together with the elimination of water molecule to produce a dipeptide. This combination can happen in two different ways ie, α-amine of glycine join with carboxylic acid group of alalnine or vice versa. So we might get two different dipeptides as shown in below figure. In each case, the linkage of the dipeptide is known as a peptide linkage.



Similarly if we join the three amino acids together, we would get a tripeptide and so on. Hence if we joined more aminoacids together, we would get a polypeptide.



By convention, when we drawn peptide chains, the -NH_2 group which hasn't been converted into a peptide link is written at the left-hand end. The unchanged -COOH group is written at the right-hand end. The end of the peptide chain with the -NH_2 group is known as the N-terminal, and the end with the -COOH group is the C-terminal. A protein chain (with the N-terminal on the left) will therefore look like as follows.



4.3 Peptide Synthesis

Peptide synthesis is important for following reasons including

1) Confirming the structure of natural peptides (e.g. for medical research etc.)

2) To investigate how peptide structure and function are controlled by the amino acid sequence

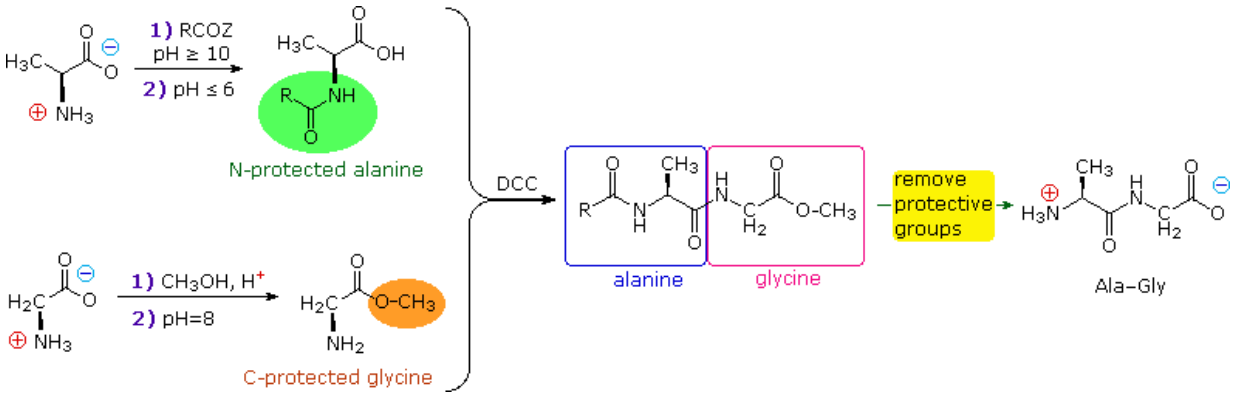
However, peptide synthesis it is not as straight forward as mixing the amino acids together to form the amides. In case if we follow general procedure to synthesize a peptide from its component amino acids, we must face two major difficulties.

The first one is statistical in nature, and is illustrated by taking dipeptide Ala-Gly as an example. Suppose if we ignore the chemistry involved in the aminde bond formation, a mixture of equal molar amounts of alanine and glycine would generate four different dipeptides. These are: Ala-Ala, Gly-Gly, Ala-Gly & Gly-Ala. In the case of tripeptides, the number of possible products from these two amino acids rises to eight. It clearly indicates that some kind of selectivity must be exercised if complex mixtures are to be avoided.

The second difficulty arises from the fact that carboxylic acids and 1º or 2º-amines do not form amide bonds on mixing, but will generally react by proton transfer to give salts (the intermolecular equivalent of zwitterion formation).

Therefore in order to control the coupling reaction to form amide bond between two amino acids, it is necessary to use protecting groups. By protecting the amine group of one amino acid component and the carboxylic acid group of the other amino acid, a specific amide bonds can be formed.

The strategy used for peptide synthesis is outlined below. The following example shows a selective synthesis of the dipeptide Ala-Gly.



Following sequential steps are performed in peptide synthesis

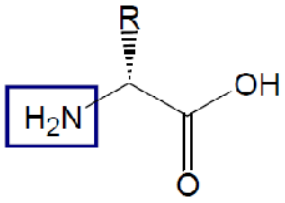
a) Protect the amino group in the N-terminal amino acid and the carboxyl group of the C-terminal amino acid

b) Couple the two amino acids by forming the new amide bond

c) Deprotect the termini of the new peptide (as and if required)

By repeating the process, polypeptides can be grown one amino acid residue at a time, or by building pieces and then joining those together.

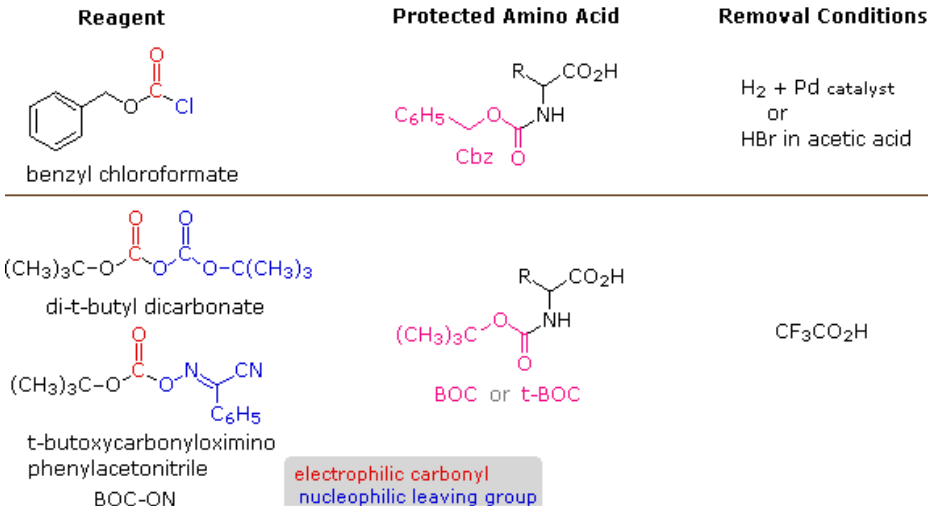
4.4 N-protecting groups used in peptide synthesis



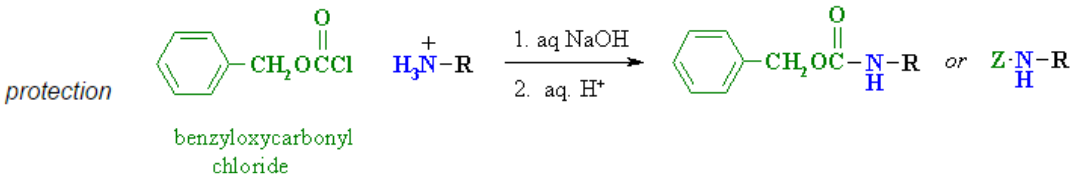
Carbamates type protection is mainly used for protection of amnio group of aminoacid in peptide synthesis. An important issue remains to be addressed here is, since the N-carbamate protection group is also an amide, conditions used for removal of this function might also cleave the just formed peptide bond. Furthermore, the conditions often required for amide hydrolysis might cause extensive racemization of the amino acids in the resulting peptide. Hence it is important to give careful attention to the design of specific N-protective groups. Following three qualities are desired for best N-protecting groups in peptide synthesis.

1) The protective amide should be easy to attach for amino acids.
2) The protected amino group should not react under peptide forming conditions.
3) The protective amide group should be easy to remove under mild conditions.

A number of protecting groups that satisfy these conditions have listed below. Among them, Benzoyloxycarbonyl (Cbz) and t-butoxycarbonyl (BOC or t-BOC), are most widely used in solution phase synthesis.

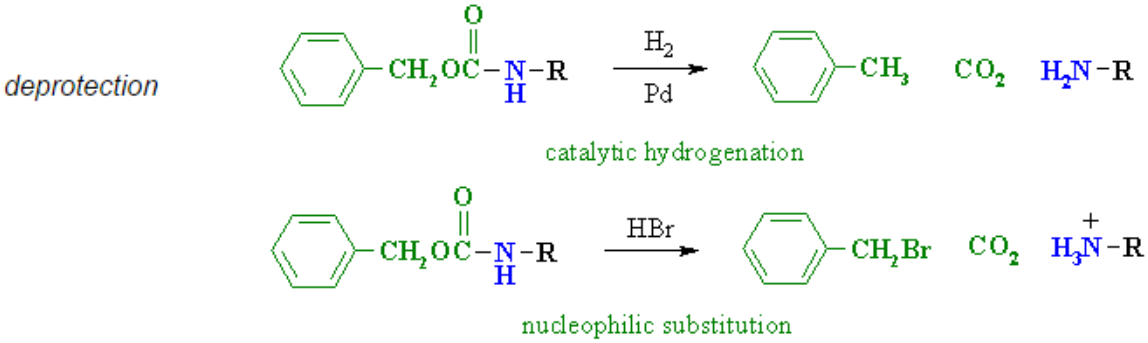


1) benzoyloxycarbonyl groups (abbreviation = Z or Cbz in older literature)

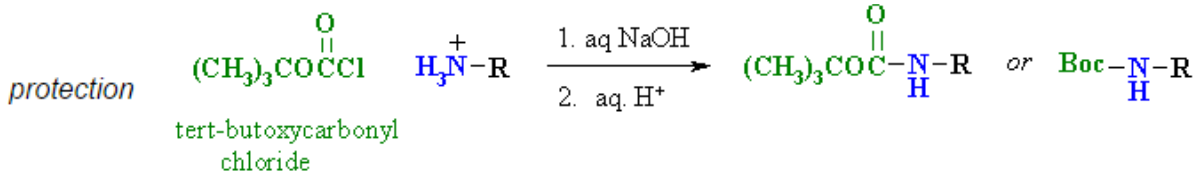


CBz protection was easily deprotected either by catalytic hydrogenation conditions in presence of palladium catalyst, this process is called hydroginalysis. or

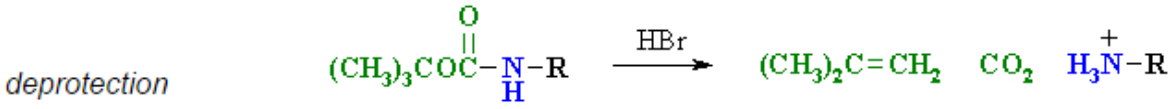
Nucleophilic substitution reaction by HBr as shown below



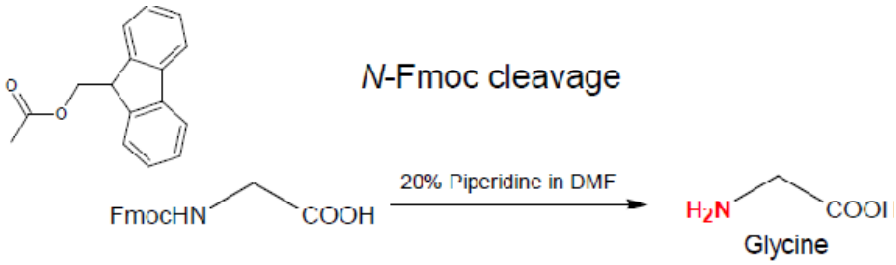
2) tert-butoxycarbonyl groups (abbreviation = Boc)



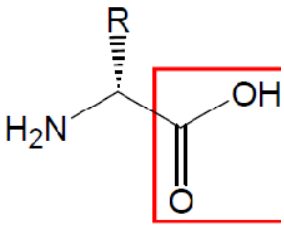
Boc protection was easily deptrotected by acid treatment. This reagent will be removed by treatment of 50% trifluoroacetic acid solution in dichloromethane solvent or just treatment by concentrated HBr



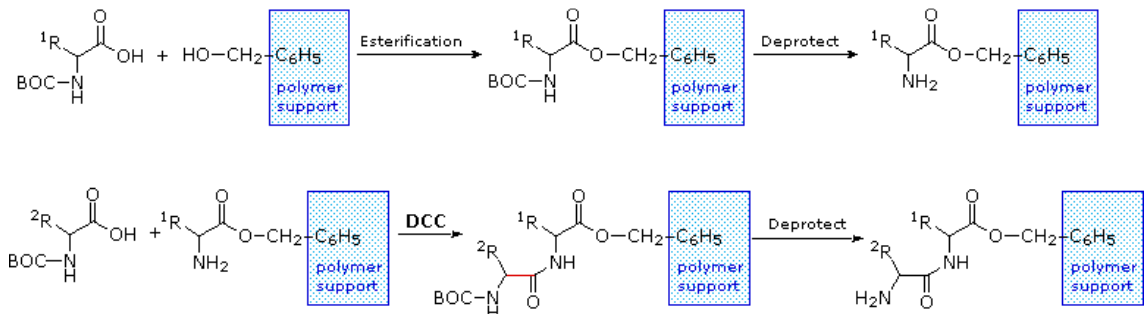
Other N-protection group extensively used in peptide synthesis is 9-fluorenylmethyl carbamate (Fmoc) group. Currently it is widely used protective group in solid phase peptide synthesis. The advantage of Fmoc is that it is cleaved under very mild basic conditions (e.g. piperidine),



4.5 Carboxy (-COOH) protection groups



The most common acid protecting groups used are the methyl or ethyl esters. They are stable in most coupling reaction conditions and deprotection reaction conditions of N-protection groups. Allyl ester, benzyl ester, t-butyl ester are other acid protecting groups used in peptide synthesis. After the desired peptide bond is formed the N-protective group can be removed under relatively mild conditions. Cleavage of the reactive benzyl or tert-butyl groups generates a common carbamic acid intermediate (HOCO-NHR) which spontaneously loses carbon dioxide, giving the corresponding free amine. This can be used for further elongation of peptide chain, if the ester at the C-terminus is left as such by reacting with another N-Protected amino acid. This sequence of reactions may be repeated, using a different N-protected amino acid as the acylating reagent to prepare tripeptide, tetrapeptide etc…. Removal of the protecting groups would then yield a corresponding free peptide.



It is important to mention here that, the synthesis of a peptide of significant length (e.g. ten residues) by this approach requires many steps, and the product must be carefully purified after

each step to prevent unwanted cross-reactions. To facilitate the tedious and time consuming purifications, and reduce the material losses that occur in handling, a clever modification of this strategy has been developed. This procedure is popularly known as the Merrifield Solid Phase peptide Synthesis.

Using this method two or more moderately sized peptides can be joined together by selective peptide bond formation, provided side-chain functions are protected and do not interfere. In this manner good sized peptides and small proteins may be synthesized in the laboratory.

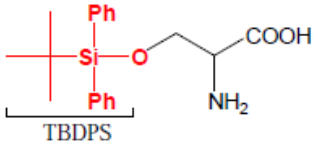
4.6 Side chain protection groups

Amino acids have reactive moieties at the N-termini, C-termini, and side chains. One category of protecting group allows temporary protection of the α-amino group. This temporary protection group is removed after each peptide bond formation during oligopeptide synthesis. Another type of protecting group can permanently protect groups by blocking the amino acids' side-chain functionalities. These permanent protecting groups withstand in the condition used for repeated cleavages of the temporary protecting group. They are removed only after the completion of oligopeptide synthesis. Untimely removal of these protecting groups should not cause the formation of undesirable by-products.

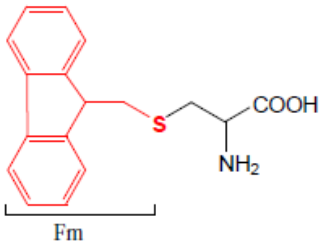
Some of the side chains (R-groups) they must be protected before coupling are hydroxyl group in Ser, Thr and Tyr, thiol group in Cys, amine group in Lys, Guavanine group in Arg, immidazole group in His, indole group in Trp and carboxylic acids group in Asp and Glu.

The selection of protection group for these side chains are depends on what type of protection group is used for α-NH_2 group. In general orthogonal protection strategy is used for α-NH_2 and side chain functional group of amino acids. Base sensitive side chain protection groups are used for different functional groups in Boc and acid sensitive side chain protection groups are used for different functional groups in Fomc chemistry and are listed below.

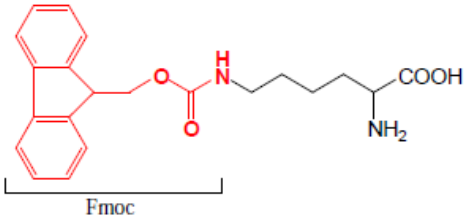
Base Sensitive Protecting Groups



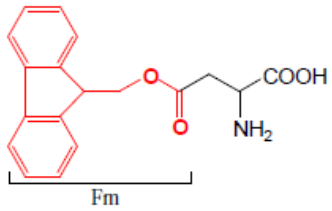
Cleavage : 2M NaOH (aq) EtOH (1:1)



Cleavage : 20% Piperidine in DMF

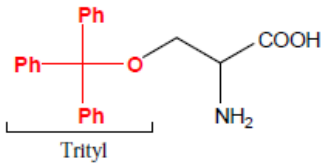


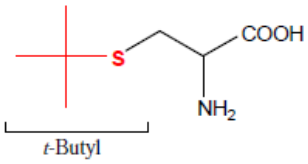
Cleavage : 20% Piperidine in DMF

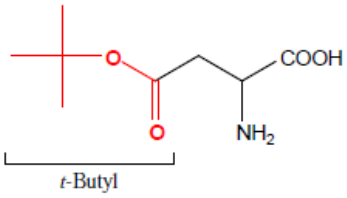


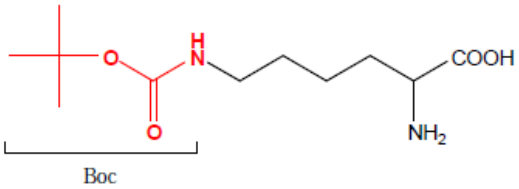
Cleavage : 20% Piperidine in DMF

Acid Sensitive Protecting Groups







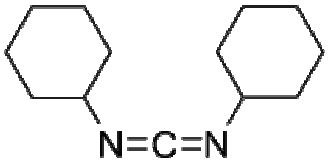


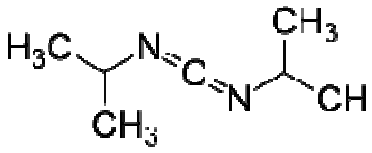
4.7 Coupling reagents

For coupling the peptides the carboxyl group is usually activated. This is important for speeding up the reaction. For this purpose activating groups are used. There are two main types of activating groups used in peptide synthesis, they are carbodiimides and aromatic oximes.

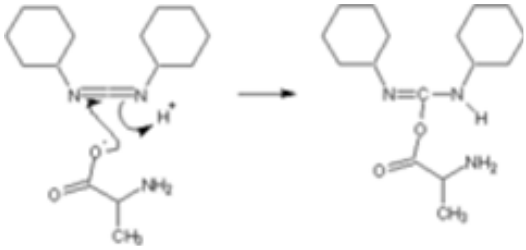
Carbodiimides

Most common carbodiimibe areganets used in peptide synthesis are dicyclohexylcarbodiimide (DCC) and diisopropylcarbodiimide (DIC).

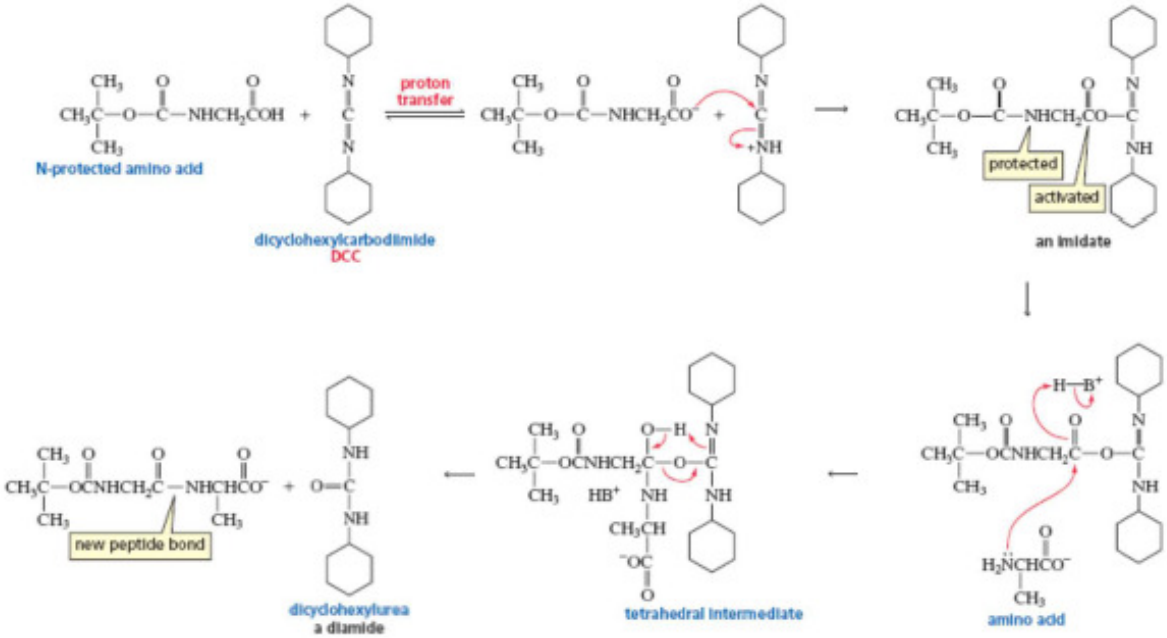




These reagents react with carboxylic acid yields a highly reactive O-acyl-urea. To enhance the electrophilicity of carboxylate group, the –OH group of carboxylic acid must first be "activated" into a better leaving group. DCC is used for this purpose. The negatively charged oxygen will act as a nucleophile, attacking the central carbon in DCC. DCC is temporarily attached to the carboxylate group.

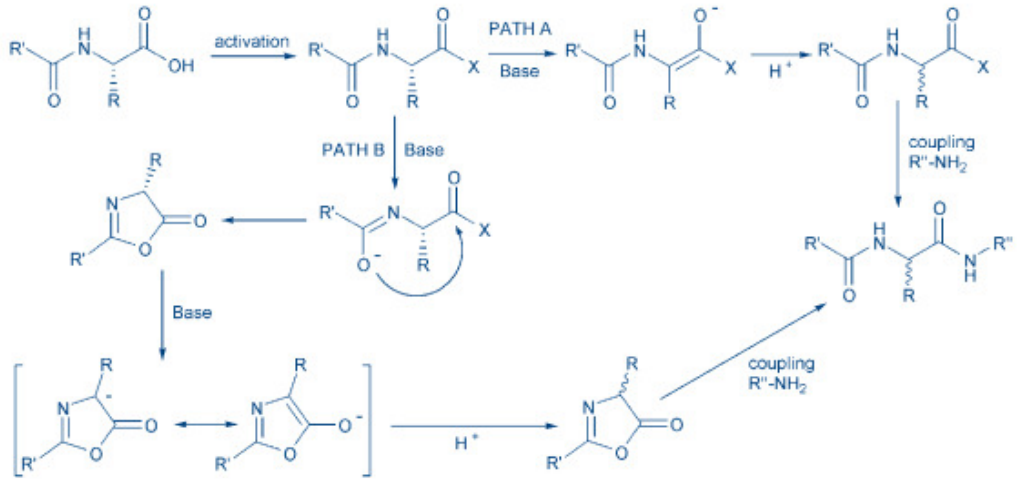


The nucleophilic attack by an amino group of another amino acid yields the dipeptide followed by releasing acyl urea as shown in below reaction scheme.



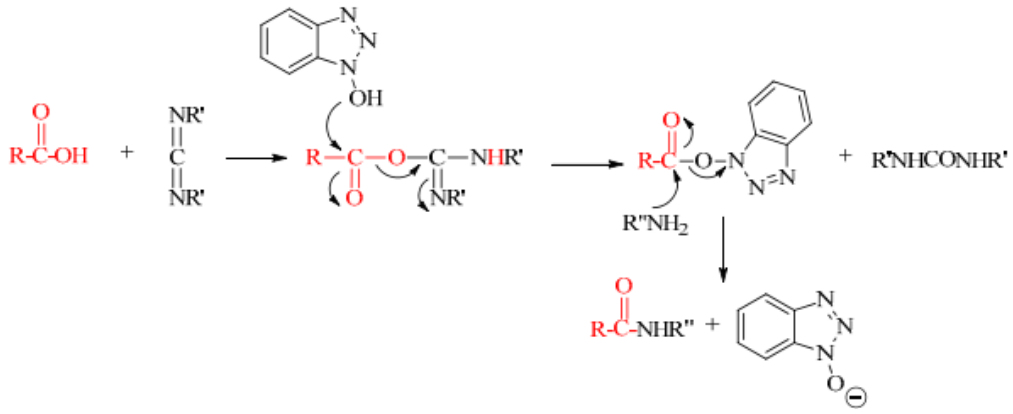
4.7 Racemization of peptide bond

A common problem in peptide synthesis is the racemization of the amino acid, usually through 5(4H)-oxazolone (azlactone) formation and other side reactions. Carbodiimide coupling reagents are too reactive and they can cause racemization of the amino acid. During the activation of carboxyl group of amino acids (except for glycine), electron-withdrawing groups bound to the α-amino moiety (acyl,) which considerably increase the tendency to racemize. The mechanism of racemization is shown in below scheme, which involves the abstraction of the α-hydrogen from the α-carbon atom of the activated amino acid, either by direct formation of an enolic intermediate (direct α-abstraction, path A) or by formation of a 5-membered oxazolinone ring (path B), which isomeric aromatic configuration is readily formed in the presence of bases.

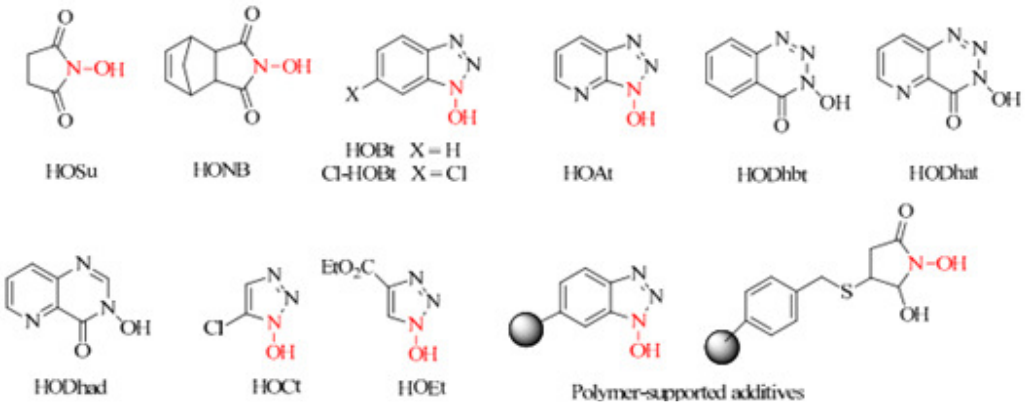


4.8 Racemisation suppressants

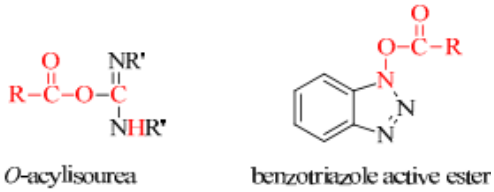
Additives are used in coupling reactions to inhibit side reactions and reduce racemization. The reactivity of peptide coupling reagents may be enhanced by the addition of additives that may also serve to reduce the extent of racemization. For instance, active species can be captured using a hydroxylamine derivative to give the corresponding active ester. Many additives have a X-OH structure that can form active esters with the carboxylic acid.



Ko¨nig and Geiger first reported the use of HOBt is used as racemisation suppressant in peptide coupling reactions with carbodiimide coupling reagents. Several common additives are: N-hydroxysuccinimide (HOSu), N-hydroxy-5-norbornene-2,3-dicarboximide (HONB), 1-hydroxybenzotriazole (HOBt), 6-chloro-1-hydroxybenzotriazole (6-Cl-HOBt), 1-hydroxy-7-azabenzotriazole (HOAt) and more recently 3-hydroxy-4-oxo-3,4- dihydro-1,2,3-benzotriazine (HODhbt) and its aza derivative (HODhat).



Addition of benzotriazoles to carbodiimide-based coupling reagents form benzotriazole active esters that are less reactive than the O-acylisoureas formed from carbodiimides, thereby reducing racemization of the protected amino acid and avoiding the formation of undesired derivatives.



4.9 Geometry of peptide bond

Typical single C-N bond length is 1.47 Å and typical double C=N bond length is 1.27 Å. The peptide bond length lies between C=O to NH is 1.33 Å. This is in between bond length observed for single and double carbon-nitrogen bond.

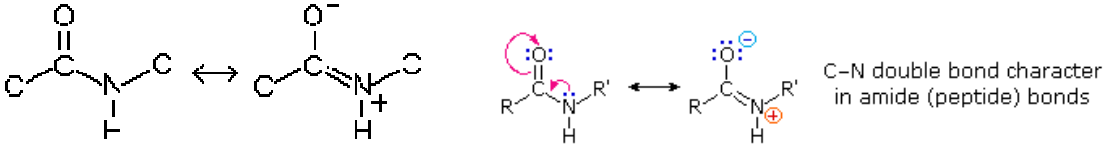


Figure : Resonance forms of the peptide bond.

This is because nitrogen electron pair delocalization into the carbonyl group results in significant double bond character between the carbonyl carbon and the nitrogen. This keeps the peptide links relatively planar and resistant to conformational change. Thus the conformational flexibility of peptide chains is limited since rotations about the bonds leading to the alpha-carbon atoms is restricted. This restriction is due to the rigid nature of the amide (peptide) bond

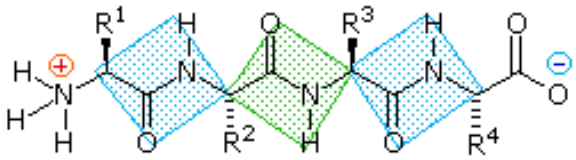
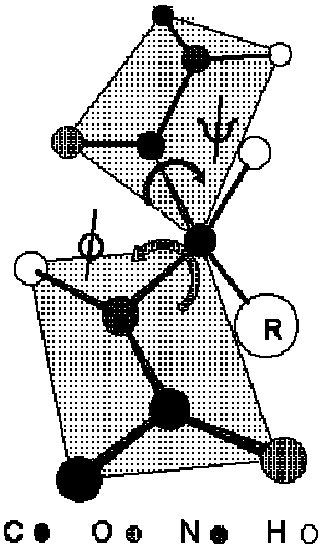


Figure : Planar peptide bond segment

The shape of the peptide chain can be defined by the three consecutive bond torsional angles mentioned below Peptide bonds are almost invariably fixed at omega = 180^{o} or trans based on the relative alignment of C atoms on either side of the peptide bond. This aspect of peptide structure is an important factor influencing the conformations adopted by proteins and large peptides.

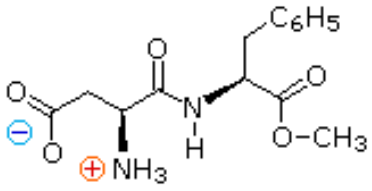


Bond	Rotation	Torsion angle defined
NH to C	free	phi (Φ)
C to C=O	free	psi (Ψ)
C=O to NH	rigid planar	omega (ω)
(peptide bond)	(due to double bond character)	

4.10 Naming the Peptides

Because the N-terminus of a peptide chain is distinct from the C-terminus, a small peptide composed of different amino acids may have a several constitutional isomers. For example, a dipeptide made from two different amino acids may have two different structures. Thus, aspartic acid (Asp) and phenylalanine (Phe) may be combined to make Asp-Phe or Phe-Asp, remember that the amino acid on the left is the N-terminus. The methyl ester of the first dipeptide is the artificial sweetener aspartame, which is nearly 200 times sweeter than sucrose. Neither of the

component amino acids is sweet (Phe is actually bitter), and derivatives of the other dipeptide (Phe-Asp) are not sweet. It is convenient that peptide sequence is written N-terminus end on left and C-terminus end on right. Thus peptide sequence is always written by N- to C-terminus from left to write.



Asp-Phe-OCH_3
Aspartame

Similarly a tripeptide composed of three different amino acids can be made in 6 different constitutions, and the tetrapeptide (composed of four different amino acids) would have 24 constitutional isomers. Thus when all twenty natural amino acids are possible components of a peptide, the possible combinations are enormous. Simple statistical probability indicates that the decapeptides made up from all possible combinations of these amino acids would total 20^{10}.

4.11 Biological importance of insulin

Insulin is a relatively small peptide hormone produced by β-cells in the pancreas. It's main job is to signal the liver, muscle and fat tissues to take up glucose from the blood and store it as glycogen.

*Insulin is composed of two peptide chains referred to as the A chain and B chain***.** A and B chains are linked together by two disulfide bonds, and an additional disulfide is formed within the A chain. In most species, the A chain consists of 21 amino acids and the B chain of 30 amino acids.

The human body requires the blood glucose level maintained in a very narrow range. Homeostasis is regulated by two hormones, insulin and glucagon which are both secreted by the endocrine pancreas. The production of insulin and glucagon by these pancreatic cells ultimately determines if a patient has diabetes or another related problem

Insulin is secreted by the beta cells of the pancreas in response to high blood sugar, a low level of insulin is always secreted by the pancreas. After a meal, the amount of insulin secreted into the blood increases as the blood glucose rises. Likewise, as blood glucose falls, insulin secretion by the pancreatic beta cells decreases. In response to insulin, cells (muscle, red blood cells, and fat cells) take glucose in from the blood, which ultimately lowers the high blood glucose levels back to the normal range. As the glucose level in the blood drops to normal, insulin release slows or stops. If it drops too low, an antagonistic hormone, called glucagon, is released which does the opposite of insulin, stimulating cells to break down glycogen and release glucose

Insulin does much more than just control of blood glucose levels. Its effects depend on the cell type that receives its signal. Fat cells, for example, don't take up or store glucose. Instead, they respond to insulin by taking the fats that enter the blood stream and turning them into fatty acids, which they store in large vacuoles. Thus insulin promotes the uptake and storage of fat in our adipose tissues.

Furthermore, insulin stimulates the body to absorb most amino acids. However, it doesn't lead to intake of tryptophan by cells. This creates an effective rise in tryptophan concentration in the blood, allowing it to pass through the blood brain barrier. In the brain, tryptophan is converted to serotonin, a neurotransmitter whose primary purpose, in this case, is to reduce appetite.

As we said earlier human insulin consists of two chains having a total of 51 amino acids, connected by di sulfide linkages. Insulin is necessary for proper utilization of carbohydrates and people with severe diabetes must take insulin injections. The amount of human insulin available is too small to meet the need, so bovine insulin (from cattle) or insulin from hogs or sheep is used instead. Insulin from these sources is similar to human insulin, but not identical. the differences are in the 8, 9 and 10 positions of the A chain and the C-terminal position (30) of the B chain.

	A chain	**B chain**
	8 9 10	30
Human	-Thr- Ser-Ile	-Thr-
Bovine	-Ala-Ser-Val-	-Ala-
Hog	-Thr-Ser-Ile-	-Ala-
Sheep	-Ala-Gly-Val-	-Ala-

The remainder of the molecule is the same in all four varieties of insulin. Despite the slight differences in the structure, all these insulin can be used by humans and perform the same function as human insulin. However none of the other three are quite as effective as human insulin. Another factor is that sometimes patients become allergic to say for example bovine insulin can switch to hog or sheep insulin without causing allergies.

4.12 Summary of the unit

Peptides are short chains of amino acid monomers linked by peptide bonds. The covalent chemical bonds are formed when the carboxyl group of one amino acid reacts with the amino group of another. The shortest peptides are dipeptides, consisting of 2 amino acid joined by a single peptide bond, followed by tripeptides, tetrapeptides etc….

C-terminus is the end of an amino acid chain, terminated by a free carboxyl group (-COOH). N-terminus refers to the start of a peptide with a free amine group (-NH2). The convention for writing peptide sequence is to put the C-terminal end on right and write the sequence from N-to C-terminus.

4.13 Key words

Peptides; Peptide Synthesis; N-protecting groups; Carboxy (-COOH) protection groups; Side chain protection groups; Coupling reagents; Racemization of peptide bond; Racemisation suppressants; Geometry of peptide bond

4.13 References for further studies

1) Amino Acids, Peptides and Proteins in Organic Chemistry, Analysis and Function of Amino Acids and Peptides; Andrew B. Hughes; *John Wiley & Sons*, **2013**.

2) Amino Acids and Peptides; G. C. Barrett, D. T. Elmore; *Cambridge University Press*, **1998**.

3) Amino Acids, Peptides And Proteins; Sp Bhutani; *Ane Books Pvt Ltd*, **2010**.

4) Lehninger Principles of Biochemistry; David L. Nelson, Michael M. Cox; *W. H. Freeman*, **2012**.

5) Introduction to Peptides and Proteins; Ulo Langel, Benjamin F. Cravatt, Astrid Graslund, N.G.H. von Heijne, Matjaz Zorko, Tiit Land, Sherry Niessen; *CRC Press*, **2009**.

6) Solid-Phase Synthesis: A Practical Guide; Fernando Albericio; *CRC Press*, **2000**.

4.14 Questions for self understanding

1) What are Peptides?

2) How Peptide bonds are formed?

3) Discuss the Different N-protecting groups used in peptide synthesis.

4) Discuss the Carboxy (-COOH) protection groups used in peptide synthesis.

5) Name and write the structure of amino acids which needs side chain protection groups.

6) Write the different side chain protection groups available for amino acids?

7) What are coupling reagents? Explain their role in peptide synthesis.

8) Write the different coupling reagents used in peptide synthesis.

9) Write the mechanism of racemization of peptide bond.

10) What are racemisation suppressants? Mention any two racemisation suppressants used in peptide synthesis.

11) Discuss the role of racemisation suppressants in peptide synthesis.

12) Explain briefly geometry of peptide bond.

13) How the peptide sequence is normally written?

14) What do you mean by N-terminal and C-terminal in peptide?

15) Explain the method adopted for naming the peptides?

UNIT-5

Structure

5.0 Objectives of the unit

After studying this unit you are able to

- Predict the differences between peptides and proteins
- Write the schematic graph of classification of proteins
- List out the biological functions of proteins
- Explain the primary structure of protein
- Explain the different methods available for N-Terminal group analysis of protein
- Explain the different methods available for C-Terminal group analysis of protein
- Explain the different secondary structure adopted by protein
- Explain the tertiary structure of protein
- Explain the quaternary structure of protein

5.1 Introduction

Protein molecules are biological tools and perform variety of functions in living organism. They serve as structural materials in all living organisms (e.g., actin and myosin in animal muscle cells). Proteins are involved in diverse functions such as catalysis, metabolic regulation, transport, and defense. This unit begins with a review of the structures and chemical properties of the amino acids. This is followed by descriptions of the structural and functional features of peptides

Molecules with molecular weights ranging from several thousand to several million daltons are called polypeptides. Those with low molecular weights, typically consisting of fewer than 50 amino acids, are called peptides. The term protein describes molecules with more than 50 amino acids. Each protein consists of one or more polypeptide chains.

Of all the molecules encountered in living organisms, proteins have the most diverse functions. Proteins can be distinguished based on their number of amino acids (called amino acid residues), their overall amino acyl composition, and their amino acid sequence.

5.2 Proteins

Chemically, proteins are polypeptide chains of amino acids. The peptide bond (amide) defines them as oligopeptide, poly peptide and protein. Oligo means few and poly means many (oligo = few, poly = many), thus oligopeptides are molecules having 2-20 peptide bonded amino acids. Similarly polypeptides are molecules having roughly 20+ peptide bonded amino acids.

Structurally the sequence —N— Cα —CO— is repeated for each residue, producing the main chain or the backbone of the protein, to which the side chains of the amino acids are attached. The backbone has two ends that are easily distinguishable, and it can be considered directed. The convention is that a protein starts at the amino group (N-terminal) and end at the carboxyl group (C-terminal).

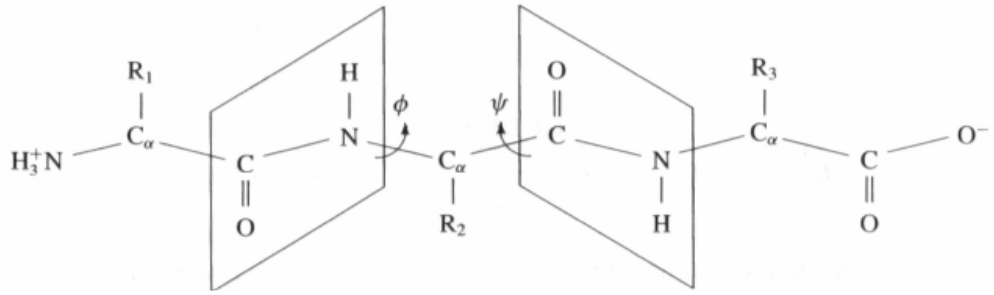


Figure 1: Amino acid chain

Protein is a more generic term that can mean any oligo/polypeptide. The difference between polypeptides and proteins lie in the function. Usually Most polypeptides shorter than about 40 amino acids in length do not fold whereas long polypeptide chains typically longer than 40 to 50 amino acids fold into a well-defined structure in a specific manner to have a specific function. This structure is known as protein. Scientists typically distinguish polypeptides and proteins based on structure. The term polypeptides are often reserved for the shorter, unstructured polypeptides that do not fold into a protein structure. *Proteins are polypeptides that fold into a fixed and well defined 3D structure*.

5.3 Classification of proteins

Approximately 30,000 proteins are found in humans, out of which only few have been adequately described. Many of them exhibit large similarities in structure as well as function. Small differences in the primary structure of proteins lead to large differences in tertiary structure and in function.

5.3.1 Traditional Classification of proteins

This classification is based on the source of protein molecule obtained. Accordingly the proteins have been divided into two well-defined groups.

They are i) Animal proteins

ii) Plant proteins.

Animal proteins are the proteins derived from animal sources. Ex: eggs, milk, meat and fish. They are also called as higher-quality proteins because they contain adequate amounts of all the essential amino acids.

Plant proteins are the proteins derived from plant sources. Ex: soybeans; dicot seeds; peas and beans; Groundnuts; Green leafy vegetables. Plant proteins are also called lower-quality proteins since they have a low content of one or more of the essential amino acids.

5.3.2 Structural classification of proteins

This classification is based on the shape of protein molecule**.** On the basis of the shape of protein molecule, the proteins have been grouped in to two categories, they are

i) Globular proteins

ii) Fibrous proteins

1. *Globular or Corpuscular Proteins*

These have an axial ratio (length: width) of less than 10 (usually not over 3 or 4) and, henceforth, possess a relatively spherical or ovoid shape. These are soluble in water or in aqueous media containing acids, bases, salts or alcohol, and diffuse readily. Globular proteins have greater variety of biological functions and are dynamic rather than static in their activities. Tertiary and quaternary structures are usually associated with this class of proteins. Ex; nearly all enzymes are globular proteins, protein hormones, blood transport proteins, antibodies and nutrient storage proteins.

A simple functional classification of globular proteins is not possible because of following two reasons

a) These proteins perform a variety of different functions.

b) Many widely-differing globular proteins perform almost similar functions.

2. Fibrous or Fibrillar Proteins.

These have axial ratios greater than 10 and, henceforth, resemble long ribbons or fibres in shape. These are mainly of animal origin and are insoluble in all common solvents such as water, dilute acids, alkalies and salts and also in organic solvents. Most fibrous proteins serve in a structural or protective role. The fibrous proteins are extremely strong and possess following two important properties which are characteristic of the elastomers.

a) They can stretch and later recoil to their original length.

b) They have a tendency to creep, i.e., if stretched for a long time, their basic length increases and equals the stretched length but, if the tension on the two ends of the fibril is relaxed, they creep to their shorter and shorter length. Ex; the proteins of connective tissues, bones, blood vessels, skin, hair, nails, horns, hoofs, wool and silk.

5.3.3 Classification based on composition and solubility

This is the most accepted system of classification of proteins. This classification is based on the proposals made by the committees of British Physiological Society and the American Physiological Society. In this classification the proteins are divided into 3 major groups, based on their composition, they are,

1) Simple

2) Conjugated

3) Derived.

a) Simple Proteins or Holoproteins

These are of globular type except for scleroproteins which are fibrous in nature. This group includes proteins containing only amino acids, as structural components. On decomposition with acids, these liberate the constituent amino acids.

These are further classified mainly on their solubility basis as follows

1. Protamines and histones. These are basic proteins and occur almost entirely in animals, mainly in sperm cells, possess simplest structure and lowest molecular weight (approximately 5,000). These proteins are soluble in water and not coagulated by heat; strongly basic in character because of high content of basic amino acids (lysine, arginine). They form salts with mineral acids

2. Albumins. These are widely distributed in nature but more abundant in seeds. Soluble in water and dilute solutions of acids, bases and salts. Coagulated by heat.

3. *Globulins.* These are of two types— pseudoglobulins and euglobulins. Euglobulins are more widely distributed in nature than the pseudoglobulins. They either soluble (pseudoglobulins) or insoluble (euglobulins) in water and coagulated by heat.

4. *Glutelins.* These have been isolated only from plant seeds. These are Insoluble in water, dilute salt solutions and alcohol solutions but soluble in dilute acids and alkalies. They are coagulated by heat.

5. Prolamines. These have also been isolated only from plant seeds. These are insoluble in water and dilute salt solutions but soluble in dilute acids and alkalies and also in 60 – 80% alcohol solutions. They are not coagulated by heat

6. *Scleroproteins or Albuminoids.* These occur almost entirely in animals and are, therefore, commonly known as the 'animal skeleton proteins'. These are insoluble in water, dilute solution of acids, bases and salts and also in 60–80% alcohol solutions. They are not attacked by enzymes.

b) Conjugated or Complex Proteins or Heteroproteins

These are also of globular type except for the pigment in chicken feathers which is probably of fibrous nature. These are the proteins linked with a separable nonprotein portion called prosthetic group. The prosthetic group may be either a metal or a compound. On decomposition with acids, these liberate the constituent amino acids as well as the prosthetic group.

They are further classified based on the nature of the prosthetic group present. They are metalloproteins, chromoproteins, glycoproteins, phosphoproteins, lipoproteins and nucleoproteins.

1. Metalloproteins. These are the proteins linked with various metals. These may be of stable nature or may be more or less labile. Based on their reactivity with metal ions, the metalloproteins are classified into 3 groups:

I. Metals strongly bound by proteins. Some heavy metals (Hg, Ag, Cu, Zn) become strongly binded with proteins like collagen, albumin, casein etc., through the —SH radicals of the side chains. Some other proteins have strong binding affinities for Fe (siderophilin) and Cu (ceruloplasmin).

II. Metals bound weakly by proteins. Ca belongs to this category. Here the binding takes place with the help of radicals possessing the electron charge.

III. Metals which do not couple with proteins. Na and K belong to this group. These form compounds with nucleic acids where apparently electrostatic bonds are present.

2. Chromoproteins. These are proteins coupled with a coloured pigment. Such pigments have also been found among the enzymes like catalase, peroxidase and flavoenzymes. Ex: chlorophyll is present in leaf cells in the form of a protein, the chloroplastin. myoglobin, hemoglobin, hemocyanin, hemoerythrin, cytochromes, flavoproteins, catalase, etc.

3. Glycoproteins and Mucoproteins. These are the proteins containing carbohydrate as prosthetic group. Glycoproteins contain small amounts of carbohydrates (less than 4%), whereas mucoproteins contain comparatively higher amounts (more than 4%). Ex; glycoproteins-egg albumin, elastase certain serum globulins and also certain serum albumins.

Mucoproteins-ovomucoid from eggwhite, mucin from saliva and Dioscorea tubers, osseomucoid from bone and tendomucoid from tendon.

4. Phosphoproteins. These are proteins linked with phosphoric acid; mainly acidic. Ex; casein from milk and ovovitellin from egg yolk.

5. Lipoproteins. Proteins forming complexes with lipids (cephalin, lecithin, cholesterol) are called lipoproteins. These are soluble in water but insoluble in organic solvents. Ex; lipovitellin and lipovitellenin from egg yolk; lipoproteins of blood.

6. Nucleoproteins. These are compounds containing nucleic acid and protein, esp., protamines and histones. These are usually the salt-like compounds of proteins since the two components have opposite charges and are bound to each other by electrostatic forces. They are present in nuclear substances as well as in the cytoplasm. These may be considered as the sites for the synthesis of proteins and enzymes. Ex, nucleoproteins from yeast and thymus and also viruses which may be regarded as large molecules of nucleoproteins; nucleohistones from nuclei-rich material like glandular tissues;

c) Derived Proteins

These are derivatives of proteins resulting from the action of heat, enzymes or chemical reagents. This group also includes the artificially-produced polypeptides.

I. Primary derived proteins. These are derivatives of proteins in which the size of protein molecule is not altered materially.

i) *Proteans*. These are Insoluble in water; appear as first product produced by the action of acids, enzymes or water on proteins. Ex, edestan derived from edestin and myosan derived from myosin.

ii) *Metaproteins or Infraproteins.* These are insoluble in water but soluble in dilute acids or alkalies; produced by further action of acid or alkali on proteins at about 30–60°C. Ex, acid and alkali metaproteins.

iii) *Coagulated Proteins.* These are insoluble in water; produced by the action of heat or alcohol on proteins. Ex, coagulated eggwhite.

II. Secondary derived proteins. These are derivatives of proteins in which the hydrolysis has certainly occurred. The molecules are smaller than the original proteins.

i) *Proteoses.* These are Soluble in water; coagulable by heat; produced when hydrolysis proceeds beyond the level of metaproteins. Ex, albumose from albumin; globulose from globulin.

ii) *Peptones.* These are soluble in water; noncoagulable by heat; produced by the action of dilute acids or enzymes when hydrolysis proceeds beyond proteoses.

iii) *Polypeptides.* These are combinations of two or more amino acid units. In fact, the proteins are essentially long chain polypeptides.

Although widely accepted, the system outlined above has certain discrepancies

1. The classification is arbitrary.
2. The criterion of solubility is not well demarcated as some globulins (pseudoglobulins) are also soluble in water.
3. Protamines and histones should have been kept under derived proteins.
4. The group metaproteins is an artificial assemblage.

5.4 Biological functions of proteins

Depending upon their physical and chemical structure and location inside the cell, different proteins perform various functions. The proteins may be grouped under following categories Based on the metabolic functions they perform.

Classification of proteins on the basis of their biological functions

Class of protein	*Function*	*Examples*
Enzymic proteins	Biological catalysts	Urease, Amylase, Catalase, Cytochrome C, Alcohol dehydrogenase.
Structural proteins	Strengthening or protecting biological structures	Collagen, Elastin, Keratin, Fibroin
Transport or carrier proteins	Transport of ions or molecules in the body	Myoglobin, Hemoglobin, Ceruloplasmin, Lipoproteins
Nutrient and storage proteins	Provide nutrition to growing embryos and store ions	Ovalbumin, Casein, Ferritin
Contractile or motile proteins	Function in the contractile system	Actin, Myosin, Tubulin
Defense proteins	Defend against other organisms	Antibodies, Fibrinogen, Thrombin
Regulatory proteins	Regulate cellular or metabolic activities	Insulin, G proteins, Growth hormone
Toxic proteins	Hydrolyze (or degrade) enzymes	Snake venom, Ricin.

i) Catalytic activity

All the chemical reactions of biomolecules are catalyzed by the enzymes. Nearly all enzymes are globular proteins. Chemically, some enzymes are simple proteins, containing only amino acid residues. Others are complex proteins, containing a major protein part (apoenzyme) and a small nonprotein part (prosthetic group) associated with the protein unit. Enzymes catalyze a variety of reactions. Urease, amylase, catalase, cytochrome C, alcohol dehydrogenase are some of the examples of enzymic proteins.

ii) Structural integrity

Some proteins maintain the native form and position of the organs. The cell wall and primary fibrous constituents of the cell have structural proteins. The structural proteins are usually inert to biochemical reactions. Ex,Collagen, which has very high tensile strength, is the most abundant protein of animals. It is found in connective tissue such as tendons, cartilage, matrix of bones and cornea of the eye. Leather is almost pure collagen. Ligaments contain elastin, a structural protein capable of stretching in two dimensions. α keratin constitutes almost the entire dry weight of hair, wool, feathers, nails, claws, quills, scales, horns, hooves, tortoise shell and much of the outer layer of skin. The major component of silk fibres and spider webs is fibroin. The wing hinges of some insects are made of resilin, which has nearly perfect elastic properties.

iii) Transport and carrier of biological factors

Certain proteins, especially in the animals, are involved in the transport of many essential biological factors to various parts of the organisms. Ex, Hemoglobin of erythrocytes binds O_2 as the blood passes through the lungs, carries it to the peripheral tissues, and there releases it to participate in the oxidation of nutrients. The blood plasma contains lipoproteins, which carry lipids from the liver to other organs. Other kinds of transport proteins are present in the plasma membranes and intracellular membranes of all organisms. Ceruloplasmin transports copper in blood.

iv) Nutrient storage

Ovalbumin is the major protein of eggwhite. The milk protein, casein stores amino acids. The seeds of many plants store nutrient proteins, required for the growth of the germinating seedlings. Ferritin, found in some bacteria and in plant and animal tissues, stores iron.

v) Contractile or motility

Some proteins endow cells and organisms with the ability to contract, to change shape, or to move about. Actin and myosin function in the contractile system of skeletal muscle and also in many nonmuscle cells. Tubulin is the protein from which microtubules are built.

vi) Defense

Many proteins defend organism against invasion by other species or protect them from injury. The antibodies (or immunoglobulins), the specialized proteins made by the lymphocytes of vertebrates, can precipitate or neutralize invading bacteria, viruses or foreign proteins from another species. Fibrinogen and thrombin are blood-clotting proteins that prevent loss of blood when the vascular system is injured.

vii) Regulation of cellular or physiological activity

Some proteins help regulate cellular or physiological activity. Among them are many hormones, such as insulin which regulates sugar metabolism, and growth hormone which is required for bone growth. The cellular response to many hormonal signal is often mediated by a class of GTP-binding proteins called G proteins. Other regulatory proteins bind to DNA and regulate the biosynthesis of enzymes and RNA molecules involved in cell division.

viii) Toxicity

Some proteins act as toxic substances, such as snake venom, bacterial toxins and toxic plant proteins like ricin. These toxic proteins also have defensive functions.

5.5 Structure of protein

Structural features of proteins are usually described at four levels of complexity they are

i) Primary structure
ii) Secondary structure
iii) Tertiary structure and
iv) Quaternary structure

The first three involve only one molecule. The quaternary structure describes how proteins interact to form complex molecular structures.

5. 6 Primary structure of protein

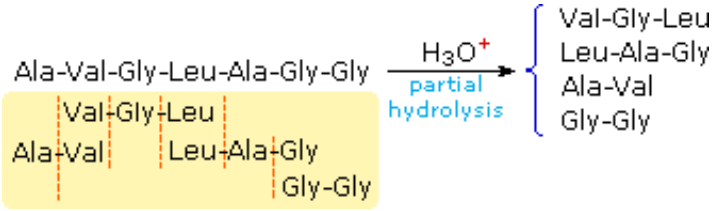
All proteins are built up from 20 amino acids which constitute the monomers. The sequence in which these amino acids are arranged differs from protein to protein. *The different amino acids that make up a peptide or protein, and the order in which they are joined together by peptide*

bonds is referred to as the primary structure. Or the sequence of amino acids that comprise a protein is called its primary structure.

For examples A complete hydrolysis of a protein or peptide, followed by amino acid analysis establishes its gross composition, but does not provide any bonding sequence information. Partial hydrolysis will produce a mixture of shorter peptides and some amino acids. If the primary structures of these fragments are known, it is sometimes possible to deduce part or all of the original structure by taking advantage of overlapping pieces.

For example, many possible primary structures can be written for a heptapeptide composed of three glycines, two alanines, a leucine and a valine.

Let us consider partial hydrolysis of the hexapeptide gave two known tripeptide and two known dipeptide fragments, as shown below.



We could identify the primary structure by simple analysis of the overlapping units. This kind of structure determination is very inefficient and unreliable. Because first we need to know the structures of all the overlapping fragments, second larger peptides would give complex mixtures which would have to be separated and examined to find suitable pieces for overlapping. However, it should be noted, that modern mass spectrometry uses this overlap technique effectively.

5.6.1 N-Terminal Group Analysis

Many techniques have been used to investigate the primary structure or amino acid sequence of small to medium peptides. The commercial instruments that automatically sequence the peptides and proteins are also available. The most important and commonly used technique is the identification of the N-terminal and C-terminal aminoacid units of a peptide chain. This provides helpful information of peptide sequence.

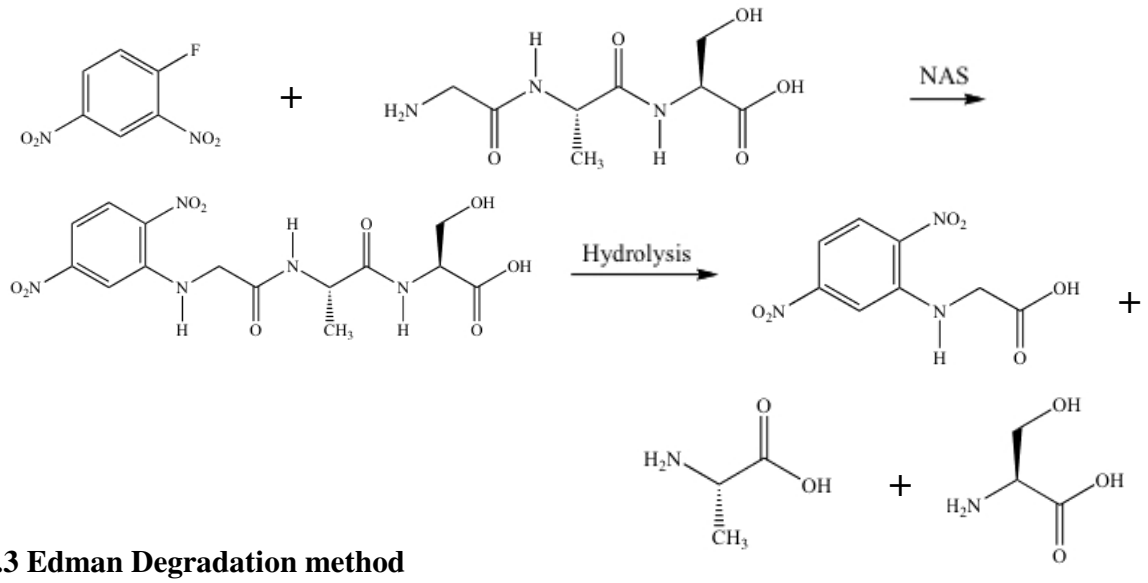
N-terminal analysis is accomplished by

i) Sanger's Method and

ii) Edman Degradation method

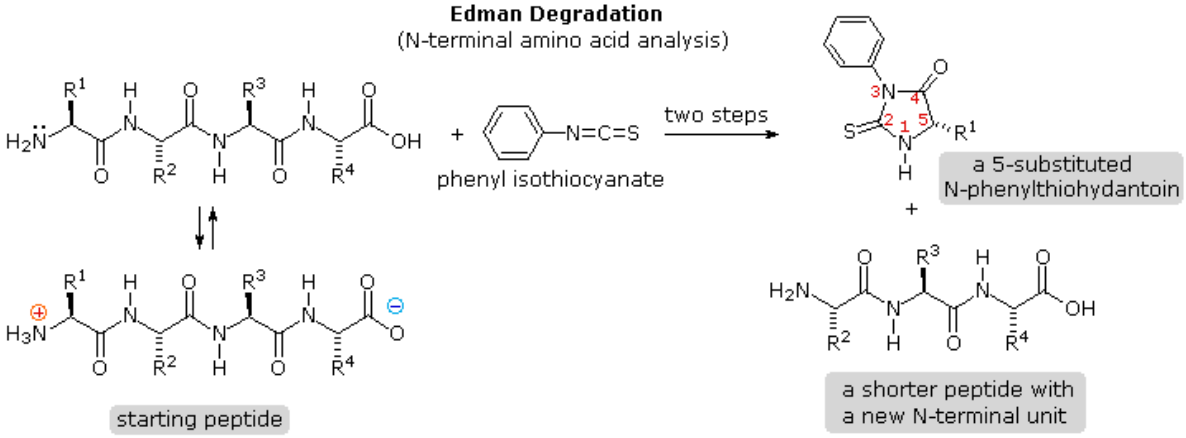
5.6.2 Sanger's method

2,4-dinitrofluoro benzene is called Sanger's reagent. In this method the poly peptide is treated with Sanger's reagent in presence of $NaHCO_3$. The N-terminal amino group of the peptide gets substituted for fluorine atom to give 2,4-dinitrophenyl (DNP) derivative. This is then compleately hydrolyzed to give a DNP aminoacid and a mixture of free aminoacids. By identifying the DNP-amino acid, the N-terminal amino acid of the polypeptide can be determined. Lysine has an ε-amino group in its side chain so that this residue forms ε-DNP if Lys is somewhere along the peptide, because its amino group will be in a peptide bond. If lysine is in the N-terminal amino acid, then it form α,ε-diDNP lysine. Thus the N-terminal lysine residue is distinguished from the non-N-terminal lysine residue.



5.6.3 Edman Degradation method

Edman Degradation method is outlined in the below diagram. It is one of the molecular method. It uses a three-stage reaction cycle for step-wise removal of amino acid residues from the N-terminus of a polypeptide chain. The three stages of the Edman cycle are (i) coupling, (ii) cyclization, and (iii) conversion. A free amine function, usually in equilibrium with zwitterion species, is necessary for the initial bonding to the phenyl isothiocyanate reagent.

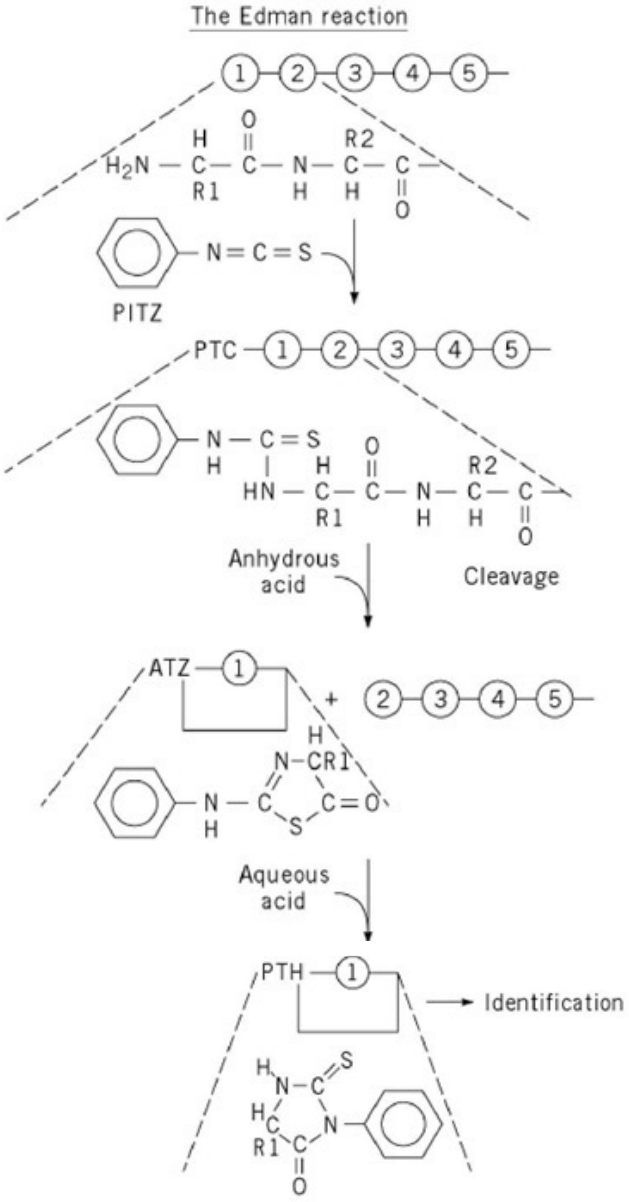


i) Coupling involves a nucleophilic attack of the N-terminal α-amino group on the thiocyanate carbon of phenylisothiocyanate (PITC), to form the phenylthiocarbamyl-peptide derivative (PTC-peptide).

ii) Cyclization constitutes the formation of the anilinothiazolinone derivative (ATZ-amino acid) in anhydrous acid, thereby liberating the N-terminal amino acid residue in a cyclic form, and leaving the remaining peptide truncated by one residue

iii) Conversion involves the rearrangement of the liberated ATZ derivative to the corresponding phenylthiohydantoin derivative (PTH-amino acid) by opening of the CS bond and re-closure with the CO bond.

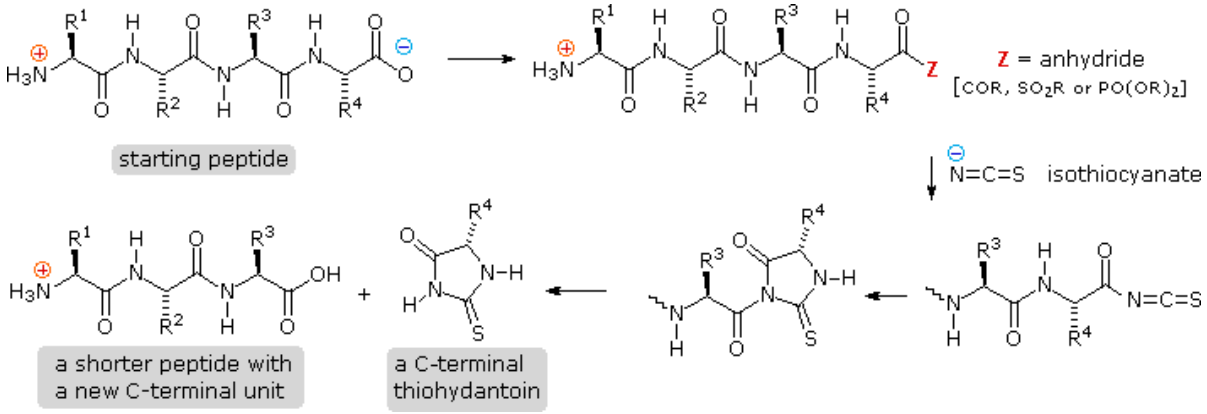
The products of the Edman degradation are a thiohydantoin heterocycle incorporating the N-terminal amino acid together with a shortened peptide chain. Amine functions on a side-chain, as in lysine, may react with the isothiocyanate reagent, but do not give thiohydantoin products. A major advantage of the Edman procedure is that the remaining peptide chain is not further degraded by the reaction. This means that the N-terminal analysis may be repeated several times, thus providing the sequence of the first three to five amino acids in the chain. A disadvantage of the procedure is that, peptides larger than 30 to 40 units do not give reliable results.



5.6.4 C-Terminal Group Analysis

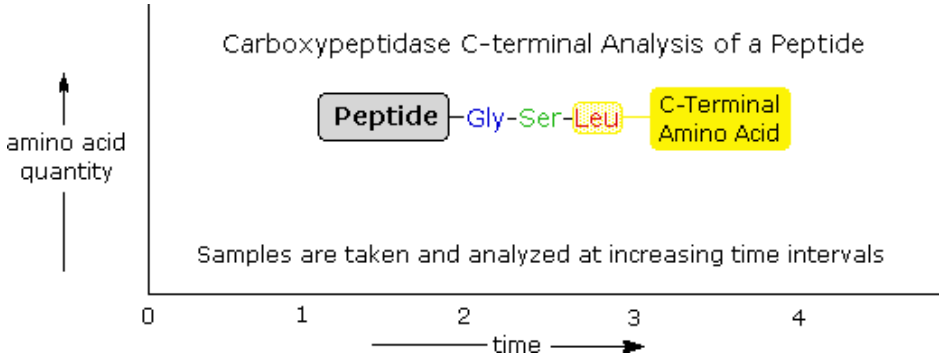
C-terminal analysis of peptide chains can be accomplished chemically or enzymatically. The chemical analysis is slightly more complex. First, side-chain carboxyl groups and hydroxyl groups must be protected as amides or esters. Next, the C-terminal carboxyl group is activated as an anhydride and reacted with thiocyanate. The resulting acyl thiocyanate immediately cyclizes to a hydantoin ring, and this can be cleaved from the peptide chain in several ways, Depending on the nature of this final cleavage, the procedure can be modified to give a C-terminal acyl thiocyanate peptide product which automatically rearranges to a thiohydantoin incorporating the

penultimate C-terminal unit. Thus, repetitive analyses may be conducted in much the same way they are with the Edman procedure.



Enzymatic Analysis

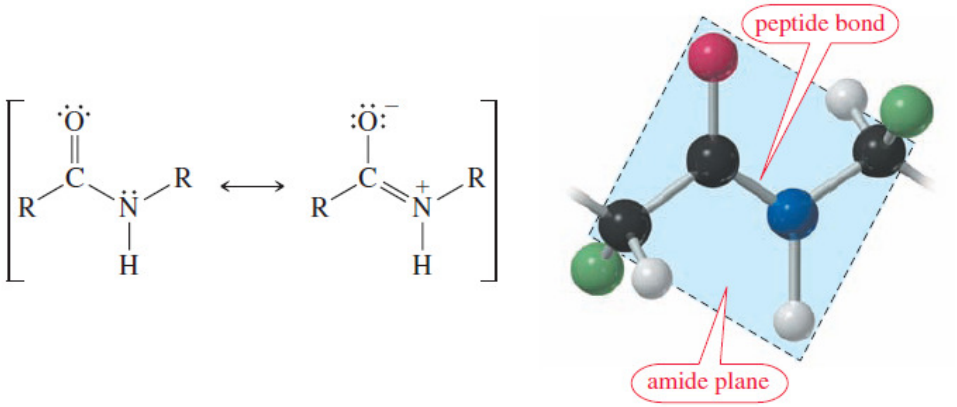
Enzymatic C-terminal amino acid cleavage by one of several carboxypeptidase enzymes is a convenient method of analysis. Because the shortened peptide product is also subject to enzymatic cleavage, care must be taken to control the conditions of reaction so that the products of successive cleavages are properly monitored. The following example illustrates this feature. A peptide having a C-terminal sequence Gly-Ser-Leu is subjected to carboxypeptidase cleavage, and the free amino acids cleaved in this reaction are analyzed at increasing time intervals. The leucine is cleaved first, the serine second, and the glycine third, as demonstrated by the sequential analysis.



5.7 Secondary structure

We know that peptide bonds are very stable. This stability is partly due to the strong resonance interaction between the nonbonding electrons on nitrogen and the carbonyl group. The amide nitrogen is no longer a strong base, and the bond has restricted rotation because of its partial double-bond character. Below figureshows the resonance forms we use to explain the partial

double-bond character and restricted rotation of an amide bond. In a peptide, this partial doublebond character results in six atoms being held rather rigidly in a plane.

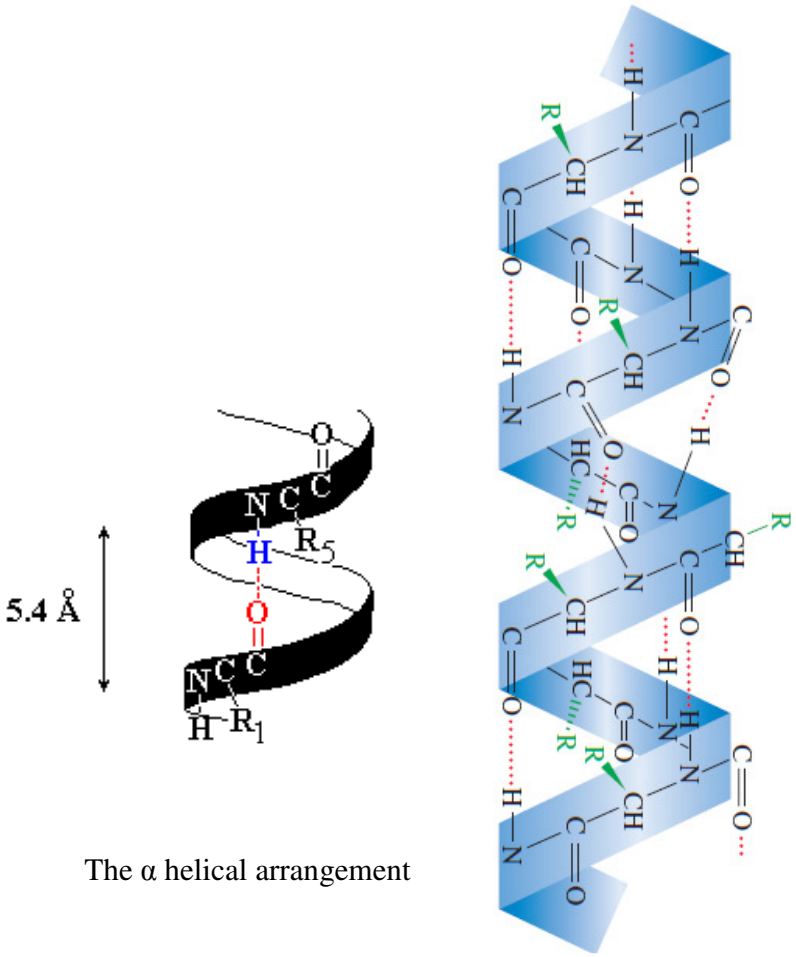


In a peptide, the amide bond is called a peptide bond. It holds six atoms in a plane: the C and O of the carbonyl, the N and its H, and the two associated α carbon atoms.

The six atoms that from the peptide groups are lie on a plane thereby forming the planar structure. Due to free rotational restriction of peptide bond, polypeptide chain consists of several repeating patterns and these patterns are called secondary structure of protein. The most commonly observed types of secondary structure are the α-helix and the β-pleated sheet. Both α-helix and β-sheet patterns are stabilized by localized hydrogen bonding between the carbonyl and N—H groups in the polypeptide's backbone.

5.7.1 α-helix structure

The α-helix is a rigid, rodlike structure that forms when a polypeptide chain twists into a right-handed helical conformation. Hydrogen bonds are formed between the N—H group of each amino acid with the carbonyl group of the amino acid four residues away. There are 3.6 amino acid residues per turn of the helix, and the pitch (the distance between corresponding points per turn) is 0.54 nm or 5.4 Å and width of the helix is about 4Å. Amino acid R groups extend outward from the helix. Since α helix has 3.6 amino acids per turn of the helix, which places the C=O group of amino acid 1 (or i) exactly in line with the H-N group of amino acid 5 (i + 4) similarly (C=O of aminoacid 2 with H-N of amino acid 6 and so on.....). This alignment of and spacing of atoms in the peptide chain is ideal for a hydrogen bonds to form between C=O:---H-N.



The α helical arrangement

The α helix formed by peptide chain is found to be exclusively right handed because for L-amino acid only right handed α-helices are allowed. A left handed arrangement has similar dimensions, but amino acid side chains places next to the C=O group. This makes the structure over-crowded. Whereas in the right handed helix, amino acid side chains places next to the much smaller N-H, which fir better with minimum steric repulsion.

In the α-helix, the C=O---H-N bonds are almost parallel with the helix axis and point in the same direction. Therefore helix as a whole has strong dipole moment such that the amino end or N-terminal has a net partial positive charge and the carboxy end or C-terminal has net negative charge. Thus the H-bond dipoles reinforce in the helix, positive end towards the start or N-terminal end of the dipole. H-bonds are skewed relative to the helix axis in the other helices, so the reinforcement of dipoles is less effective.

Among 20 standard amino acids, α-helices are preferentially made by certain amino acid residues these residues are called helix propensity amino acids or helix-formers, they are alanine, glutamine, lecucine, lysine and methionine.

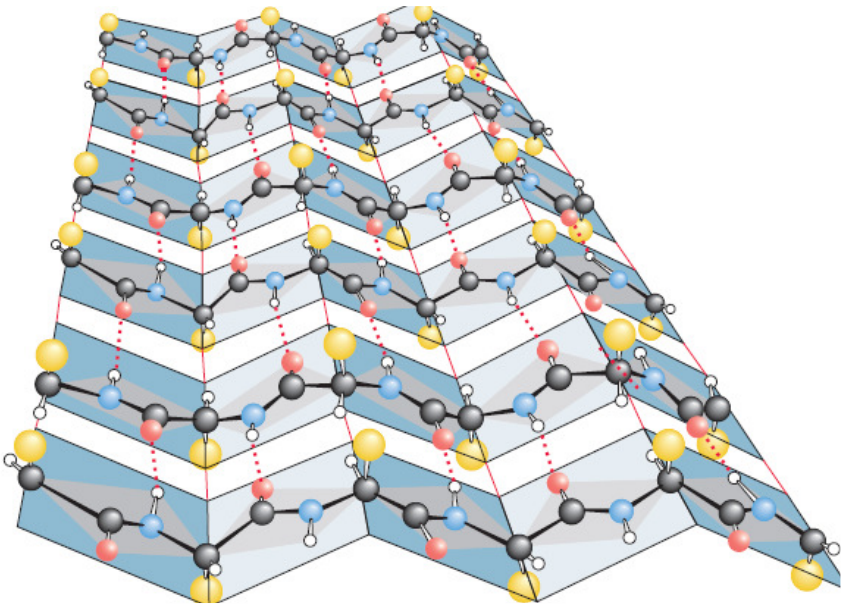
Proline amino acid residue does not fit in the middle of the helical conformation hence it is called as helix- broken residue. Eventhough proline will easily fit within the 3 residues of α-helical structure at N-terminal end. Glycine, tyrosine and serine are the other amino acid residues occur rarely in helices. An α-helix occurs on the surface of a protein, it is made up of hydrophobic residues on the side facing the interior of the protein and hydrophilic residues on the side.

5.7.2 β- Strand structure

β-strands are highly extended polypeptide chains. If extended strands are lined up side by side, H-bonds bridge from strand to strand. Thus β-strands interact with other β-strands to form β-sheets.

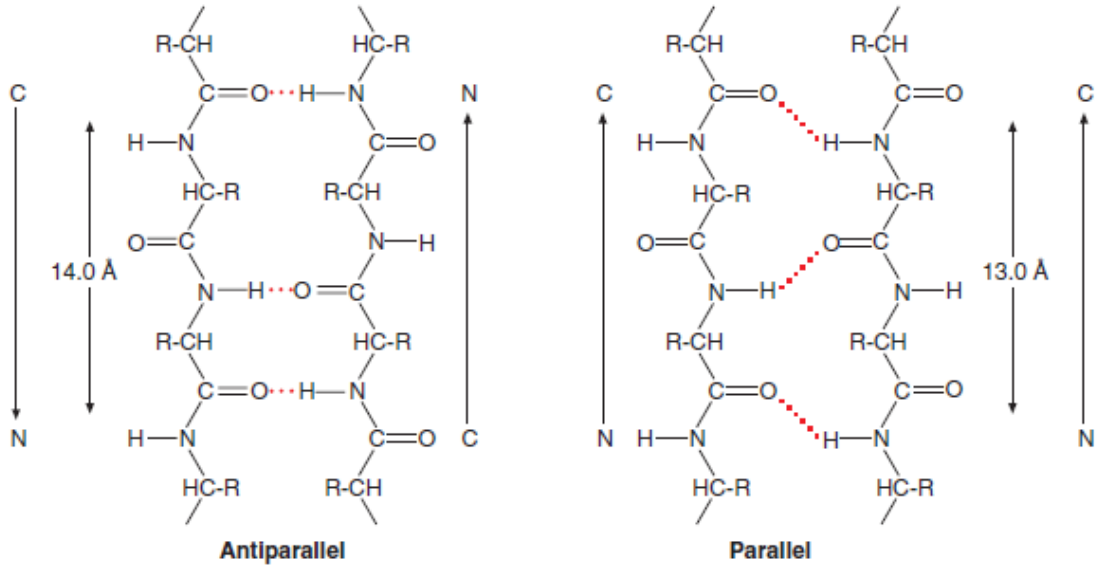
Each individual segment is referred to as a β-strand. Rather than being coiled, each β-strand is fully extended. β-sheets are stabilized by hydrogen bonds that form between the polypeptide backbone N—H and carbonyl groups of adjacent chains. This interaction can occurs in two ways.

i) All the interacting β-strands could run in the same direction, ie, the amino termini of all strands are on the same side. In this case a parallel β-sheet is formed in which somewhat distorted H-bonds are formed between the amino groups of one strand and the carboxy groups of the other.

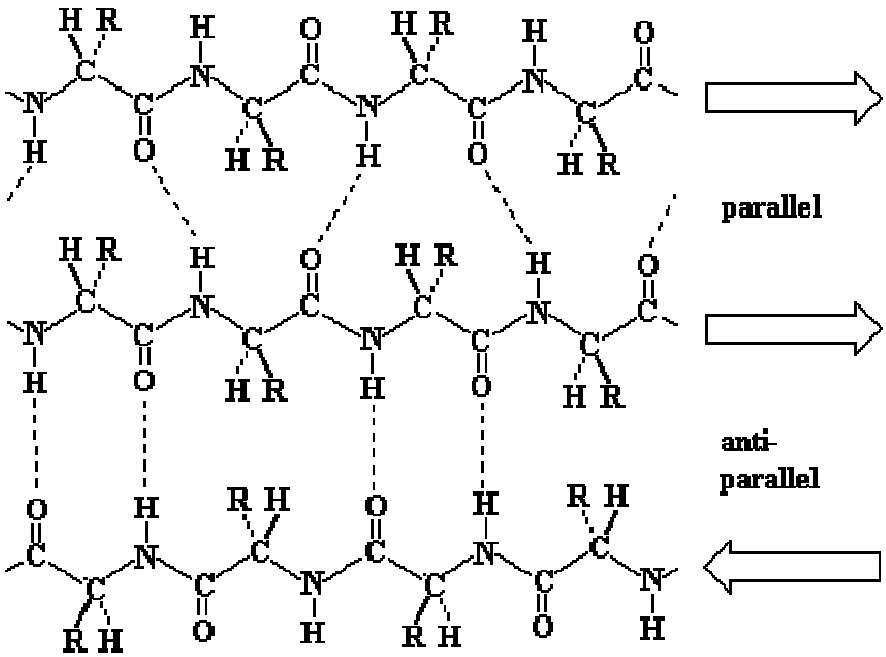


ii) In this arrangement the strands are run in opposite directions, resulting in an antiparallel sheet. Here again N-H-----O hydrogen bonds stabilizes the structure.

In both cases the amino acid side chain project out of the sheet on either side. The sheets are usually not flat but have a twist and always in a right handed sense. In parallel β-sheet structures, the hydrogen bonds in the polypeptide chains are arranged in the same direction. In antiparallel β-sheet structures these bonds are arranged in opposite directions. Occasionally, mixed parallel antiparallel β-sheets are also observed.



Antiparallel beta-sheet is significantly more stable due to the well aligned H-bonds.



5.8 Super-Secondary Structure

Some proteins contain an ordered organization of secondary structures that form distinct functional domains or structural motifs. Examples the helix-turn-helix domain of bacterial proteins that regulate transcription and the leucine zipper, helix-loop-helix and zinc finger domains of eukaryotic transcriptional regulators. These domains are termed super-secondary structures.

5.9 Amino acid preferences for different secondary structure

α-helix may be considered as the default state for secondary structure. Although the potential energy is not as low as for beta sheet, H-bond formation is intra-strand, so there is an entropic advantage over beta sheet, where H-bonds forms from strand to strand, with strand segments that may be quite distant in the polypeptide sequence.

The main criterion for α-helix preference is that the amino acid side chain should cover and protect the backbone H-bonds in the core of the helix. Most amino acids do this with some key exceptions,

α-helix preference: Ala, Leu, Met, Phe, Glu, Gln, His, Lys, Arg

The extended structure leaves the maximum space free for the amino acid side chains, as a result, amino acids with large bulky side chains prefer to form β -sheet structures.

Just plain large: Tyr, Trp, (Phe, Met)

Bulky and awkward due to branched beta carbon: Ile, Val, Thr

Large S atom on beta carbon: Cys

The remaining amino acids have side chains which disrupt secondary structure, and are known as secondary structure breakers

Gly: side chain H is too small to protect backbone H-bond

Pro: side chain linked to alpha N, has no N-H to form H-bond.

Asp, Asn, Ser: H-bonding side chains compete directly with backbone H-bonds.

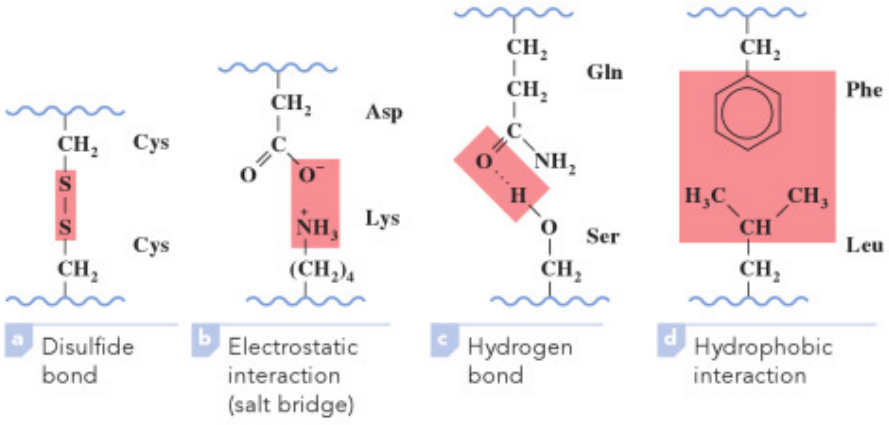
Clusters of breakers give rise to regions known as loops or turns which mark the boundaries of regular secondary structure, and serve to link up secondary structure segments.

5.10 Tertiary Structure

Tertiary structure of protein is more complicated than secondary structure. Tertiary structure refers to the complete three-dimensional structure of the polypeptide units of a given protein. As we know that secondary structures are made of α-helices and β- sheets, both these are local

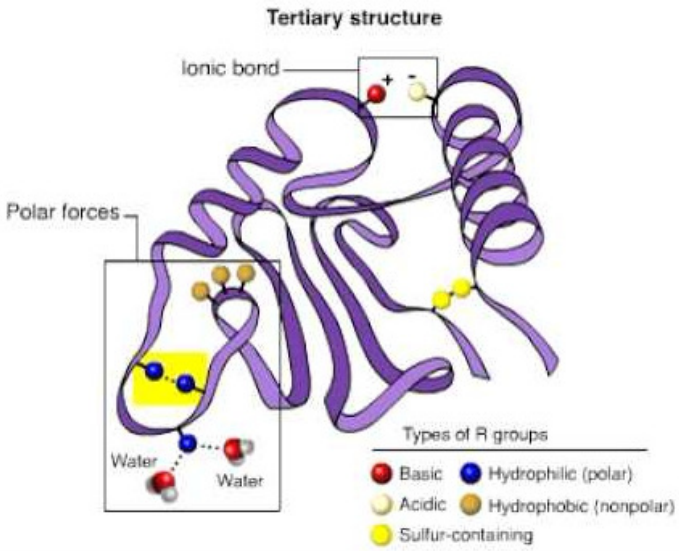
structures. *Tertiary structures is due to the spatial relationship of different secondary structures to one another within a polypeptide chain and how these secondary structures themselves fold into the three-dimensional form of the protein. Tertiary structures involve packaging of the secondary structures into compact globular regions called protein domains*. A protein can have one or more domains. The important factor for tertiary structure is that it contains only one polypeptide. *Tertiary protein domains are formed by combinations of disulfide bonds, hydrogen bonds, ionic bonds, and non-polar hydrophobic interactions. The type of side chain on the amino acid determines the type of interaction*.

.



The four types of stabilizing interactions between amino acid R groups that contribute to the tertiary structure of proteins.

In tertiary structure interaction taking place between the amino acid residues that are not necessarily close together in the primary structure.

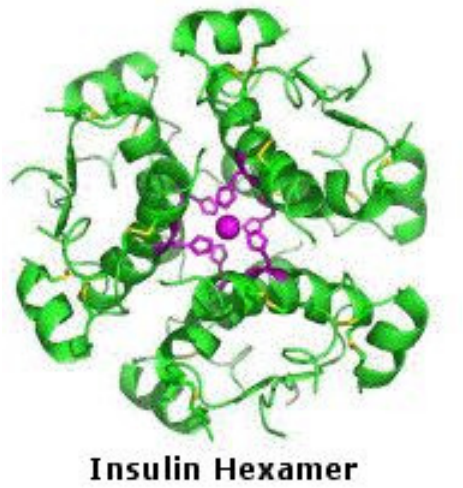


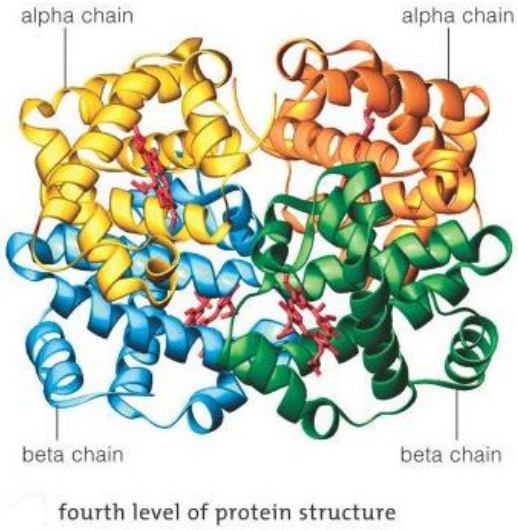
Thus tertiary structure is a result of the bonds, not between the peptide chain's atoms, but between the side chains of amino acids. They involve alpha helices and beta sheets, but also regions unique to each protein.

For example, hydrogen bonds often form between sidechains of adjacent regions quite far apart in the actual amino acid sequence. A disulphide bridge formed between the sulphur atoms of neighboring cysteine side chains. It is a strong covalent bond, which provides stability to the protein structure. Ionic bonds are formed between polar or charged side chains and reduce the flexibility. Non-polar hydrophobic side chains tend to cluster closely in the interior of the protein, the weak van der Waals forces holds these clusters.

5.11 Quaternary structure

The quaternary structure is the arrangement of more than one protein molecule in a multi-subunit complex. A single polypeptide chain is considered as a protein if it can function on its own. However, many proteins are actually comprised of several polypeptide chains. In this case, the individual peptide chains are called protein subunits, and they cannot function on their own.





fourth level of protein structure

Example for a protein with quaternary structure is hemoglobin. In hemoglobin, one protein binds to oxygen while another binds carbon dioxide. This is how one protein can serve two functions.

Assembly of multiple polypeptide chains in one integral unit gives rise to a stable structure. The Subunits may be identical or different. The subunits are held together by both hydrophobic interactions and ionic interactions between polar/charged amino side chains. Understanding protein-protein interaction is key to understanding and controlling the formation of protein complexes. Common shorthand used for describing such proteins is to use Greek letters for each

type of subunit, and subscript numeral to specify numbers of units. Ex, A protein designated $\alpha_2\beta$ consists of two α units and one β. Quaternary structure adds stability by decreasing the surface/volume ratio of smaller subunit. Large structure provides rigidity necessary to orient the substrate and key amino acids to enable function.

5.12 Denaturation of Proteins

Denaturation of proteins involves the disruption and possible destruction of both the secondary and tertiary structures. Since denaturation reactions are not strong enough to break the peptide bonds, the primary structure (sequence of amino acids) remains the same after a denaturation process. Denaturation disrupts the normal α-helix and β-sheets in a protein and uncoils it into a random shape.

Denaturation occurs because the bonding interactions responsible for the secondary structure (hydrogen bonds to amides) and tertiary structure are disrupted. In tertiary structure there are four types of bonding interactions between "side chains" are normally exists they are hydrogen bonding, salt bridges, disulfide bonds, and non-polar hydrophobic interactions which may be disrupted. A variety of reagents and conditions can cause denaturation. The most common observation in the denaturation process is the precipitation or coagulation of the protein.

Heat can be used to disrupt hydrogen bonds and non-polar hydrophobic interactions. This occurs because heat increases the kinetic energy and causes the molecules to vibrate so rapidly and violently that the bonds are disrupted. Ex, the proteins in eggs denature and coagulate during cooking.

Hydrogen bonding occurs between amide groups in the secondary protein structure. Hydrogen bonding between "side chains" occurs in tertiary protein structure in a variety of amino acid combinations. All of these are disrupted by the addition of alcohol. Ex, a 70% alcohol solution is used as a disinfectant on the skin. This concentration of alcohol is able to penetrate the bacterial cell wall and denature the proteins and enzymes inside of the cell. A 95% alcohol solution merely coagulates the protein on the outside of the cell wall and prevents any alcohol from entering the cell. Alcohol denatures proteins by disrupting the side chain intramolecular hydrogen bonding. New hydrogen bonds are formed between the added alcohol molecule and the protein side chains.

Salt bridges result from the neutralization of an acid and amine on side chains. The final interaction is ionic between the positive ammonium group and the negative acid group. Any

combination of the various acidic or amine side chains will have this effect. Acids and bases disrupt salt bridges held together by ionic charges. A type of double replacement reaction occurs where the positive and negative ions in the salt change partners with the positive and negative ions in the new acid or base added. Ex, this reaction occurs in the digestive system, when the acidic gastric juices cause the curdling (coagulating) of milk.

Heavy metal salts denature the proteins in the same manner as acids and bases. Heavy metal salts usually contain Hg^{+2}, Pb^{+2}, Ag^{+1} Tl^{+1}, Cd^{+2} and other metals with high atomic weights. Since salts are ionic they disrupt salt bridges in proteins. The reaction of a heavy metal salt with a protein usually leads to an insoluble metal protein salt. Heavy metals may also disrupt disulfide bonds because of their high affinity and attraction for sulfur and will also lead to the denaturation of proteins.

Disulfide bonds are formed by oxidation of the sulfhydryl groups on cysteine. Different protein chains or loops within a single chain are held together by the strong covalent disulfide bonds. Oxidizing agents cause the formation of a disulfide bond. The reducing agents act on disulfide bonds and split it apart. Reducing agents add hydrogen atoms to make the thiol group, -SH.

5.13 Renaturation of protein

The original structure of some proteins can be regenerated upon removal of the denaturing agent and restoration of conditions favouring the native state. This process, called renaturation,

The original structure of a protein is a three-dimensional structure. The process of returning a denatured protein structure to its original structure and normal level of biological activity is known as reconstitution of protein. This reconstitution of a protein structure is also known as renaturation of protein. The renaturation of proteins is technically the opposite of denaturation of proteins.

In a renatured protein, the primary structure of the biopolymer remains the same, but the protein which had been denatured gets restored back to its former native structure and is able to function as effectively as before. Renaturation is depends on the protein, and the renaturation process. Lot of factors contributes to an individual protein's to refold. Some of them are size, sequence, secondary structure, amount and type of inter-amino acid links like disulfide bonds, number of subunits, how it was denatured before etc….. Smaller proteins will refold more easily than larger ones. Hydrophilic proteins tend to refold better than more hydrophobic ones, especially

membrane-bound proteins. Multi-subunit proteins need some help to properly reassemble. Adding chaperones/heat shock proteins will help the process and will give better results.

5.14 Summary of the unit

On the basis of their chemical composition, proteins are divided into two classes: simple and complex.

Simple proteins are also known as homoproteins, they are made up of only amino acids. Examples are plasma albumin, collagen, and keratin.

Conjugated proteins: Sometimes also called heteroproteins, they contain in their structure a non-protein portion. Three examples are glycoproteins, chromoproteins, and phosphoproteins.

On the basis of their shape, proteins may be divided into two classes: fibrous and globular.

Fibrous proteins: They have primarily mechanical and structural functions, providing support to the cells as well as the whole organism. These proteins are insoluble in water as they contain many hydrophobic amino acids. The presence on their surface of hydrophobic amino acids facilitates their packaging into very complex supramolecular structures.

Globular proteins: Most of the proteins belong to this class. They have a compact and more or less spherical structure, more complex than fibrous proteins. In this regard, motifs, domains, tertiary and quaternary structures are found, in addition to the secondary structures. They are generally soluble in water but can also be found inserted into biological membranes (transmembrane proteins), thus in a hydrophobic environment. Unlike fibrous proteins, that have structural and mechanical functions, they act as enzymes; hormones; membrane transporters and receptors.

Proteins are macromolecules and have four different levels of structure – primary, secondary, tertiary and quaternary. Sequences with fewer than 50 amino acids are generally referred to as peptides, while the terms protein or polypeptide are used for longer sequences. A protein can be made up of one or more polypeptide molecules. The end of the peptide or protein sequence with a free carboxyl group is called the carboxy-terminus or C-terminus. The terms amino-terminus or N-terminus describe the end of the sequence with a free α-amino group.

The amino acids differ in structure by the substituent on their side chains. These side chains confer different chemical, physical and structural properties to the final peptide or protein. the amino acid sequence makes up the primary structure of the protein.

Stretches or strands of proteins or peptides have distinct characteristic local structural conformations. These conformations are called secondary structure and which dependee on hydrogen bonding pattern. The two main types of secondary structure are the α-helix and the ß-sheet.

The α-helix is a right-handed coiled strand. The side-chain substituents of the amino acid groups in an α-helix extend to the outside. Hydrogen bonds form between the oxygen of the C=O of each peptide bond in the strand and the hydrogen of the N-H group of the peptide bond four amino acids below it in the helix. The hydrogen bonds make this structure especially stable. The side-chain substituents of the amino acids fit in beside the N-H groups.

The hydrogen bonding in a β-sheet is between strands (inter-strand) rather than within strands (intra-strand). The sheet conformation consists of pairs of strands lying side-by-side. The carbonyl oxygens in one strand from hydrogen bond with the amino hydrogens of the adjacent strand. The two strands can be either parallel or anti-parallel depending on whether the strand directions (N-terminus to C-terminus) are the same or opposite. The anti-parallel ß-sheet is more stable due to the more well-aligned hydrogen bonds.

The overall three-dimensional shape of an entire protein molecule is the tertiary structure. The protein molecule will bend and twist in such a way as to achieve maximum stability or lowest energy state. Although the three-dimensional shape of a protein may seem irregular and random, it is fashioned by many stabilizing forces due to bonding interactions between the side-chain groups of the amino acids.

Many proteins are made up of multiple polypeptide chains, often referred to as protein subunits. These subunits may be the same (as in a homodimer) or different (as in a heterodimer). The quaternary structure refers to how these protein subunits interact with each other and arrange themselves to form a larger aggregate protein complex. The final shape of the protein complex is once again stabilized by various interactions, including hydrogen-bonding, disulfide-bridges and salt bridges

5.15 Key words

Proteins; Classification of proteins; *Simple Proteins or Holoproteins*; *Conjugated or Complex Proteins or Heteroproteins*; *Derived Proteins*; Biological functions of proteins; Primary structure of protein; N-Terminal Group Analysis; Sanger's method; Edman Degradation method; C-Terminal Group Analysis; Enzymatic Analysis; Secondary structure; α-helix structure; β- Strand

structure ; Super-Secondary Structure; Tertiary Structure; Quaternary structure; Denaturation of Proteins; Renaturation of protein

5.16 References for further studies

1) Proteins: Structure and Function; David Whitford; *Wiley*, **2005**.

2) Fundamentals of Protein Structure and Function; Engelbert Buxbaum; *Springer*, **2015**.

3) Molecular Biology of the Cell, 6th Ed. Bruce Alberts, Alexander Johnson, Julian Lewis, David Morgan, Martin Raff, Keith Roberts, Peter Walter. *Garland Science*, **2014**.

4) Introduction to Protein Structure; Carl Ivar Branden, JohnTooze; *Garland Science*, **1999**.

5) Biophysics; 2nd Ed. Vasanthapattabhi, N. Gautham; *Narosa publishing house*; **2010**.

5.17 Questions for self understanding

1) What are Proteins? Write the common factors and differences between peptides and proteins.

2) Discuss the classification of proteins on traditional and structural basis with examples.

3) Discuss the classification of proteins based on composition and solubility.

4) Discuss the different biological functions of proteins.

5) Discuss the primary structure of protein.

6) Explain the N-Terminal group Analysis of protein chain using Sanger's method.

7) Explain the N-Terminal group Analysis of protein chain using Edman degradation method.

8) Discuss the advantage of Edman degradation method.

9) Explain the C-Terminal croup analysis of protein sequence by enzymatic analysis.

10) What are the different secondary structures commonly found in protein?

11) What is α-helix structure? Explain its significance.

12) What is β- Strand structure? Explain its significance

13) Write a note on Super-Secondary Structure of protein

14) Discuss the amino acid preferences for different secondary structure

15) Write a note on tertiary structure of protein

16) Write a note on Quaternary structure protein

17) What is denaturation of Proteins?

18) What is renaturation of protein?

UNIT-6

Structure

6.0 Objectives of the unit

After studying this unit you are able to

- Identify the location of ψ (psi) angle in peptide bond
- Identify the location ϕ (phi) angle in peptide bond
- Explain the procedure for drawing Ramachandran plot
- Identify the regions for different protein secondary structures in Ramachandran plot
- Recognize the regions of right and left handed α-helix in Ramachandran plot
- Recognize the regions of β-strand Ramachandran plot
- Explain the Loops region in protein and their significance
- Identify the different β-turns in protein structure

6.1 Introduction

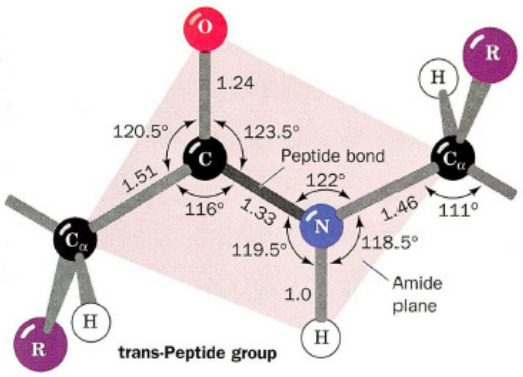
The peptide bond is formed between amine group of one amino acid and carboxy group of another amino acid. The C-N bond has a partial double bond character, which means that the molecule cannot rotate about it. The six atoms that form the peptidegroup are constrained to lie on a plane thereby forming a planar peptide group. Ramachandran and his colleague realized that this property of peptide bond simplified the geometrical analysis of the polypeptide chain. If the rotational possibilities of longer side chains are ignored and back bone of the protein chain alone is considered, each residue is described by two torsion angles ϕ and ψ. The freedom of rotation about these bonds is restricted by possible steric hindrances.

The secondary structures that polypeptides can adopt in proteins are governed by hydrogen bonding interactions between the electronegative carbonyl oxygen atoms and the electropositive amide hydrogen atoms in the backbone chain of the molecule. These hydrogen-bonding interactions can form the framework that stabilizes the secondary structure. Many secondary structures with reasonable hydrogen bonding networks could be proposed but we see only a few possibilities in polypeptides composed of L-amino acids (proteins). Most of the possible secondary structures are not possible due to limits on the configuration of the backbone of each amino acid residue. Understanding of these limitations will help to understand the secondary structures of proteins.

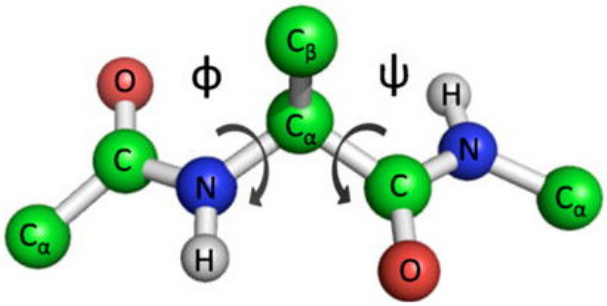
6.2 Key facts about a polypeptide chain

In peptide bond the oxygen and nitrogen atoms of the amide are linked to same sp^2 carbon atom and both have lone pair electrons. The lone pair electron on the nitrogen undergoes delocalizion

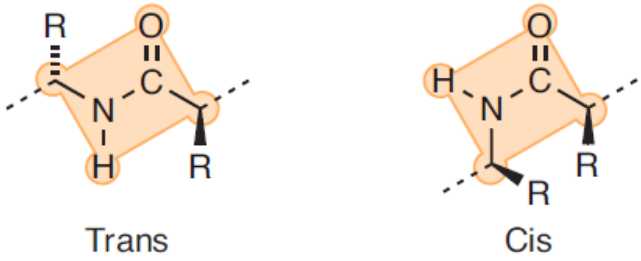
due to high electronegativity of the oxygen atom which stabilizes the negative charge on it. Hence the bond between Nitrogen and carbonyl group has partial double bond character .i.e, the peptide bond has partial double-bond character. Therefore it is shorter and rigid and has restricted rotation. Other bonds in the peptides are single bonds but they also experiences restriction of rotation due to steric hi



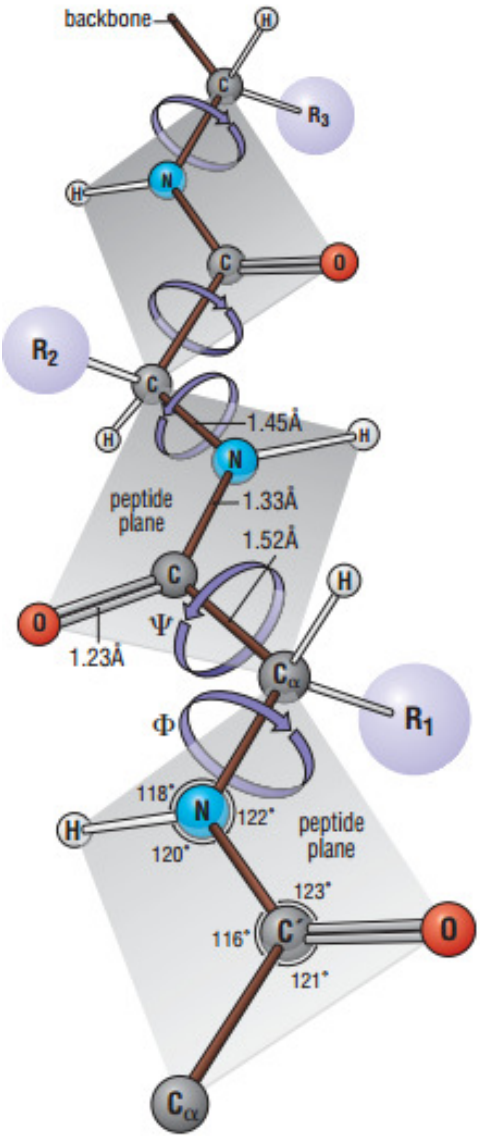
The folding pattern of a polypeptide chain can be described in terms of the angles of rotation around the amide bond (generally referred as main chain bond). The angle between amide nitrogen, α-carbon atom and side chain residue is designated as Phi (ϕ) and the angle between side chain residue, α-carbon atom and carbonyl carbon of the amide group is referred as psi (ψ). These ϕ and ψ angles describe the main chain conformation of the peptides and are called Ramachandran angles



The angle between nitrogen atom and carbonyl group is referred as omega (ω). Due to partial double bond character the peptide bond has either *trans* conformation, corresponds to the omega (ω) = 180^0 or *cis* conformation, corresponds to the omega (ω) = 0^0 conformation.



Except for Pro, *trans* is the more stable conformation for peptides formed by all amino acids.



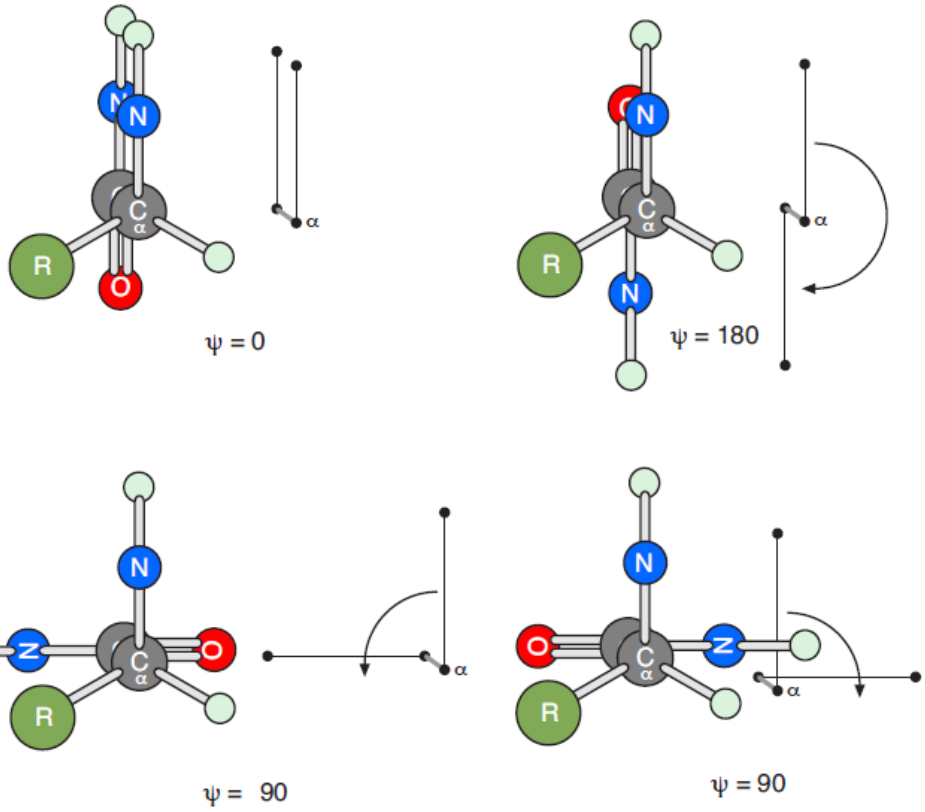
6.3 The ψ (psi) Angle

The bond from the α-carbon to the carbonyl group (at the C-terminus) of the amino acid residue can rotate and turn the whole plane of the amide group, which includes the carbonyl carbon, in a 360-degree range. This angle is measured by looking along that bond with the carbon of the carbonyl group in the rear and the α-carbon to the front. We measure the apparent angle between the two bonds to nitrogen that wecan see coming out of the axis of the Cα→C(C=O) bond. This angle is labeled ψ (psi) and is measured from –180° to +180° with the positive direction being when we turn the rear group clockwise so that the rear nitrogen bond is clockwise of the front

nitrogen bond (or when we turn the front group counterclockwise so that the rear nitrogen bond is clockwise of the front.)

ψ = 0° ψ = 180°

The ψ (psi) angle in an amino acid residue (In these examples ϕ =180°)

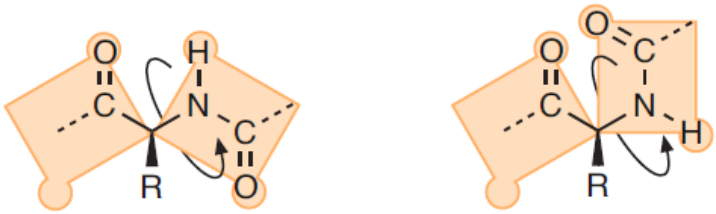


The ψ (psi) angle seen along the Cα→C(C=O) axis. (In these examples ϕ=180°)

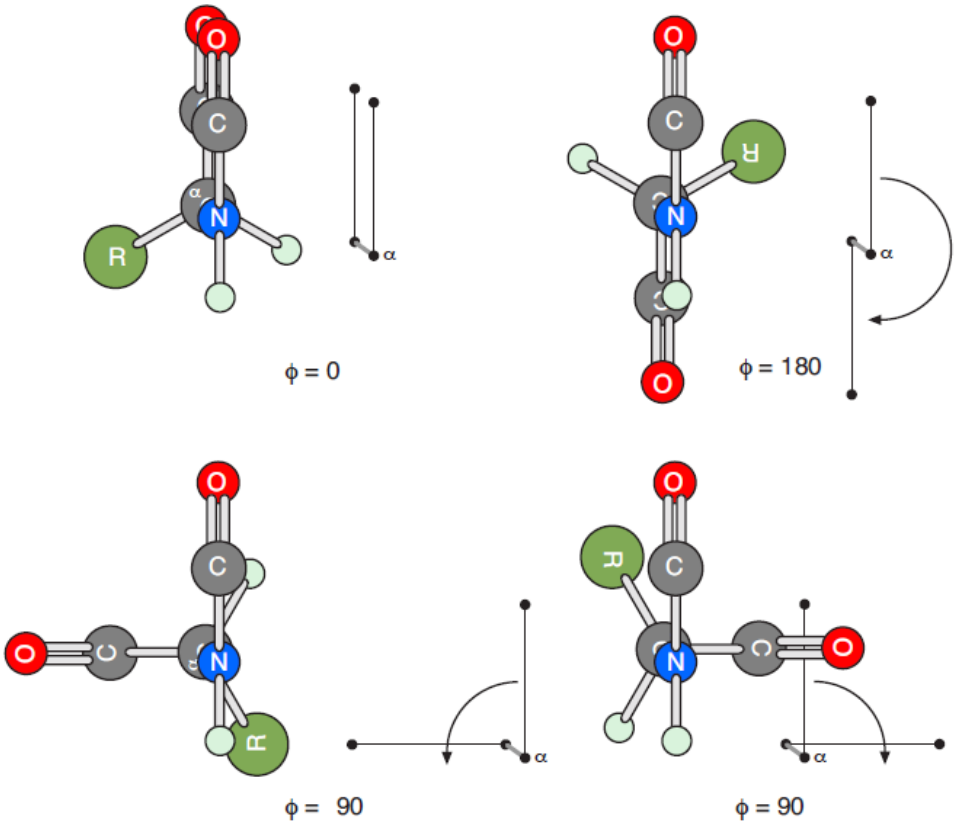
6.4 The ϕ (phi) Angle

The bond from the nitrogen (at the N-terminus) to the α-carbon of the amino acid residue can rotate and turn the whole plane of the other amide group, which includes the nitrogen, in a 360-degree range. This angle is measured by looking along that bond with the nitrogen atom in front and the α-carbon to the rear. We measure the apparent angle between the two bonds to the carbonyl carbons that we can see coming out of the axis of the N→Cα bond. This angle is

labeled ϕ (phi) and is measured from –180° to +180° with the positive direction being when we turn the rear group clockwise so that the rear carbonyl bond is clockwise of the front carbonyl bond (or when we turn the front group counterclockwise so that the rear carbonyl bond is clockwise of the front.)



The ϕ (phi) angle in an amino acid residue. (In these examples ψ=180°)



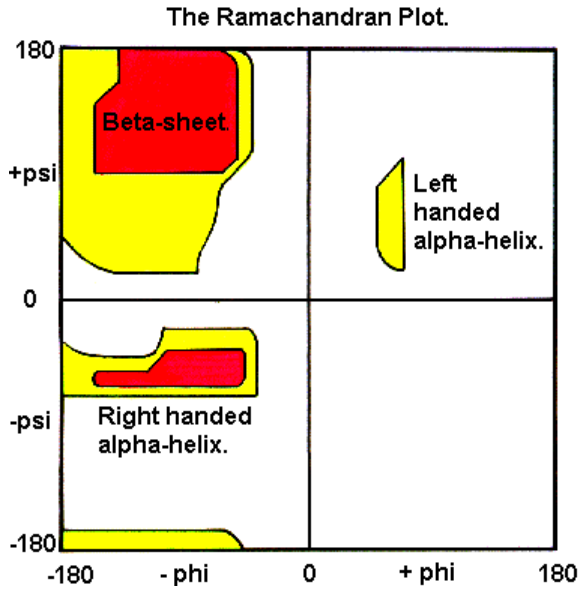
The ϕ (phi) angle seen along the N→Cα axis. (In these examples ψ =180°)

6.5 The Ramachandran Plot

Angle ψ can be varied from –180° to 180° and ϕ from –180° to 180° (that is 360° of rotation for each). But many combinations of these angles are almost never seen and others are very much

common in proteins. Ramachandran showed that serious steric interferences occurred between C=O groups and amino acid side chain centred at $\phi = 120^{o}$, and serious interferences between peptide backbone C=O and N-H occurred at $\phi = \psi = 0^{o}$.

G N Ramachandran used models of small polypeptides to systematically vary ϕ and ψ angles to find the stable conformations. The structure was examined for each conformation by close contacts between atoms. Atoms were treated as hard spheres with dimensions corresponding to their van der Waals radii. Therefore, ϕ and ψ angles which cause spheres to collide correspond to sterically disallowed conformations of the polypeptide backbone. *Ploting of ϕ values on the horizontal axis and the ψ values on the vertical from + 180^{0} to -180^{0} scale of a polypeptide is called Ramachandran plot*. This plot provides the ϕ and ψ values for each amino acid in a protein. *The regions on the plot with the highest density of dots are the so-called allowed regions of the Ramachandran plot, also called low-energy regions*. Some values of ϕ and ψ are forbidden since some atoms will come too close to each other, resulting in a "steric clash" because when two atoms are too close to each other the energy of the system gets too high.



This is a convenient presentation and allows clear distinction of the characteristic regions of α-helices and β-sheets. In the above diagram the white areas correspond to conformations where atoms in the polypeptide come closer than the sum of their van der Waals radi. *These regions are sterically disallowed for all amino acids except glycine which is unique in that it lacks a side chain*. The dark regions correspond to conformations where there are no steric clashes, ie these are the allowed regions namely the α-helical and beta-sheet conformations. The faint areas are

also the allowed regions if slightly shorter van der Waals radi are used, ie the atoms are allowed to come a little closer together. These are called extended regions. This brings out an additional region which corresponds to the left-handed α-helix.

L-amino acids usually did not fall on the extended regions, but occassionally individual residues fall in this region. The residues are usually glycine and asparagine or aspartate where the side chain forms a hydrogen bond with the main chain and therefore stabilises conformations corresponding to these region, otherwise this is an unfavourable conformation.

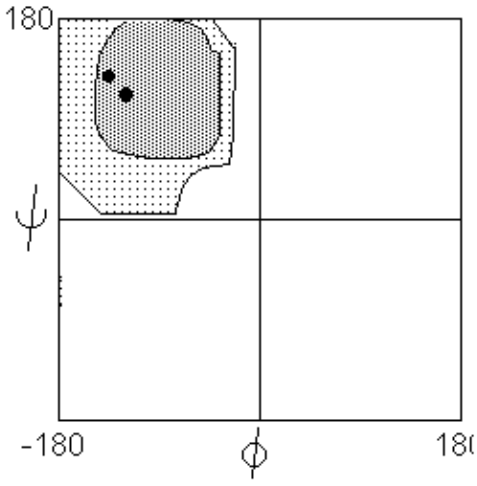
Disallowed regions generally involve steric hindrance between the side chain C_β methylene group and main chain atoms. Glycine has no side chain and therefore can adopt ϕ and ψ angles in all four quadrants of the Ramachandran plot. Hence it frequently occurs in turn regions of proteins where any other residue would be sterically hindered.

6.6 Ramachandran plots and protein Secondary structure

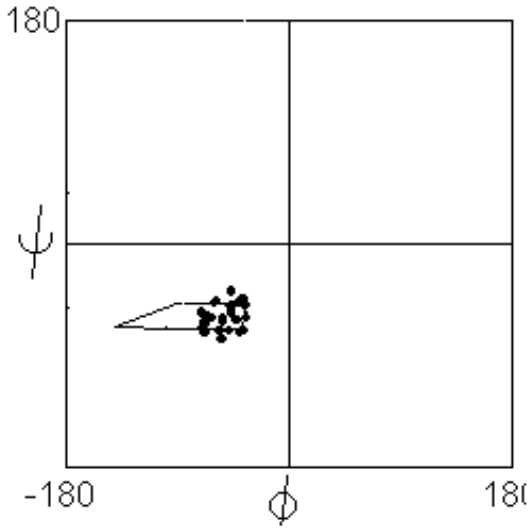
Ramachandran plots show the relationship between the ϕ and ψ angles of a protein referring to dihedral angles between the N and the Cα and the Cα and the C_β. A Ramachandran plot can be used in two different ways. One is to show in theory which values, or conformations, of the ϕ and ψ angles are possible for an amino-acid residue in a protein. A second is to show the empirical distribution of data points observed in a single structure

In a complete 360^0 rotation, there are two regions where interactions are favorable: they are

i) The region of highest stability is a broad plateau centered at $\phi = -120^0$ and $\psi = +135^o$. This region includes the parallel and antiparallel β sheet. The breadth of this region indicates that a degree of variability is allowed to the β structure.



ii) A slightly lower but still energetically favored region is represented by a strip running from ϕ = -60° to -120° and ψ = -60°. The α-helical conformation lies in this region, while 3_{10} and π helices flank the favored region.

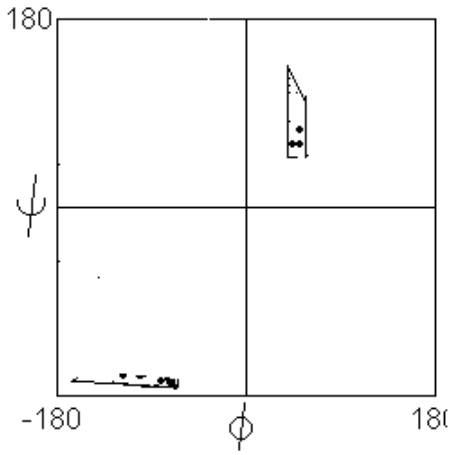


In addition to the two favored regions, there are two regions calculated as allowed but not favored. These are

a) The mirror image of the α-helix strip (left handed α- helix)

b) The mirror image of the strongly favored β-sheet turns out to be strongly disfavored, except by polyglycine (which has a side chain consisting only of a single H atom).

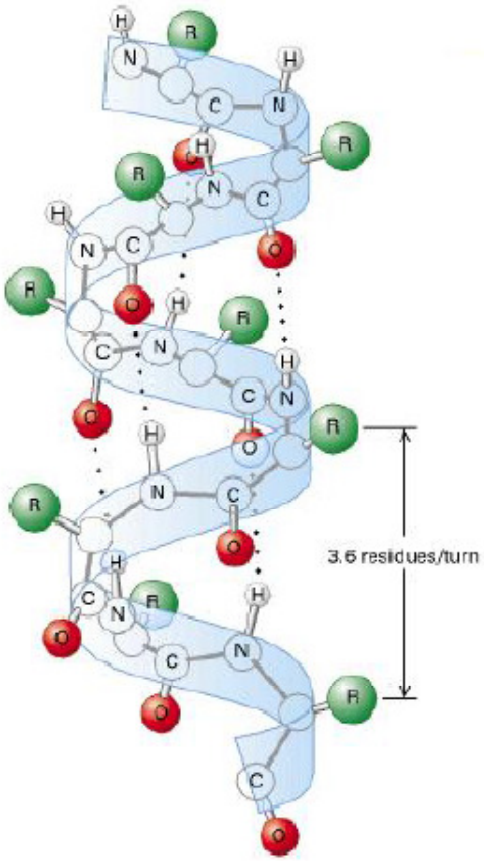
Statistical analysis of observed phi and psi angles in proteins show that most lie in or close to the two favored regions, with the exception of glycine, which may adopt conformations corresponding to the mirror image of the favored states (distribution below right).



The remainder of the polypeptide exists either in the form of tight turns or loops, connecting elements that link one segment of secondary structure to the next. Tight turns consist of 4-5 amino acids in a fairly well defined structure.

6.7 The α-helix

Characteristics of α-helix are, helical residues have negative ϕ and ψ angles, typical values being -60 degrees and -50 degrees. Every main chain C=O and N-H group is hydrogen-bonded to a peptide bond 4 residues away (i.e. O_i to N_{i+4}). This gives a very regular, stable arrangement. 3.6 residues per turn 5.4 Å repeat along the helix axis. Each residue corresponds to a rise of approxymately 1.5 Å.



Other types of helices rarely found in polypeptide chain conformations are

The 2.2_7 helix,

The 3_{10} helix and

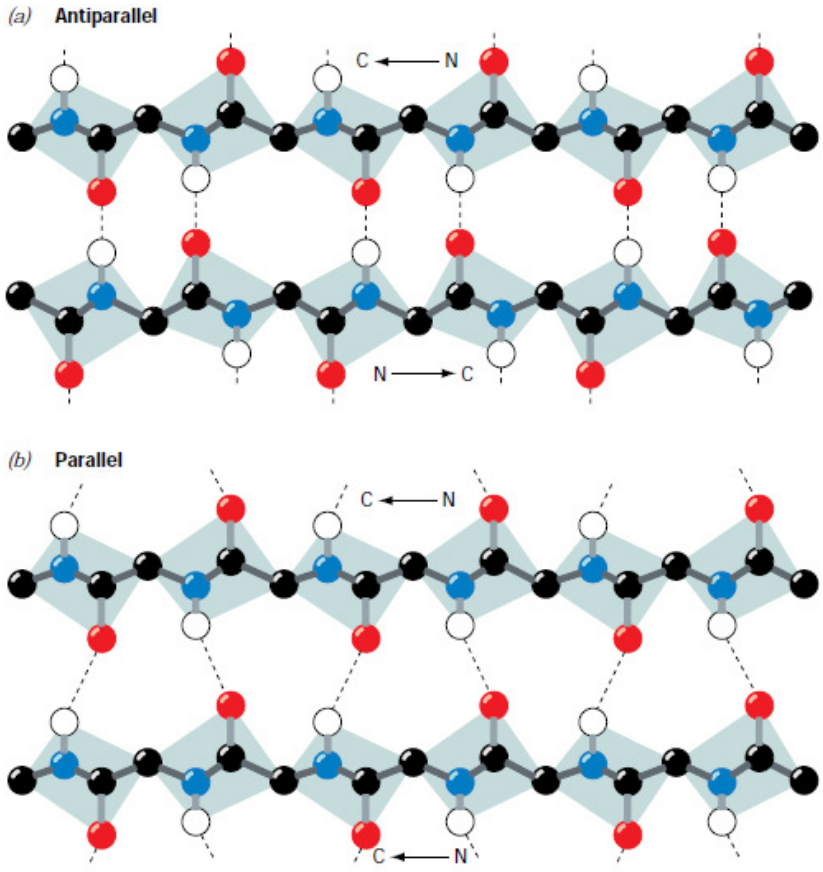
The pi helix.

3_{10} refers to a helix of 3 amino acids per turn, with hydrogen bonds from C=O of 1st amino acid to H-N of 4th amino acid. Thus the hydrogen bond closes a loop of 10 atoms.

2.2_7 is a tight helix of 2.2 amino acids per turn, and a 7 atom loop being closed by the H-bond by C=O of 1st amino acid to H-N of 3rd amino acid. The alpha helix would be 3.6_{13} in this nomenclature; the pi helix is 4.4_{16}. Although all these helices could be modelled, experimental data was only consistent with alpha helix.

6.8 β-strand

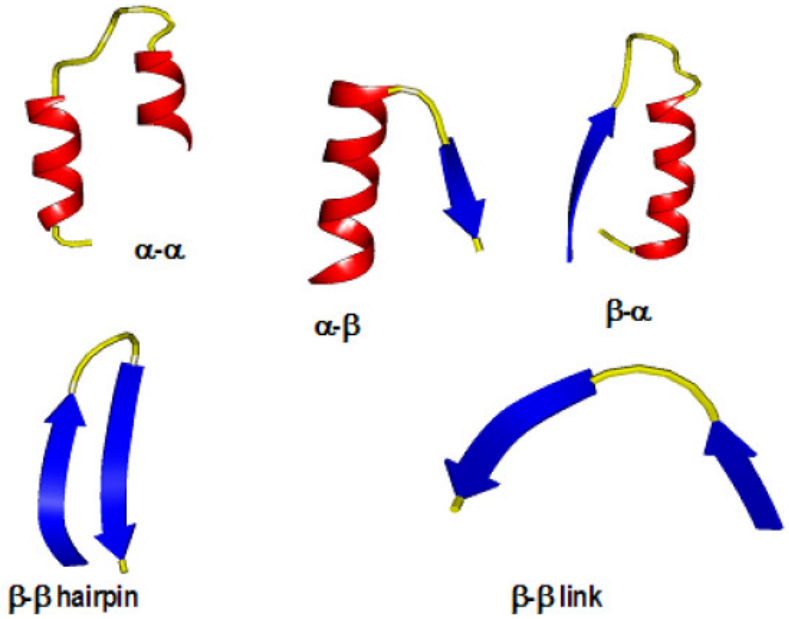
Characteristics of β-strand are**,** Positive ψ angles, typically +130 degrees, and negative ϕ values, typically -140 degrees. No hydrogen bonds amongst backbone atoms from the same strand. β-strands can form parallel or antiparallel beta-sheets. Characteristics: Stabilized by hydrogen bonds between backbone atoms from adjacent chains. The axial distance between adjacent residues is 3.5 Å. There are two residues per repeat unit which gives the β-strand a 7 Å pitch



6.9 Loops

Two dominant regions in protein structure are α-helices and β-sheets, these regions have repetitive conformation of ϕ and ψ dihedral angles and conserved hydrogen bond patterns. There is a second type of region in the protein structure with no-repetitive patterns connecting α-helices and β-strands. These regions are called loops. Generally loops were considered as an irregular

conformation or sometime referred as random coils but in modern biomolecules study loops are considered as an additional structural class. Thus loop is the general name for a mobile part of the polypeptide chain with no fixed secondary structure. Due to their flexibility and non-periodic nature, loops are not classified.



Different types of loops commonly found in protein structure

Loops play important roles in protein function, stability and folding. They generally have hydrophilic residues and are found mainly at the surface of the proteins. Functional differences between the members of the same family are usually a consequence of structural differences on the protein surface in a given fold, structural variability is a result of substitutions, insertions and deletions of residues between members of the family. Such changes frequently correspond to exposed loop regions that connect elements of secondary structure in the protein fold. Thus loop often determine the functional specificity of a given protein and contribute to active and binding sites.

6.10 Turns

Loops that have only 4 or 5 amino acid residues are called turns when they have internal hydrogen bonds. Reverse turns are a form of tight turn where the polypeptide chain makes a 180° change in direction. Reverse turns are also called β turns because they usually connect adjacent β strands in a β sheet.

There are different types of tight turns depending upon the number of atoms forming the turn. These are as follows

Delta(δ) turn: It is the smallest tight turn which involves only two amino acid residues and the intraturn hydrogen bond for a delta-turn is formed between the backbone NH(i) and the backbone CO(i+1).

Gamma (γ) turn: It involves three amino acid residues and the intraturn hydrogen bond for a gamma-turn is formed between the backbone CO(i) and the backbone NH(i+2).

Beta (β) turn : A beta-turn involves four amino acid residues and may or may not be stabilized by the intraturn hydrogen bond between the backbone CO(i) and the backbone NH(i+3).

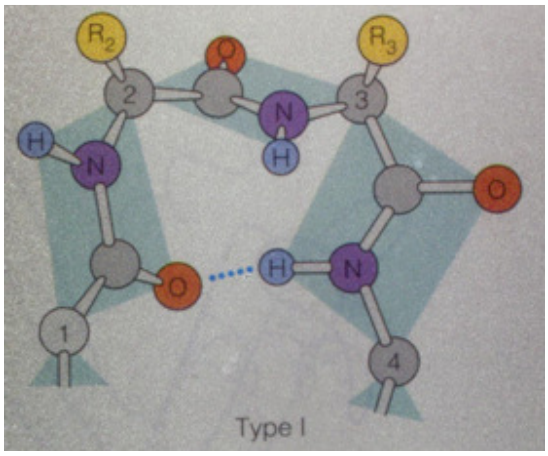
Alpha (α) turn: An alpha-turn involves five amino acid residues where the distance between the Cα(i) and the Cα(i+4) is less than 7Å and the pentapeptide chain is not in a helical conformation.

Pi (π) turn : It is the largest tight turn which involves six amino acid residues.

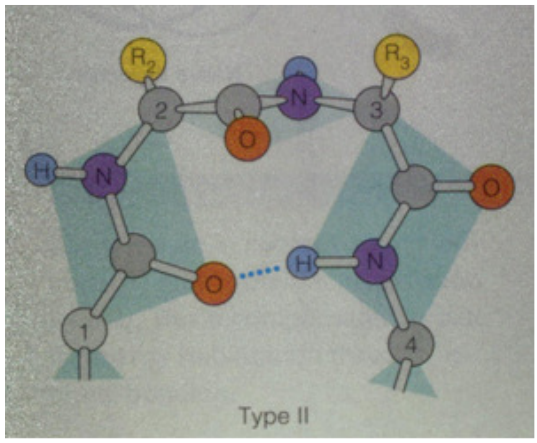
6.10 β-turn

A β-turn is defined by four consecutive residues i, i+1, i+2 and i+3 that do not form a helix and have a Cα(i)-Cα(i+3) distance less than 7Å and the turn lead to reversal in the protein chain. β-turns are common secondary structure elements that link successive runs of α helix or β sheet conformation where the polypeptide chain reverses its direction in space. Therefore β-turns are commonly located at the surface of a globular protein. Common β-turns connect the ends of two adjacent strands of an antiparallel β- sheet.

The conformation of β-turn is defined in terms of ϕ and ψ of two central residues, i+1 and i+2 and can be classified into different types on the basis of ϕ and ψ. The most common types of β turns are Type I and Type II β turns contain four amino acids in which the first and fourth residues in the turn are hydrogen bonded to one another. The second and third residues in the β turn commonly hydrogen bond to water at the surface of the protein.



Type I



Type II

The key point about turns is that they are highly ordered structures stabilized by internal hydrogen bonds. This is why they are counted as the third form of secondary structure along with the α helix and β strand.

6.11 Validation of Ramachandran plot

In 1963, Gopalasamudram Narayana Ramachandran and his coworkers were able to predict which conformations of the polypeptide backbone are possible by using simple electric desk calculators. In retrospect, these predictions were fully in line with the experimental protein structure determination of myoglobin in 1958.

Following few clever assumptions made the original predictions possible.

i) There are four covalent bonds in the protein backbone. One of them is the carbonylic C=O double bond and is irrelevant from a stereochemical perspective. Rotations around it are impossible and even if they did occur they would not affect the shape of the polypeptide backbone.

ii) The relevance of the bond between the carbonylic C atom and the amidic N atom of the next amino acid is also minor, since the partial double-bond character of this carbon–nitrogen bond. This bond can be in a *cis* or a trans conformation, which means that the dihedral angle ω can only assume a value of 0° (cis) or a value of 180° (trans) with minor distortions.

iii) In contrast, the other two covalent bonds of the polypeptide backbone are much more interesting The rotation around the N—Cα bond can be monitored by the dihedral angle Ci−1—N—Cα—C, which is named φ, while the rotation around the Cα—C bond can be measured by the dihedral angle N—Cα—C—Ni+1, which is termed ψ (where Ci−1 and Ni+1 indicate the carbonylic C atom of the preceding residue and the amidic N atom of the following residue,

respectively). Both of these covalent bonds are single and as a consequence there is possibility of modifying the conformation of the molecule by rotating around them. However, the rotations are not completely free because of interatomic clashes that can occur during the rotation. Using a hard-sphere atomic model grounded on basic quantum-mechanics principles, atomic co-penetrations are impossible, i.e. `forbidden'.

G. N. Ramachandran and coworkers ideated a simple yet surprisingly efficient method to explore the energy landscape associated with this rotation using a small model compound: N-acetyl-L-alanine-methylamide. The chemical groups conjugated to the N- and the C-termini of alanine mimicked two residues, one preceding and the other following alanine. All possible φ and ψ value combinations were computationally generated and for each it was verified whether interatomic clashes occurred, assuming a hard-sphere atomic model. The φ/ψ space that can be populated by a peptide is only about one quarter of the theoretically available space.

6.12 Summary of the unit

In an α-helix the polypeptide backbone follows a helical path. There are 3.6 amino acid residues per turn of the helix. Some protein domains assume other helical structures, but the α-helix is most common. An α-helix is stabilized by hydrogen bonds between backbone amino and carbonyl groups and those in the next turn of the helix, represented as N-H.....O=C. The hydrogen and oxygen atoms are attracted to one another because the H atom carries a partial positive charge and the O atom carries a partial negative charge, due to unequal sharing of electrons in N-H and O=C bonds. In an α-helix, the amino acid R-groups protrude out from the helically coiled polypeptide backbone. The surface of an α-helix largely consists of the R-groups of amino acid residues. Some amino acids have a greater tendency to be found within an α-helix. The amino acid proline tends to interrupt an α-helix. Its fused ring, which includes the α-carbon and the peptide-forming amino N, prevents the polypeptide backbone from assuming a conformation compatible with an α-helix in the vicinity of a proline

The 3(10) helix occurs close to the upper right of the alpha-helical region and is on the edge of allowed region indicating lower stability. A Ramachandran plot is a way to visualize backbone dihedral angles ψ against ϕ of amino acid residues in protein structure.

In a β-sheet, strands of protein lie adjacent to one another, interacting laterally via H bonds between backbone carbonyl oxygen and amino H atoms. The strands may be parallel (N-termini of both strands at the same end) or antiparallel. R groups of amino acids in a β-strand alternately

point to one side or the other of a β-strand. Hence every other amino acid is exposed on one side or the other of a β-sheet.

In a polypeptide the main chain N-Cα and Cα-C bonds relatively are free to rotate. These rotations are represented by the torsion angles phi (ϕ) and psi (ψ), respectively. A Ramachandran plot is basically a way to visualize the the dihedral angles ψ and φ of a protein backbone. Due to steric hindrances that occur between adjacent atoms within a protein structure, the φ and ψ values are usually constrained within specific areas of the plot, particularly for ordered structures such as helices and sheets. The Ramachandran plot shows the phi-psi torsion angles for all residues in the structure except those at the chain termini. Glycine residues are separately identified as these are not restricted to the regions of the plot appropriate to the other sidechain types.

6.13 Key words

Polypeptide chain; The ψ (psi) angle; The ϕ (phi) angle; The Ramachandran Plot; The α-helix; β-strand; Loops; Turns; β-turn; Validation of Ramachandran plot

6.14 References for further studies

1) Proteins: Structure and Function; David Whitford; *Wiley*, **2005**.

2) Fundamentals of Protein Structure and Function; Engelbert Buxbaum; *Springer*, **2015**.

3) Molecular Biology of the Cell, 6th Ed. Bruce Alberts, Alexander Johnson, Julian Lewis, David Morgan, Martin Raff, Keith Roberts, Peter Walter. *Garland Science*, **2014**.

4) Introduction to Protein Structure; Carl Ivar Branden, JohnTooze; *Garland Science*, **1999**.

5) Biophysics; 2nd Ed. Vasanthapattabhi, N. Gautham; *Narosa publishing house*; **2010**.

6) Biomolecular Forms and Functions: A Celebration of 50 Years of the Ramachandran Map IISc Press-WSPC publication; Manju Bansal, N. Srinivasan; *World Scientific*, **2013**.

7) Structural Bioinformatics; Jenny Gu, Philip E. Bourne; *John Wiley & Sons*, **2011**.

8) Biopolymer Chemistry; Olav Smidsrød, Størker Moe, Størker T.Moe; *Tapir Academic Press*, **2008**.

6.15 Questions for self understanding

1) Write the structueal key facts about a polypeptide chain.

2) What is the ψ (psi) angle? Explain its significance for secondary structure.

3) What is the ϕ (phi) angle? Explain its significance for secondary structure.

4) What is the Ramachandran Plot?

5) Discuss the Ramachandran plots and the protein secondary structure.

6) Discuss the structure of α-helix in terms of ψ and ϕangles and the parameters of helix.

7) Explain why proline is not fit for α-helix structure.

8) Discuss the structure of β-strand terms of ψ and ϕangles.

9) What are Loops? Explain their structural feature and their significance.

10) What are Turns? What are the difference between turns and loops?

11) What is β-turn? What are the different types of β-turn exists? Explain the differences between them.

12) Discuss the validation of Ramachandran plot in predicting and conforming the secondary structure of protein.

UNIT-7

Structure

7.0 Objectives of the unit

After studying this unit you are able to

- Overview the Protein secondary structure
- Explain the factors affecting the stability and orientation of the helix
- Explain the factors affecting the stability and orientation of the β-sheets
- Identify the forces controlling protein structure
- Explain the role of salt bridge in protein stability
- Explain the role of disulfide bond (linkage) in stabilizing the protein structure

7.1 Introduction

In principle, a protein could have a nearly limitless number of shapes (structures) due to the fact that free rotation is allowed about many of its covalent bonds. However, the great majority of proteins have a specific chemical or structural function, which suggests that each has a unique 3D structure. This idea is supported by the finding that most proteins can be crystallized. Nonetheless, most proteins display at least a moderate degree of flexibility which is needed in their performance of function. Interestingly, parts of many proteins have no fixed structure. The lack of definable structure can be crucial to function. The spatial arrangement of atoms in a protein or any part of a protein is called its conformation. The functional conformation of the protein is called its native state. The native state is usually the conformation that is thermodynamically most stable. A protein's conformation is stabilized largely by multiple weak noncovalent interactions. These include hydrogen bonds, ionic interactions, van der Waals interactions, and the hydrophobic effect of nonpolar amino acid side-chains in the interior of the protein. Disulfide bonds also contribute to structural stabilization. In the native state, the number of weak noncovalent interactions is maximal.

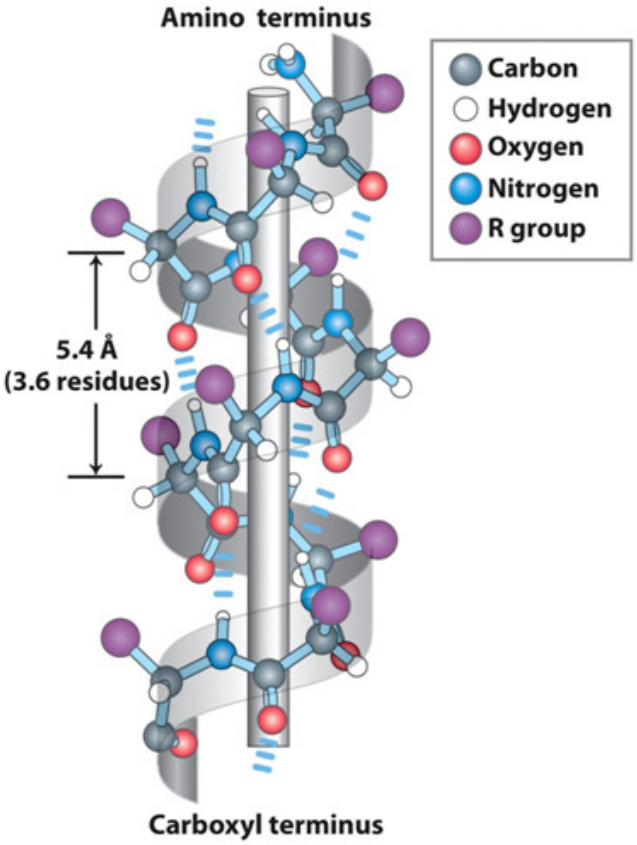
7.2 Overview of Protein Secondary Structure

Secondary structure refers to stable, short-range, periodic folding elements that are common in proteins. A regular secondary structure occurs when each dihedral angle, ϕ and ψ, remains the same or nearly the same throughout the element. There are a few types of secondary structures that occur widely in proteins. These include the α-helix, the β- sheet, and β-turns.

7.3 Factors affecting the stability and orientation of the helix

In the α-helix the peptide backbone adopts a cylindrical spiral structure in which there are 3.6 amino acids per turn (5.4 Å). The R groups point out from the α helix axis, and mediate contacts

to other structure elements in the folded protein. The α helix is stabilized by hydrogen bonds between backbone carbonyl oxygen and amide nitrogen atoms that are oriented parallel to the helix axis. In fact the structure maximizes the use of internal hydrogen bonding. Hydrogen bonds occur between residues located in the n and n + 4 positions relative to one another. *The α helix forms more readily than many other conformations in part because of its optimal use of internal hydrogen bonding.*



The following factors contributes to the formation and stability of α-helical structure

i) Electrostatic repulsion and attraction between helix atoms

ii) Side chain: interactions between residue side chains spaced 3-4 residues apart stabilize or destabilize depending upon their bulky nature

iii) Presence of proline introduces a 'kink' to the helix and breaks hydrogen bonds thus destabilize the α-helical structure. Although this depends upon preceding residues. Proline can also help to form a helix when present at N-terminal.

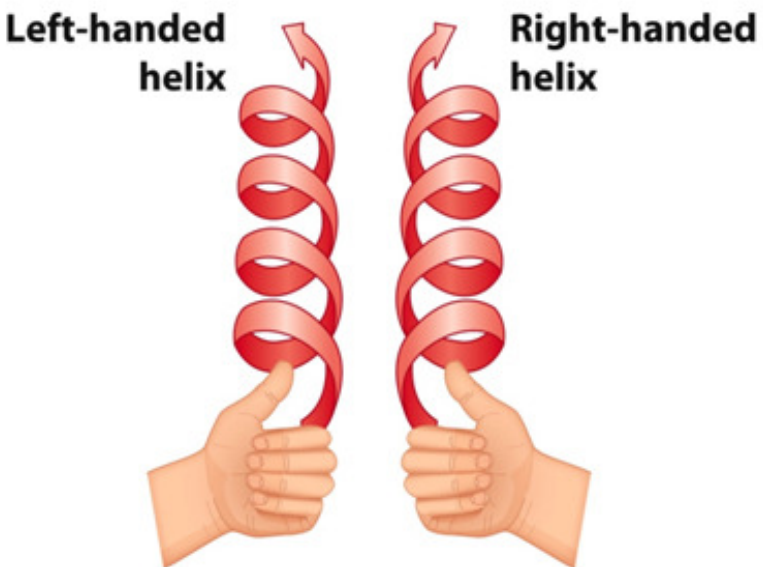
iv) Interaction between residues at the end of the helix and the dipole stabilize or destabilize depending upon the residues

v) Interaction of carbonyl group (C=O) with solvent tends to bend the helix slightly along its longitudinal axis there by stabilize α-helical structure

vi) Interaction with other secondary structure elements within protein, stabilizing effect is observed especially when in four-helix bundles.

vii) Helices can be 'capped' either at the N-terminal, the C-terminal, or both ends of the helix by residues which form extra hydrogen bonds to both main chain (polypeptide backbone) and side chain atoms. This has a tendency to stabilize the helix.

Protein structure studies have determined that right-handed α helices occur in proteins. Extended left-handed α helices have not been observed in proteins, presumably because they are theoretically less stable. Experiments have shown that an α helix can form in a polypeptide consisting all residues must be of one stereoisomeric form ie, either L- or D-amino acids otherwise the α helix will be disrupted. The most stable form of an α helix formed by D-amino acids is left-handed.

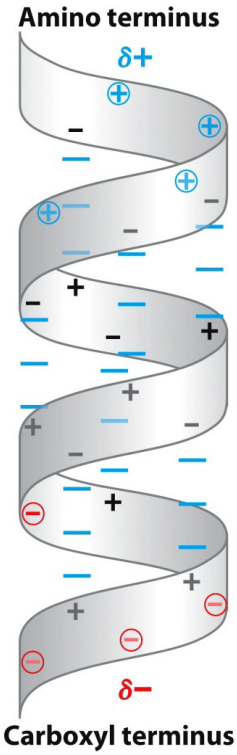


The properties of the R group (side chain) strongly affect the capacity of the backbone atoms to take up the characteristic ϕ and ψ angles of an α helix. Alanine with its small methyl group in its side chain shows the greatest propensity to form an α helix under most conditions. In contrast, amino acids such as threonine and asparagine, with bulky groups attached to the β carbon of the amino acid show a reduced propensity to occur within an α helix.

Proline and glycine show little tendency to occur in α helices. In proline the N atom is part of a rigid ring, and rotation about the N-Cα bond is not possible. This places a destabilizing kink into an α helix. In addition, the N atom when in a peptide bond linkage has no substituent hydrogen

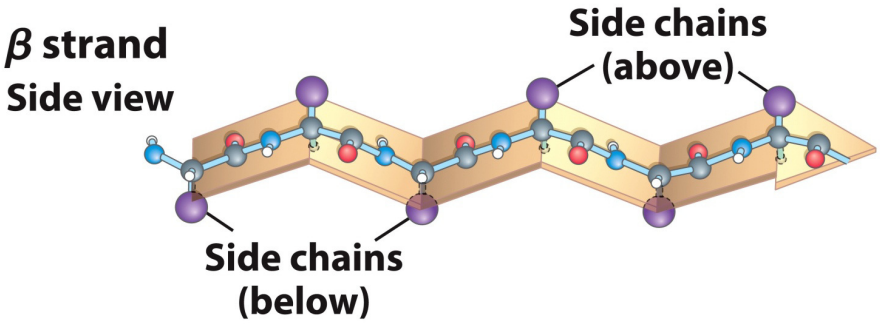
that can participate in hydrogen bonding to other residues. Glycine is highly conformationally flexible due to having a H atom for a side chain. This is disruptive to the stability of the α helix.

A small electrical dipole exists in each peptide bond. These dipoles are aligned through the hydrogen bonds of the helix, resulting in a net dipole along the helical axis that increases with helix length. For this reason, the helix is stabilized when amino acids with negatively charged side chains are located at the amino terminus of the helix, and vice versa. Placement of an amino acid with a positively charged side chain near the amino terminus of an α helix destabilizes it, as does the placement of an amino acid with a negatively charged side chain near the carboxyl terminus of the helix.

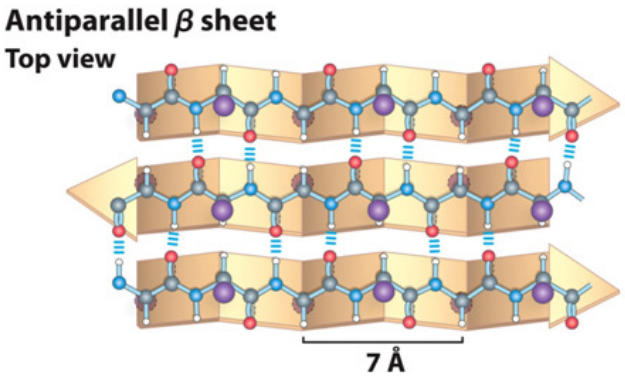


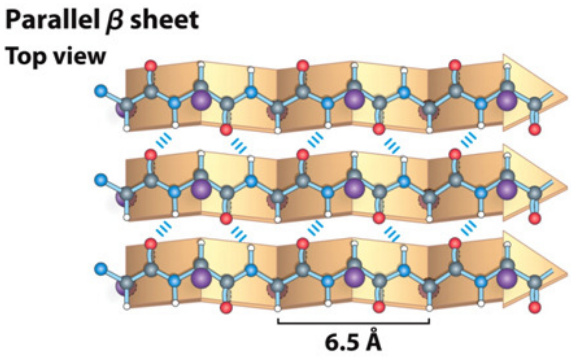
7.4 Factors affecting the stability and orientation of the β-sheets

A second type of secondary structure, the ß conformation, is very common in proteins. In this structure, the polypeptide backbone is nearly fully extended into a zigzag strand rather than a helical structure. The R groups of consecutive amino acids in a ß strand are oriented on opposite sides of the strand. As expected, the ψ and ϕ angles for the β-conformation are distinctly different from those observed in the α- helix The R groups of adjacent amino acids protrude in opposite directions from the zigzag structure, creating an alternating pattern.



The arrangement of several β strands side-by-side forms a planar type structure called a β sheet. Hydrogen bonds between adjacent strands of the sheet stabilize the structure. The adjacent strands in a β sheet can be either parallel or antiparallel, that is, having the same or opposite N-to-C terminal orientations, respectively. The hydrogen-bonding patterns are slightly different for the antiparallel and parallel β sheets, with hydrogen bonds being more perfectly aligned in the former.





When two or more pleated sheets are layered closely together within a protein, the R groups of the amino acid residues on the contact surfaces must be relatively small. The side chain groups lining one surface of a β-sheet frequently exhibit uniform chemical properties, e.g., nonpolar, polar, charged, etc. Because the β-sheet naturally defines two surfaces and can simultaneously make stable interaction with quite different physical environments.

Close internal packing of the backbone atoms in β-beta sheet structures serves to optimize van der Waals interactions and minimize energetically unfavorable hydrophobic interactions between non-polar protein groups and water molecules in the environment. Collectively, these factors help reduce the net free energy of the β-beta sheet thereby increasing its stability relative to other structures

7.5 Forces Controlling Protein Structure

a) Hydrogen Bonding

Polypeptides contain numerous proton donors and acceptors both in their backbone and in the R-groups of the amino acids. The environment in which proteins are found also contains the ample H-bond donors and acceptors of the water molecule. H-bonding, therefore, occurs not only within and between polypeptide chains but with the surrounding aqueous medium.

b) Hydrophobic Forces

Proteins are composed of amino acids that contain either hydrophilic or hydrophobic R-groups. It is the nature of the interaction of the different R-groups with the aqueous environment that plays the major role in shaping protein structure. The spontaneous folded state of globular proteins is a reflection of a balance between the opposing energetics of H-bonding between hydrophilic R-groups and the aqueous environment and the repulsion from the aqueous environment by the hydrophobic R-groups. The hydrophobicity of certain amino acid R-groups tends to drive them away from the exterior of proteins and into the interior. This driving force restricts the available conformations into which a protein may fold.

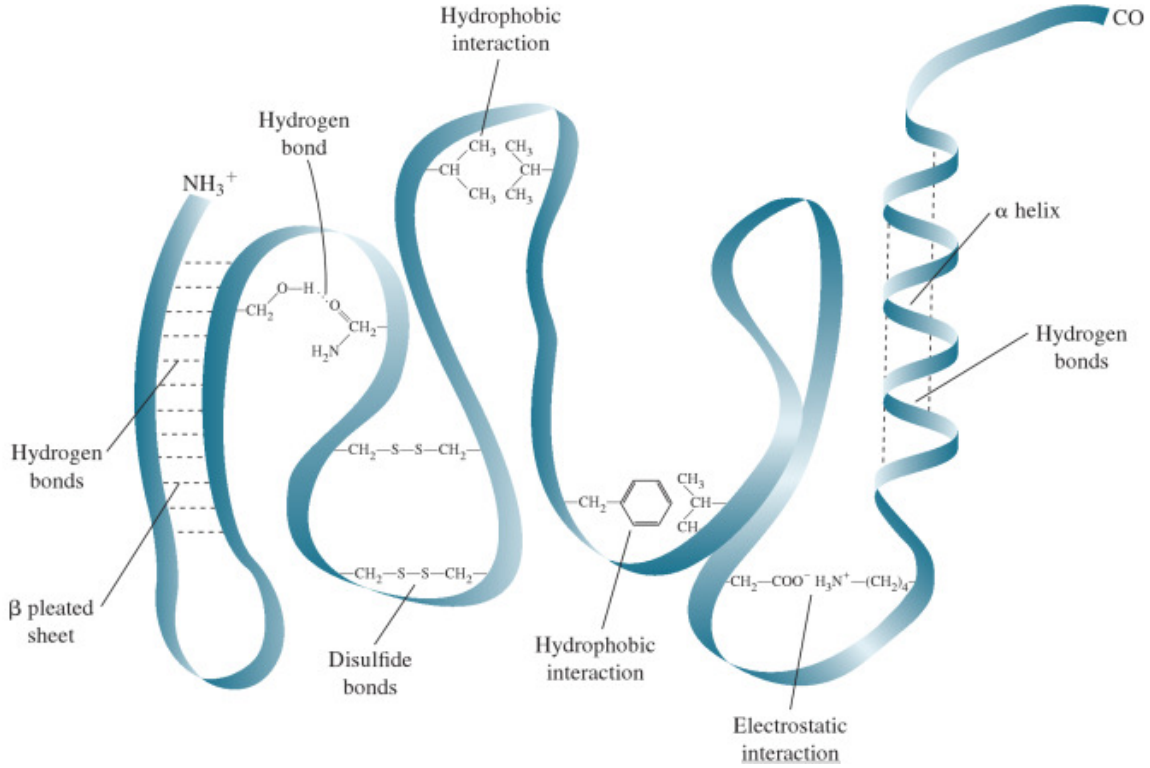
c) Electrostatic Forces

Electrostatic forces are mainly of three types; charge-charge, charge-dipole and dipole-dipole. Typical charge-charge interactions that favour protein folding are those between oppositely charged R-groups such as K (Lys) or R (Arg) and D(Asp) or E (Glu). A substantial component of the energy involved in protein folding is charge-dipole interactions. This refers to the interaction of ionized R-groups of amino acids with the dipole of the water molecule. The slight dipole moment that exists in the polar R-groups of amino acid also influences their interaction with water. It is, therefore, understandable that the majority of the amino acids found on the exterior surfaces of globular proteins contain charged or polar R-groups.

d) van der Waals Forces

There are both attractive and repulsive van der Waals forces that control protein folding. Attractive van der Waals forces involve the interactions among induced dipoles that arise from fluctuations in the charge densities that occur between adjacent uncharged non-bonded atoms. Repulsive van der Waals forces involve the interactions that occur when uncharged non-bonded atoms come very close together but do not induce dipoles. The repulsion is the result of the electron-electron repulsion that occurs as two clouds of electrons begin to overlap. Although van der Waals forces are extremely weak, relative to other forces governing conformation, it is the

huge number of such interactions that occur in large protein molecules that make them significant to the folding of proteins.



Disulfide bonds, electrostatic interaction, hydrogen bonds and hydrophobic interactions are all stabilizing influences that contribute to the tertiary structure of a protein.

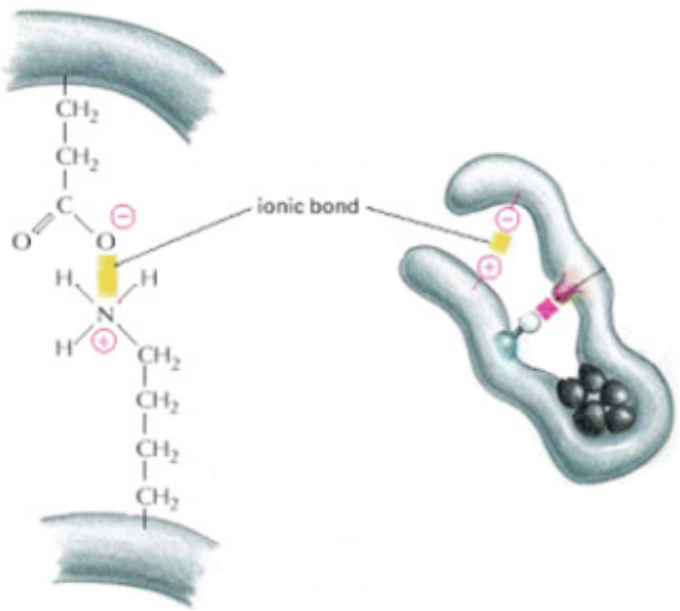
7.6 Salt bridge

Salt bridges result from the neutralization of an acid and amine on side chains. A salt bridge is a non-covalent interaction between two ionized sites. It has two components: a hydrogen bond and an electrostatic interaction. In a salt bridge, a proton migrates from a carboxylic acid group to a primary amine or to the guanidine group in Arg. Typical salt bridges involve Lys or Arg as the bases and Asp or Glu as the acids. Of all the non-covalent interactions, salt bridges are among the strongest.

Proteins fold so that positively charged side chains are often located adjacent to negatively charged side chains. The salt bridge or ionic bond between the charged functional groups helps stabilize the tertiary structure.

A salt bridge is actually a combination of two noncovalent interactions: hydrogen bonding and electrostatic interactions. This is most commonly observed to contribute stability to the entropically unfavorable folded conformation of proteins.

The salt bridge most often arises from the anionic carboxylate ($RCOO^-$) of either aspartic acid or glutamic acid and the cationic ammonium (RNH_3^+) from lysine or the guanidinium ($RNHC(NH_2)^{2+}$) of arginine. Although these are the most common, other residues with ionizable side chains such as histidine, tyrosine, and serine can also participate, depending on outside factors perturbing their pKa's. The distance between the residues participating in the salt bridge is also cited as being important. The distance required is less than 4 Å (400 pm). Amino acids greater than this distance do not qualify as forming a salt bridge. Due to the numerous ionizable side chains of amino acids found throughout a protein, the pH at which a protein is placed is crucial to its stability. This is largely due to alterations in half of what makes a salt bridge, electrostatic interactions. At extremes pH, two amino acids participating in a salt bridge could lose their ability to do so as one will lose its charge.



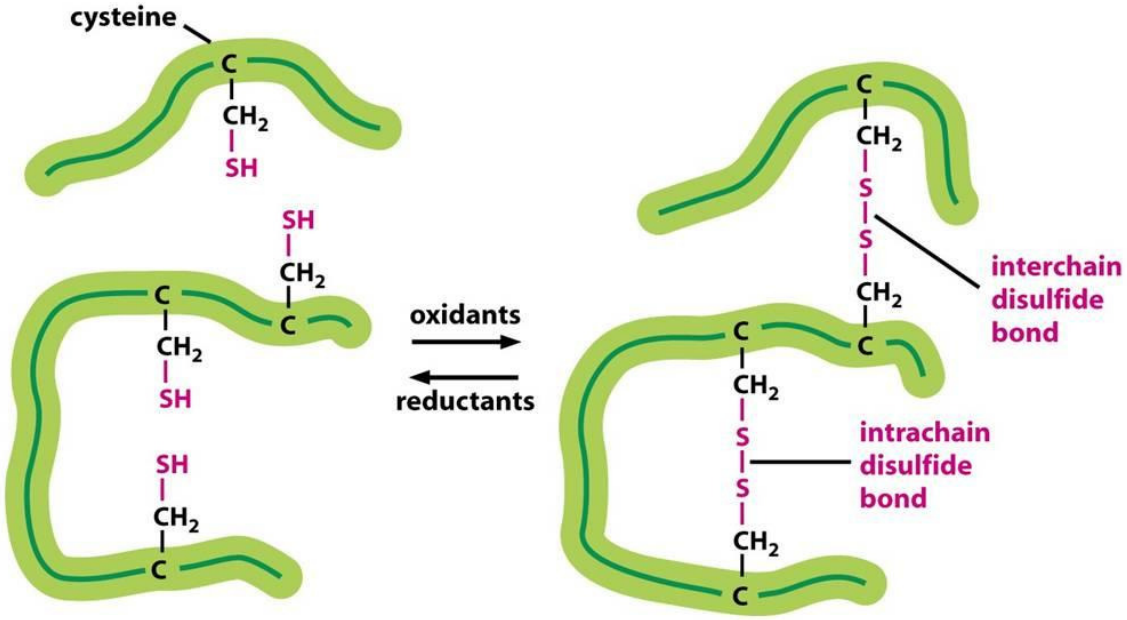
Ionic bonding (Salt bridges)

Covalent Bonds: Some proteins are cross-linked with covalent bonds. The most common is a disulfide bond between the sulfur atoms of two cysteines. These are most common in extracellular proteins.

7.7 Disulfide bond (linkage)

Cysteine is the sole amino acid whose side chain can form covalent bonds, yielding disulfide bridges with other cysteine side chains --CH_2-S-S-CH_2--. Disulfide bonds are formed by oxidation of the sulfhydryl groups on cysteine. Different chains or loops within a single protein are held together by the strong covalent disulfide bonds.

Covalent bonds between cysteine side chains can be important determinants of protein structure.



the major proteins in egg white, are ovalbumin, conalbumin, ovomucoid, and lysozyme, all contain significant amounts of cysteine, either alone or linked by a disulfide bond, in addition, the albumen protein conalbumin has a notably high affinity to bind di- and trivalent metal ions.

Boiling of an egg causes massive uncoiling and then recoiling of these proteins. As a result of denaturation, individual proteins lose their organized spatial structure. Through aggregation, various proteins gather together, with the resulting aggregates being maintained through hydrogen bonds, disulfide bridges, and an assortment of ionic and hydrophobic interactions; coagulation leads to random clustering of proteins that have already succumbed to denaturing.

Plunging hard-cooked eggs into cold water helps keep the yolk from turning green. The green-gray color (and the whiff of sulfur smell that often accompanies it) comes from the reaction of iron in the egg yolk and sulfur in the egg white. When heated, the two can combine to make green-gray ferrous sulfide and hydrogen sulfide gas. To avoid getting a green yolk, cook your

eggs just long enough to reach the desired hardness and quickly plunge the cooked eggs into cold water to stop the cooking process and minimize the iron-sulfur reaction.

7.8 Summary of the unit

Under physiologic conditions, the hydrophobic side-chains of neutral, non-polar amino acids such as phenylalanine or isoleucine tend to be buried on the interior of the protein molecule thereby shielding them from the aqueous medium. The alkyl groups of alanine, valine, leucine and isoleucine often form hydrophobic interactions between one-another, while aromatic groups such as those of phenylalanine and tryosine often stack together. Acidic or basic amino acid side-chains will generally be exposed on the surface of the protein as they are hydrophilic.

The formation of disulfide bridges by oxidation of the sulfhydryl groups on cysteine is an important aspect of the stabilization of protein tertiary structure, allowing different parts of the protein chain to be held together covalently. Additionally, hydrogen bonds may form between different side-chain groups. As with disulfide bridges, these hydrogen bonds can bring together two parts of a chain that is some distance away in terms of sequence. Salt bridges, ionic interactions between positively and negatively charged sites on amino acid side chains, also help to stabilize the tertiary structure of a protein.

Due to the nature of the weak interactions controlling the three-dimensional structure, proteins are very sensitive molecules. The term native state is used to describe the protein in its most stable natural conformation in situ. This native state can be disrupted by a number of external stress factors including temperature, pH, removal of water, presence of hydrophobic surfaces, and presence of metal ions. The loss of secondary, tertiary or quaternary structure due to exposure to a stress factor is called denaturation. Denaturation results in unfolding of the protein into a random or misfolded shape.

7.9 Key words

Protein secondary structure; Orientation of the helix; Orientation of the β-sheets; Forces controlling protein structure; Salt bridge; Disulfide bond

7.10 references for further studies

1) Proteins: Structure and Function; David Whitford; *Wiley*, **2005**.

2) Fundamentals of Protein Structure and Function; Engelbert Buxbaum; *Springer*, **2015**.

3) Molecular Biology of the Cell, 6th Ed. Bruce Alberts, Alexander Johnson, Julian Lewis, David Morgan, Martin Raff, Keith Roberts, Peter Walter. *Garland Science*, **2014**.

4) Introduction to Protein Structure; Carl Ivar Branden, JohnTooze; *Garland Science*, **1999**.

5) Biophysics; 2nd Ed. Vasanthapattabhi, N. Gautham; *Narosa publishing house*; **2010**.

6) Biochemistry; Mary Campbell, Shawn Farrell; *Cengage Learning*, **2007**.

7.11 Question for self understanding

1) Discuss the factors affecting the stability and orientation of the helix

2) Discuss the factors affecting the stability and orientation of the β-sheets

3) Explain the different forces controlling protein structure

4) What are salt bridges? How they are formed? Explain their role in protein stability

5) What is disulfide bond or disulfide linkage? Explain its significance.

UNIT-8

Structure

8.0 Objectives of the unit

After studying this unit you are able to

- Explain the general structural features of carbohydrates
- Write the classification chart of carbohydrates
- Write the structure of monosaccharides
- Name the acyclic and cycli monosaccharide
- Explain the cause and consequences of anomeric effect
- Explain the methods used for determination of the ring size
- Draw the glycosidic linkage between two monosaccharides
- Explain the Mutarotation and Epimerization

8.1 Introduction

Carbohydrates are the most abundant organic compounds in the plant world. They act as storehouses of chemical energy (glucose, starch, glycogen). They are components of supportive structures in plants (cellulose), crustacean shells (chitin), and connective tissues in animals (acidic polysaccharides) and are essential components of nucleic acids (D-ribose and 2-deoxy-D-ribose). Carbohydrates make up about three fourths of the dry weight of plants. Animals (including humans) get their carbohydrates by eating plants, but they do not store much of what they consume. Less than 1% of the body weight of animals is made up of carbohydrates. Carbohydrates are the most abundant class of organic compounds found in living organisms. They originate as products of photosynthesis, an endothermic reductive condensation of carbon dioxide requiring light energy and the pigment chlorophyll.

$$nCO_2 + nH_2O + \text{energy} \xrightarrow[\text{chlorophyll}]{\text{sunlight}} C_nH_{2n}O_n + nO_2$$

The name carbohydrate means hydrate of carbon and derives from the formula $C_n(H_2O)_m$. Following are two examples of carbohydrates with molecular formulas that can be written alternatively as hydrates of carbon.

Glucose (blood sugar): $C_6H_{12}O_6$, or alternatively $C_6(H_2O)_6$

Sucrose (table sugar): $C_{12}H_{22}O_{11}$, or alternatively $C_{12}(H_2O)_{11}$

Not all carbohydrates, have this general formula. Some contain too few oxygen atoms to fit this formula, and some others contain too many oxygens. Some also contain nitrogen. The term

carbohydrate has become so firmly rooted in chemical nomenclature that, although not completely accurate, it persists as the name for this class of compounds.

At the molecular level, most carbohydrates are polyhydroxyaldehydes, polyhydroxyketones, or compounds that yield either of these after hydrolysis. Therefore, the chemistry of carbohydrates is essentially the chemistry of hydroxyl groups and carbonyl groups, and of the acetal bonds formed between these two functional groups.

8.2 Carbohydrates

Carbohydrates are either polyhydroxy aldehydes or keynotes or compounds that yield polyhydroxy aldehydes or ketones on hydrolysis. Carbohydrates are also known as Saccharides. They mainly consist of carbon, hydrogen, and oxygen. The simple unit of carbohydrate is called monosacaride or sugar. They are bio polymers formed by the condensation polymerization of monomer units called monosaccharides. Hence monosaccharide is called as simplest units carbohydrates consist of primary and secondary alcoholic group with carbonyl group.

8.3 Structure and classification of carbohydrates

There are mainly two types of carbohydrates. They are

a) Simple carbohydrates and

b) Complex carbohydrates

Simple carbohydrates are also called as sugar and exist in either a natural or refined form. Natural sugars are found in fruits and vegetables. While refined sugars are found in biscuits, honey, jams, chocolate, brown and white cane sugar, pizzas and soft drinks.

Lactose, Maltose, Sucrose, Glucose etc…. are the examples of simple sugars.

Complex carbohydrates are also known as polysaccharides, ex: they formed by longer saccharide chains. They take longer to break down, oxidized slowly to release energy.

Starch, cellulose tec… are the examples of complex carbohydrates

Based on the number of sugar units produced up on hydrolysis, carbohydrates are classified as

1) Monosaccharides

2) Disaccharides

3) Oligosac and

3) Polysaccharides.

Monosacharides are single sugar unit. They are also called as monomers of carbohydrates

Disaccharides yields two sugar units up on hydrolysis

Oligosacharides gives 3 to 10 sugar units up on hydrolysis and

Polysaccharides gives more than 10 units up on hydrolysis

8.4 Monosaccharides

The general formula for monosacarides is $C_nH_{2n}O_n$, where the integer n is at least 3 and rarely more than 6 and with one of the carbons being the carbonyl group of either an aldehyde or a ketone. Thus monosaccharides contain the hydroxyl (alcohol, OH) and Carbonyl (Aldehyde or ketone) functional groups. They cannot be broken down into simple sugar.

Based on the nature of carbonyl group present in the monosacarides, they are classified as aldose and ketose.

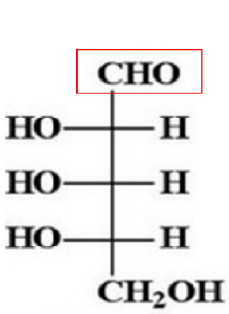
Aldoses are sugars having an aldehyde function or an acetal equivalent group.

Example, glucose.

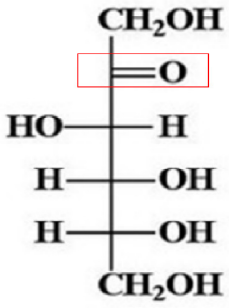
Ketoses are sugars having a ketone function or an acetal equivalent group.

Example, fructose

Thus Monosaccharides containing an aldehyde group are classified as aldoses and those containing a ketone group are classified as ketoses. A ketose can also be indicated with the suffix ulose, thus, a five- carbon ketose is also termed a Pentulose.



This sugar has an aldehyde functional



This sugar has a ketone functional

8.5 Naming acyclic monosaccharide

Three carbon monosaccharides are called trioses, four carbon are called tetroses, five carbon are called pentoses, six carbon are hexoses, and so on. The number of carbon atoms in a molecule is used in the suffix of carbohydrate naming. The suffix-ose indicates that a molecule is a

carbohydrate, and the prefixes tri-, tetr-, pent-, and so forth indicate the number of carbon atoms in the chain.

Size	Triose $C_3H_6O_3$	Tetrose $C_4H_8O_4$	Pentose $C_5H_{10}O_5$	Hexose $C_6H_{12}O_6$	Heptose $C_7H_{14}O_7$	Octose $C_8H_{16}O_8$

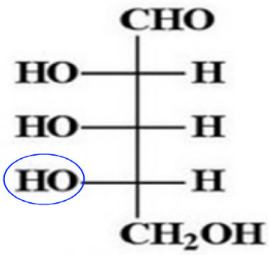
Acyclic monosaccharides have following three different characteristics

a) The number of carbon atoms it contains

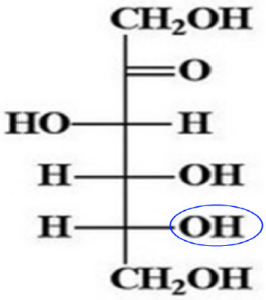
b) Its D or L configuration and

c) The placement of its carbonyl group (aldehyde or ketone).

These characteristics are combined to name monosaccharide carbohydrates. A monosaccharide is given D configuration if the hydroxyl group is to the right of the last stereocenter in a Fischer projection, whereas L configuration is given if the OH is to the left of the last stereocenter carbon. D or L is usually put in the beginning of the carbohydrate when naming the molecule. For example

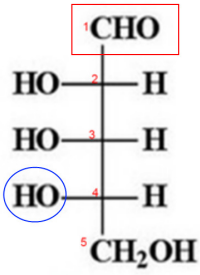


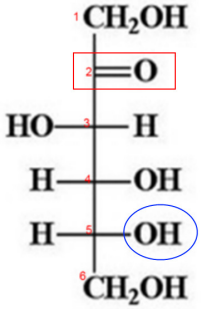
This sugar has L configuration because OH is to the left of the last stereocenter.



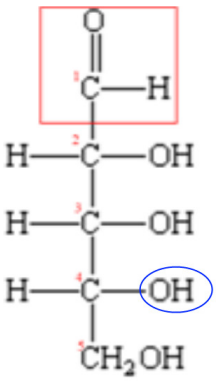
This sugar has D configuration because OH is to the right of the last stereocenter

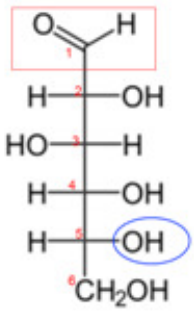
Putting all the concepts together the naming of monosacarides are illustrated in the below examples

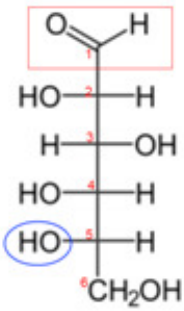




L-aldopentose D-ketohexose



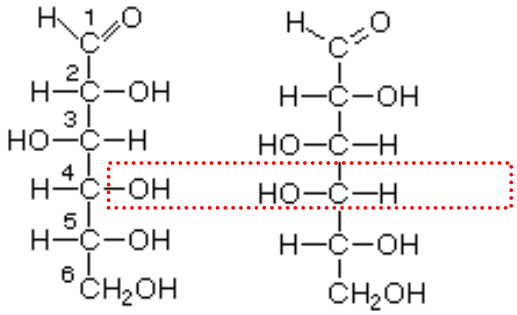




Monosaccharides contain the hydroxyl (alcohol, OH) functional group, whereas disaccharides and polysaccharides contain both the hydroxyl (alcohol, OH) functional group and a glycosidic bond between two simple sugar molecules

Monosaccharides are generally represented in open chain form. Sometimes their chemical and physical properties cannot be explained by using open chain form and for that purpose they can represent in cyclic forms called as Haworth projections. Monosaccharide like glucose, fructose, mannose and galactose can show different isomerism. For example, glucose and fructose are functional isomer. Glucose contains aldehyde group and fructose contains ketonic group. They can also show stereoisomerism due to the presence of chiral carbon atoms. *The stereoisomers which are differing in its configuration at only one chiral carbon atom are called Epimers*. More generally *Diastereomers which differ in only one stereocenter (out of two or more) are called epimers*.

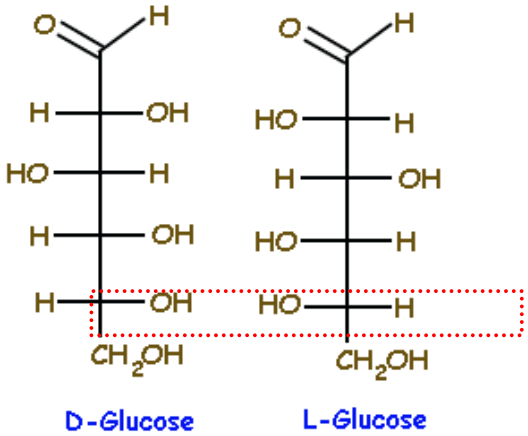
For example, glucose and galactose are Epimers of each other, as they differ in only in the position of hydroxyl group at C_4.



D-Glucose D-Galactose

D-glucose and D-galactose can therefore be referred to as epimers as well as diastereomers.

It is worth to mention here that D and L configurations of glucose are as follows.

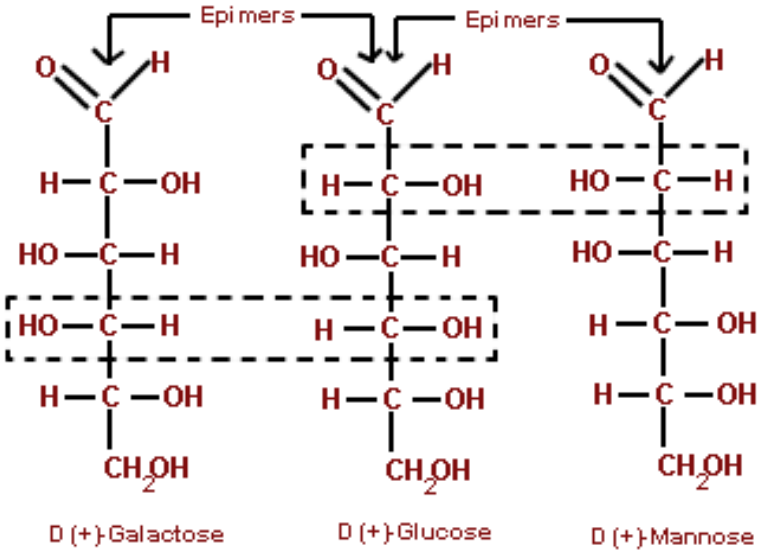


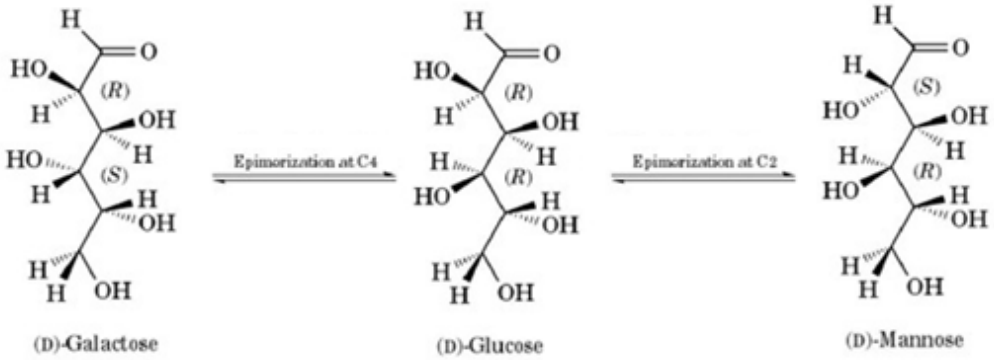
Both configurations of glucose are assigned on the basis of last chiral carbon atom that is C_5.

L-(-) glucose and D(-) glucose are enantiomers since they are mirror images to each other and are not sumpermiposible and are not epimers because they are related by inversion of more than two chiral centers . Here + ve or - ve sign indicates its optical rotation, i.e. dextrorotatory and levorotatory.

However it important to remember that, *D & L forms are not related with their optical rotation but they indicate their configurations*. Hence D-configuration can be dextrorotatory or levorotatory and vice versa. *The D and L configuration assigned on the basis of configuration of glyceraldehyde molecule which contains one chiral atom.*

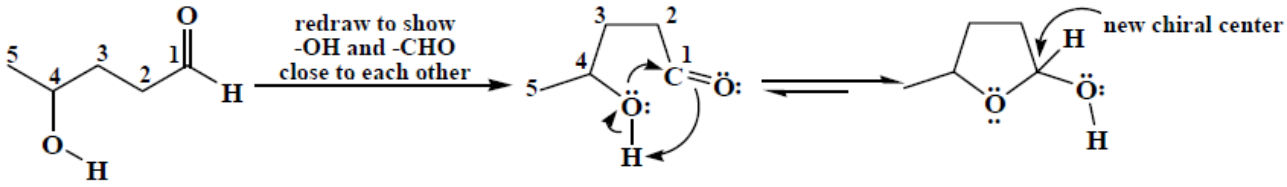
Glucose, galactose and mannose are epimers of each other as glucose and mannose are differ in the position of hydroxyl group at C_2 position whereas glucose differs from galactose at C_4 position.





8.6 The cyclic structure of monosaccharides

We know that aldehydes and ketones react with alcohols to form hemiacetals. The cyclic hemiacetals form very readily when hydroxyl and carbonyl groups are part of the same molecule and their interaction can form a five- or six-membered ring. For example, 4-hydroxypentanal forms a five-membcred cyclic hemiacetal.



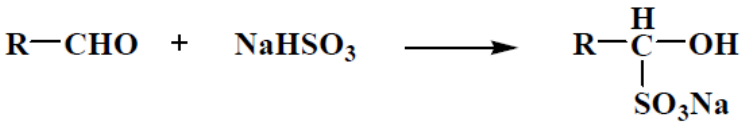
A Cyclic hemiacetal

Note that 4-hydroxpentanal contains one chiral center and that a second chiral center is generated at carbon 1 as a result of hemiacetal formation.

8.7 The cyclic structure of Glucose

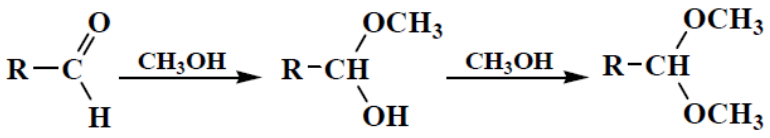
For the glucose it is apparent that the open chain structure proposed by E. Fischer dose not account for following reactions,

1) Glucose does not give the characteristic reagent of aldehydes such as the colour with Schiff's reagent and the formation of stable addition product with sodium bisulphit.



2) On acetylation glucose yields 2 different pentacetates (designated α and β).
3) Glucose forms crystalline products when refluxed with methanolic hydrogen chloride, they are methyl α and methyl β-D-glucosides with α-isomer (α)D + 158°, m.p. 166° and β-isomer (α)D - 34°, m.p. 108°

These glucosides have no reducing properties. In the D-series the sugar isomer with the most positive rotation is the α-isomer, the isomer with the lower rotation being called β-isomer. In case of a normal aldehyde a hemi-acetal then an acetal are formed with methanolic hydrogen chloride.



whereas only one (OCH_3) group is introduced into glucose to form the glucoside (hemi-acetal).

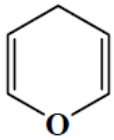
4) Corresponding with the glucosides two α- and β-modifications of D-glucose itself were isolated: α-(α)D + 112°, m.p. 146° and β-(α)D + 18.7°, m.p. 156°

α-D-glucose crystallizes from water below 35° or from cold ethanol. similarlly β-D-glucose crystallizes from water above 98° or from hot pyridine or hot acetic acid, or may be prepared by heating α-D-glucose at 105° for some time.

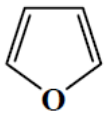
The two forms are interconvertible in solution when either form is dissolved in water, the rotation gradually changes until an equilibrium value (+52.7°) is reached.

All these observations are not explained by open chain structure of glucose. As we know the monosaccharides have hydroxyl and carbonyl groups in the same molecule. As a result, they too exist almost exclusively as five- and six-memhered cyclic hemiacecals.

Glucose can exist in both six membered and five membered ring forms. The six-membered ring structure of glucose is described as a pyranose ring (similarity to pyran) while the five-membered ring is described as a furanose ring.



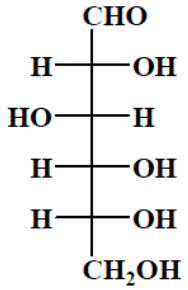
pyran



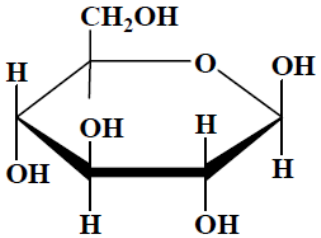
furan

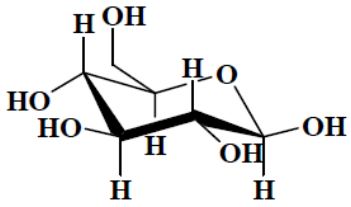
The glycosides of the pyranose sugars are described as pyranosides and those of the furanose sugars as furanosides.

A monosaccharide existing as a five-membered ring is a furanose; one existing as a six-membered ring is a pyranose. A pyranose is most commonly drawn as either a Haworth projection or a chair conformation

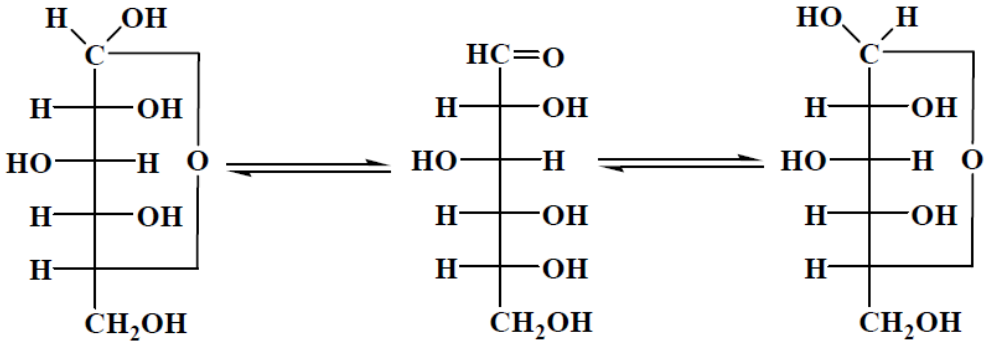


D-Glucose



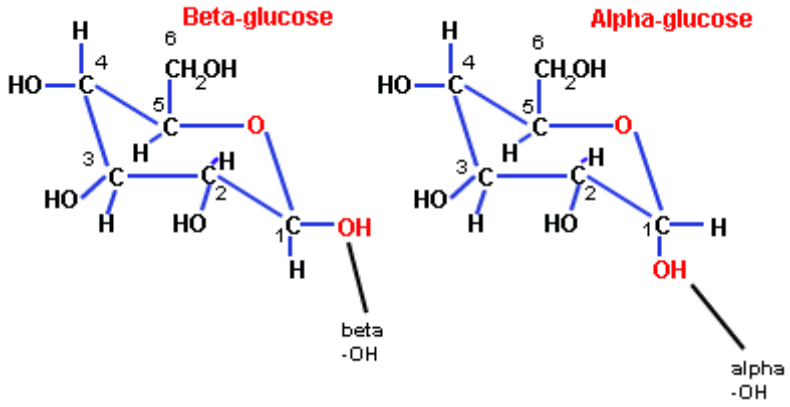


β-D-Glucopyranose (β-D-Glucose)

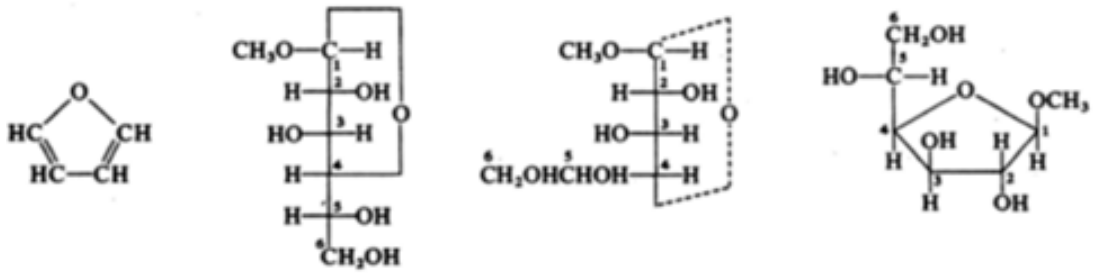


α-D-Glucopyranose D-Glucose β-D-Glucopyranose

Both cyclic forms of glucose, α-glucopyranose and β-glucopyranose are also epimers. They are differing from each other at anomeric carbon atom. They are also known as anomers of each other.



The α- and β-forms are differing in configuration at C1 and they are known as anomers (ano, upper). The hydroxyl group on C1 is cis to the hydroxyl on C2 in the α-form and trans in the β-form. Five and six-membered rings are favored over other ring sizes because of their low angle and eclipsing strain In six membered ring structure C5-OH group form a hemiacetal linkage with C1 carbonyl groupof glucose. Whereas in five membered structures C4-OH group form a hemiacetal linkage with C1 carbonyl group glucose.

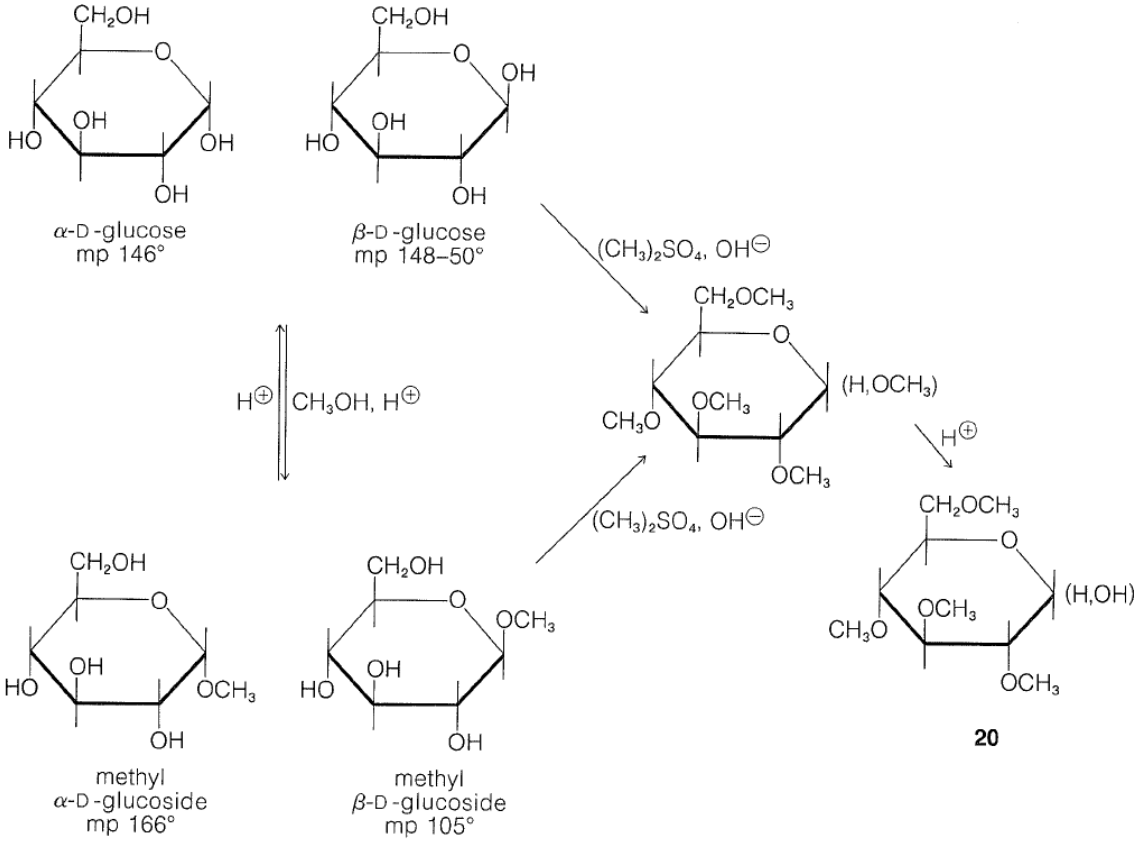


8.8 Determination of the ring size

Now we have powerful spectroscopic methods available to determine the rings sizes formed by the simple monosaccharides. The way in which this was done chemically in early days (glucose for example) highlights the difference in reactivity between ether and alcohol functions.

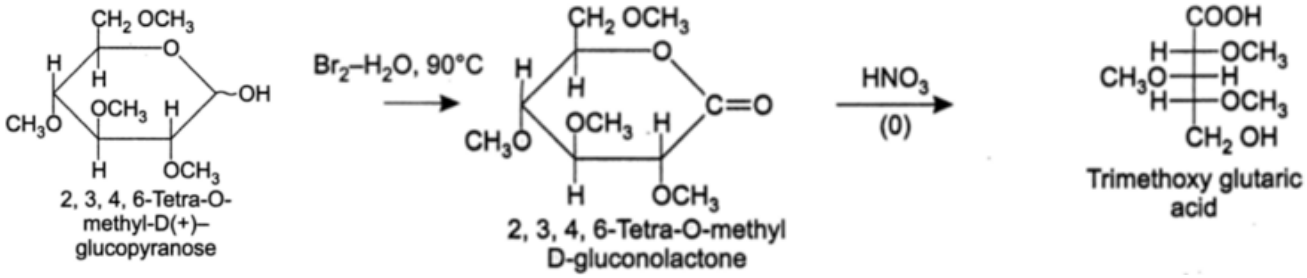
8.9 Evidence for pyranose structure of D(+)-Glucose

The acid-catalyzed methylation of glucose with methanol gave two distinct glucosides, methyl α D-glucoside and methyl β-D-glucoside, corresponds to displacement of the hemiacetal hydroxyl by methoxyl to form an acetal. The remaining four hydroxyl groups can be methylated in basic solution by dimethyl sulfate or by methyl iodide and silver oxide in N,N-dimethylmethanamide, $HCON(CH_3)_2$ solution. Hydrolysis of either of these pentamethyl glucose derivatives with aqueous acid affects only the acetal linkage and leads to a tetramethylated glucose.



Oxidation of a tetramethylated glucose with bromine at 90^0C yield a lactone, this lactone on further oxidation with HNO_3 affords trimethoxy glutaric acid.

The pyranose ring structure of D-glucose gave a tetra-O-methyl-D-glucose and showing that this substance actually was 2,3,4,6-tetra-O-methyl-D-glucose. The key feature is the fact that all but the two carbons involved in hemiacetal formation are protected from oxidation by being substituted with O-methyl groups in place of hydroxy groups.

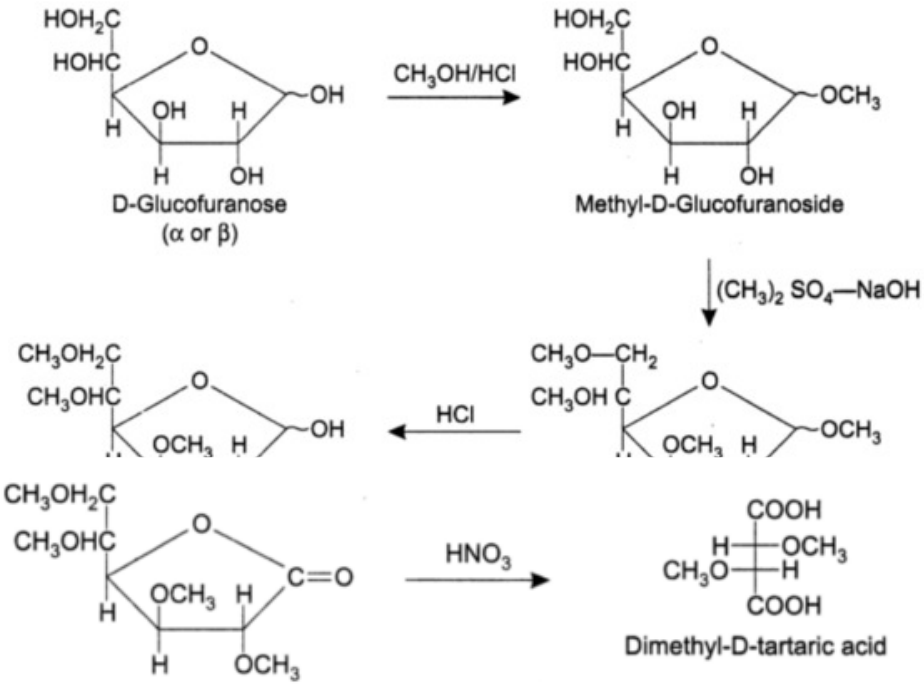


One –COOH group of this acid comes from the lactone group that is combined with the hydroxyl in the formation of lactone ring in the gluconolactone. The second carboxyl is derived from the non-methylated carbon, i.e, from the –CHOH group that is involved in the ring formation of the sugar. So there are three methoxyl groups in the lactone ring and thus C-5 hydroxyl must be involved in the ring formation. Therefore the lactone is 2,3,4,6-tetra-O-methyl gluconolactone. Working backwards in this way D(+)-glucose must have structure as D(+)-glucopyranose. This is true for both α and β isomer.

8.10 Evidence for Furanose structure of D(+)-Glucose

Fischer pepared the methyl glucoside in a different way, first he dissolved the glusoce in methanol. Then added 1% hydrochloric acid and mixture was allowed to stand at 0^0C. He obtained syrup instead of a crystalline product. This syrup was subsequently shown to a mixture of α and β glucofuranosides

The methylglucoside so obtained is methylated with exess of dimetylsulphate in the presence of sodium hydroxide to give methyl-tetra-O-methyl- D-Glucoside. This on hydrolysis with dilute hydrochloric acid gives tetra-O-methyl-D-Glucose. This tetra-O-methyl-D-Glucose when treated with bromine at 90^0C obtains a lactone which on further oxidation with HNO_3 offered a dimethoxy tartaric acid.

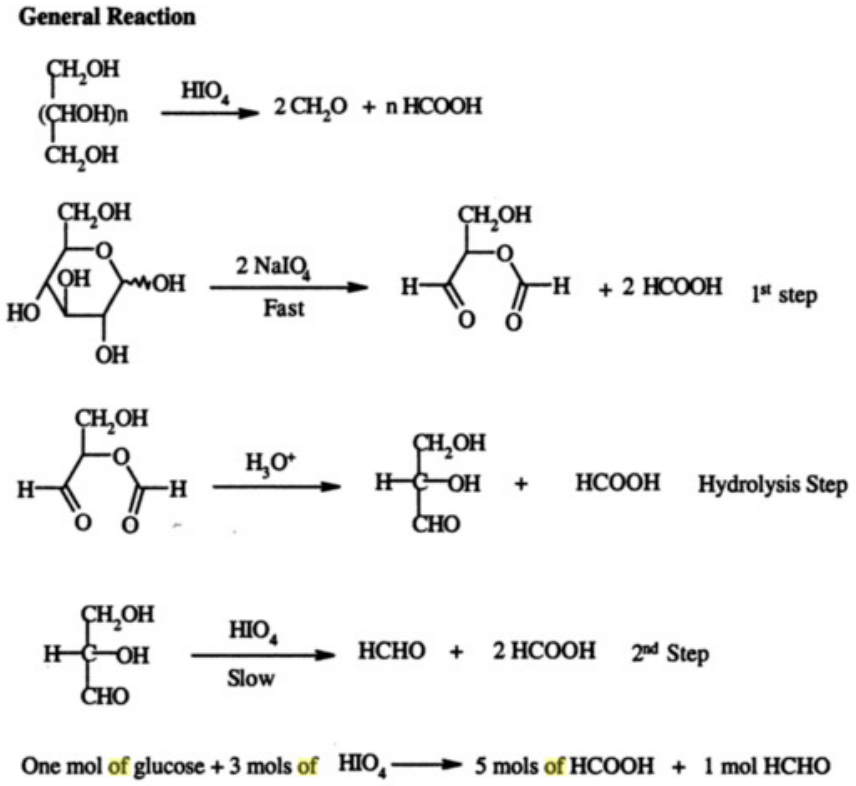


The formation of dimethoxy-D-tartaric acid indicates that there are two methoxy groups within the ring i.e., lactone is a γ-lactone. Therefore its structure is given as 2,3,5,6-tetra- O-methyl gluconolactone and hence the glucose has got furanose structure.

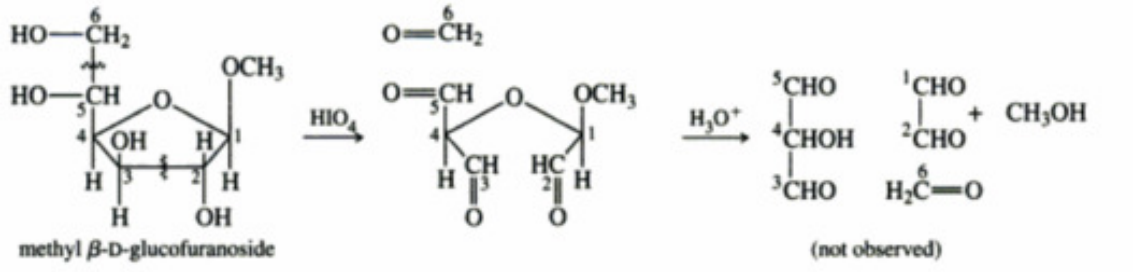
8.11 Determination of ring size by oxidation with periodic acid

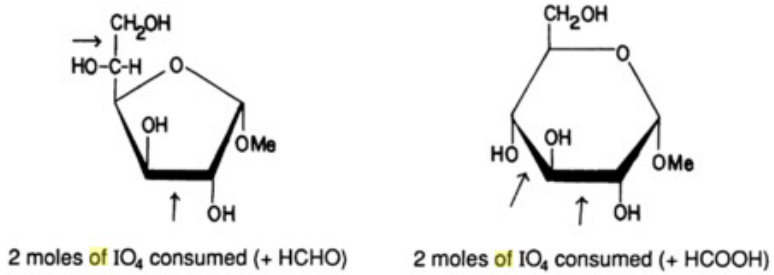
The methylation method mentioned earlier for determination of ring size of glucose is quite laborious and structures are deduced in an indirect manner. In contrast the periodic acid oxidation method procedures are facile and direct. The consumption of oxidant and liberation of products such as CO_2, formaldehyde and formic acid can be determined quantitatively.

Periodic acid cleaves the C-C bonds of α-glycols and converte CH_2-OH group in to one mole of formaldehyde and CH-OH group into one mole of formaic acid. The cis-glycols are cleaved faster than the trans-glycols.



The reaction in the case of sugars is known to take place in two steps namely selective oxidation and over-oxidation. The first step involves the selective oxidation resulting in the formation of formic acid and aldehydes.



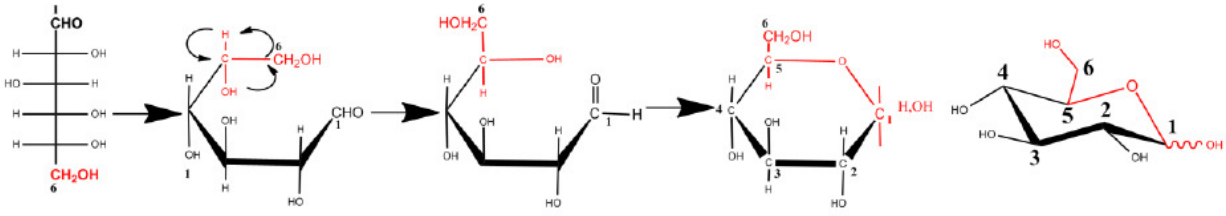


One mole of periodic acid is consumed for each pair of vicinal hydroxyl groups. Thus the quantity of periodic acid consumed and the quantity of formic acid and formaldehyde formed can reveals the number of free vicinal hydroxyl group in the sugar molecule

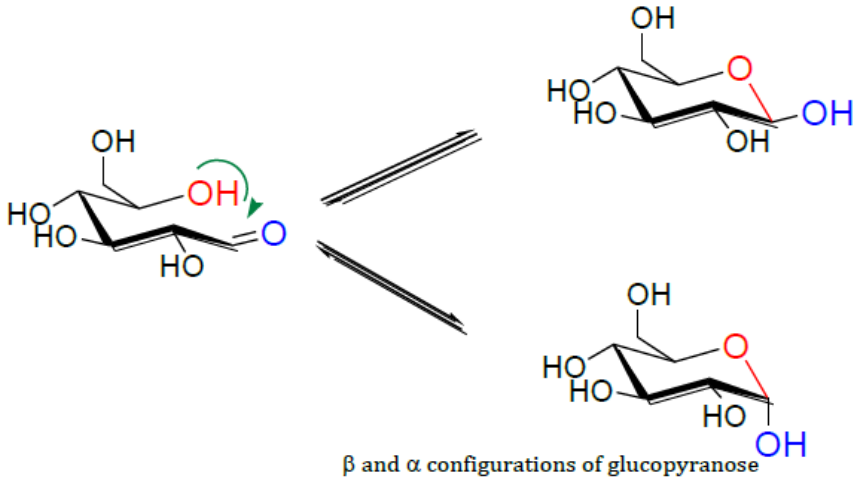
8.12 Conformation and configurations of monosaccharides

We know that in aqueous medium, monosaccharides with a suitable carbon-chain length, having both hydroxyl and carbonyl functions, undergo intramolecular (cyclic) hemiacetal formations. The formation of cyclic structure is accelerated under weak acidic or alkaline conditions. Stable five- (4 C and 1 O atom) and six (5 C and 1 O atom) membered rings are formed. The drawing of the cyclic forms of the Fischer projection formulas does not provide a realistic representation of –OH groups stereochemistry of glucose molecule.

More realistic drawings of the cyclic forms were introduced by Haworth in the 1920s, and are referred to as Haworth representations. A perspective drawing of the ring offers a simplified model. The ring is oriented almost perpendicular to the plane of the paper, but viewed from slightly above so that the edge closer to the viewer is drawn below the more distant edge, with the intracyclic oxygen behind and the anomeric carbon at the right-hand end. To define the perspective, the ring bonds closer to the viewer are often thickened. Below figure is a schematic representation of a pyranose ring closure in D-glucose that shows the reorientation at C5 necessary to allow ring formation.



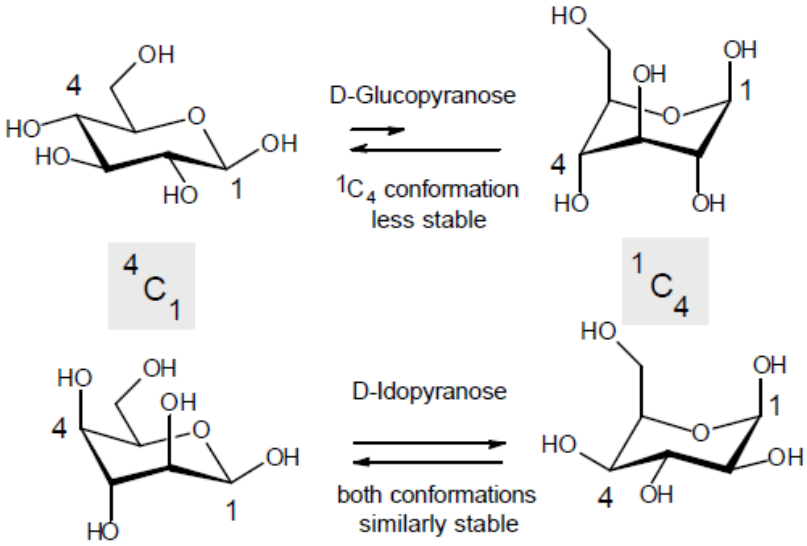
In the case of D-glucose, the hydroxyl group at C5 reacts intra-molecularly with the aldehyde group at C1. As the carbonyl carbon atom C1 of the open-chain form becomes an additional asymmetric carbon in the hemiacetal formation, two pyranose rings are formed. To describe the stereochemistry around C1 (denoted by the anomeric carbon atom), the terms α and β have been chosen. In the α-anomer, the exocyclic oxygen atom at the anomeric center is formally cis in the Fisher projection, to the oxygen attached to the anomeric reference atom; in the β anomer these oxygen atoms are formally trans.



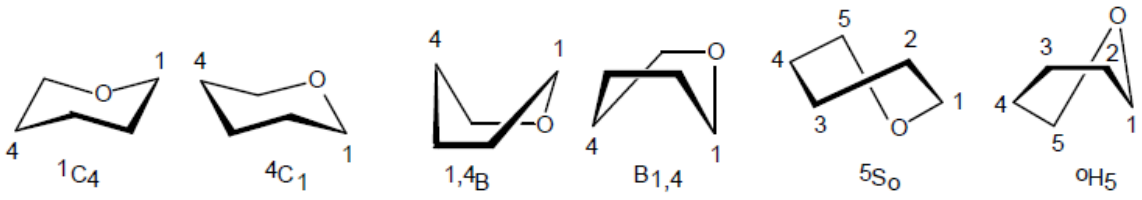
β and α configurations of glucopyranose

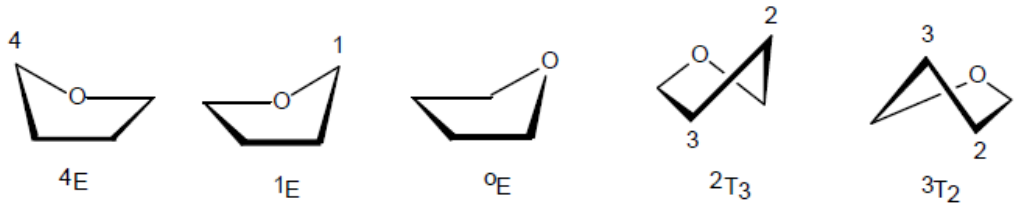
Furanose ring structures occur in envelope (E) and twist (T) conformations which can be represented on a pseudo-rotational wheel. The difference in energy between the different conformations on the wheel is generally low. Because furanoses can adopt several low energy conformations, the Haworth projection still appears to be the simplest means to avoid the complexity of structural representation.

Six-membered ring structures of monosaccharides can occur in two chair (C), six boat (B), six skew (S), and twelve half-chair (H) conformations. Like cyclohexane, the 6–membered ring of monosaccharides also exists in two isomeric chair conformations, which are specified as $^{1}C_{4}$ and $^{4}C_{1}$, respectively, where the letter C stands for 'chair' and the numbers indicate the carbon atoms located above or below the reference plane of the chair, made up by C–2, C–3, C–5 and the ring oxygen. The two chair conformations have the lowest energy, and strongly dominate. The preference for these low energy conformations is dictated by the relative orientations of the hydroxyl groups. In the case of D-glucopyranoses, only the $^{4}C_{1}$ conformation is of importance, whereas the $^{1}C_{4}$ conformation dominates in α-D-idopyranose. The $^{1}C_{4}$ conformation of β–D–glucopyranose is unfavored compared to its $^{4}C_{1}$ conformation because the van der Waals repulsion of the 1,3–diaxially positioned ring substituents result in a free energy difference between the two chairs of approximately 25 kJ/mol. Consequently only one, the $^{4}C_{1}$ conformation of β–D– glucose is observed by NMR spectroscopy. In β-D-arabinopyranose both chair conformations are in equilibrium.



Other principal conformations of pyranoses are half–chair (H), boat (B), and skew (S) conformation, which are named as indicated. The chair is by far the most stable and only the skew conformation has an energy minimum in a similar range, but this is still some 20 kJ higher than the chair. Principal conformations of the furanose ring are the envelope forms (^1E, E^1, ^2E, E^2, ^3E, E^3, ^4E, E^4, ^0E, E^0) and the twist forms ($^0T_1$, $^1T_0$, $^1T_2$, $^2T_1$, $^2T_3$, $^3T_2$, $^3T_4$, $^4T_3$, ^4T_O, $^OT_4$).





In addition to intramolecular van der Waals interactions, carbohydrate conformations are determined by some other factors, such as electrostatic interactions as well as intramolecular hydrogen bond formation and especially the anomeric effect.

8.12 Anomeric effect

We know that the equatorially positioned substituents are the most energetically favored compared to their axial counterparts in every molecule with a chair conformation for steric reasons.

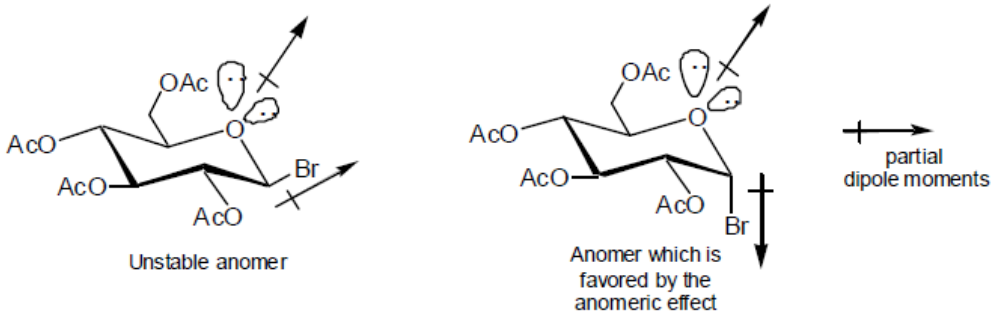
However, the anomerically bound groups in carbohydrates do not follow this rule completely. For example, in D–pyranoses, D–pyranosides and especially carbohydrate derivatives with electronegative groups at the anomeric center the anomeric α–configured derivatives with the anomeric group located in an axial position are often more stable than would be predicted from the steric interactions they have with adjacent substituents. An aqueous solution of D–glucose, for example, contains the α– and the β–form in a ratio of 36:64, and the effect is even pronounced for D–mannose, where the α:β–ratio is 69:31.

The unusual preference of the sterically unfavored axial position over the equatorial position at the anomeric center in the six memberd pyronosides is referred as the 'anomeric effect'

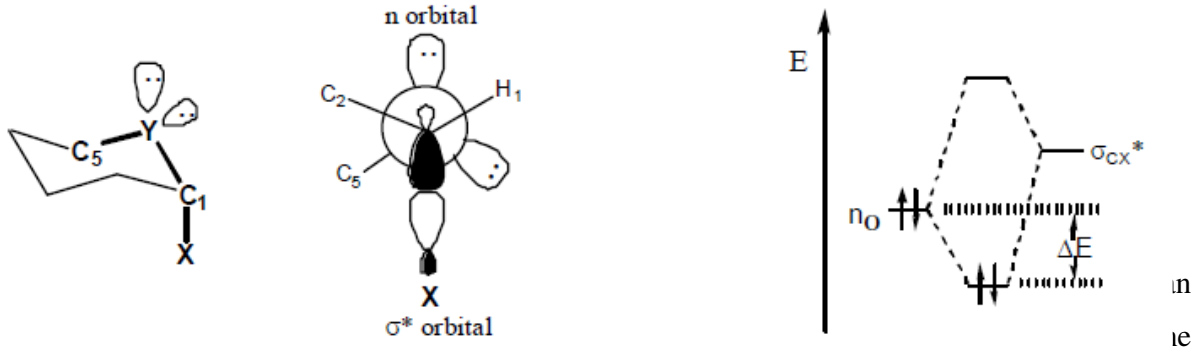
The anomeric effect was discovered in the case of carbohydrates but has been found to be of general importance for molecules, where two heteroatoms are bound to a tetrahedral center. Thus, the essential group for the appearance of an anomeric effect is

–C–**Y**–C–**X** where **Y** = N, O, S
and **X** = Br. Cl. F. N. O. S

The anomeric effect can be explained in several ways. It partly involves a dipole-dipole effect based on intramolecular electrostatic interactions of two dipoles next to the anomeric center. One of the two dipoles arises from the two lone pair electrons of the endocyclic carbohydrate ring oxygen. The other dipole points along the polarized bond between the anomeric carbon atom and its bound atom X. *Anomeric configurations, where the two dipoles partially neutralize each other are favored over the diastereomers where the anomeric configuration leads to partial intramolecular addition of the two dipoles.*

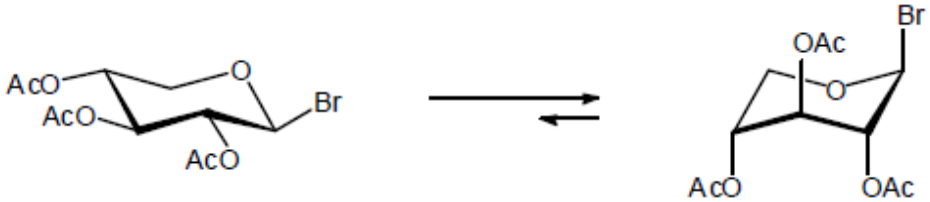


The anomeric effect virtually ensures the axial configuration of an electronegative substituent at the anomeric center such as in the case of acetobromoglucose, where the β–anomer is unknown. Also unfavorable lone pair–lone pair interactions have been used to explain the anomeric effect. Most importantly the anomeric effect is a stereoelectronic effect, in which a lone pair of electrons located in a n–molecular orbital of the atom Y overlaps with the antibonding σ*–orbital of the C–X bond. This favorable $n_X \rightarrow \sigma^*$ delocalization of nonbonding electrons ('negative hyperconjugation') is only possible with an anti-periplanar arrangement of the involved orbitals as found in the axial anomer. This interaction is also reflected by bond length changes, slightly shortening the Y–C1 bond while lengthening the C1–X bond.

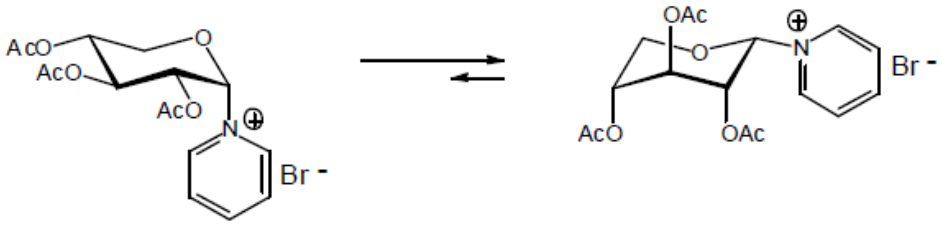


σ_{CX}^* orbital is of low energy.

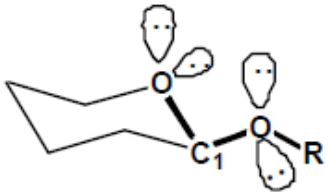
The anomeric effect is strongly influenced by the substituent at C–2. When this is equatorial, as in glucose and galactose, the anomeric effect is weakened, and is enhanced in the case of an axial C–2–substituent as in mannose. Moreover, the nature of the anomeric group is of crucial influence for the anomeric effect, as it is proportional to the electronegativity of the anomerically bound atom. Solvents also influence the anomeric effect, such that increased polarity of the solvent used decreases the influence of the anomeric effect on the equilibration of the two alternative conformers in solution. The anomeric effect may lead to conformational changes as for β–xylopyranosyl bromide which prefers the sterically unfavored $^1C_4$ conformation due to the strong anomeric effect of the bromo atom.



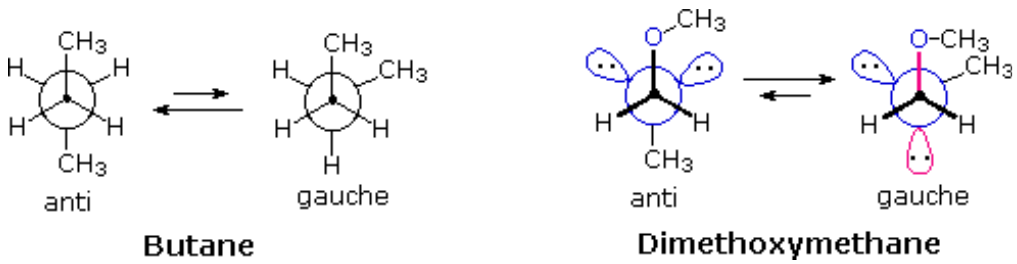
If the substituent at the anomeric center is electropositive compared to the anomeric carbon, such as a positively charged nitrogen atom, the same electrostatic considerations as stated for the anomeric effect lead to the stabilization of the anomer with the equatorially positioned anomeric group. This effect, which is causally no different from the anomeric effect, has been termed the 'reverse anomeric effect'. It is assisted by the fact that an equatorial ring position is energetically favored due to steric reasons, especially in the case of a large substituent like a pyridinium group. An anomeric pyridinium group, for example, leads to a reverse dipole at the anomeric carbon and consequently a 'reverse anomeric effect' is observed.



In alkyl glycopyranosides the anomeric effect operates not only along the endocyclic C–1 oxygen bond but also along the exocyclic C–1 oxygen bond. The anomeric effect leading to prefered conformations of the exocyclic alkoxy group is called ‘exo -anomeric effect’. Again it is an anti-periplanar arrangement of a lone pair on the aglycon oxygen and the C1 – O5 bond which determines the favored conformation. In axially configured acetals the exo - anomeric effect is less important because it operates in the opposite direction than the ‘endo-anomeric effect’. However, in an equatorial acetal the exo -anomeric effect is dominant and dictates the prefered conformation of the aglycon alkoxy group.



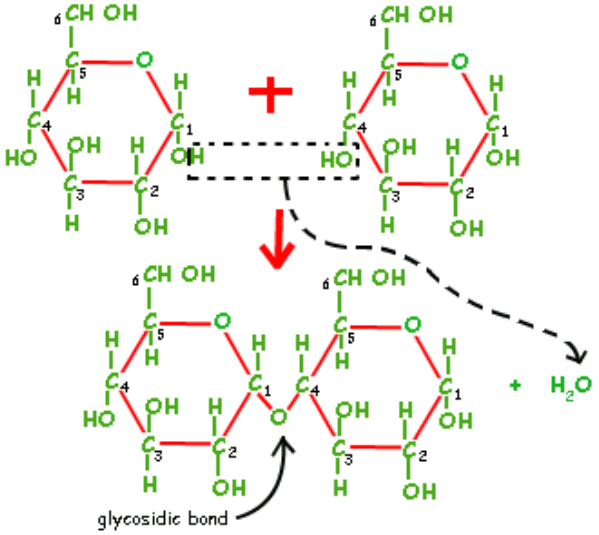
The anomeric effect is also observed in acyclic systems, as the following comparison of butane and dimethoxymethane illustrates.



The favored conformation changes from anti in the case of butane to gauche for the diether.

8.13 Glycosidic linkage

During polymerization, carbonyl group of one monosaccharide unit gets condenses with alcoholic unit of another monosaccharide to form disaccharide unit. Further as the number of monosaccharide units increases, the polymeric chain becomes longer. These polymeric chains are called as polysaccharides.



A covalent bond which bonded a carbohydrate molecule to another molecule is called as glycosidic bond. A glycosidic bond can be formed between the hemiacetal group of a saccharides and the hydroxyl group of some organic compound like an alcohol. *If the glycosidic linkage is in between a carbohydrate residue and another molecule is not saccharide, then it is called a* ***glycone***.

The presence of glycosidic bond between two saccharide units, results a structure with reducing end. The bond of an amino group or other nitrogen-containing group with the sugar is also called as a glycosidic bond. For example, in a nucleoside, the sugar- base bond is also a glycosidic bond and glcosides are substance which contains a glycosidic bond.

1) If glycosidic oxygen linked with aglycone or reducing end sugar, it is called as O-glycosidic bonds.

2) Similarly S-glycosidic bonds are found in thioglycosides, where the sulfur atom bonded in place of oxygen in glycosidic bond and in N-glycosidic bonds, nitrogen is bonded in place of nitrogen, such compounds are called as glycosylamines.

3) C-glycosyl bonds have the glycosidic carbon in place of oxygen. Glycosidic linkage is assigned on the basis position of carbon atom involves in linkage as it is an acetal linkage between carbon atoms of two monomer units.

4) The most common monosaccharide is glucose which contains carbonyl group (aldehyde group) with five hydroxyl groups. The open chain form of glucose contains carbonyl group at C-1 position, four hydroxyl groups at C-2, 4, 5, 6 on right side and on C-3 at left side.

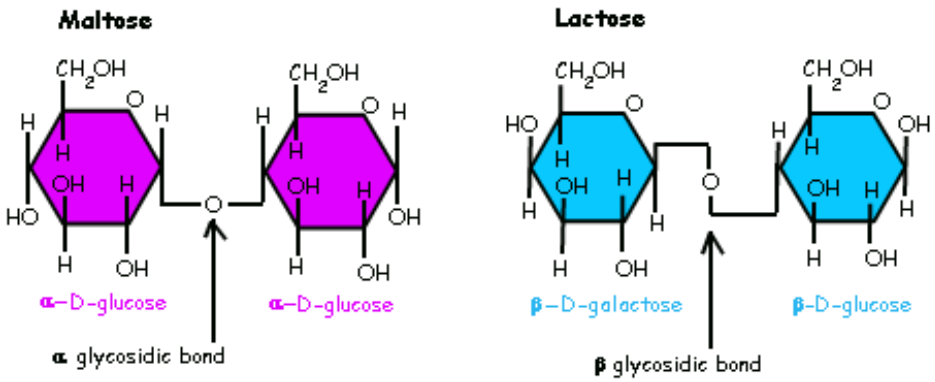
5) The C-1 of glucose is very close to C-5 in three dimensional space arrangements, hence it can arrange in ring form instead of open chain by the bond formation between H-OH of C-5 with C-1 by hemiacetal linkage and form a cyclic structure.

6) In cyclic structure of glucose, the OH groups on carbons 2, 4, and 5 which were on the right in open chain form, will be in the down positions, and the hydroxyl group (OH) lies above the plane of C-3 position.

8.14 Alpha1-4 Glycosidic Linkage

Glycosidic linkage named after the anomers of monosaccharide units (alpha and beta units) involve in this linkage. The position of alpha and beta glycosidic bonds can be identifying on the basis of the relative stereochemistry of the anomeric position and the stereo center furthest from C-1 in the saccharide.

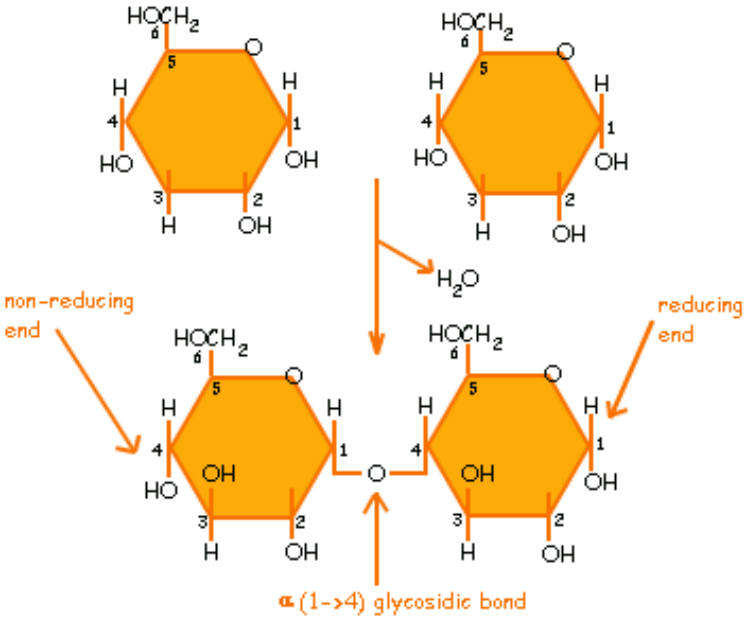
Hence the alpha and beta designations of the glycosidic bonds are based on the configuration of the first-position carbon of the monosaccharide which are linked in a glycosidic bond. For example, maltose is formed by an alpha glycosidic bond, whereas lactose consists of a beta glycosidic bond.



In D-hexose sugars like D-glucose in their pyranose forms, glycopyranose, the α-glycosidic bond is formed in an axial orientation and a α-glycosidic bond is oriented equatorially.

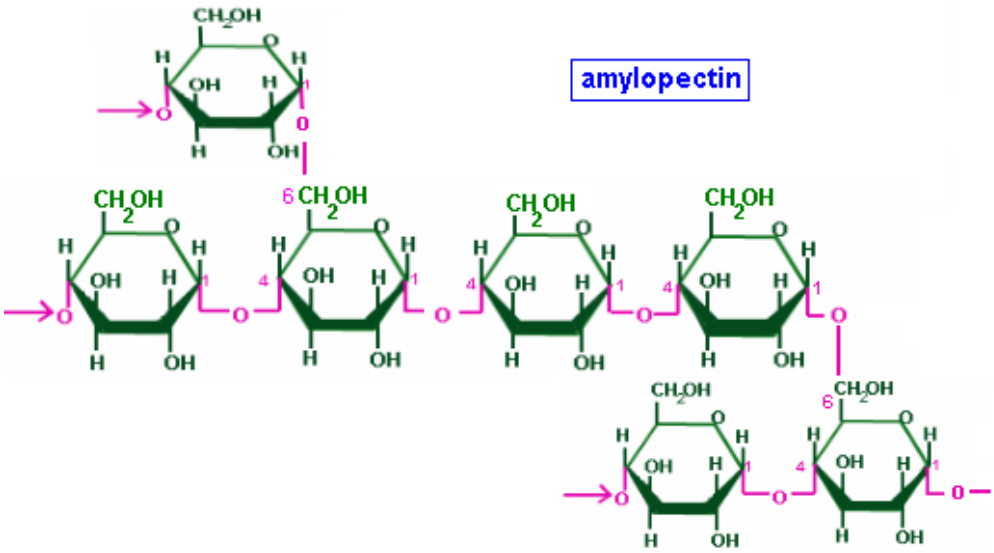
If alpha-form of monosaccharides involves in glycosidic linkage, it is known as alpha-glycosidic linkage which can be represented by the position of carbon atoms involve in bond formation.

For example in maltose sugar, two glucose units get condense and bonded with glycosidic linkage. The hydroxyl group on C-1 of one α -glucopyranose is bonded with the hydroxyl group of C-4 of another molecule with the elimination of water. Since C-1 and C-4 of two units involve in glycosidic linkage, it termed as alpha1-4-glycosidic linkage.



8.15 Alpha 1- 6 Glycosidic linkage

Similar to alpha1-4-glycosidic linkage, monosaccharide units can form alpha 1-6 linkage in which C-1 of one unit gets bonded with C-6 of another unit through glycosidic linkage.

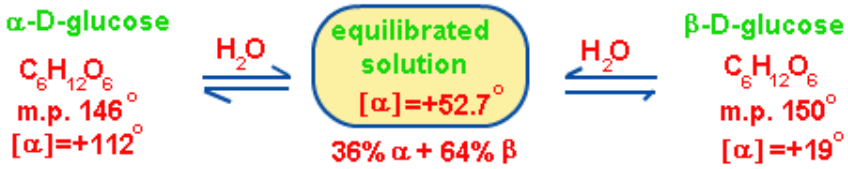


For example, in the amylopectin unit of starch is a huge branches polymeric chain of glucose containing numerous amylase-like chains which are bonded with alpha (1-4) bonds, and branch points are connected with alpha-1-6 glycosidic linkage.

8.16 Mutarotation

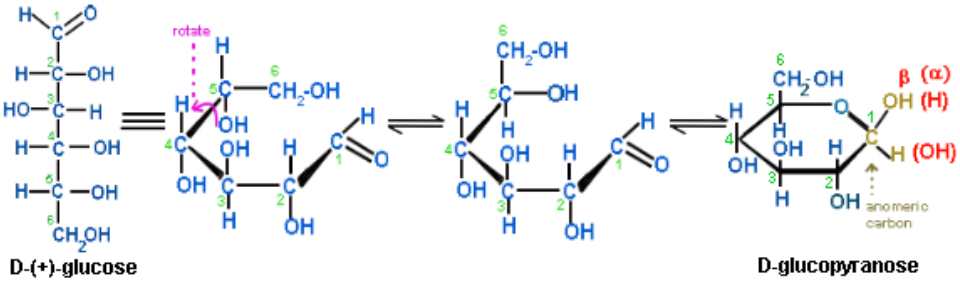
Carbohydrates act as the most important source of energy for our body and one of the main types of nutrients. In living systems carbohydrates mainly convert to glucose which is also called as blood sugar and uses this sugar as energy source. The excess of glucose stores in the form of starch (in plants) and glycogen (in animal). Depending on their chemical structures, carbohydrates can be simple or complex compounds.

As we discuss earlier in this unit, the D (+) glucose exists in two cyclic stereoisomeric forms α-glucose and β-glucose. In aqueous acid solution, the specific rotation of aqueous solution of α-D(+)glucose falls gradually from +111° to 52.5° and that of β-D(+)-glucose increases from +19.2° to 52.5°.



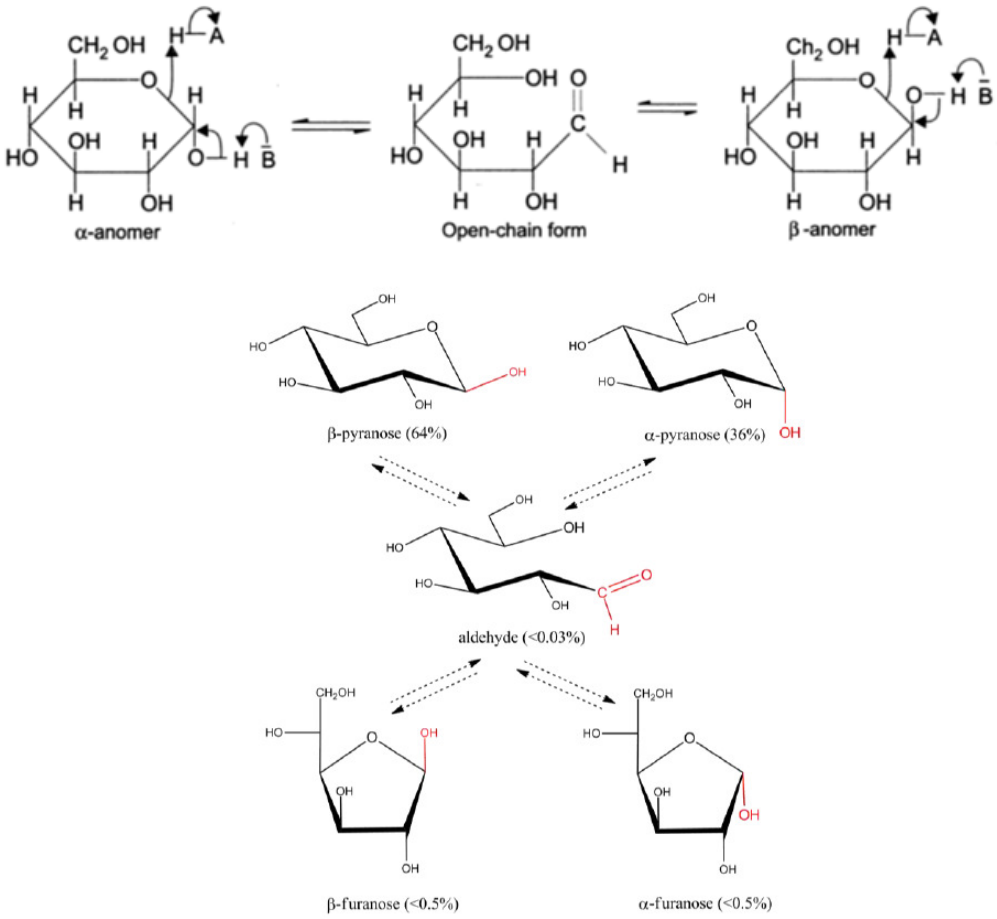
The spontaneous change in specific rotation of an optically active compound with time to an equilibrium value is called as mutarotation.

The cyclic structure of glucose involves the hemi-acetal formation between C_5-OH of glucose combines with C_1-aldehyde group. Because of that C_1 position becomes chiral and show two possible arrangements of H and OH group around it.



The phenomenon of mutarotation indicated that the two forms of D (+)-glucose are interconvertable and they have a common intermediate. This change presumably takes place via open-chain structure of glucose. ,When mutarotation takes place, the ring of one form opens up and then recloses in the inverted position to give the second form. At equilibrium about 63% of

the β-anomer and 37%of the α-anomer are present. The amount of free aldehyde exist in very small (~0.01%).



8.17 Epimerization

Epimers are diastereomers that are related by the inversion of configuration at a single chiral center. The diastereomers that are related by the inversion of more than a single chiral center are not epimers. Thus, D-glucose and D-mannose similarly D-glucose and D-galactose are epimers. Whereas D-mannose and D-galactose are not epimers, because they are related by inversion at two chiral centers, C2 and C4

The chemical conversion of one epimer to another is called epimerization. If this interconversion is catalyzed by an enzyme, the enzyme is an epimerase. For example, UDP-glucose-4-epimerase catalyzes the epimerization of the C4 carbon of glucose. In the reaction, UDP-glucose is epimerized to UDP-galactose.

8.18 Summary of the unit

Carbohydrates (also called saccharides) are molecular compounds made from just three elements, carbon, hydrogen and oxygen. Monosaccharides (e.g. glucose) and disaccharides (e.g. sucrose) are relatively small molecules. They are often called sugars. Other carbohydrate molecules are very large (polysaccharides such as starch and cellulose).

Carbohydrates are a source of energy for the body e.g. glucose and a store of energy, e.g. starch in plants. They are the building blocks for polysaccharides (giant carbohydrates), e.g. cellulose in plants and glycogen in the human body and the components of other molecules eg DNA, RNA, glycolipids, glycoproteins, ATP.

Monosaccharides are the simplest carbohydrates and are often called single sugars. They are the building blocks from which all bigger carbohydrates are made. Monosaccharides have the general molecular formula $(CH_2O)_n$ where n can be 3, 5 or 6. They can be classified according to the number of carbon atoms in a molecule

n = 3 trioses, e.g. glyceraldehyde

n = 5 pentoses, e.g. ribose and deoxyribose ('pent' indicates 5)

n = 6 hexoses, e.g. fructose, glucose and galactose ('hex' indicates 6)

Pentoses and hexoses can exist in cyclic and non-cyclic forms. In the non-cyclic form their structural formulae show they contain either an aldehyde group or a ketone group. Monosaccharides containing the aldehyde group are classified as aldoses, and those with a ketone group are classified as ketoses. Aldoses are reducing sugars and ketoses are non-reducing sugars. However, in water pentoses and hexoses exist mainly in the cyclic form, and it is in this form that they combine to form larger saccharide molecules.

Glucose is the most important carbohydrate fuel in human cells. Its concentration in the blood is about 1 gdm^{-3}. The small size and solubility in water of glucose molecules allows them to pass through the cell membrane into the cell.

There are two forms of the cyclic glucose molecule: α-glucose and β-glucose.

Monosaccharides are rare in nature. Most sugars found in nature are disaccharides. These form when two monosaccharides react. A condensation reaction takes place releasing water. This process requires energy. A glycosidic bond forms and holds the two monosaccharide units together.

The three most important disaccharides are sucrose, lactose and maltose. They are formed from the α forms of the appropriate monosaccharides. Sucrose is a non-reducing sugar. Lactose and maltose are reducing sugars.

Disaccharide	Monosaccharides
sucrose	α-glucose + α-fructose
maltose	α-glucose + α-glucose
α-lactose	α-glucose + β-galactose

Lactose also exists in a beta form, which is made from β-galactose and β-glucose

Disaccharides are also soluble in water, but they are too big to pass through the cell membrane by diffusion. They are broken down in the small intestine during digestion to give the smaller monosaccharides that pass into the blood and through cell membranes into cells. This is a hydrolysis reaction and is the reverse of a condensation reaction. It releases energy.

Monosaccharides are converted into disaccharides in the cell by condensation reactions. Further condensation reactions result in the formation of polysaccharides. These are giant molecules which are too big to escape from the cell. These are broken down by hydrolysis into monosaccharides when energy is needed by the cell.

8.19 Key words

Carbohydrates; Classification of carbohydrates; Monosaccharides; Cyclic structure of Glucose; Anomeric effect; Glycosidic linkage; Epimerization

8.20 References for the further studies

1) Biochemistry; Reginald Garrett, Charles Grisham; Cengage Learning, 2016.
2) Carbohydrates; S. P. Bhutani; Ane Books Pvt Ltd; 2010.
3) Principles of Organic Chemistry; Robert J. Ouellette, J. David Rawn; Academic Press, 2015.
4) Organic Chemistry; Joseph M. Hornback; Cengage Learning, 2005.
5) Chemistry of Biomolecules; S.P. Bhutani; Ane Books Pvt Ltd, 2009.

8.21 Questions for self understanding

1) What are Carbohydrates? Write their general formula.
2) Discuss the structure and classification of carbohydrates.
3) What are monosaccharides?
4) Draw the cyclic structure of Glucose.

5) Explain the procedure of determination of the ring size by methylation method.

6) Discuss the evidence for pyranose structure of D(+)-Glucose.

7) Discuss the evidence for furanose structure of D(+)-Glucose.

8) Explain the determination of ring size by oxidation with periodic acid.

9) Discuss the conformation and configurations of monosaccharides.

10) What is anomeric effect? Discuss the cause and consequences of anomeric effect.

11) What is Glycosidic linkage? Give example

12) Write the structure of

a) Alpha1-4 Glycosidic linkage

b) Alpha 1- 6 Glycosidic linkages

13) What is mutarotation?

14) What is epimerization?

UNIT-9.0

Structure

9.0 Objectives of the unit

After studying this unit you are able to

- Write the structural constituents of lipids
- Classify the lipids based on their structure
- Explain the difference between Fats and oils
- Explain the structural difference between fatty acids and waxes
- List out the chemical properties of triacylglycerols
- Write the biological importance of Phospholipids
- Write the properties of Fats and Oil
- Write the biological importance of lipids

9.1 Introduction

In the last two blocks we looked at what constitutes the major energy source for most people. We now turn to a group of compounds that is not only a major portion of our diet, but also a major constituent of our own bodies. Lipids include fats, oils, and cholesterol, as well as other compounds are less well known, but very important for the operation of our bodies.

Lipids are a class of compounds that are defined by how they are isolated, rather than by their composition. This makes them different from all of the other molecules we've seen so far.

9.2 Lipids

Lipids are molecules that can be extracted from plants and animals by low polarity solvents such as ether, chloroform, or even acetone. They are not appreciably soluble in water. Therefore lipids are defined as *a naturally occurring molecule from a plant or animal soluble in nonpolar organic solvents*. Fats (and the fatty acids from which they are made) belong to this group, as do oils, waxes, and steroids.

9.3 Classifications of lipids

Lipids are classified based on their physical properties at room temperature, on polarity, or on their essentiality for humans, but the preferable classification is based on their structure.

Based on structure, they are classified in to three major groups. They are Simple lipids, Complex lipids and Derived lipids

a) Simple lipids

Simple lipids consist of two types of structural moieties. They are

i) Glycerol esters that is esters of glycerol and fatty acids: e.g. triacylglycerols, mono- and diacylglycerols;

ii) cholesteryl esters that is, esters of cholesterol and fatty acids.

b) Complex lipids

Complex lipids consist of more than two types of structural moieties.

They are

i) Phospholipids that is glycerol esters of fatty acids; phosphoric acid, and other groups containing nitrogen;

ii) Phosphatidylcholine that is phosphatidic acid linked to choline, also called lecithin;

iii) Phosphatidyl acylglycerol in which more than one glycerol molecule is esterified to phosphoric acid: e.g. cardiolipin and diphosphatidyl acylglycerol;

iv) glycoglycerolipids that is 1,2-diacylglycerol joined by a glycosidic linkage through position -3 with a carbohydrate moiety. Etc..

c) Derived lipids

They occur as such or are released from the other two major groups because of hydrolysis ie, these are the building blocks for simple and complex lipids.

They include, fatty acids and alcohols; fat soluble vitamins A, D, E and K; hydrocarbons and sterols.

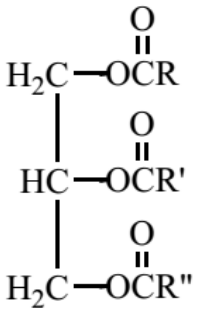
Lipids are also classified as Hydrolyzable lipids and Nonhydrolyzable lipids

Hydrolyzable lipids are those that contain a functional group that will react with water. The functional group is usually an ester and the list of compounds includes neutral fats, waxes, phospholipids, and glycolipids.

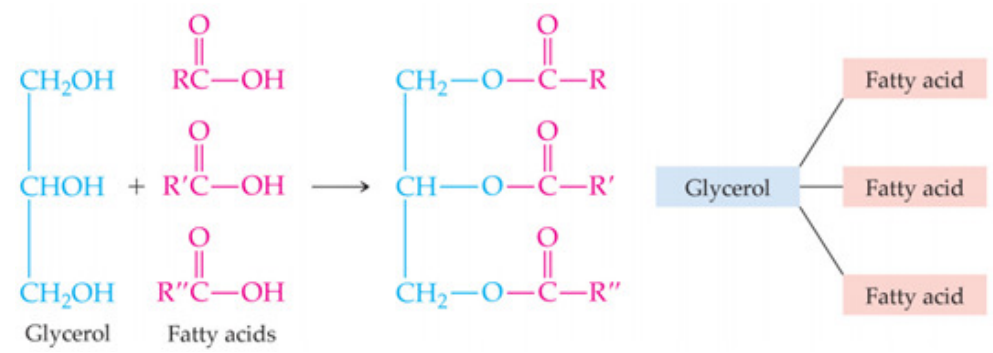
Nonhydrolyzable lipids lack such functional groups and include steroids and fat-soluble vitamins (e.g. A, D, E, and K).

9.4 Fats and oils

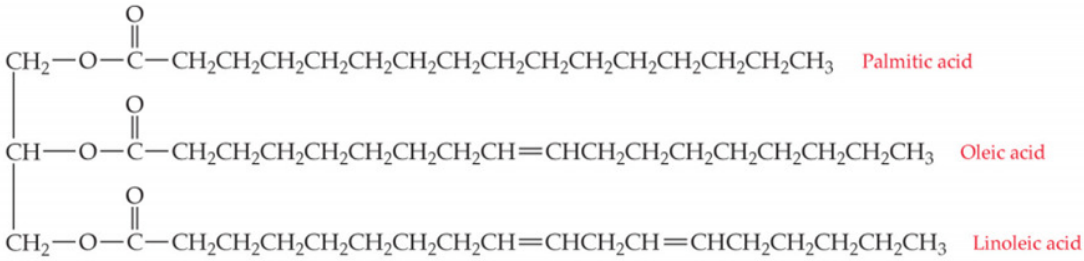
Fats and oils are composed of triacylglycerols. These are compounds prepared by the union of glycerol (1,2,3-trihydroxypropane) and 3 fatty acids to form a triester. These are the neutral fats and a generic example is shown below.



The old name for these molecules was triglycerides and that name is still used by physicians in blood tests. Complete hydrolysis of triacylglycerols yields three fatty acids and a glycerol molecule. Typically each triacylglycerol contains 3 fatty acids. These fatty acids may be the same or different.



Example



9.5 Fatty acids

Fatty acids are long chain carboxylic acids (typically 16 or more carbon atoms) which may or may not contain carbon-carbon double bonds. Fatty acids almost always contain an even number of carbon atoms and are usually unbranched.

Oleic acid, $CH_3(CH_2)_{16}CO_2H$ is the most abundant fatty acid in nature. It is monounsaturated.

9.6 Classification of fatty acids

Fatty acids are mainly two types they are saturated fatty acids and unsaturated fatty acids. Fatty acids with no carbon-carbon double bonds are called saturated fatty acids. Fatty acids with at least one double bond are called unsaturated fatty acids. Further those with one double bond are called monounsaturated, and those with two or more double bonds are called polyunsaturated. Saturated fats are typically solids and are derived from animals, while unsaturated fats are liquids and usually extracted from plants. In all unsaturated fats the double bond assumes a cis geometry. This prevents the molecules from packing as efficiently as they do in saturated molecules and this causes the lower boiling points of unsaturated fats.

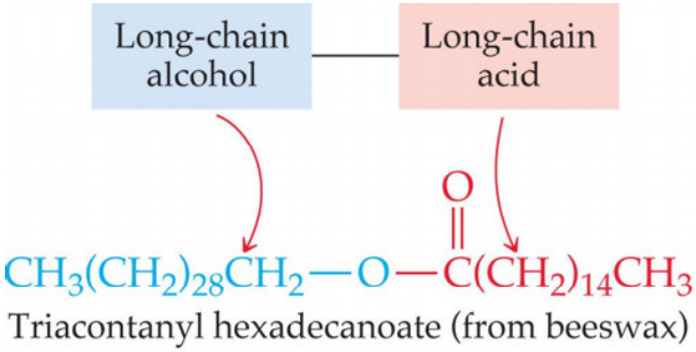
Fatty acids may also be classified as essential or nonessential fatty acids.

Essential fatty acids are those that our body cannot synthesize and must be obtained from diets.

Nonessential fatty acids are those that we can make in our body because we have the proper enzymes present.

9.7 Waxes

Waxes are simple esters with very long hydrocarbon chains 'RCOOR' ,with long, straight hydrocarbon chains in both R groups; they are secreted by sebaceous glands in the skin of animals and perform mostly external protective functions.



Short chain esters generated pleasant odors (e.g. $CH_3(CH_2)_2CO_2CH_2CH_3$ = pineapple). One of the reasons they give a pleasant smell is that with short hydrocarbon chains they are slightly water-soluble so can be dispersed throughout fruits and vegetables and being small and lightweight they have high vapor pressures (evaporate readily).

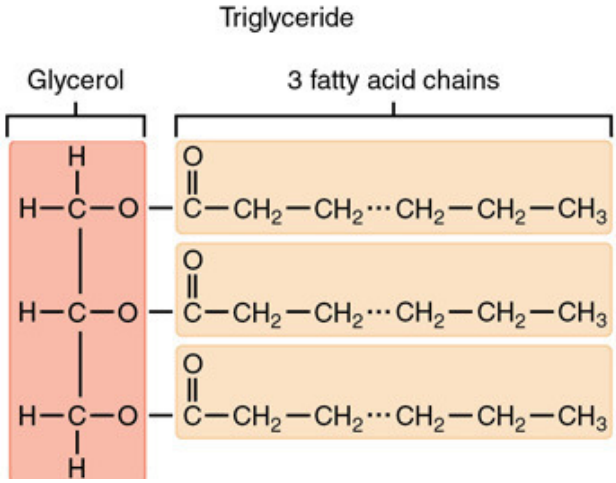
In contrast, the wax $CH_3(CH_2)_{24}CO_2(CH_2)_{29}CH_3$ (melissyl cerotate, a major component of beeswax) is a water- insoluble, solid at room temperature. The chain, including ester oxygen, is 57 atoms long. Lanolin is another wax.

In plants, waxes cover the outside of fruits, vegetables, and leaves to prevent excessive loss of water and to protect against attack by parasites.

In animals, waxes can make feathers water repellent or keep skin soft, among other things. Whale oil, $CH_3(CH_2)_{14}CO_2(CH_2)_{15}CH_3$ (cetyl palmitate) was once used as fuel, in ointments cosmetics and in candles.

9.8 Triacylglycerols

Triacylglycerols are carboxylic acid triesters of glycerol, a three-carbon trialcohol. They make up the fats stored in our bodies and most dietary fats and oils. They are a major source of biochemical energy.

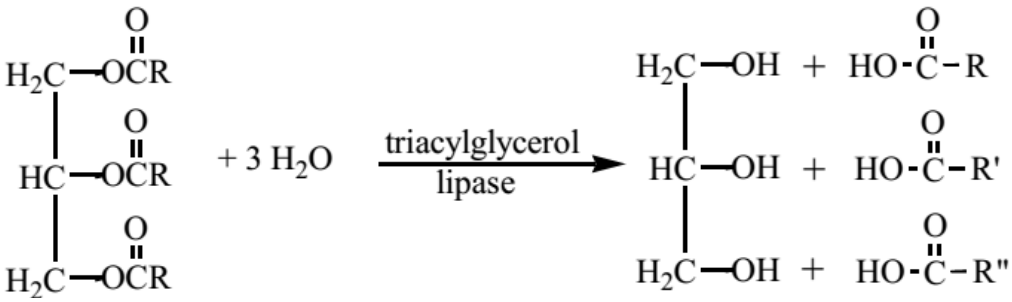


Triacylglycerols are the form in which fat energy is stored in adipose tissue. The various dietary plant oils, such as olive oil and corn oil, are also triacylglycerols.

The presence of unsaturated fatty acids in a triacylglycerol molecule makes it more fluid. This is due to the presence of the kinks in unsaturated fatty acids, which keeps the fatty acid chains from aligning uniformly. This is why plant oils, which contain primarily unsaturated fatty acids, are more liquid than animal fats. As noted above, when *trans* fatty acids are made, the trans configuration of the double bond also produces a straighter chain of carbons and thus the resulting triacylglycerols tend to be more solid. Triacylglycerols are sometimes referred to as triglycerides.

9.9 Chemical properties of triacylglycerols

The hydrolysis of a triacylglycerol yields a glycerol molecule and 3 (usually different) fatty acid molecules. Inside the body the enzyme triacylglycerol lipase employs the catalyze the hydrolysis of fatty acid.



Our body retain fat as an energy storage mechanism. When a fat molecule is hydrolyzed some amount of energy is released. Because the three new O-H bonds formed are stronger than the three C-O bonds that are broken during hydrolysis. This difference in the bond strengths accounts for most of the energy released. When fats are ingested, the body cleaves off one or two of the acid groups in the intestines. This is because fat molecules are too large

to pass through cell membranes. The monoacylglycerol (or diacylglycerol) and fatty acid then pass into the cell. Later the molecules are recombined into fat molecules. Fat molecules travel through the body by hitching a ride on proteins traveling in the bloodstream.

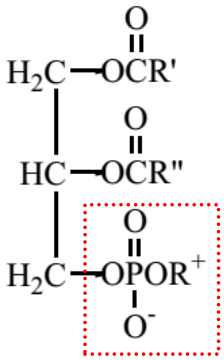
We have seen the label "partially hydrogenated vegetable oil" in the ingredient list of some food. Many times we ask ourselves what does it means?. From earlier study we learnt that hydrogen could be added to alkenes (carbon-carbon double bonds) to make alkanes. This process is called hydrogenation. As we just mention in our above discussion, vegetable oils contain very high percentages of unsaturated acids. By hydrogenating some (not all) of the C=C double bonds in the oil, its melting point slowly rises until it becomes a solid at room temperature. This is the major reason for making partially hydrogenated fats.

Solid fats are easier to store. For example, "natural" peanut butter. When we open a bottle of natural peanut butter there will be a significant oil layer on top that must be stirred back in to soften it enough to spread on a piece of bread. After a day or two days the oil begins to come out again. Standard commercial brands remove some of the peanut oil, partially hydrogenate it, and then put it back. This prevents separation and helps maintain texture.

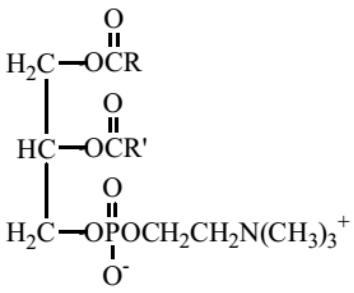
In terms of caloric content all fats are the same, 9.0 Calories per gram. Because they are liquids and won't form solid deposits in your veins and arteries, fats with very high percentages of monounsaturated and polyunsaturated fatty acids are healthier than those with high percentages of saturated fatty acids. Partially hydrogenated vegetable oil differs from animal fat only in that it lacks cholesterol.

9.10 Phospholipids

There are two classes of phospholipids. The first are the glycerophospholipids, which are themselves subdivided into two groups. The first group, phosphatides, is molecules composed of glycerol substituted with two fatty acid esters (just like in fats) and at the third position a phosphate unit connects to an alcohol. It is seen in the below structure that there is a charge on the "phospho" part of the molecule. This is because there are no examples of neutral phospholipids found in nature.

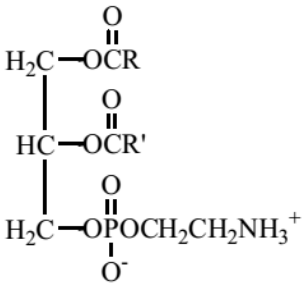


Three alcohols that form phosphatides are choline, ethanolamine, and serine. These compounds are important to the body and are transported as the following phosphatides. All three phosphatides are components of cell membranes



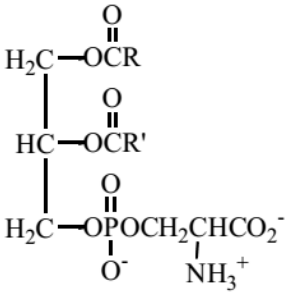
phosphatidylcholine

lecithin



phosphatidylethanolamine

cephalin

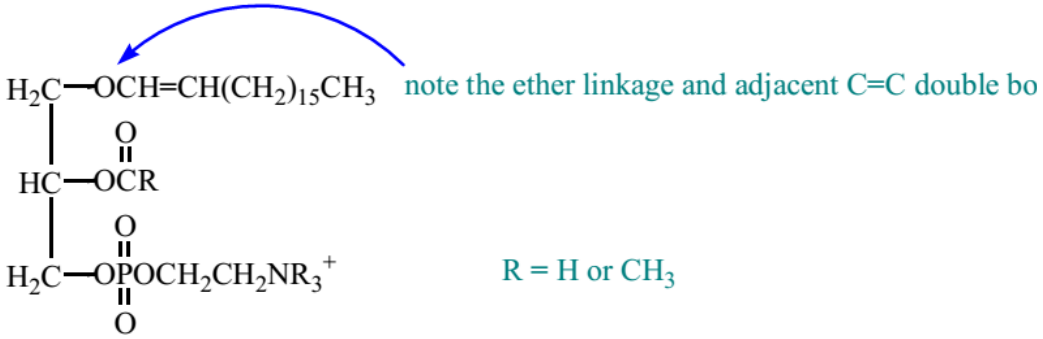


phosphatidylserine

The enzymes either cut the molecule free when it is needed or convert it to some other necessary material. The phosphate group and the organic chain attached to it carry electrical charges.

Choline is a water-soluble vitamin (usually classified as a B vitamin) used to make complex lipids. Phosphatidylcholine is the principal phospholipid of cell membranes. It is also converted to acetyl choline ($CH_3CO_2CH_2CH_2N(CH_3)_3^+$), which is an important neurotransmitter. It carries electrical charges from one nerve cell to another. Choline helps break down of homocysteine, a cardiovascular disease risk factor. A lack of this vitamin leads to fatty livers and/or hemorrhagic kidney disease.

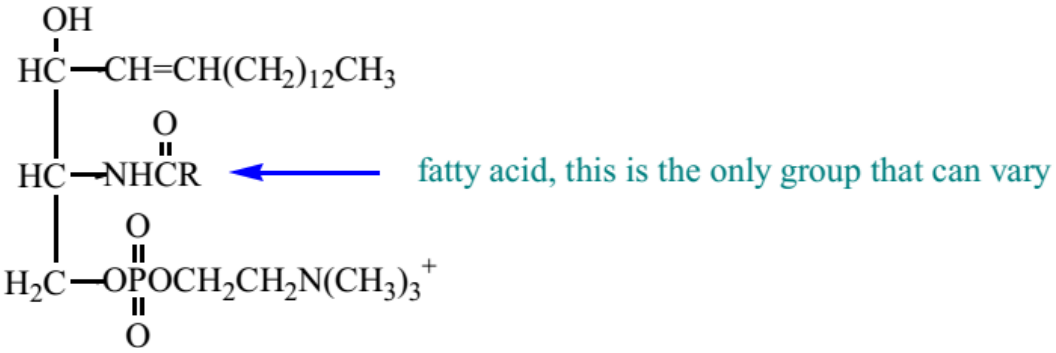
Serine is the parent of a family of amino acids that also includes glycine and cysteine. This enzyme converts serine (as part of phosphatidylserine) to glycine and cysteine. Serine is also involved in the generation of ethanolamine, which is in turn converted to choline. Interestingly, phosphatidylethanolamine is deficient in Alzheimer's patients. They also act as a histamine blocker in the body.



The other subclass of glycerophospholipids is the plasmalogens. These differ from triacylglycerols by even more than the phosphatides. A generic plasmalogen would look like as shown in above diagram

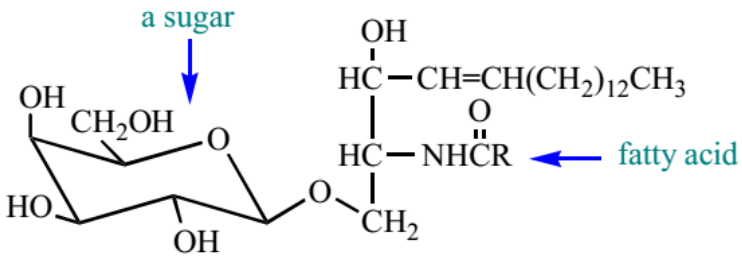
The compound with R = CH_3 is called platelet activating factor. It is a strong bronchoconstrictor. It also stimulates other cells to increase their functional and metabolic activities.

The second major class of phospholipids are the sphingolipids. Sphingolipids include the sphingomyelins and cerebrosides. Both are based on the molecule sphingosine. Sphingomyelins have following basic formula



As the name suggests this lipid is affiliated with the myelin sheath surrounding the cells of the central nervous system. Sphingomyelins comprise about 25% of the lipids in the myelin sheath and their key role is to brain function and electrical transmission through our nervous system.

The other types of sphingolipids concerned are cerebrosides, which are not phospholipids. These compounds are based on attachments to a sphingosine molecule.



These molecules are called glycolipids (glycosides are acetals of sugars). Most of these molecules incorporate β- D-galactose sugars. Cerebrosides are found most commonly in cell membranes in the brain. Cerebroside found outside the brain, is a glucocerebroside, is found in the membranes of macrophages (cells that destroy foreign microorganisms).

Several disorders are associated with malfunctioning of sphingolipid metabolism. The best known is Tay-Sachs disease, which strikes infants and is typically fatal by age 3. Niemann-Pick disease also strikes infants and is fatal early in life. Gaucher's disease and Fabry's disease strike later in life and are generally less devastating.

9.11 Properties of Fats and Oils

Triacylglycerols in natural fats and oils are nonpolar, hydrophobic molecules with no ionic charges.

Oil: A mixture of triacylglycerols that is liquid because it contains a high proportion of unsaturated fatty acids.

Fat: A mixture of triacylglycerols that is solid because it contains a high proportion of saturated fatty acids.

	SATURATED FATTY ACIDS (%)				UNSATURATED FATTY ACIDS (%)	
SOURCE	C_{12} LAURIC	C_{14} MYRISTIC	C_{16} PALMITIC	C_{18} STEARIC	C_{18} OLEIC	C_{18} LINOLEIC
Animal Fat						
Lard	—	1	25	15	50	6
Butter	2	10	25	10	25	5
Human fat	1	3	25	8	46	10
Whale blubber	—	8	12	3	35	10
Vegetable Oil						
Corn	—	1	8	4	46	42
Olive	—	1	5	5	83	7
Peanut	—	—	7	5	60	20
Soybean	—	—	7	4	34	53

The hydrocarbon chains in saturated acids are flexible and uniform in shape, allowing them to nestle together. By contrast, the carbon chains in unsaturated acids have rigid kinks wherever they contain cis double bonds. The kinks make it difficult for such chains to fit next to each other in the orderly fashion necessary to form a solid. The more double bonds there are in a triacylglycerol, the harder it is for it to solidify.

9.12 Biological importance of lipids

Lipids have several roles in the body, these include acting as chemical messengers, storage and provision of energy etc..

Chemical messengers

All multicellular organisms use chemical messengers to send information between organelles and to other cells. Since lipids are small molecules insoluble in water, they are excellent

candidates for signalling. The signalling molecules further attach to the receptors on the cell surface and bring about a change that leads to an action.

The signalling lipids, in their esterified form can infiltrate membranes and are transported to carry signals to other cells. These may bind to certain proteins as well and are inactive until they reach the site of action and encounter the appropriate receptor.

Storage and provision of energy

Storage lipids are triacylglycerols. These are inert and made up of three fatty acids and a glycerol. A free fatty acids are released from triacylglycerols during fasting to provide a source of energy and to form the structural components for cells. Dietary fatty acids of short and medium chain size are not esterified but are oxidized rapidly in tissues as a source of 'fuel". Longer chain fatty acids are esterified first to triacylglycerols or structural lipids.

Maintenance of temperature

Layers of subcutaneous fat under the skin also help in insulation and protection from cold. Maintenance of body temperature is mainly done by brown fat as opposed to white fat. Babies have a higher concentration of brown fat.

Membrane lipid layer formation

Membrane lipids are made of polyunsaturated fatty acids. Polyunsaturated fatty acids are important as constituents of the phospholipids, where they appear to confer several important properties to the membranes. One of the most important properties are fluidity and flexibility of the membrane.

Cholesterol formation

Much of the cholesterol is located in cell membranes. It also occurs in blood in free form as plasma lipoproteins. Lipoproteins are complex aggregates of lipids and proteins that make travel of lipids in a watery or aqueous solution possible and enable their transport throughout the body. Cholesterol maintains the fluidity of membranes by interacting with their complex lipid components, specifically the phospholipids such as phosphatidylcholine and sphingomyelin. Cholesterol is also the precursor of bile acids, vitamin D and steroidal hormones.

Prostaglandin formation and role in inflammation

The essential fatty acids, linoleic and linolenic acids are precursors of many different types of eicosanoids, including the hydroxyeicosatetraenes, prostanoids (prostaglandins, thromboxanes and prostacyclins), leukotrienes (and lipoxins) and resolvins etc. these play an important role in pain, fever, inflammation and blood clotting.

9.13 Summary of the unit

The lipids are a large and diverse group of naturally occurring organic compounds that are related by their solubility in nonpolar organic solvents. The common feature of these lipids is that they are all esters of moderate to long chain fatty acids. Acid or base-catalyzed hydrolysis yields the component fatty acid. Fatty acids are long-chain carboxylic acids. Natural fatty acids may be saturated or unsaturated. The higher melting points of the saturated fatty acids reflect the uniform rod-like shape of their molecules. The cis-double bond(s) in the unsaturated fatty acids introduce a kink in their shape, which makes it more difficult to pack their molecules together in a stable repeating array or crystalline lattice. Two polyunsaturated fatty acids, linoleic and linolenic, are designated "essential" because their absence in the human diet has been associated with health problems, such as scaley skin, stunted growth and increased dehydration. These acids are also precursors to the prostaglandins, a family of physiologically potent lipids present in minute amounts in most body tissues. The triesters of fatty acids with glycerol (1,2,3-trihydroxypropane) compose the class of lipids known as fats and oils. These triglycerides (or triacylglycerols) are found in both plants and animals, and compose one of the major food groups of our diet. Triglycerides that are solid or semisolid at room temperature are classified as fats, and occur predominantly in animals. Those triglycerides that are liquid are called oils and originate chiefly in plants, although triglycerides from fish are also largely oils. As might be expected from the properties of the fatty acids, fats have a predominance of saturated fatty acids, and oils are composed largely of unsaturated acids. Triglycerides having three identical acyl chains, such as tristearin and triolein (above), are called "simple", while those composed of different acyl chains are called "mixed". If the acyl chains at the end hydroxyl groups (1 & 3) of glycerol are different, the center carbon becomes a chiral center and enantiomeric configurations must be recognized. Waxes are esters of fatty acids with long chain monohydric alcohols (one hydroxyl group). Natural waxes are often mixtures of such esters, and may also contain hydrocarbons. Phospholipids are the main constituents of cell membranes. They resemble the triglycerides in being ester or amide derivatives of glycerol or sphingosine with fatty acids and phosphoric acid. The phosphate moiety of the resulting phosphatidic acid is further esterified with ethanolamine, choline or serine in the phospholipid itself. The following diagram shows the structures of some of these components.

9.14 Key words

Lipids; Simple lipids; Complex lipids; Derived lipids; Fats and oils; Fatty acids; Waxes; Triacylglycerols; Phospholipids

9.15 References for further studies

1) Lipid Biochemistry: An Introduction; Michael I. Gurr, Professor John L. Harwood, Keith N. Frayn; *John Wiley & Sons*, **2008**.

2) Lipids; Markus R. Wenk, Gilbert Di Paolo*; Academic Press*, **2012**.

3) Fatty Acid and Lipid Chemistry; F.D. Gunstone; *Springer US*, **2012**.

4) Biochemistry of Lipids, Lipoproteins and Membranes; J.E. Vance, Dennis E. Vance; *Elsevier*, **2008**.

9.16 Questions for self understanding

1) What are lipids?

2) Discuss the classifications of lipids.

3) What are a) Simple lipids, b) Complex lipids, c) Derived Lipids?

4) What are Fats and oils? How they differs each other?

5) What are fatty acids?

6) Explain the classification of fatty acids.

7) What are waxes?

8) Explain the Chemical properties of triacylglycerols

9) What are phospholipids? Write the biological significance of phospholipids.

10) Discuss the properties of fats and oil.

UNIT-10

Structure

10.0 Objectives of the unit

After studying this unit you are able to

- Explain the structural components of terpenes
- Write the classification of terpenes
- Write the structure of isoprene unit
- Write the isoprene rule and gem dialkyl rule
- List out the general properties of terpenoids

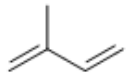
10.1 Introduction

In addition to alkaloids another large group of substances obtained from plants are the essential oils consists of a mixture of hydrocarbons and their oxygenated derivatives. These volatile oils usually have characteristic pleasant odor. Conifer wood, balm trees, citrus fruits, coriander, eucalyptus, lavender, lemon grass, lilies, carnation, caraway, peppermint species, roses, rosemary, sage, thyme, violet and many other plants or parts of those (roots, rhizomes, stems, leaves, blossoms, fruits, seed) are well known for their pleasant smell or spicy taste or exhibiting specific pharmacological activities. This is due to the presence of volatile organic compounds called essential oils. These essential oils are called terpenes. These are widely used in perfumery, as food, as flavourings, as medicines, and as solvents. The examples of typical essential oils are those obtained from cloves, roses, lavender, citronella, eucalyptus, peppermint, camphor, sandalwood, cedar, and turpentine. The components of the essential oils can be regarded as derived from isoprene

10.2 Terpenes

Terpenes are a class of hydrocarbons formed from combinations of isoprene units or their oxygenated derivatives. The hydrocarbon derivaties are called terpenes and oxygenated derivatives are called terpenoids. The term terpene originates from turpentine (lat. balsamum terebinthinae). Turpentine, the "resin of pine trees", is the viscous pleasantly smelling balsam which flows upon cutting or carving the bark and the new wood of several pine tree species (Pinaceae). Turpentine contains the "resin acids" and some hydrocarbons, which were originally referred to as terpenes. Traditionally, all natural compounds built up from isoprene subunits and for the most part originating from plants are denoted as terpenes,

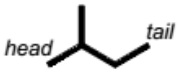
At present about 30000 terpenes are known. *Their basic structure is build up by the multiples of carbon skeleton 2-Methylbutane residues, usually referred as isoprene units* $(C_5)_n$. *Therefore, terpenes are also denoted as isoprenoids.*



2-Methyl-1,3-butadiene
(Isoprene)

In nature, terpenes occur predominantly as hydrocarbons, alcohols and their glycosides, ethers, aldehydes, ketones, carboxylic acids and esters.

The isopropyl part of 2-methylbutane is defined as the head, and the ethyl residue as the tail



2-Methylbutane

10.3 Classification of terpenes

Based on the number of 2-methylbutane (isoprene) subunits presnt, the terpenes are calcified as

Hemi terpenes (C_5) → one isoprene unit,

Mono terpenes (C_{10}) → two isoprene units,

Sesqui terpenes (C_{15}) → three isoprene units,

Di terpenes (C_{20}) → four isoprene units,

Sester terpenes (C_{25}) → five isoprene units,

Tri terpenes (C_{30}) → six isoprene units,

Tetraterpenes (C_{40}) → eight isoprene units and

Polyterpenes $(C_5)_n$ with n > 8 units

Each class can be further subdivided into subclasses according to the number of rings present in the structure.

Acyclic Terpenoids: They contain open structure.

Monocyclic Terpenoids: They contain one ring in the structure.

Bicyclic Terpenoids: They contain two rings in the structure.

Tricyclic Terpenoids: They contain three rings in the structure.

Tetracyclic Terpenoids: They contain four rings in the structure.

10.3.1 Monoterpenes

Monoterpenes are a class of terpenes that consist of two isoprene units and have the molecular formula $C_{10}H_{16}$. Monoterpenes may be linear (acyclic) or contain rings. Many monoterpenes are readily recognized by their characteristic flavors or odors (limonene in lemons; citronellal in roses and perfumes; menthol used in cough drops)

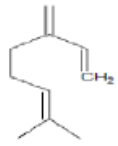
Monoterpens are further divided in to

i) Acyclic Monoterpenoids

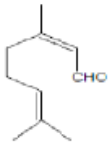
ii) Monocyclic monoterpenoids

iii) Bicyclic monoterpenoids

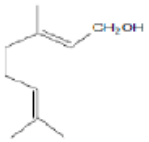
i) Acyclic Monoterpenoids



Myrcene

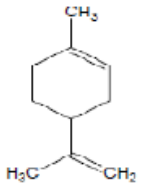


Citral

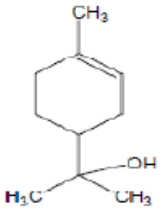


Geraniol

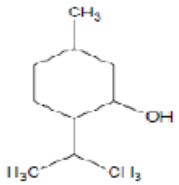
ii) Monocyclic monoterpenoids



Limonene



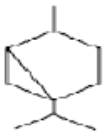
α-Terpineol



Menthol

iii) Bicyclic monoterpenoids: These are further divided into three classes.

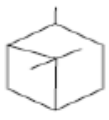
a) Containing -6+3-membered rings



Thujane

Carane

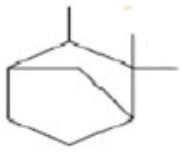
b) Containing -6+4- membered rings



Pinane

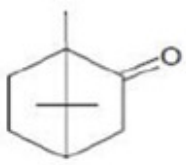
c) Containing -6+5-membered rings

Bornane (Camphane)



non bornane (iso camphane)

Some bicyclic monoterpenes are



Camphor

α-pinene

10.3.2 Sesquiterpenes

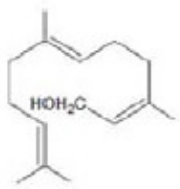
A class of terpenes that consist of three isoprene units are called sesquiterpenes and they have empirical formula $C_{15}H_{24}$. Like monoterpenes, sesquiterpenes are also exists as acyclic or contain rings, with many unique combinations.

Sesquiterpenoids are further divided in to

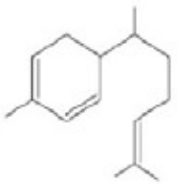
i) Acyclic sesquiterpenoids

ii) Monocyclic sesquiterpenoids

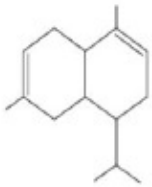
iii) Bicyclic sesquiterpenoids



Farnesol



Zinziberene

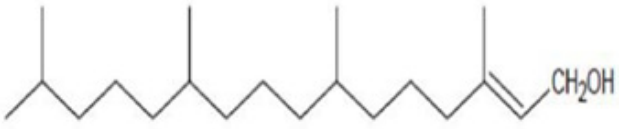


Cadinene

10.3.3 Diterpenes

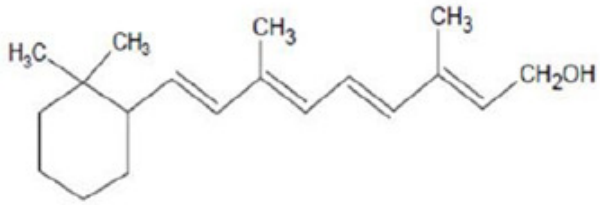
Diterpenes are a class of chemical compounds composed of of four isoprene units with the molecular formula $C_{20}H_{32}$. Diterpenoids are structurally highly complex, but their biosynthetic routes are relatively simple. Diterpenoids with their twenty-carbon backbone constitute roughly a fourth of all known plant terpenoids. Diterpenes include retinal (the visual pigment in rhodopsin), and phytol (found in chlorophyll). Gibberellic acid is a plant hormone.

i) Acyclic diterpenoids



Phytol

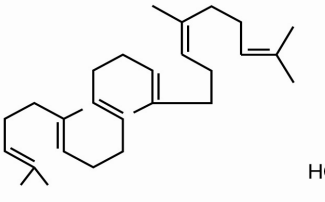
ii) Mono cyclic diterpenoids



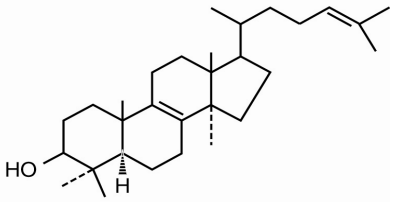
Vitamin A

The triterpene lanosterol is a constituent of wool fat and is also a precursor to cholesterol and the other steroids. Lycopene is a carotenoid found in ripe fruit, especially tomatoes.

TRITERPENES

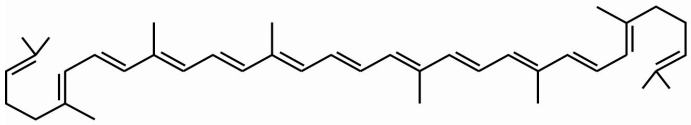


Squalene



Lanosterol

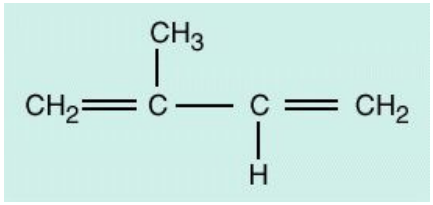
TETRATERPENES

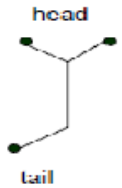


Lycopene

10.4 Isoprene unit

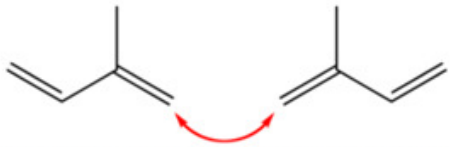
The five-carbon unit 2-Methyl-1,3-butadiene is an unsaturated hydrocarbon with a branched chain, which in the plant and animal kingdom is used as the basis for the formation of isoprenoids



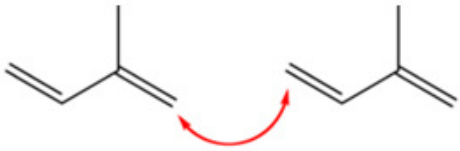


The branched chain double bond side is called head and the unbranche bouble bond side is called tail. Formally in biosynthesis of terpenes, two or more isoprene molecules are linked to one another. Linking between two isoprene molecules could occur in three ways, given that the head and the tail of the molecule are primarily involved in the linking:

Type-1. The head of one isoprene molecule could link with the head of another isoprene molecule.



Type-2. The head of one isoprene molecule could link with the tail of another isoprene molecule.

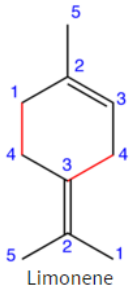


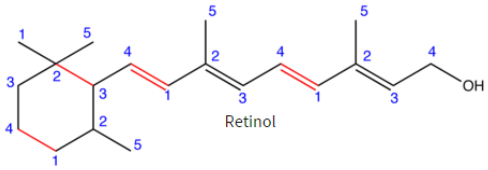
Type-3. The tail of one isoprene molecule could link with the tail of another isoprene molecule.

Cyclic terpenes also contain links that are neither 1-1, 1-4, nor 4-4, which are called crosslinks.

10.5 Isoprene rule

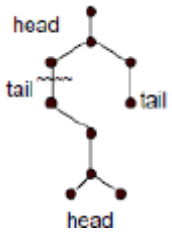
Ingold suggested that isoprene units are joined in the terpenoid via 'head to tail' fashion. Isoprene rule states that, in most naturally occurring terpenes, there are number of 1-1 or 4-4 links. Iisoprene rule states that the terpenoid molecule is constructed of two or more isoprene units joined in a 'head to tail' fashion. Ex

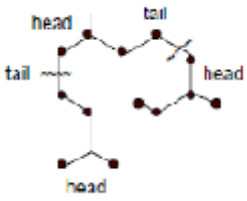




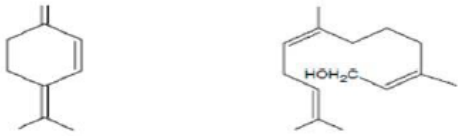
In applying isoprene rule we look only for the skeletal unit of carbon. The carbon skeletons of open chain monoterpenoids and sesquiterpenoids are,

.



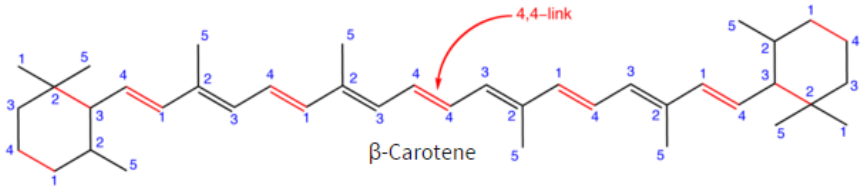


Examples

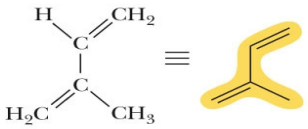


Myrcene (monoterpene) **Farnesol (Sesquiterpene)**

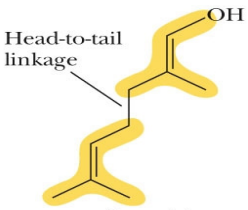
Isoprene rule can only be used as guiding principle and not as a fixed rule. For example carotenoids are joined tail to tail at their central and there are also some terpenoids whose carbon content is not a multiple of five. A terpene that does not obey the isoprene rule is called an irregular terpene.



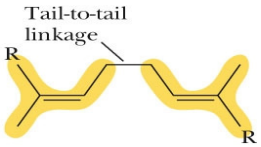
In mono-, sesqui-, di- and sester terpenes the isoprene units are linked to each other from head-to-tail fusion. Whereas tri- and tetra terpenes the one isoprene unit linked from tail-to-tail connection in the centre.



Isoprene

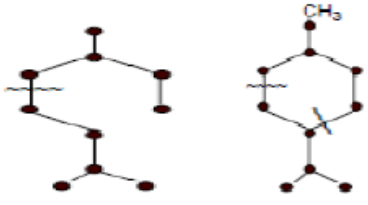


Geraniol



10.6 Gem dialkyl rule

Ingold pointed that a gem alkyl group affects the stability of terpenoids. He summarized these results in the form of a rule called ‘gem dialkyl rule’ which stated as Gem dialkyl group tends to render the cyclohexane ring unstable where as it stabilizes the three, four and five member rings.” This rule limits the number of possible structure in closing the open chain to ring structure. Thus the monoterpenoid open chain gives rise to only one possibility for a monocyclic monoterpenoid i.e the p-cymene structure.



P-cymene structure

Bicyclic monoterpenoids contain a six member and a three member ring. Thus closure of the ten carbon open chain monoterpenoid gives three possible bicyclic structures.

Camphor(6+5) system | Pinane(6+4) system | Carane(6+3) System

10.7 Isolation of Mono and Sesquiterpenoids

Both mono and sesquiterpenoids have common source i.e essential oils. Their isolation is carried out in two steps

ii) Isolation of essential oils from plant parts

ii) Separation of Terpenoids from essential oils.

ii) *Isolation of essential oils from plant parts*

The plants having essential oils generally have the highest concentration at some particular time. Therefore better yield of essential oil plant material have to be collected at this particular time. E.g. From jasmine at sunset. There are four methods of extractions of oils.

a) Expression method

b) Steam distillation method

c) Extraction by means of volatile solvents

d) Adsorption in purified fats

Steam distillation is most widely used method. In this method macerated plant material is steam distilled to get essential oils into the distillate form these are extracted by using pure organic volatile solvents. If compound decomposes during steam distillation, it may be extracted with ether at 50^{o}C. After extraction solvent is removed under reduced pressure.

ii) Separation of Terpenoids from essential oil

A number of terpenoids are present in essential oil obtained from the extraction. Definite physical and chemical methods can be used for the separation of terpenoids. They are separated by fractional distillation. The terpenoid hydrocarbons distill over first followed by the oxygenated derivatives. More recently different chromatographic techniques have been used both for isolation and separation of terpenoids.

10.8 General properties of terpenoids

1. Most of the terpenoids are colorless, fragrant liquids which are lighter than water and volatile with steam. A few of them are solids e.g. camphor. All are soluble in organic solvent and usually insoluble in water. Most of them are optically active.

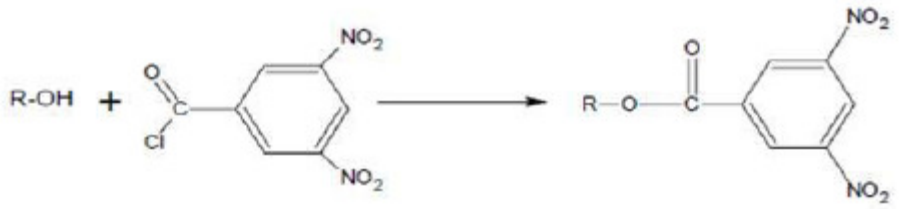
2. They are open chain or cyclic unsaturated compounds having one or more double bonds. Consequently they undergo addition reaction with hydrogen, halogen, acids, etc. A number of addition products have antiseptic properties.
3. They undergo polymerization and dehydrogenation
4. They are easily oxidized nearly by all the oxidizing agents. On thermal decomposition, most of the terpenoids yields isoprene as one of the product.

10.9 General methods of structure elucidation of terpenoids

i) Molecular formula: molecular formula is determined by usual quantitative analysis and mol.wt determination methods and by means of mass spectrometry. If terpenoid is optically active, its specific rotation can be measured.

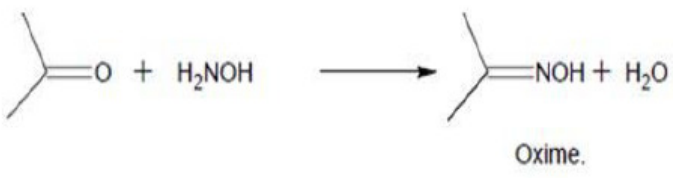
ii) Nature of oxygen atom present: If oxygen is present in terpenoids its functional nature is generally as alcohol, aldehyde, ketone or carboxylic groups.

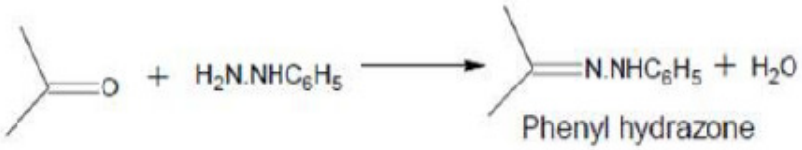
a) Presence of oxygen atom: presence of –OH group can be determined by the formation of acetates with acetic anhydride and benzoyate with 3.5-dinitirobenzoyl chloride.

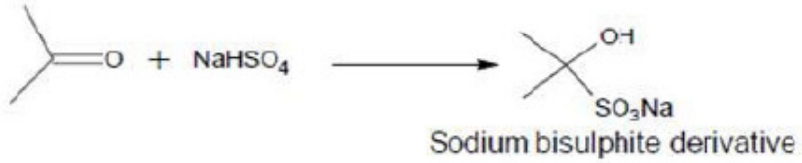


Primary alcoholic group undergo esterification more readily than secondary and tertiary alcohols.

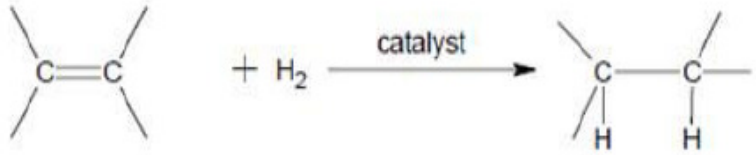
Presence of >C=O group: Terpenoids containing carbonyl function form crystalline addition products like oxime, phenyl hydrazone and bisulphite etc. If carbonyl function is in the form of aldehyde it gives carboxylic acid on oxidation without loss of any carbon atom whereas the ketone on oxidation yields a mixture of lesser number of carbon atoms.







iii) Unsaturation: The presence of olefinic double bond is confirmed by means of bromine, and number of double bonds present in the molecule is determination by analysis of the bromide or by quantitative hydrogenation or by titration with monoperpthalic acid. Presence of double bond also confirmed by means of catalytic hydrogenation or addition of halogen acids. Number of moles of HX absorbed by one molecule is equal to number of double l



Addition of nitrosyl chloride (NOCl) (Tilden's reagent) and epoxide formation with peracid also gives idea about double bonds present in terpenoid molecule.

iv) Dehydrogenation: On dehydrogenation with sulphur, selenium, polonium or palladium terpenoids converted to aromatic compounds. Examination of these products the skelton

structure and position of side chain in the original terpenoids can be determined. For example α-terpenol on Se-dehydrogenation yields pcymene.

Thus the carbon Skelton of terpenol is as follows.

v) *Oxidative degradation:* Oxidative degradation has been the parallel tool for elucidating the structure of terpenoids. Reagents for degradative oxidation are ozone, acid, neutral or alkaline potassium permanganate, chromic acid, sodium hypobromide, osmium tetroxide, nitric acid, lead tetra acetate and peroxy acids. Since oxidizing agents are selective, depending on a particular group to be oxidized, the oxidizing agent is chosen with the help of structure of degradation products.

10.10 Determination of relative stereochemistry

The problems associated with establishing the relative stereochemistry of the terpenoids can be divided into those relating to the stereochemistry of the substituent's and those associated with the stereochemistry of the ring junctions. The assignment of the stereochemistry of the menthols **(a)** utilized the comparative rates of their esterification with *p*-nitrobenzoyl chloride to determine the configuration of the alcohols at C-3. The two neomenthols were esterified more slowly than either menthol or isomenthol. This was attributed to steric hindrance by the isopropyl group and hence this group was placed *cis* to the hydroxy in the neomenthols. The much slower rate of hydrolysis of methyl podocarpate (b) compared to methyl dehydroabietate (c) was used to reveal the greater steric hindrance in the former arising from interactions with the methyl group at C-10. The ester is an axial substituent in methylpodocarpate.

10.11 Absolute stereochemistry of terpenoid secondary alcohols

A number of methods have been developed for establishing the absolute stereochemistry of a secondary alcohol. Studies on the addition of Grignard reagents to the α-keto-esters of optically active terpenoid alcohols showed that asymmetric induction could be used to

establish the stereochemistry of a chiral alcohol. The chirality of the atrolactic acid that was formed reflected that of the original secondary alcohol. Horeau's method was based on the kinetic resolution by the alcohol in its esterification by 2-phenylbutyric anhydride. The method is based on using a racemic mixture of the anhydride. One enantiomer will react faster with the chiral secondary alcohol than the other. The optical activity of the remaining 2-phenylbutyric acid will thus reflect the chirality of the secondary alcohol. An NMR method based on derivatization with (R)- and (S)- α-methoxy-α-trifluoromethylphenylacetic acid (MTPA) has become a popular strategy for establishing the absolute stereochemistry of terpenoid alcohols. Although it was originally based on an analysis of the ^{19}F NMR spectrum, a more reliable version uses the high field ^{1}H NMR spectrum. A correlation was established between the absolute stereochemistry of terpenoid secondary alcohols and the $\Delta\delta$ (δS-δR) values of adjacent protons for the (R)- and (S)-MPTA esters. This correlation was successfully applied in establishing the absolute stereochemistry of a number of marine terpenoids.

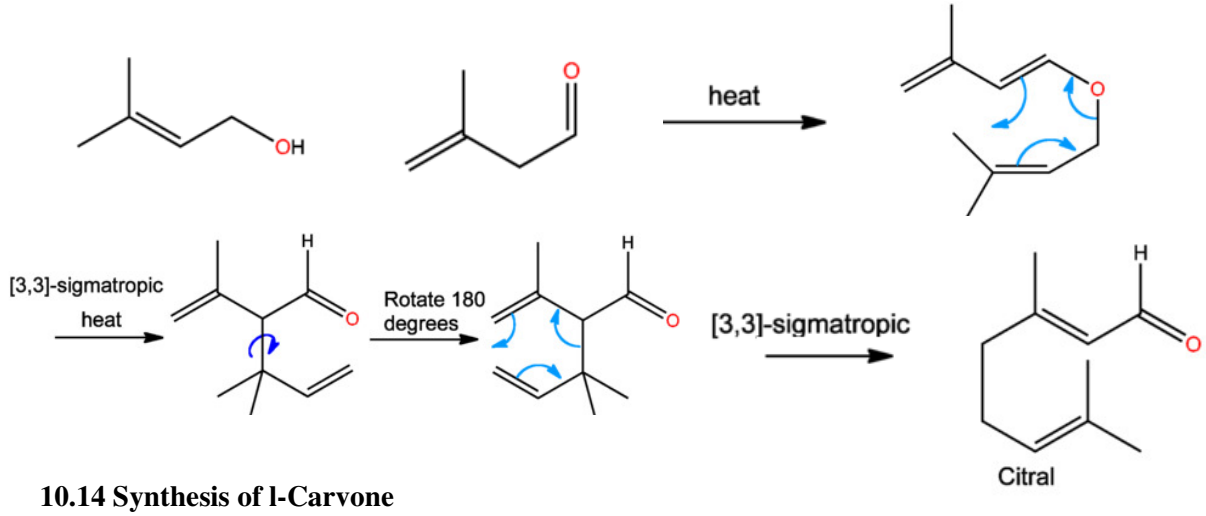
10.12 Biological application of terpenes

Terpenes have many biological applications also they have been used for many purposes in human societies. Pharmaceutical and food industries have exploited them for their potentials and effectiveness as medicines and flavour enhancers. The most widely known terpene is rubber, which has been used extensively by humans. Rubber is a polyterpene, composed of repeating subunits of isoprene. The addition of sulfur to rubber led to vulcanized rubber, which yields various degrees of pliability depending on the mixture ratio. Terpenes are also used as insecticides, cleaners, antiallergenic agents.

10.13 Synthesis of citral

Citral, 3,7-dimethyl-2,6-octadien-1-al , is found in a large number of essential oils, and occurs to the extent of 80% in lemon grass oil. It is the most important of the acyclic monoterpenes, and finds much industrial application. Hence, many synthetic methods of Citral have been reported to date. Citral is a compound which can be used in the synthesis of vitamin A. The industrial synthesis of this molecule is done by two different [3,3]-sigmatropic rearrangements in sequence.

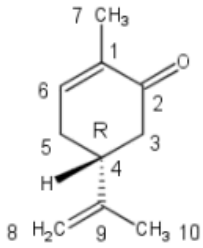
The first step is a Claisen rearrangement which involves a oxygen in the chair-like transition state. The second part is a Cope-rearrangement, in which only carbons are involved in the chair-like transition state. These two reactions happen in succession when the two starting materials are heated together, and the reaction is driven by formation of a conjugated carbonyl group in the product.



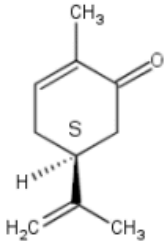
10.14 Synthesis of l-Carvone

l-Carvone is a naturally-occurring organic compound commonly found in spearmint oil extracted from mint. Spearmint oil is a complex mixture of terpene compounds, typically being comprised of 60–70% l-(–)-carvone [4*R*-(–)-carvone]. The other carvone enantiomer , d-(+)-carvone [4*S*-(+)-carvone], is extracted from the oils of caraway and dill seeds . Both carvone isomers are a substituted cyclohexenone , p-mentha-6,8-dien-2-one . The C4 atom is an asymmetric point center and causes carvone to exhibit optical activity

Absolute Configurations of *ℓ*-(-)-Carvone and *d*-(+)-Carvone



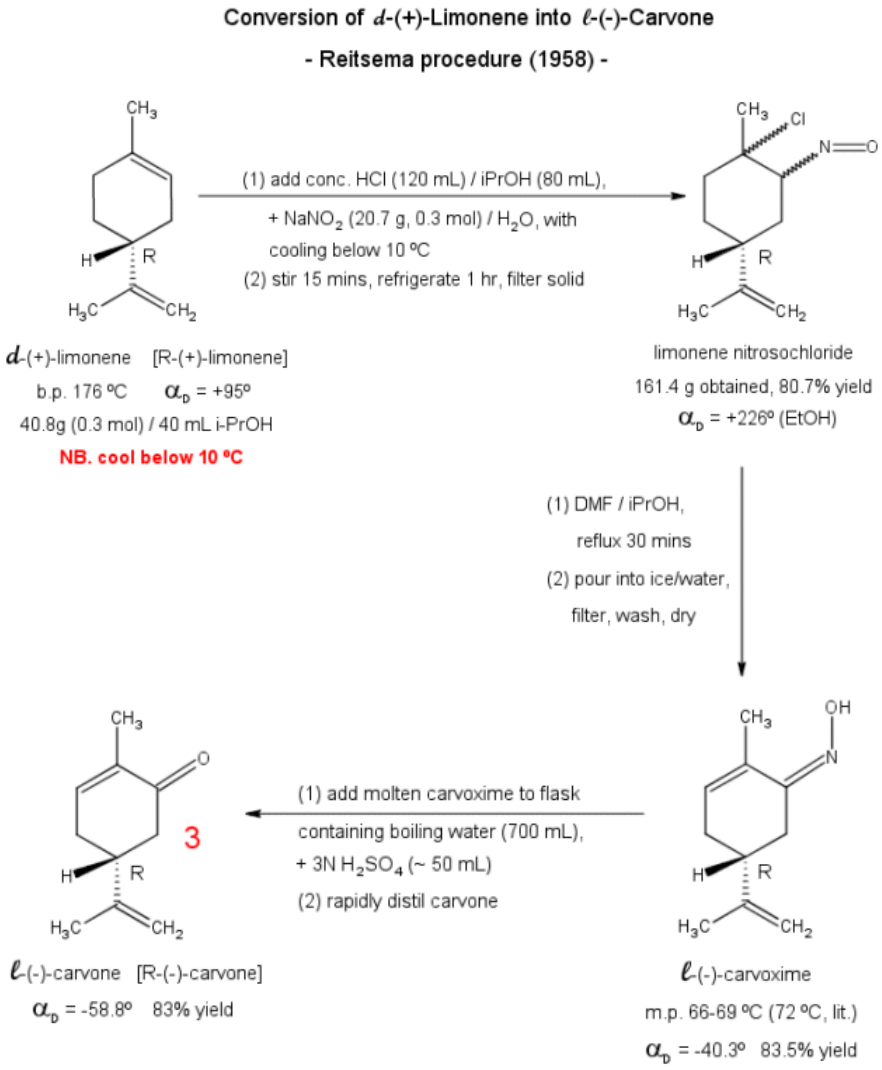
***ℓ*-(-)-carvone** [4R-(-)-carvone]
(p-mentha-6,8-dien-2-one)
occurs naturally in oil of spearmint (60-70%)
from mentha spicata and mentha viridis
b.p. 230 °C (88-90 °C @ 4 mm) ;
α_D^{20} = - 62.46°



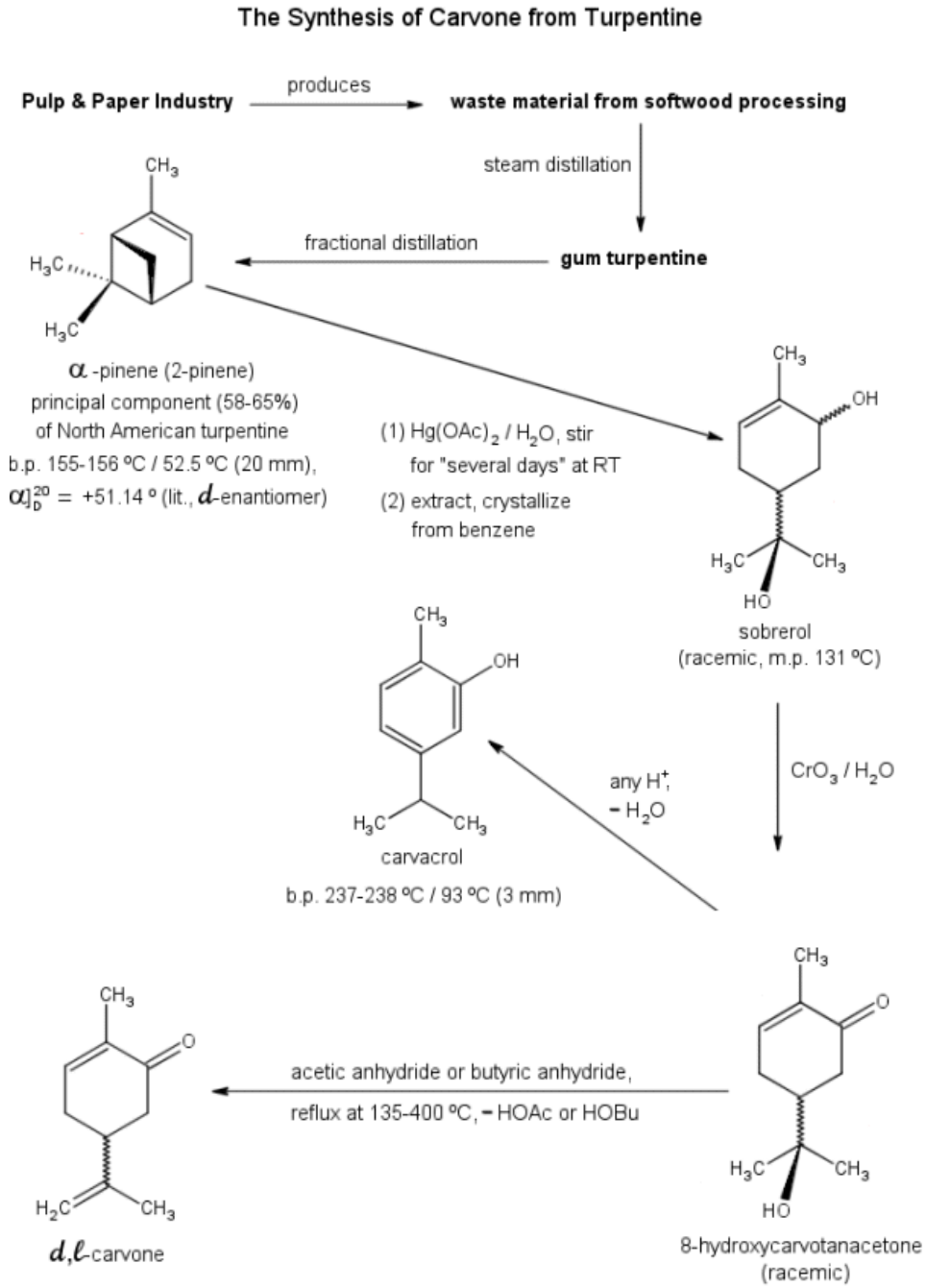
***d*-(+)-carvone** [4S-(+)-carvone]
occurs naturally in oil of caraway
seed and oil of dill seed
b.p. 230 °C (91 °C @ 5 mm) ;
α_D^{20} = + 61.2°

10.14.1 Commercial Production of l-Carvone

The citrus industry produces large quantities of waste (orange peels) from the processing of fruit into juices. This waste can be extracted with a hydrocarbon solvent such as naphtha to obtain substantial amounts of the monoterpene d-(+)-limonene , which can be converted into l-(–)-carvone via its nitrosochloride addition compound as shown below.

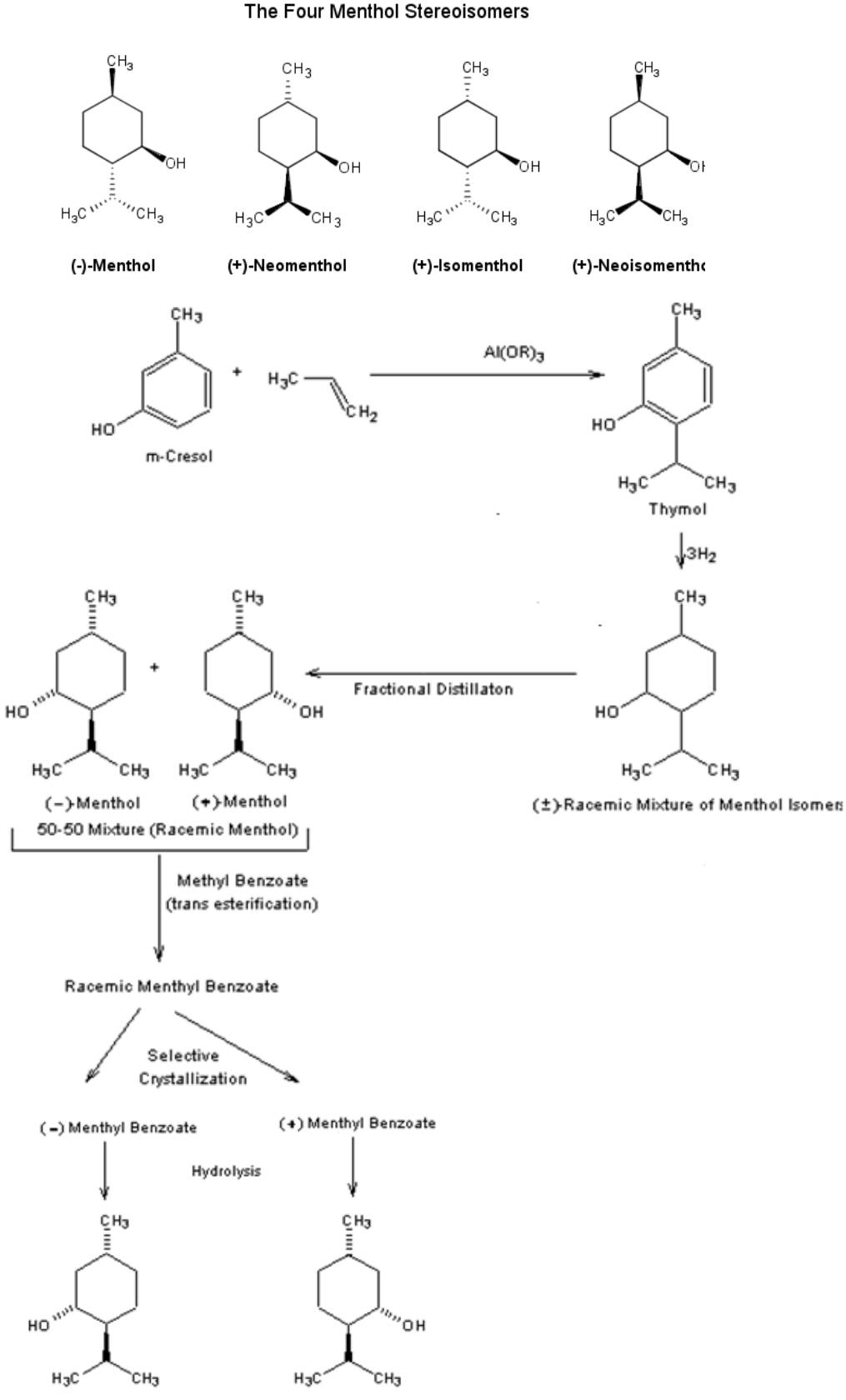


The pulp and paper industry produces large quantities of gum turpentine from the processing of deciduous (softwood) stock and logging waste by steam distillation. The principal monoterpene component present in this waste is α-pinene (2-pinene), which can be converted into sobrerol by treatment with aqueous mercuric acetate. Oxidation of sobrerol produces 8-hydroxycarvotanacetone. This can be converted to carvone by refluxing it with acetic or butyric anhydride (dehydration).

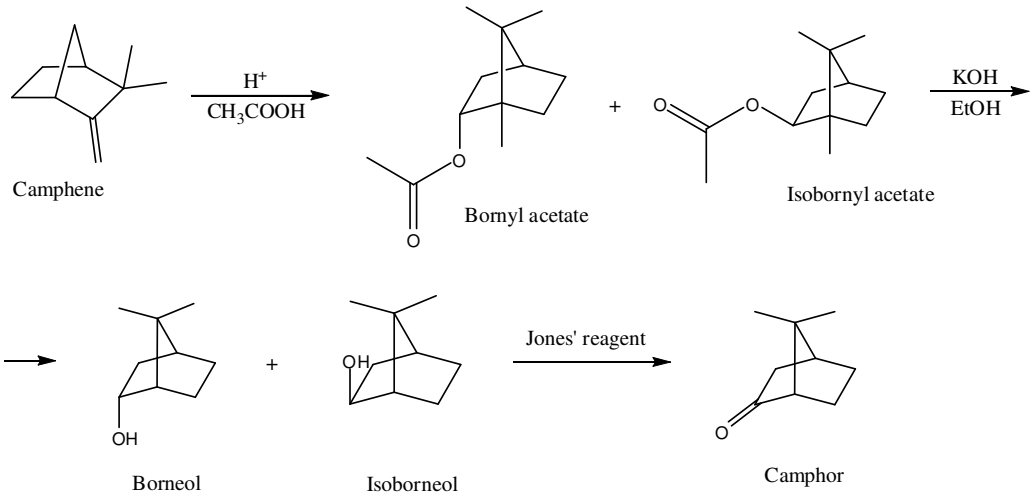


10.15 Synthesis of Menthol

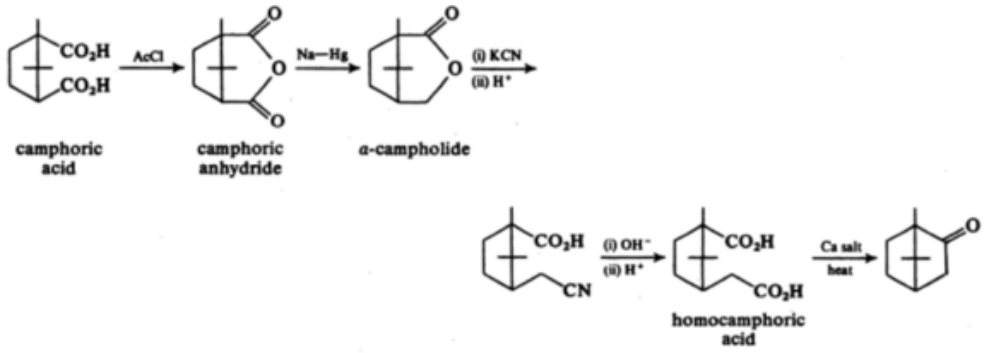
Menthol is also called 2-isopropyl-5-methyl-cyclohexanol is amongst the most important perfume/flavour chemical, extensively used in pharmaceuticals, cosmetics, toothpastes, chewing gums and toiletries. There are four stereoisomers of 2-isopropyl-5-methyl-cyclohexanol they are i) menthol itself, ii) isomenthol, iii) neomenthol and iv) neoisomenthol.


The Four Menthol Stereoisomers
(-)-Menthol
(+)-Neomenthol
(+)-Isomenthol
(+)-Neoisomenthc
m-Cresol
Al(OR)3
Thymol
3H2
Fractional Distillaton
(−)-Menthol
(+)-Menthol
50-50 Mixture (Racemic Menthol)
(±)-Racemic Mixture of Menthol Isomers
Methyl Benzoate
(trans esterification)
Racemic Menthyl Benzoate
Selective
Crystallization
(−) Menthyl Benzoate
(+) Menthyl Benzoate
Hydrolysis

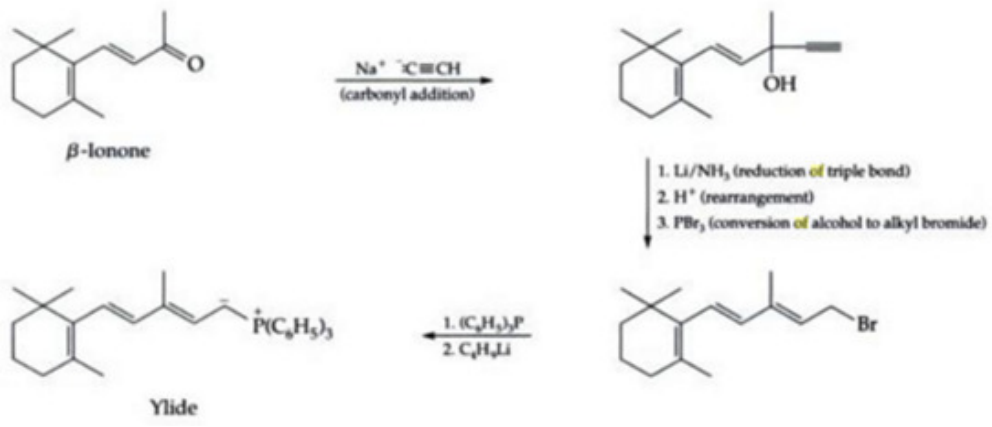
10.16 Synthesis of camphor

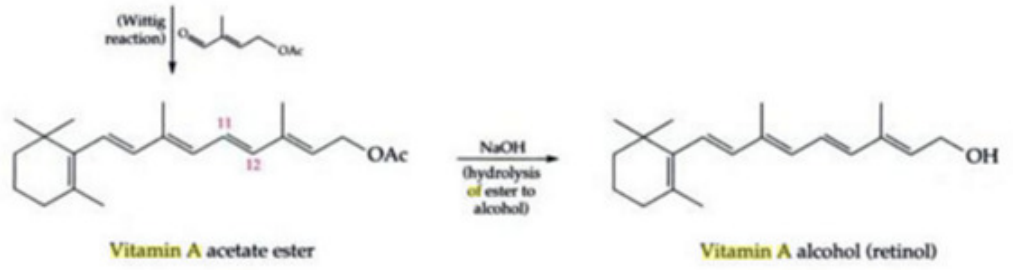


Synthesis of camphor from camphoric acid was reported by Haller et.al, in 1896. The reaction scheme they used is given below



10.17 Synthesis of vitamin A





10.18 Summary of the unit

By the modern definition "Terpenoids are the hydrocarbons of plant origin of the general formula $(C_5H_8)_n$ as well as their oxygenated, hydrogenated and dehydrogenated derivatives." Thermal decomposition of terpenoids gives isoprene as one of the product. Otto Wallach pointed out that terpenoids can be built up of isoprene unit. Isoprene rule states that the terpenoid molecules are constructed from two or more isoprene unit. Most natural terpenoids hydrocarbon have the general formula $(C_5H_8)_n$. They can be classified on the basis of value of n or number of carbon atoms present in the structure. Plant terpenoids are used extensively for their aromatic qualities. They play a role in traditional herbal remedies and are under investigation for antibacterial, antineoplastics, and other pharmaceutical functions. Plants do not only accumulate terpenes for herbivore defense, but also emit volatile blends in response to herbivory and many other biotic and abiotic stresses. These terpenecontaining volatiles attract natural enemies of the attacking herbivores but due to the complexity of these volatile blends, it is difficult to attribute a specific function to a particular terpene. The importance of terpenes in both nature and human application is difficult to overstate. Basic knowledge of terpene and isoprene biosynthesis and chemistry has accelerated the pace at which scientists have come to understand many plant biochemical and metabolic processes.

10.19 Key words

Terpenes; Monoterpenes; Sesquiterpenes; Diterpenes; Isoprene unit; Isoprene rule; Gem dialkyl rule; Citral; l-Carvone; Menthol; Camphor; Vitamin A

10.20 References for further studies

1) Organic Chemistry, Volume 2: Stereochemistry and The Chemistry Natural Products, 5/E Volume 2 of Organic Chemistry; I. L. Finar; *Pearson Education India*, **1956**.
2) Vitamins; Wilhelm Friedrich; *Walter de Gruyter*, **1988**.
3) Terpenes: Flavors, Fragrances, Pharmaca, Pheromones; Eberhard Breitmaier; John Wiley & Sons, **2006**.
4) Mechanism and Synthesis; Peter Taylor; *Royal Society of Chemistry*, **2002**.
5) Terpenoids; V.K Ahluwalia; Ane Book Pvt. Ltd. **2009**.

10.21 Questions for self understanding

1) What are terpenes

2) Discuss the classification of terpenes

3) What are monoterpenes? Give two examples.

4) What are sesquiterpenes? Give two examples.

5) What are diterpenes? Give two examples.

6) What is isoprene unit?

7) Write the isoprene rule.

8) Write the Gem dialkyl rule

9) Discuss about method of isolation of mono and sesquiterpenoids

10) What are the general properties of terpenoids

11) Discuss briefly about general methods of structure elucidation of terpenoids

12) Write a note on determination of relative stereochemistry in terpenes

13) How absolute stereochemistry of terpenoid secondary alcohols can be determined?

14) Discuss few biological applications of terpenes

15) Write the method of synthesis of citral

16) How l-Carvone is synthesized chemically?

17) Discuss the chemical synthesis of Menthol

18) Discuss the chemical synthesis of camphor

19) Discuss the chemical synthesis of vitamin A

UNIT-11

Structure

11.0 Objectives of the unit

After studying this unit you are able to

- Identify the different sources of steroids
- Explain the configuration of steroid nucleus
- List out the types of steroids
- Explain the structure and biological importance of cholesterol
- Explain the structure and biological importance of Ergosterol
- Write the Irradiation products of ergosterol
- Explain the structure and biological importance of Bile acids

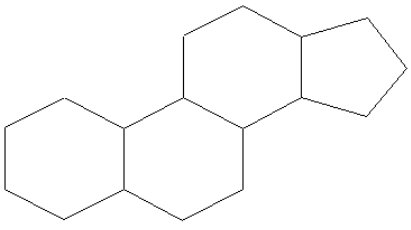
11.1 Introduction

Steroids form an important group of compounds based on the fundamental saturated tetracyclic hydrocarbon 1,2-cyclopentanoperhydrophenanthrene (sterane or gonane). This nucleus, partially or completely hydrogenated, is generally substituted by methyl groups at C_{10} and C_{13}. A chemical group (ketone, hydroxyl...) or an alkyl side-chain may also be present at C_{17}. Steroids may possess a nucleus derived from sterane by one or more C-C bond scissions or ring expansions or contractions.

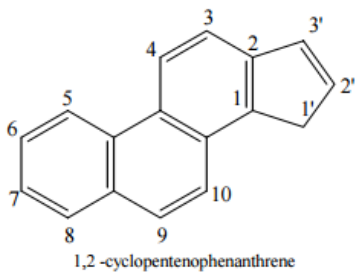
The term "steroids" was first introduced by Callow R. K et al. "for the group of compounds comprising the sterols, bile acids, heart poisons, saponins, and sex hormones". Natural steroids are derived from squalene by cyclization, unsaturation and substitution, they may be considered as modified triterpenes. Fatty acid esters of steroids are found mainly in the blood but their exact role does not completely understood yet.

11.2 Steroids

Steroids are class of natural or synthetic organic compounds characterized by a molecular structure of 17 carbon atoms arranged in four rings. This parent structure is named as gonane (also known as the steroid nucleus), may be modified in a practically unlimited number of ways by removal, replacement, or addition of a few atoms at a time. The structures of steroids are based on the 1, 2-cyclopentenophenanthrene skeleton

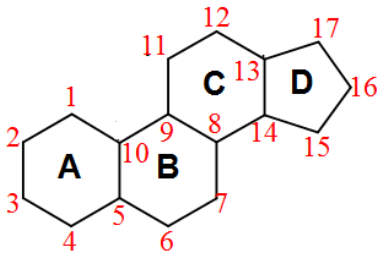


Gonane

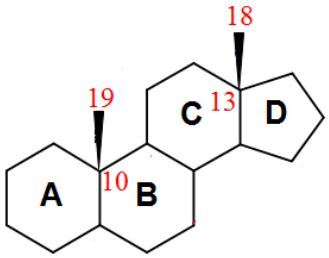


1,2 -cyclopentenophenanthrene

Rings are labeled A, B, C and D. The ring carbon atom of steroid skeletal nucleus is numbered as follows



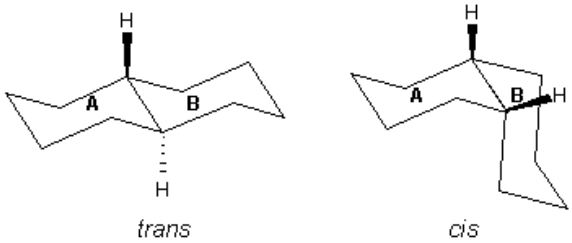
Natural steroids have two methyls at C_{10} and C_{13}. These carbons are numbered as 19 and 18 respectively. A/B rings have carbon 19, C/D rings have carbon 18.



In addition to 2 methyl groups a side chain is frequently found at C_{17}.

11.3 Configuration of Steroid nucleus

There are four rings in a steroid skeleton and hence there are three fusion points. A/B, B/C and C/D rings share two carbons each. Substituent located on the same side of the molecule as the angular methyl groups are assigned β-configuration. Substituents on the opposite side are assigned α-configuration. For two-ring system, the structures of cis - and trans-fused rings look like this

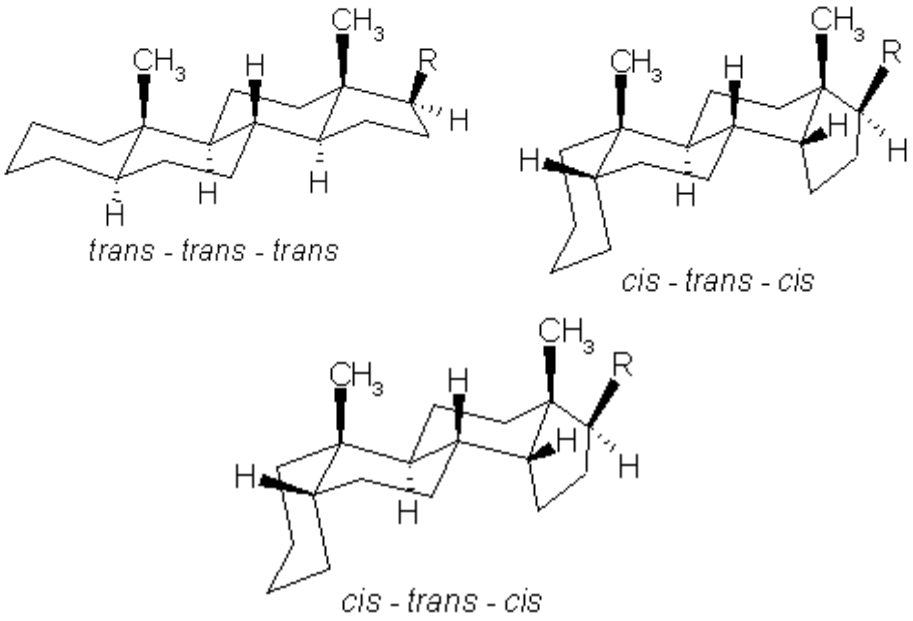


The three fusion centres, The structures most likely feasible are

i) *trans-trans-trans* (most natural and synthetic steroids have this skeleton, e.g., 5α-dihydrotestosterone)

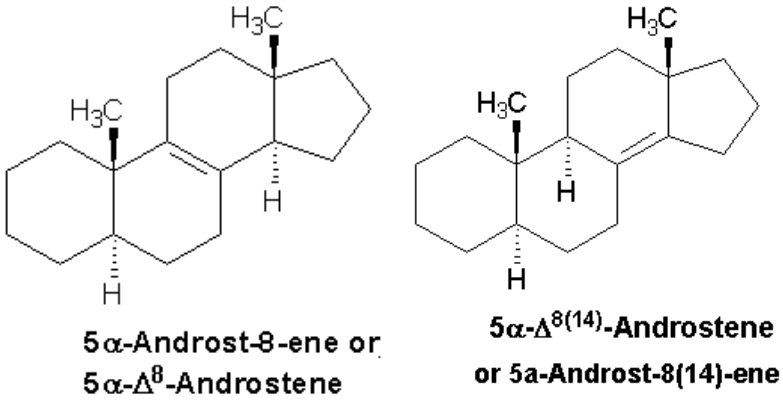
ii) *cis-trans-trans* (some natural steroids have this skeleton, e.g., cholic acids)

iii) *cis-trans-cis* (few natural steroids have this skeleton, e.g. cardiac glycosides)



In steroid nomenclature substituents (and hydrogens) which are below the plane of the rings are designated as α (alpha) and those which are above the plane of the rings are designated as β (beta). The stereochemistry of the hydrogen at C-5 (if present) must always be included in the name of the steroid (eg. 5α-estrane) since the geometry at this ring fusion determines the shape of the steroid.

Example



11.4 Types of steroids

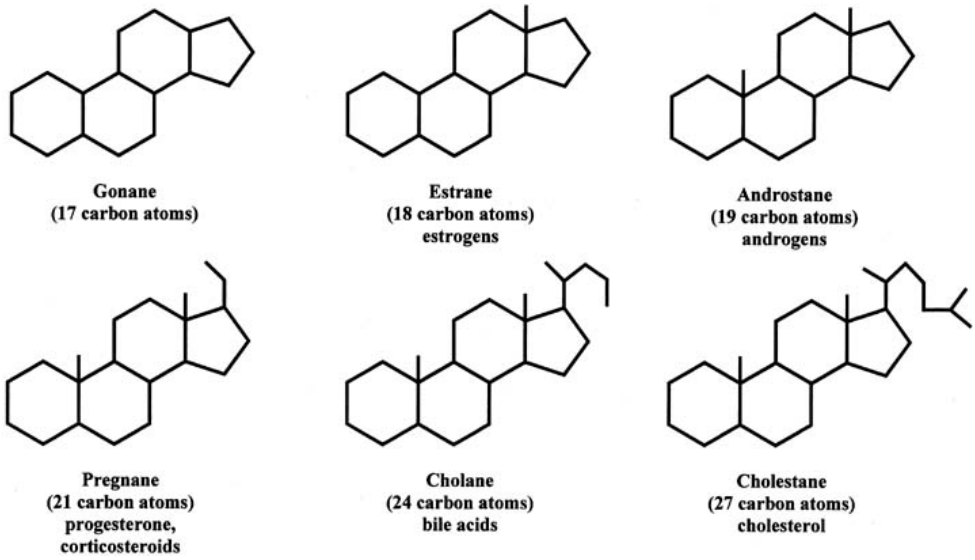
Steroids are divided into different groups of parent compounds, based on the number of carbons that they contain. In addition to gonanes, which consist of 17 carbons,

estranes consist of 18 carbons (C18 steroids) and include estrogens.

Androstanes have 19 carbons (C19 steroids) and include androgens.

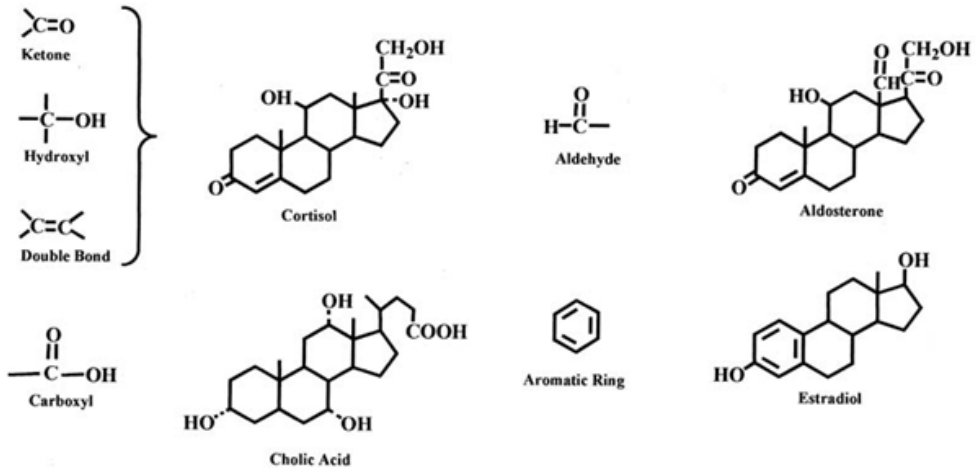
Pregnanes contain 21 carbons (C21 steroids) and include progesterone and corticosteroids. *Cholanes* have 24 carbons and include bile acids, and

cholestanes have 27 carbons and include cholesterol as well as cholesterol-like compounds. The compounds in this group are also referred to as sterols.



Classification of steroids, based on the number of carbons in the molecule

In each group of parent steroids, compounds differ in their characteristics because of the presence of different functional groups on the molecules. Common functional groups presents are the ketone group, hydroxyl group, and double bond. Other functional groups rarely presents are the carboxyl and aldehyde groups, which are present in the molecules of bile acids and aldosterone respectively. An important characteristic of the C18 steroids is the presence of an aromatic ring that is found in estrogens



Functional groups present in chemical structures of steroids

Based on the function steroids are classified in to

i) Anabolic Steroids: These are interact with androgen receptor and enhance muscle mass

ii) Glucocorticoids: These regulate metabolism and immune function; anti-inflammatory activity

iii) Mineralocorticoids: These maintain blood volume and renal excretion

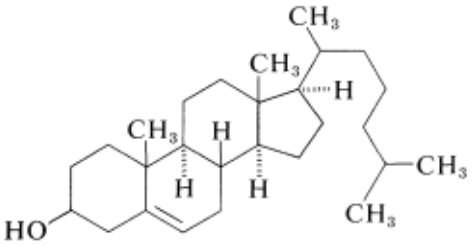
iv) Progestins: These involve in development of female sex organs and characteristics

v) Phytosteroids: These are Plant steroids

vi) Ergosteroids: These are Steroids of the fungi; vitamin D related

11.5 Structure and biological importance of cholesterol

Cholesterol has a molecular formula of $C_{27}H_{45}OH$. This molecule is composed of three regions: a hydrocarbon tail, a ring structure region with 4 hydrocarbon rings and a hydroxyl group.



cholesterol

The hydroxyl (OH) group is polar, which makes it soluble in water. The 4-ring region of cholesterol is the signature of all steroid hormones such as testosterone and estrogen. The combination of the steroid ring structure and the hydroxyl (alcohol) group classifies cholesterol as a "sterol." Cholesterol is the animal sterol. Plants only make trace amounts of cholesterol, but make other sterols in larger amounts.

The last region is the hydrocarbon tail. Like the steroid ring region, this region is composed of carbon and hydrogen atoms. Both the ring region and tail region are non-polar, which means they dissolve in fatty and oily substances but will not mix with water. Because cholesterol contains both a water-soluble region and a fat-soluble region, it is called amphipathic.

However, cholesterol is not water-soluble enough to dissolve in the blood. Along with fats and fat-soluble nutrients, it travels in the blood through lipoproteins such as LDL and HDL.

11.6 Biological functions of cholesterol

Cholesterol is a structural component of cell membranes. The cholesterol content of a membrane varies with the tissue and with specific membrane function. The ratio of cholesterol to polar lipids affects the stability, permeability, and protein mobility of a

membrane. Membranes with high ratios have high stability and relatively low permeability. Their major function is a protective barrier. Membranes of intracellular organelles such as mitochondria have low cholesterol ratios and are consequently fluid and permeable. They serve primarily in synthetic and degradative reactions and in energy production. The outer membranes of most cells have intermediate cholesterol-polar lipid ratios and have both protective and metabolite-transport functions.

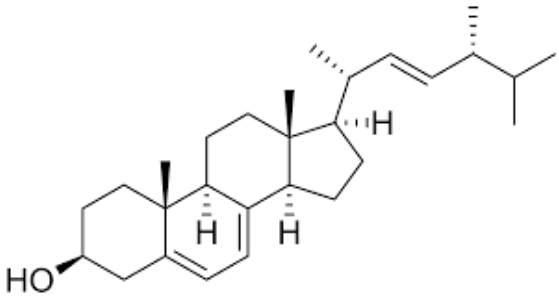
In addition to its role in membrane structure cholesterol has following other important functions.

Cholesterol is stored in the adrenals, testes, and ovaries, chiefly as the fatty acid ester, and converted to steroid hormones. These hormones include the male and female sex hormones (androgens and estrogens) as well as the adrenal corticoids (cortisol, corticosterone, aldosterone, and others).

In the liver cholesterol is the precursor of the bile acids, 24 steroid carboxylic acids that aid in the digestion of foods, especially lipids, and, when linked with the amino acids glycine or taurine, form the bile salts.

11.7 Structure and biological importance of Ergosterol

Ergosterol, or provitamin D_2, is a compound that belongs to the steroid family and is related to cholesterol. It is found in fungi such as Saccharomyces and Candida. This compound is an essential sterol component of fungal and protozoal cell membranes, so it is an important target of antifungal drugs and anti-trypanosomal drugs

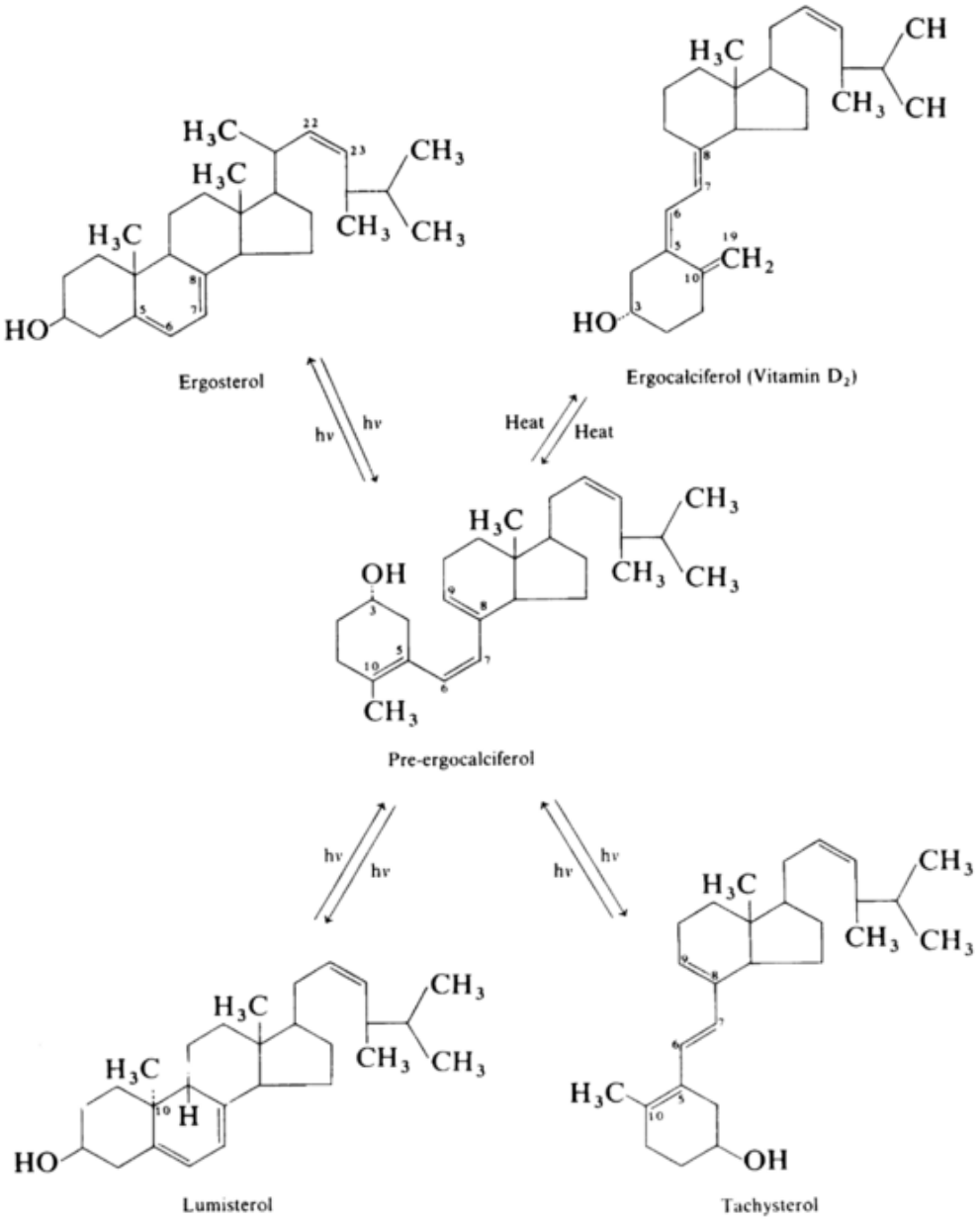


Principal structural difference between ergosterol and cholesterol is the extra methyl group on the side chain in ergosterol and this difference affects the preferential binding of polyene to ergosteral.

Ergosterol is a biological precursor to Vitamin D_2. Ergosterol, a sterol that is found in lower eukaryotic membranes, is converted into viosterol by ultraviolet light, and is then changed into ergocalciferol, which is a form of vitamin-D. This compound is structurally similar to cholesterol and when subjected to ultraviolet light, it is turned into vitamin D_2. This sterol can be easily transformed to peroxide of ergosterol by photo-oxidation with singlet oxygen.

11.8 Irradiation products of ergosterol

The transformation of ergosterol (provitamin D2) to the corresponding vitaminD begins immediately on exposure to ultraviolet light of suitable wave length. This process occurs in overlapping stages, with the formation of a series of products. The first stage is the photochemical conversion of ergosterol to pre-ergocalciferol. This is a reversible reaction. The next stage is the thermal conversion of pre-ergocalciferol to ergocalciferol (vitaminD_2). During the irradiation, tachysterol is formed via a reversible photochemical side reaction of pre-ergocalciferol.



Over irradiation of ergosterol yields toxisterol, superasterol I and superasterol II, these are biologically inactive products. An isomer mixture rich in lumisterol is obtained by irradiating ergosterol with ultraviolet light of wave length exceeding 280mμ.

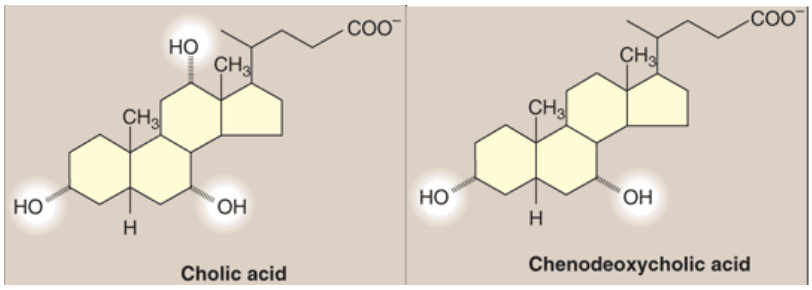
11.9 Bile acids

Bile acids are the end products of cholesterol metabolism. The bile acids contain 24 carbons, with two or three hydroxyl groups and a side chain that terminates in a carboxyl group. The carboxyl group has a pKa of about six and, therefore, is not fully ionized at physiologic pH, hence, they called "bile acid."

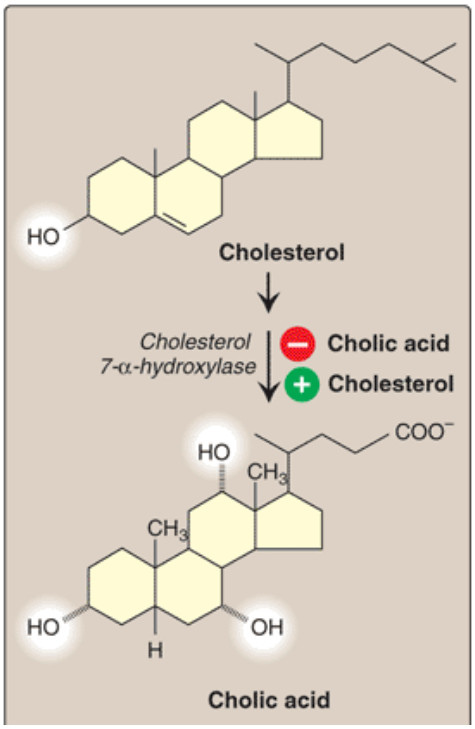
The bile acids are amphipathic in nature because, the molecules have both a polar and a nonpolar face, and can act as emulsifying agents in the intestine, helping prepare dietary triacylglycerol and other complex lipids for degradation.

11.10 Structure

All bile acids consist of two connecting units, a rigid steroid nucleus and a short aliphatic side chain and the steroid nucleus of bile acids has the saturated tetracyclic hydrocarbon perhydrocyclopentanophenanthrene, containing three six-member rings (A, B and C) and a fivemember ring (D). In addition, there are angular methyl groups at positions C-18 and C-19.

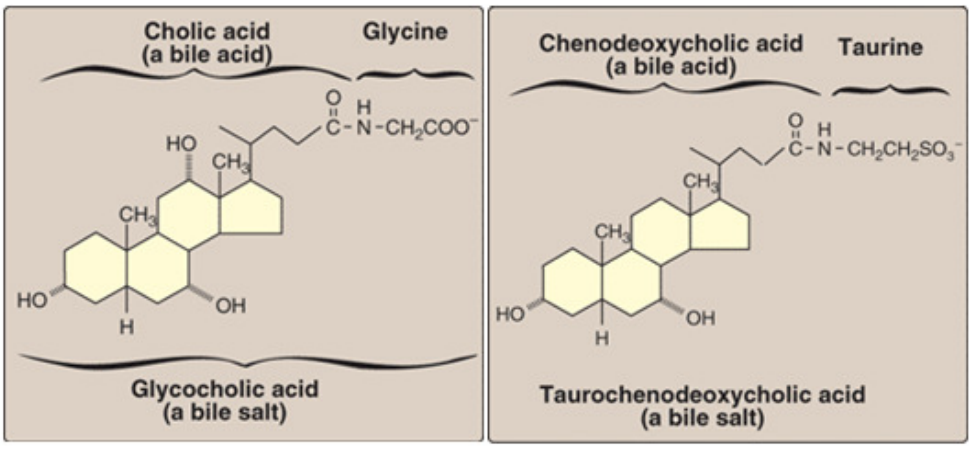


Bile acids are synthesized from cholesterol in the liver by involving a number of complex steps in both the steroid nucleus and the side chain. The hydroxyl groups are inserted at specific positions on the steroid structure, the double bond of the cholesterol B ring is reduced, and the hydrocarbon chain is shortened by three carbons, introducing a carboxyl group at the end of the chain. The most common resulting compounds, cholic acid (a triol) and chenodeoxycholic acid (a diol, are called "primary" bile acids). The rate-limiting step in bile acid synthesis is the introduction of a hydroxyl group at carbon 7 of the steroid nucleus by cholesterol-7-α-hydroxylase, an ER-associated cytochrome P450 (CYP) enzyme found only in liver. The enzyme is down-regulated by cholic acid and up-regulated by cholesterol.



Before the bile acids leave the liver, they are conjugated to a molecule of either glycine or taurine (an endproduct of cysteine metabolism) by an amide bond between the carboxyl group of the bile acid and the amino group of the added compound.

These new structures are called bile salts and include glycocholic and glycochenodeoxycholic acids, and taurocholic and taurochenodeoxycholic acids



The ratio of glycine to taurine forms in the bile is approximately 3:1. Addition of glycine or taurine results in the presence of a carboxyl group with a lower pKa(from glycine) or a sulfonate group (from taurine), both of which are fully ionized (negatively charged) at physiologic pH. Bacteria in the intestine can remove glycine and taurine from bile salts, regenerating bile acids.

They can also convert some of the primary bile acids into "secondary" bile acids by removing a hydroxyl group, producing deoxycholic acid from cholic acid and lithocholic acid from chenodeoxycholic acid.

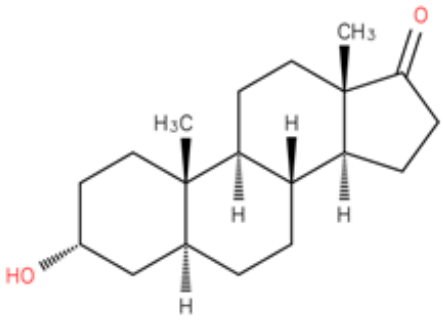
Bile salts secreted into the intestine are efficiently reabsorbed (greater than 95%) and reused. The mixture of primary and secondary bile acids and bile salts is absorbed primarily in the ileum. They are actively transported from the intestinal mucosal cells into the portal blood, and are efficiently removed by the liver parenchymal cells. Since dile acids are hydrophobic, they require a carrier in the portal blood. Albumin carries them in a noncovalent complex, just as it transports fatty acids in blood. The liver converts both primary and secondary bile acids into bile salts by conjugation with glycine or taurine, and secretes them into the bile.

11.11 Androsterone

Androsterone is one of the androgens, which are steroid hormones that affect the masculinization of the fetus and child and that maintain or create masculine traits in adults. Testosterone is the most active and abundant of the androgens. Androsterone was originally isolated from male urine in very small amounts.

Structure

A molecule of androsterone is very similar to testosterone. It consists of three six-member carbon rings and one five member ring joined along common sides. The molecular formula is $C_{19}H_{30}O_2$. Its chemical structure is 3α-hydroxy-5α-androstan-17-one. It is a weak androgen with a potency about 1/7th that of testosterone.



It has a beta-isomer called epiandrosterone, in which an hydroxyl group occupies a different geometric position. The ratio of androsterone to epiandrosterone is thought to be important in how masculine behavior is interpreted by others. In addition, it can be converted to dihydrotestosterone (DHT) from 3α-hydroxysteroid dehydrogenase. Androsterone is known to be an inhibitory androstane neurosteroid, acting as a positive allosteric modulator of the GABAA receptor, and possesses anticonvulsant effects. The unnatural enantiomer of androsterone is more potent as a positive allosteric modulator of GABAA receptors and as an

anticonvulsant than the natural form. Androsterone's 3β-isomer is epiandrosterone, and its 5β-epimer is etiocholanolone.

11.12 Biological significance

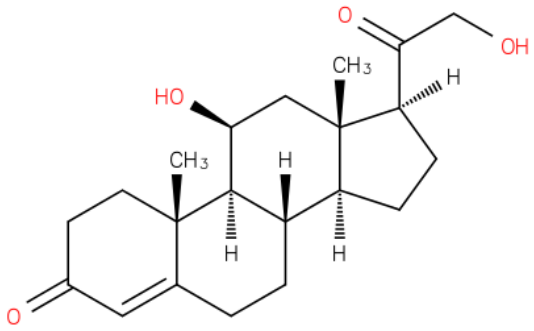
Androsterone is generally considered to be an inactive metabolite of testosterone, which when conjugated by glucuronidation and sulfation allows testosterone to be removed from the body, but it is a weak neurosteroid that can cross into the brain and could have effects on brain function.

11.13 Corticosterone

Corticosterone is a 21-carbon steroid hormone of the corticosteroid type produced in the cortex of the adrenal glands.

Structure

A corticosterone molecule is composed of 21 carbon, 30 hydrogen, and four oxygen atoms. Corticosterone is 21-hydroxy steroid that consists of pregn-4-ene substituted by hydroxy groups at positions 11 and 21 and oxo groups at positions 3 and 20.



11.14 Biological significance

Corticosteroids are involved in many processes in the body. They are an integral part of the immune response, similar in structure to cortisol. Corticosterone, also called Kendall's compound B, is a steroid hormone secreted by one of the outer layers of the adrenal cortex in humans. The hormone is used by the human body in response to stressors such as allergens and other environmental factors. Unlike the other steroid hormones produced by the body, corticosterone is not used for anti-inflammatory purposes. It is an antagonist for insulin use and is vital in the synthesis of carbohydrates and in protein degradation.

11.15 Summary of the unit

Steroids are a group of compounds which are soluble in organic fat solvents but cannot be hydrolyzed by NaOH. Steroids containing alcoholic group at C-3 positions and a side chain of 8 to 10 carbon atoms at C-17 are called sterols.

Cholesterol is the best example of sterols. The name cholesterol originates from the Greek word chole- and stereos, and the suffix for alcoholic group. It is most abundant in brain, nerves, and adrenals and in skin. It is prepared by liver. It helps in formation of different hormones. But it is absent in prokaryotic cells. It is an important component for bile juice formation.

Cholesterol is essential for all animal life, cholesterol is necessary for building and maintenance of membranes. Other sites of high synthesis rates include the intestines, adrenal glands, and reproductive organs. It helps to regulates membrane fluidity over the range of physiological temperatures.

Cholesterol also functions in intracellular transport, cell signaling and nerve conduction. Human breast milk also contains significant quantities of cholesterol.

High level of cholesterol: Cholesterol is very important for all animals but a high level of serum cholesterol may lead to heart disease. Cholesterol is prepare in the body by liver. It helps in preparation of certain adrenal hormones.

Low level of cholesterol: If the cholesterol level inside our body becomes low, then we might suffer from depression, anxiety and even lead to a case of suicide.

Bile acids are physiological detergents that generate bile flow and facilitate intestinal absorption and transport of lipids, nutrients, and vitamins. Bile acids also are signaling molecules and inflammatory agents that rapidly activate nuclear receptors and cell signaling pathways that regulate lipid, glucose, and energy metabolism. The final and longest segment of the small intestine. It is specifically responsible for the absorption of vitamin B12 and the reabsorption of conjugated bile salts. Bile salts are composed of the salts of four different kinds of free bile acids (cholic, deoxycholic, chenodeoxycholic, and lithocholic acids); each of these acids may in turn combine with glycine or taurine to form more complex acids and salts. Bile salts and acids can be synthesized from cholesterol or extracted from the bloodstream by the liver.

11.16 Key words

Steroids; Steroid nucleus; Cholesterol; Ergosterol; Ergosterol; Bile acids; Androsterone; Corticosterone

11.17 References for further studies

1) Organic Chemistry, Volume 2: Stereochemistry and The Chemistry Natural Products, 5/E Volume 2 of Organic Chemistry; I. L. Finar; *Pearson Education India*, **1956**.

2) Natural Products Chemistry, Volume 2; Koji Nakanishi, Toshio Goto, Shô Itô; *Elsevier*, **2013**.

3) Strategies for Organic Drug Synthesis and Design; Daniel Lednicer; *John Wiley & Sons*, **2009**.

4) Steroid Chemistry at a Glance; Daniel Lednicer; *John Wiley & Sons*; **2011**.

5) Hormones; Anthony W. Norman, Helen L. Henry*; Academic Press*, **2014**.

11.18 Questions for self understanding

1) What are steroids? What is the general structure of steroids?

2) How are carbons numbered in steroid nucleus?

3) How are rings oriented in steroids with respect to each other?

4) How are steroids named?

5) How many configurational isomers are possible for steroidal skeleton? What is α- and β-orientation?

6) What are the different classes of steroids? Which steroid falls under what class?

7) Explain the structure and biological importance of cholesterol.

8) Explain the structure and biological importance of ergosterol.

9) Discuss the irradiation products of ergosterol.

10) What are bile acids?

11) Discuss the structure and significance of bile acids.

12) Explain the structure and biological significance androsterone

13) Explain the structure and biological significance corticosterone

UNIT-12

Structure

12.0 Objectives of the unit

After studying this unit you are able to

- Explain the important sources of vitamins
- Explain the classification of vitamins
- List out the functions of vitamins
- Explain the deficiency of vitamins
- Explain the side effects and overdose

12.1 Introduction

A vitamin is an organic compound required as a nutrient in tiny amounts by an organism. A compound is called a vitamin when it cannot be synthesized in sufficient quantities by an organism, and must be obtained from the diet. Thus, the term is conditional both on the circumstances and the particular organism. For example, ascorbic acid functions as vitamin C for some animals but not others, and vitamins D and K are required in the human diet only in certain circumstances. Therefore, the term vitamin does not include other essential nutrients such as: Dietary minerals, Essential fatty acids, or Esential amino acids, nor does it encompass the large number of other nutrients that promote health but are otherwise required less often

12.2 vitamins

Vitamins are the nutrients our bodies need in order to maintain functions such as immunity and metabolism. By definition *vitamins are a group of organic compounds which are essential for normal growth and nutrition and are required in small quantities in the diet because they cannot be synthesized by the body.*

12.3 Classification of vitamin

Vitamins are classified mainly on the basis of their biological and chemical activity. Thus, each "vitamin" may refer to several vitamer compounds that all show the biological activity associated with a particular vitamin.

For example, "vitamin A," which includes the compounds retinal, retinol, and many carotenoids.

Vitamins are classified as either water-soluble or fat soluble.

In humans there are 13 vitamins are presents. Out of them 4 are fat-soluble, they are vitamin A, D, E and K. And 9 are water-soluble, they are 8 B vitamins and vitamin C.

12.3.1 Fat-soluble vitamins

Fat-soluble vitamins are absorbed through the intestinal tract with the help of lipids (fats). Because they are more likely to accumulate in the body, they are more likely to lead to hyper vitaminosis than water-soluble vitamins.

12.3.2 Water-soluble vitamins

Water-soluble vitamins dissolve easily in water, and in general, are readily excreted from the body, to the degree that urinary output is a strong predictor of vitamin consumption. Because they are not readily stored, consistent daily intake is very important.

12.4 Functions of vitamins

Vitamins are functions as Hormones (e.g. vitamin D), Antioxidants (e.g. vitamin E), mediators of cell signalling, regulators of cell, tissue growth and differentiation (e.g. vitamin A). Precursors for enzyme cofactor bio-molecules (coenzymes), that help act as catalysts and substrates in metabolism (e.g. B complex vitamins) . e.g. biotin is part of enzymes involved in making fatty acids. Vitamins also act as coenzymes to carry chemical groups between enzymes. e.g. folic acid carries various forms of carbon group – methyl, formyl and methylene - in the cell.

12.5 Deficiency of vitamins

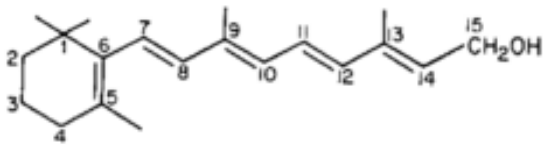
A primary deficiency occurs when an organism does not get enough of the vitamin in its food. A secondary deficiency may be due to an underlying disorder that prevents or limits the absorption or use of the vitamin, due to a “lifestyle factor”, such as smoking, excessive alcohol consumption, or the use of medications that interfere with the absorption or use of the vitamin.

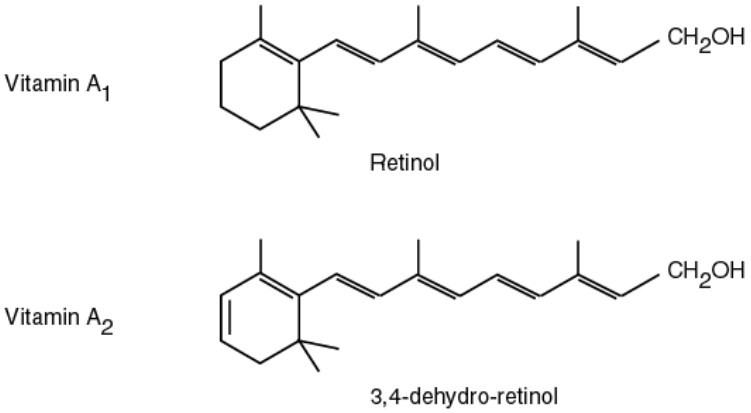
12.6 Side effects and overdose

Vitamin poisoning, hyper-vitaminosis or vitamin overdose refers to a condition of high storage levels of vitamins, which can lead to toxic symptoms. At high enough dosages, some vitamins cause side effects such as nausea, diarrhea, and vomiting

12.7 Vitamin A

Vitamin A, also called retinol or vitamin A_1. Vitamin A is a fat-soluble vitamin which is important as a component of the protein rhodopsin, a light absorbing pigment found in the retina of the eyes. There are two forms of vitamin A. vitamin A_1 is common in marine fishes, whilst vitamin A2 is common





The only difference between vitamin A_1 and A_2 is of a bond in the ring. Vitamin A_2 contain additional bouble bond between C3 and C4

Vitamin A is most abundant in fatty fish and especially in fish-liver oils it is also found in milk fat, eggs, and liver. Vitamin A is not present in plants, but many vegetables and fruits contain one or more of a class of pigments that can be converted to vitamin A in the body. beta-carotene (provitamin A) is an excellent source of vitamin activity

12.7 Biological functions of vitamin A

Vitamin A is readily destroyed upon exposure to heat, light or air. The vitamin A functions directly in vision, it is converted into retinaldehyde, a component of a light-sensitive pigment called rhodopsin (visual purple), which is present in the retina of the eye. Vitamin A is combined with specific proteins in the form of retinoic acid. It functions in the regulation of embryonic development and growth. Retinoic acid is also essential for maintenance of the epithelial tissues (the skin and the mucous membranes lining the internal body surfaces), for sperm formation, and for proper functioning of the immune system.

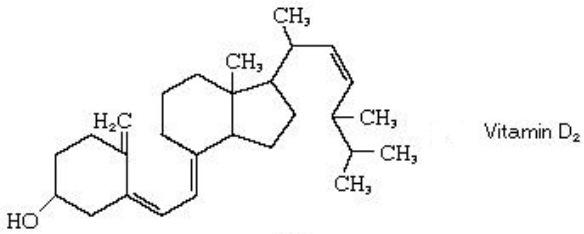
12.8 Difference of vitamin A

Vitamin A is important in cells other than the retina since deficiency causes damage in epithelial, bony and connective tissues. β-Carotene is often used as a dietary source of vitamin A.

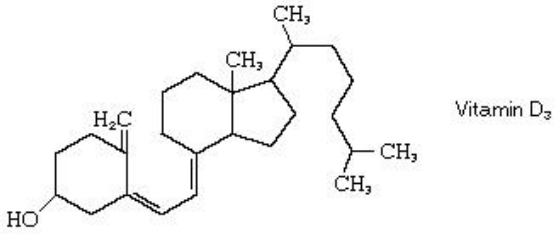
12.9 Vitamin D

The vitamin D refers to group of compounds consists of several molecules. The most important are vitamin D_2 (ergocalciferol) and vitamin D_3. All forms of vitamin D belong to a family of lipids called secosteroids. Secosteroids are very similar in structure to steroids except that two of the B-ring carbon atoms of the typical four steroid rings are not joined, whereas in steroids they are. Vitamin D is a precursor to 1,25-dihydroycholecalciferol a

hormone important in regulating calcium and phosphate levels in the serum. Vitamin D synthesis occurs in most animals by ultraviolet radiation of 7- dehydrocholesterol.

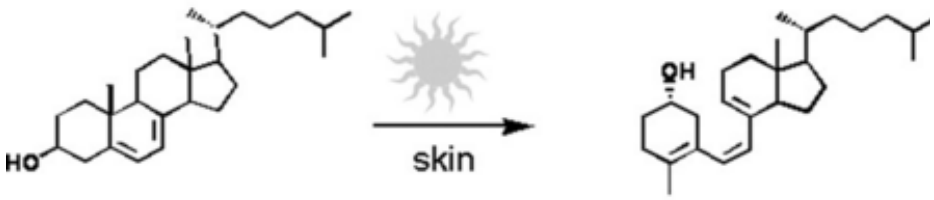


Vitamin D_2

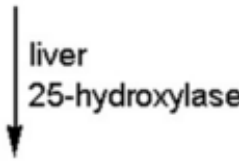


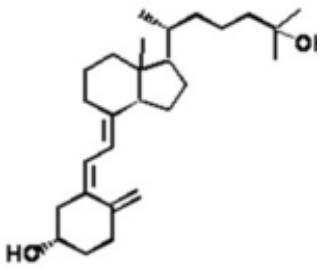
Vitamin D_3

Conversion of 7- dehydrocholesterol to vitamin D_3 is taking place in two steps. In the first step conrotatory photochemical ring opening of ring B is taking place. In the second step [1,7] sigmatropic antarafacial hydrogen shift occurs in thermal mode. The active form of vitamin D_3 is 1,25-dihydroxyvitamin D_3

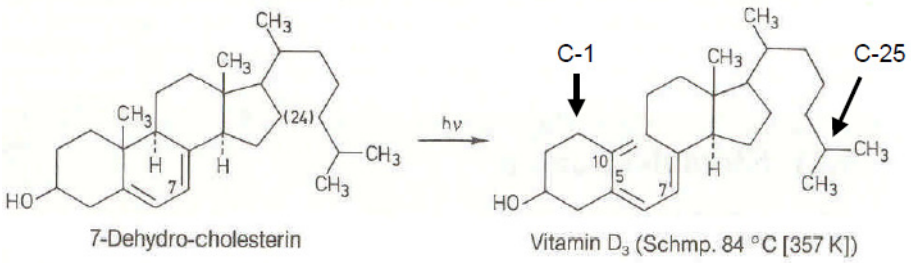


7,8-dehydrocholesterol pre-vitamin D_3





25-hydroxyvitamin D_3



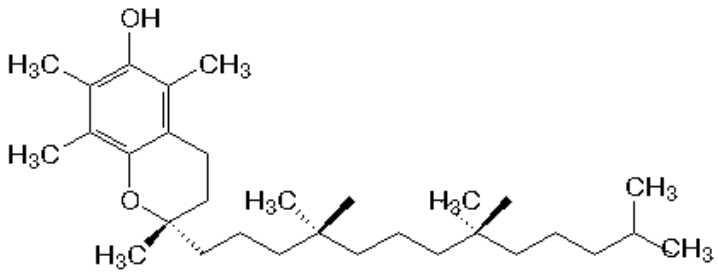
7-Dehydro-cholesterin Vitamin D_3 (Schmp. 84 °C [357 K])

12.10 Physiological functions of Vitamin D

Vitamin D is well known as a hormone involved in mineral metabolism and bone growth. Its most dramatic effect is to facilitate intestinal absorption of calcium, although it also stimulates absorption of phosphate and magnesium ions. In the absence of vitamin D, dietary calcium is not absorbed at all efficiently. Vitamin D stimulates the expression of a number of proteins involved in transporting calcium from the lumen of the intestine, across the epithelial cells and into blood. The best-studied of these calcium transporters is calbindin, an intracellular protein that ferries calcium across the intestinal epithelial cell.

12.11Vitamin E

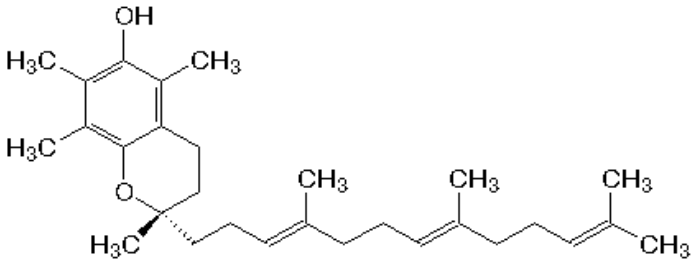
Vitamin E is also known as tocopherol. The vitamin E exists in two types of structures, the tocopherol and tocotrienol structures. Both structures are similar except the tocotrienol structure has double bonds on the isoprenoid units. There are many derivatives of these structures due to the different substituents possible on the aromatic ring at positions 5, 6, 7, and 8.



Vitamin E (α-tocopherol)

There are three chiral centers, at positions 2', 4', and 8', in the tail. Therefore there is a possibility of eight stereoisomers. The most abundant of the naturally-occurring forms is the *R*,*R*,*R* form.

The tocotrienols share the same ring structure, but have an unsaturated tail.

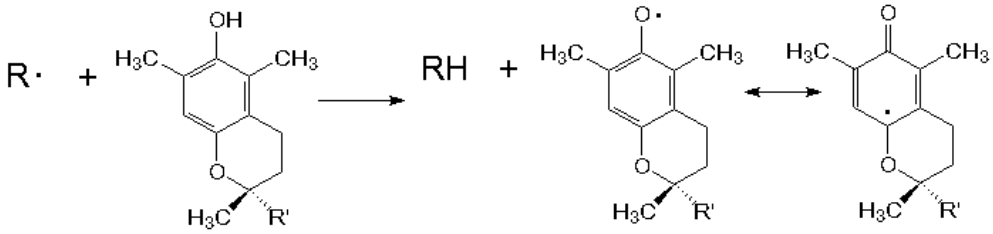


Tocotrienol Structure

12.12 Physiological functions of vitamin E

Vitamin E acts as an antioxidant, particularly protecting polyunsaturated fatty acids. High dietary levels of poly-unsaturated fatty acids increase the requirement for dietary vitamin E. An interaction exists between vitamin E and selenium, a metallic antioxidant; vitamin E requirement are greater in selenium-depleted fish.

Tocopherols (Vitamin E) interrupt free radical chain reactions by capturing the free radical, this imparts to them their antioxidant properties. The free hydroxyl group on the aromatic ring is responsible for the antioxidant properties. The hydrogen from this group is donated to the free radical, resulting in a relatively stable free radical form of the vitamin.

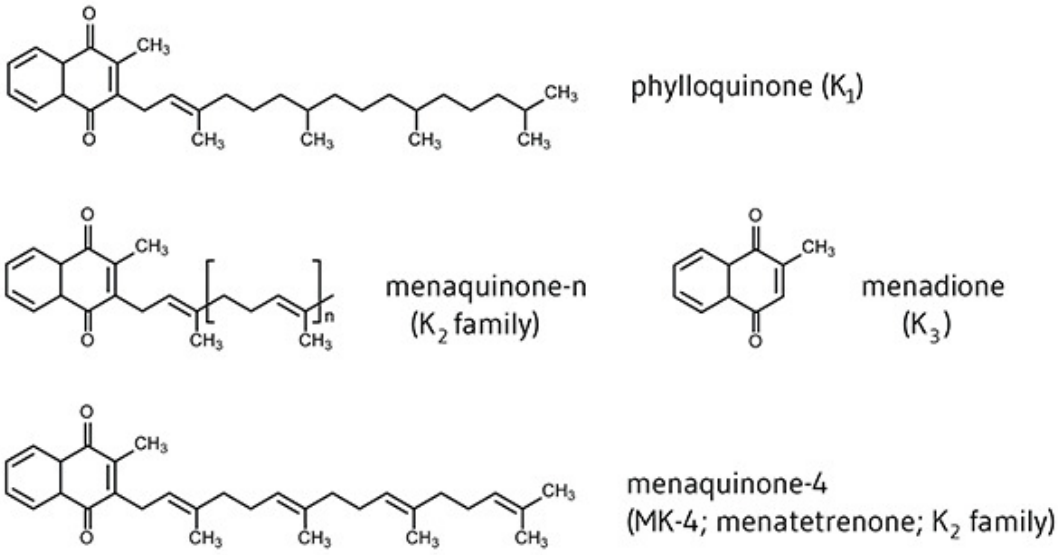


Resonance-stabilized radical

12.13 Vitamin K

Vitamin K is the family name for a series of fat-soluble compounds, which have a common 2-methyl-1, 4-naphthoquinone nucleus but differ in the structures of a side chain at the 3-position. They are synthesised by plants and bacteria. In plants the only important molecular form is phylloquinone (vitamin K_1), which has a phytyl side chain. Bacteria synthesise a family of compounds called menaquinones (vitamin K_2), which have side chains based on repeating unsaturated 5-carbon (prenyl) units. These are designated menaquinone-n (MK-n) according to the number (n) of prenyl units. Some bacteria also synthesise menaquinones in which one or more of the double bonds is saturated. The compound 2-methyl-1,4-

naphthoquinone (common name menadione) may be regarded as a provitamin because vertebrates can convert it to MK-4 by adding a 4-prenyl side chain at the 3-position.

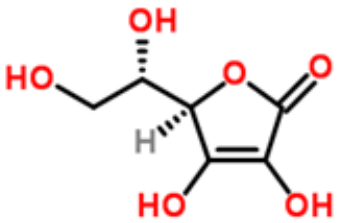


12.14 Physiological functions of vitamin K

Vitamin K is important in the synthesis of prothrombin, a protein which is important in blood clotting. Vitamin K is required by all animals, including fish, for normal blood clotting. Vitamin K functions as a cofactor for the enzyme, γ-glutamylcarboxylase (GGCX), which catalyzes the carboxylation of the amino acid glutamic acid (Glu) to γ-carboxyglutamic acid (Gla). Vitamin K-dependent γ-carboxylation occurs only on specific glutamic acid residues in vitamin K-dependent proteins (VKDP) which is critical for their ability to bind calcium. The ability to bind calcium ions (Ca^{2+}) is required for the activation of the several vitamin K-dependent clotting factors, or proteins, in the coagulation (clotting) cascade.

12.15 Vitamin C

Vitamin C is also known as ascorbic acid and ascorbate, is a six-carbon lactone which is synthesised from glucose by many animals. Vitamin C is synthesised in the liver in some mammals and in the kidney in birds and reptiles.



Vitamin C is an electron donor (reducing agent or antioxidant), and all of its biochemical and molecular functions is accounted for by this function.

12.16 Physiological functions of vitamin C

The first recognized function of ascorbic acid is its role in hydroxylating the proline to hydroxyproline for use in cartilage synthesis. It is involved in carnitine synthesis and in the detoxification of pesticides and other toxicants in processes involving cytochrome P450. Vitamin C contributes to the health of teeth and gums, preventing haemorrhaging and bleeding. It also improves the absorption of iron from the diet. Vitamin C is also needed for the metabolism of bile acids which may have implications for bloodcholesterol levels and gallstones. Moreover, vitamin C plays an important role in the synthesis of several important peptide hormones, neurotransmitters and carnitine

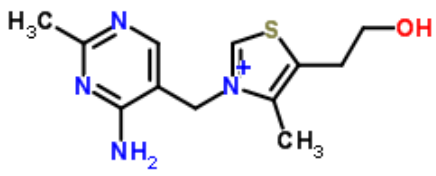
It is a highly labile vitamin easily destroyed by cooking or lengthy or improper storage of food. It is usually added to five to 10-fold excess to allow for degradation during storage and to provide some shelf-life to the feed.

12.17 Vitamin B

The vitamin B family is made up of eight B vitamins. Although they are commonly recognized as a group and often work together in the body, each of the B vitamins performs unique and important functions.

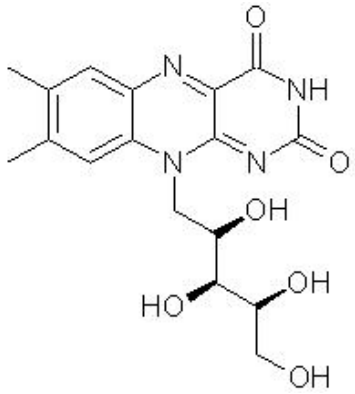
12.17.1 Vitamin B_1

Vitamin B_1 is also known as thiamin. The active form of thiamin is thiamine pyrophosphate. Thiamine pyrophosphate is a coenzyme for reactions which are involved in carbohydrate metabolism. Thiamin is needed to produce cellular energy from the foods you eat, and is also required for the synthesis of DNA and RNA. Vitamin B_1 play important role in the conversion of pyruvate to acetyl CoA.



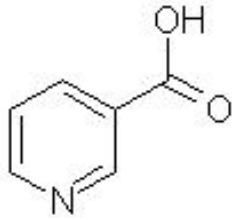
12.17.2 Vitamin B_2

Vitamin B_2 is also known as riboflavin. Riboflavin is a yellow pigmented molecule which in the body forms a component of the molecule flavin adenine dinucleotide (FAD). FAD is a cofactor in a number of oxidation-reduction reactions and acts as an energy currency similar to ATP. It is particularly important in the degradation of pyruvate, fatty acids and amino acids, and in the process of electron transport. Riboflavin is a basic building block for normal growth and development. It is needed for healthy cellular energy production and also supports the antioxidant activity in the body.



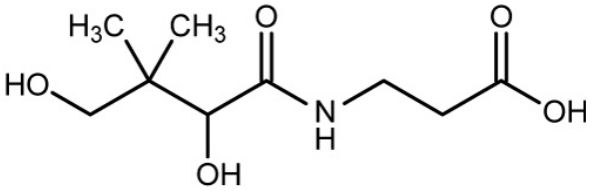
12.17.3 Vitamin B_3

Vitamin B_3 is also known as nicotinic acid, or niacin. It is a component of two high energy molecules-nicotinamide adenine dinucleotide (NAD) and nicotinamide adenine dinucleotide phosphate (NADP). NAD and NADP are, like FAD, important in a number of oxidation and reduction reactions that occur within cells. Nicotinic acid can be synthesized by most animals from the amino acid tryptophan, and nicacin deficiencies can be avoided by feeeding a diet high in tryptophan. Tryptophan, however, is only present in very small amounts in most animals meals, and supplementation of the diet with nicotinic acid is therefore advisable. Vitamin B_3, support over 200 chemical reactions in the body including cellular energy production and fatty acid synthesis.



12.17.4 Vitamin B_5

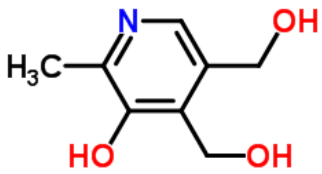
Vitamin B_5 also known as Pantothenic acid or coenzyme A, is the coenzyme of acetyl, acyl or propionic CoA. This molecule serves as a carrier of various carbohydrate groups and is involved in reactions in fatty acid oxidation, fatty acid synthesis, pyruvate oxidation and acetylations. The presence of coenzyme A within a cell is fundamental to the transfer of energy throughout the various reactions.



12.17.5 Vitamin B_6

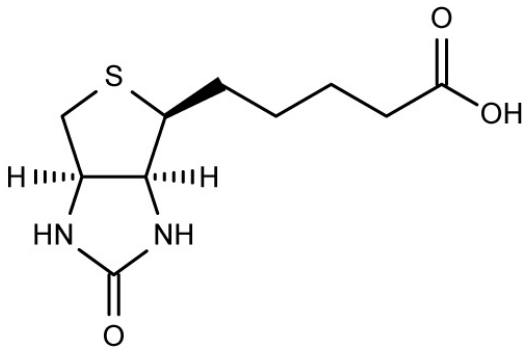
Vitamin B_6, also known as pyridoxine, is needed to metabolize amino acids and glycogen (the body's storage form of glucose), and is also necessary for normal nervous system,

hormone and red blood cell function. Pyridoxine acts as an important coenzyme in transamination reactions.



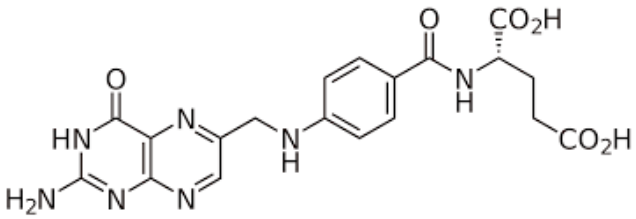
12.17.6 Vitamin B_7

Vitamin B_7 is also known as Biotin, it acts to facilitate carbon dioxide transfer in reactions which require the addition of CO_2 to another molecule. Biotin is made by intestinal bacteria. A glycoprotein component of egg white called avidin binds biotin in the gut such that it cannot be absorbed by the intestinal mucosa and thus can be used to induce biotin deficiency. Heat denatures avidin, allowing it to be digested and so any biotin complexed can be absorbed.



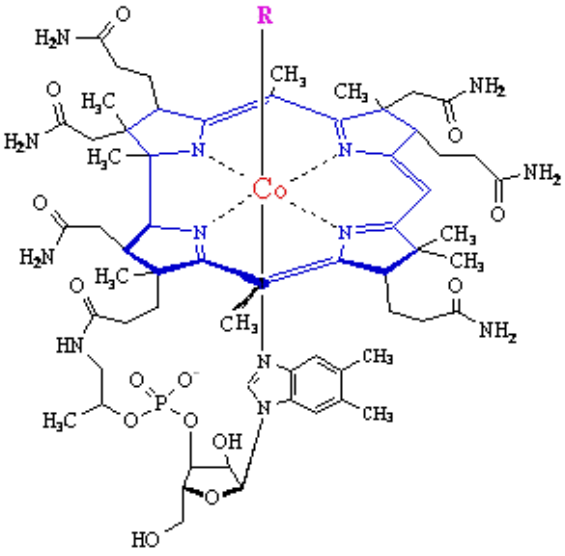
12.17.7 Vitamin B_9

Vitamin B_9 is also known as Folic acid. It is a cofactor in the transfer of single carbon entities to other molecules in the same way that biotin is a carrier for CO_2. The folic acid molecule is composed of three separate parts called pterodine, p-aminobenzoic acid and glutamic acid. In some organisms, folic acid can be synthesized if p-aminobenzoic acid is provided in the diet. folic acid is needed for DNA synthesis, the formation of red blood cells and for the metabolism of amino acids. Folic acid is most commonly known for its role in fetal health and development as it is critical for the formation of a baby's spinal cord and nervous system. This important developmental process occurs during the initial weeks of pregnancy, and so adequate folic acid intake is especially important for all women of child-bearing age.



12.17.8 Vitamin B_{12}

Vitamin B_{12} is essential for normal maturation and development. It is required for the synthesis of choline and the metabolism of single carbon fragments. The requirement for vitamin B_{12} in the diet is only as a trace, and it is difficult to induce a dietary deficiency because vitamin B_{12} is synthesized by microorganisms in the gut. The usual problem with vitamin B_{12} deficiency occurs because a carrier mucoglycoprotein called intrinsic factor is lacking in the gut. The absence of intrinsic factor means that the vitamin B_{12} present in the intestinal contents will not be absorbed by the animal and deficiency results.



12.18 Summary of the unit

A vitamin is an organic compound that occurs as a natural component of foods and must be supplied exogenously in small amounts to maintain growth, health, and reproduction of an organism. Hence, a vitamin is essential because it cannot be made in the organism that requires it, yet it functions, usually after metabolic alteration, in indispensable ways. Since only small quantities, usually microgram to milligram amounts per human adult per day, are required to avoid a deficiency disease, vitamins are classified as micronutrients along with the trace elements. Vitamins are classified by their biological and chemical activity, not on their structure. Thus, each "vitamin" may refer to several vitamer compounds that all show the biological activity associated with a particular vitamin. For example, "vitamin A," which includes the compounds retinal, retinol, and many carotenoids. Vitamins are classified as either water-soluble or fat soluble. In humans there are 13 vitamins: 4 fat-soluble (A, D, E and K) and 9 water-soluble (8 B vitamins and vitamin C). Fat-soluble vitamins are absorbed through the intestinal tract with the help of lipids (fats). Because they are more likely to

accumulate in the body, they are more likely to lead to hyper-vitaminosis than water-soluble vitamins. Water-soluble vitamins dissolve easily in water, and in general, are readily excreted from the body, to the degree that urinary output is a strong predictor of vitamin consumption. Because they are not readily stored, consistent daily intake is very important.

12.19 Key Word

Vitamins; Fat-soluble vitamins; Water-soluble vitamins; Vitamin A; Vitamin D; Vitamin E; Vitamin K; Vitamin C; Vitamin B; Vitamin B_1 ;Vitamin B_2; Vitamin B_3; Vitamin B_5; Vitamin B_6; Vitamin B_7; Vitamin B_9; Vitamin B_{12}

12.20 References for further study

1) Vitamins and hormones, Volume 17; Gerard Meurant; *Academic Press*, **1959**.

2) Handbook of Vitamins; Janos Zempleni, John W. Suttie, Jesse F. Gregory III, Patrick J. Stover; *CRC Press*, **2013**.

3) The Vitamins: Chemistry, Physiology, Pathology, Methods; W. H. Sebrell, Robert S. Harris; *Academic Press*, **2014**.

4) Natural Products Chemistry, Volume 2; Koji Nakanishi, Toshio Goto, Shô Itô; *Elsevier*, **2013**.

5) Organic Chemistry, Volume 2: Stereochemistry and The Chemistry Natural Products, 5/E Volume 2 of Organic Chemistry; I. L. Finar; *Pearson Education India*, **1956**.

12.21 Questions for self understanding

1) What are vitamins?

ii) How vitamins are classified?

iii) Give two examples for fat-soluble vitamins and water-soluble vitamins

iv) Discuss the functions of different vitamins

v) Explain the consequences deficiency of vitamins in human

vi) Discuss the side effects and effect of overdose

vii) Write the structure of vitamin A. What are the common sources of this vitamin?

viii) Discuss the biological functions of vitamin A

ix) Write a note on deficiency of vitamin A

x) Write the structure of Vitamin D. What are the common sources of this vitamin?

xi) Explain the physiological functions of Vitamin D

xii) Write the structure of vitamin E. What are the common sources of this vitamin?

xiii) Explain the Physiological functions of vitamin E

xiv) Write the structure of vitamin K. What are the common sources of this vitamin?

xv) Explain the Physiological functions of vitamin K

xvi) Write the structure of vitamin C. What are the common sources of this vitamin?

xvii) Explain the Physiological functions of vitamin C

xviii) Write the structure of vitamin B. What are the common sources of this vitamin?

xix) Write the structure and functions of following vitamins

a) Vitamin B_1

b) Vitamin B_2

c) Vitamin B_3

d) Vitamin B_5

e) Vitamin B_6

f) Vitamin B_7

g) Vitamin B_9

h) Vitamin B_{12}

UNIT – 13

Structure

13.0 Objectives of the unit

After studying this unit you are able to

- Identify the origin of Alkaloids
- List the Characteristics of alkaloids
- Explain the general method of isolation of alkaloids from plants
- Write the name of alkaloids according to nomenclature of alkaloids
- Classify the alkaloids based on their structure and function
- Explain the different qualitative chemical tests for alkaloids detection
- Elucidate the structure of atropine

13.1 Introductions

Natural products such as carbohydrates, proteins, lipids, and nucleic acids are produced by a living organism. All of which play an important role in metabolic reactions. These are called primary metabolites. Primary metabolites are usually found in all living organisms such as plants and animals. They are widely distributed in living systems and are usually involved in essential life processes. They form the fundamental building block of living material e.g. mevalonic acids and nucleotides.

There are other naturally occurring products isolated from plants or are produced by microorganisms. These are not primary metabolites and have extraordinary complexity in their structure and are called secondary metabolites. The secondary metabolites are chemicals synthesised by plants but they are not directly used by them. They are used indirectly by man as a source of pharmaceutical preparations. Secondary metabolites are generally built from primary metabolites. These metabolites have restricted distribution and are characteristics of individual genera or species. Secondary metabolites are essential to the existence of the organism but play an important role to the survival. They do not seem to have any obvious metabolic or evolutionary function. In fact, some compounds may be formed as the result of a "metabolic accident" or by-products of the synthesis machinery of the cellular enzymes. Regardless of their utility to the parent organism, their value to man as drugs, herbs, flavorings, poisons, dyes, and so on is undisputed.

13.2 Alkaloids

A precise definition of the term ‘alkaloid’ (alkali-like) is somewhat difficult because there is no clear-cut boundary between alkaloids and naturally occurring complex amines. Typical alkaloids are derived from plant sources, they are basic, they contain one or more nitrogen atoms usually present in a heterocyclic ring and they usually have a marked physiological action on man or other animals. The term alkaloid is applied to nitrogen containing molecules

belonging to one of the largest and most diverse families of naturally occurring compounds. Thus, *Alkaloids are naturally-occurring nitrogen based organic compounds and are usually heterocyclic in nature. The nitrogen is enclosed in a heterocyclic ring system.*

13.3 Characteristics of alkaloids

Most alkaloids are well-defined crystalline substances. They form salts with acids. In the plant they may exist in free states, as salts or as N-oxides. In addition to the elements carbon, hydrogen and nitrogen, most alkaloids contain oxygen.

1) They are basic in nature due to the presence of nitrogen in their ring.

2) They have complex structures.

3) They have bitter in taste.

4) They are mostly obtained from plant materials.

5) They have high pharmacological and physiological activities

Most of the alkaloids contain oxygen and optically active. The optical activity of alkaloids are very useful for resolving racemic acids. The alkaloids form insoluble precipitates with solutions of phosphotungstic acid, phosphomolybdic acid, picric acid, potassium mercuri-iodide, etc. Many of these precipitates have defined crystalline shapes and so may be used to help in the identification of an alkaloid. Some of these reagents are also used as a means of detecting alkaloids in paper and thin layer chromatography.

13.4 General method of isolation of alkaloids from plants

Alkaloids are usually found in seeds, root, leaves, or bark of the plant, and generally occur as salts of various plant acids, e.g., acetic, oxalic, citric, malic , tartaric, etc. A common method of isolation of alkaloids is as follows. The plant is dried, finely powdered, then isolation of alkaloids by extraction of refluxing with suitable solvent. The most commonly used solvent is methanol. The solvent is removed by distillation and the residue extracted with petroleum ether to remove the plant oil. The remaining residual fraction is treated with inorganic acids. The alkaloids are soluble in inorganic acids thus they are isolated as their salts. Finally free alkaloids are isolated by treatment of their salt solution with sodium carbonate followed by extraction with a suitable solvent like ether, chloroform etc...

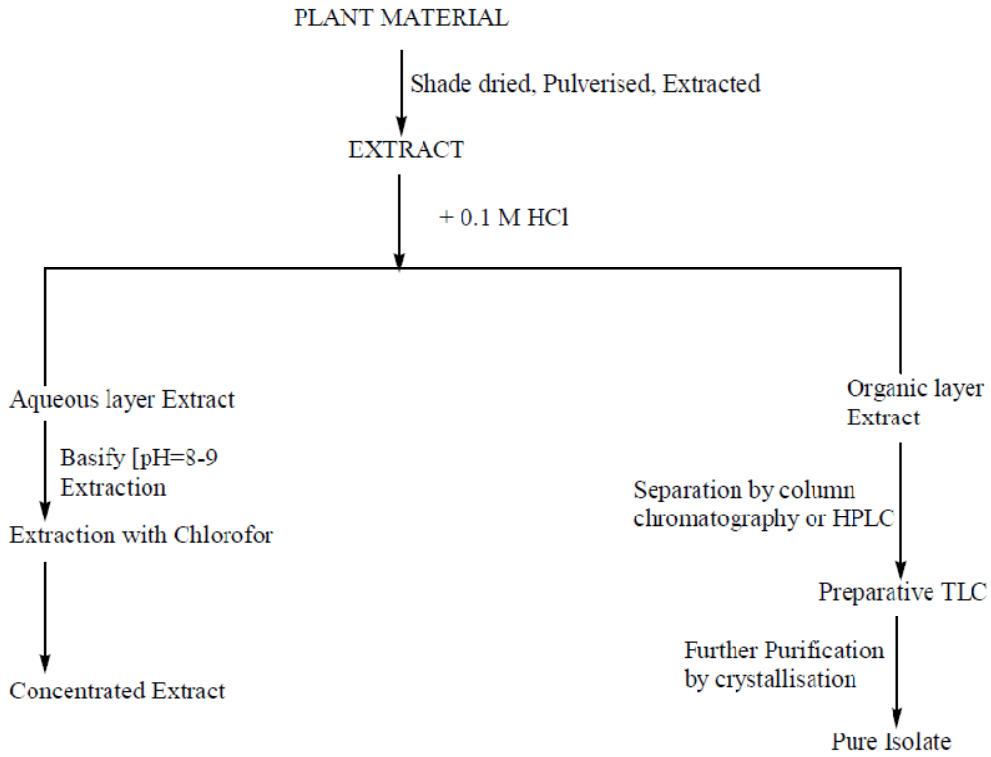
Individual alkaloids are separated from each other through difference in the solubilities of the bases and their salts, or by fractional crystallization, or by column chromatography.

In order to isolate pure alkaloid from plant source, the procedures to be followed are summarized as follows

Step 1: Collection of plant material, air-drying the plant material to remove water, pulverization of the air-dried plant material and solvent extraction protocol of the powdered plant materials.

Step 2: Purification of the alkaloid-This is done to separate alkaloid from the solution. This is done by running chromatography on the syrupy form of the extract.

Step 3: involves the crystallization and further purification of the isolates.



13.5 Nomenclature of Alkaloids

There is no systemic nomenclature method for alkaloids due to the complexity in their structures. Because of the complex structure and of historical reasons no attempt has been made for the systematic nomenclature for alkaloids. By agreement, chemical rules designate that names of alkaloids should end with the suffix (-ine).

The names of alkaloids are obtained in various ways:

(a) From the generic name of the plant yielding them (such as atropine).

(b) From the specific name of the plant yielding them (such as cocaine and belladonine).

(c) From their physiological activity (such as emetine).

(d) From their physical property (such as hygrine, while hygro = moist).

(e) Occasionally of discoverer (such as pelletierine, after pelletier).

(f) From the common name of the drug yielding them (such as ergotamine).

13.6 Classification of alkaloids

Alkaloids show great variety in their botanical and biochemical origin, in chemical structure and in pharmacological action. Consequently, many different systems of classification are possible. Alkaloids are classified into two broad classes. They are a) Classification based on the nature of the ring systems and b) Classification on the basis of plant sources.

13.6.1 *Classification based on plant source*

In this case alkaloids are classified on the basis of the plant source such as the family and the genus. Structural overlap may occur in this classification. For example, morphine alkaloid is from Apocynaceae family. Opium alkaloids such as morphine, codeine, nicotine and papaverine are derived from opium plant.

Rauwolfia alkaloidsnis reserpine, derived from Rauwolfia family.

Reserpine is an antihypertensive alkaloid. It equally act as tranquilizer.

Chintonia Alkaloids: Quinine, from Cinchonine, etc.

Coca Alkaloods: from Erythroxyllum species

Solonaceae alkaloids: from Solanaceae family, etc.

13.6.2 *Classification based on the chemical structure of alkaloids*

a) *Heterocyclic Alkaloids*:

This class includes pyrollidine nucleus, Pyridine nucleus, Piperidine nucleus, Pyridine-Piperidine nucleus, etc

b) *Isoquinoline type of alkaloids*

Isoquinoline is chemically known as benzo [c] pyridine or 2-benzanine. The alkaloids that possess isoquinoline skeletons are known as isoquinoline alkaloids. Isoquinoline groups may be further subdivided into several groups such as simple isoquinoline, benzylisoquinoline, bisbenzylisoquinoline, protobarbene, aporphine, oxoaporphine, phenantrene and miscellaneous isoquinoline type alkaloids.

Alkaloids are grouped in three main categories based on both of knowledge and speculation about their biogenesis

a) True Alkaloid

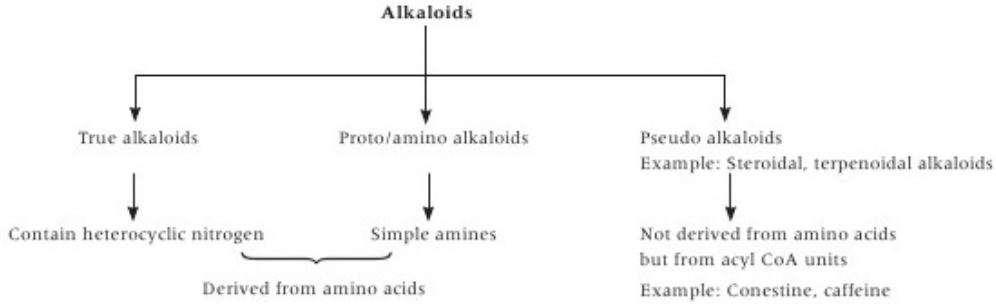
The true alkaloids are compound in which the nitrogen-containing heterocyclic system is derived from a biogenetic amine, formed by decarboxylation from an amino acid. They are usually found as salts in plant

b) Pseudo Alkaloid

Pseudo alkaloids are apparently unrelated to amino acid. They are nitrogen containing molecules but they have carbon skeletons derived from monoterpenes and other acetate derivatives and aliphatic polyketoacids

c) Proto Alkaloid

These compound like true alkaloids, are derived from amino acid or biogenetic amines but they do not contain any heterocyclic system. They are represented in nature by biogenetic amine themselves and their methylated derivatives



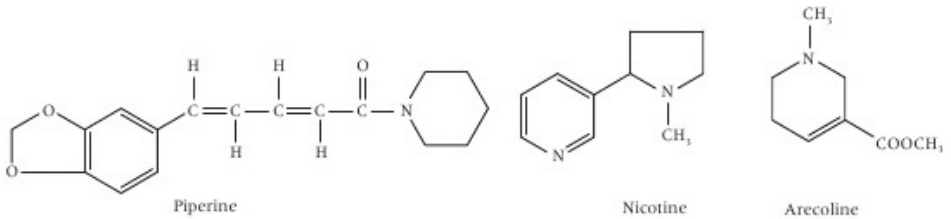
Based on the chemical nature, alkaloids are further classified into major groups as mentioned below.

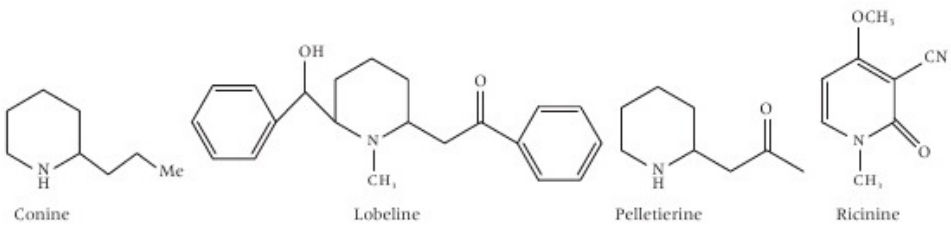
1. Heterocyclic or typical alkaloids
2. Nonheterocyclic or atypical alkaloids [protoalakloid or biological amines]

They are further subdivided as follows

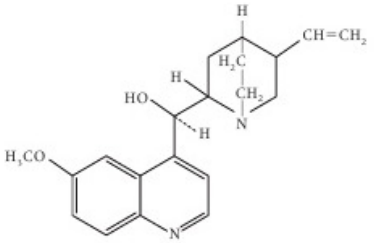
Heterocyclic Alkaloids

1. Pyridines and piperidines

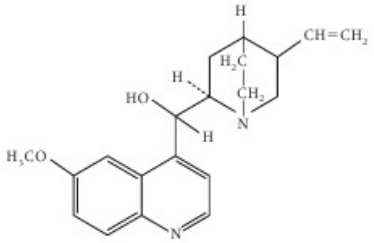




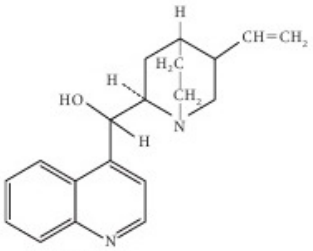
2. Quinolines



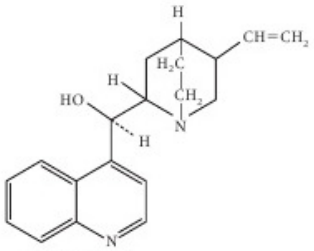
Quinine



Quinidine

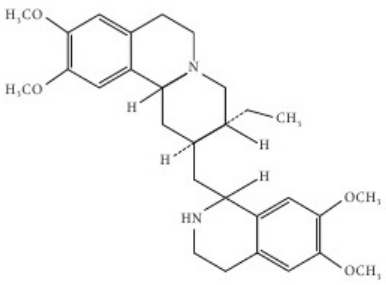


Cinchonine

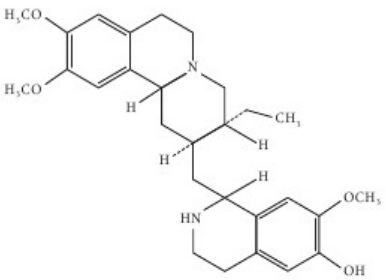


Cinchonidine

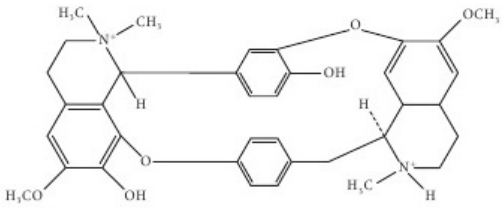
3. Isoquinolines



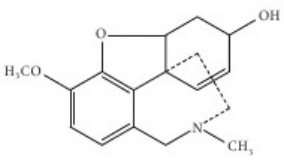
Emetine



Cephaline

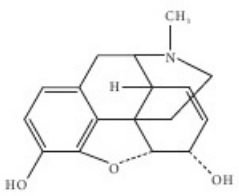


(±)-Tubocurarine

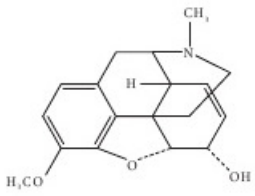


Galanthamine

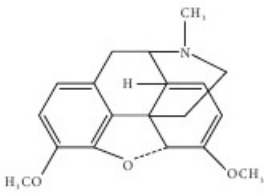
4. Phenanthrenes



Morphine

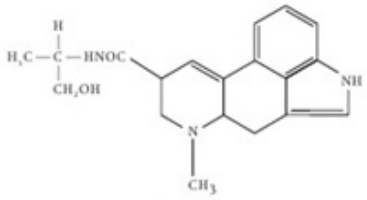


Codeine

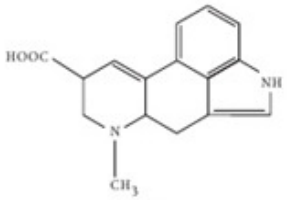


Thebaine

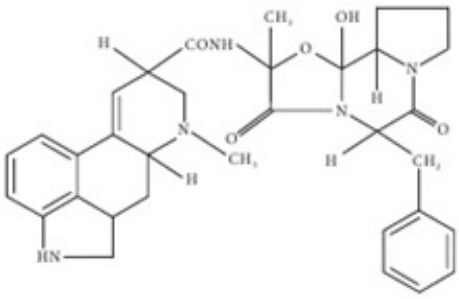
5. Indole alkaloids



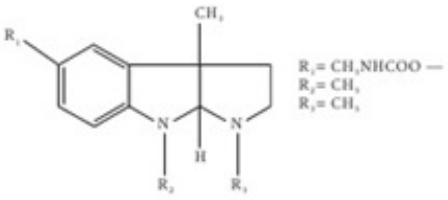
Ergometrine



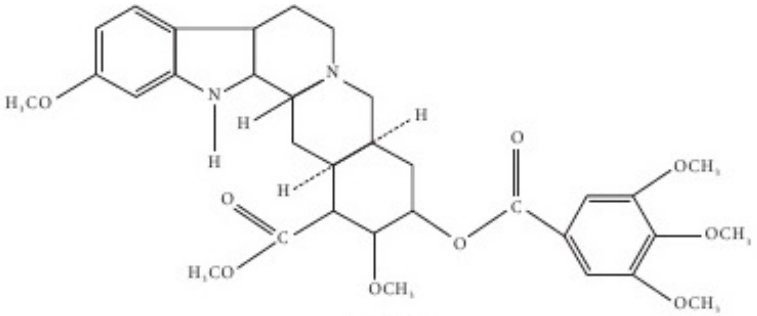
Lysergic acid



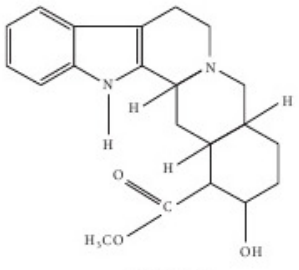
Ergotamine



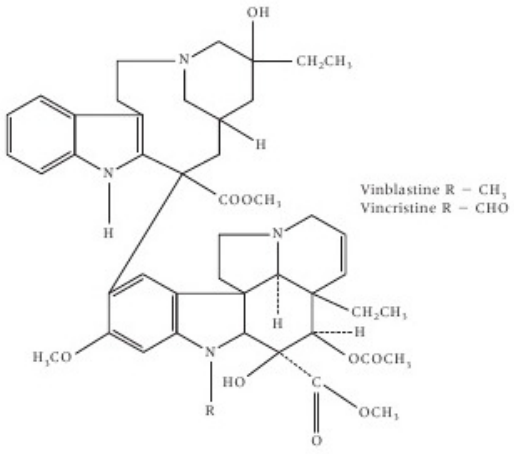
Physostigmine

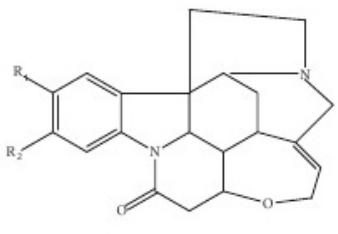


Reserpine



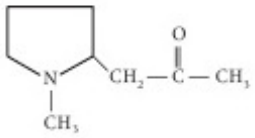
Yohimbine



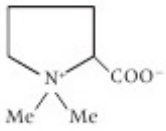


Strychnine $R_1 = R_2 = -H$
Brucine $R_1 = R_2 = -OCH_3$

6. Pyrrole and pyrrolidines

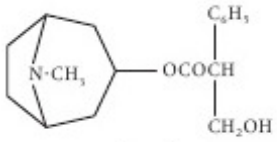


Hygrine

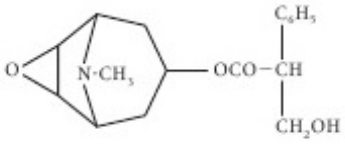


Stachydrine

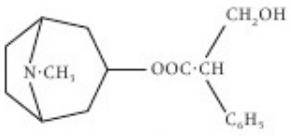
7. Tropane alkaloids



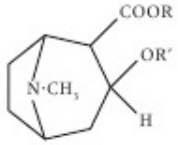
Atropine



Hyoscine

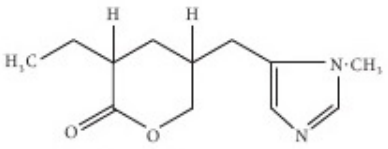


Hyoscyamine



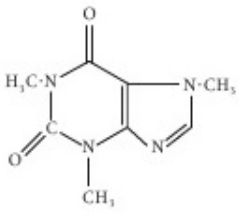
R= CH_3
R′= C_6H_5CO (Benzoyl)
Cocaine

8. Imidazole or glyoxalines

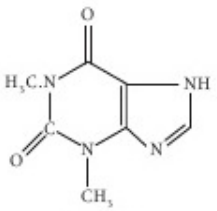


Pilocarpine

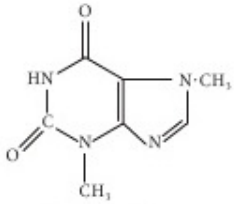
9. Purines



Caffeine

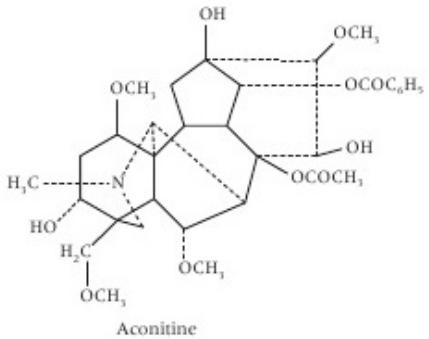


Theophylline



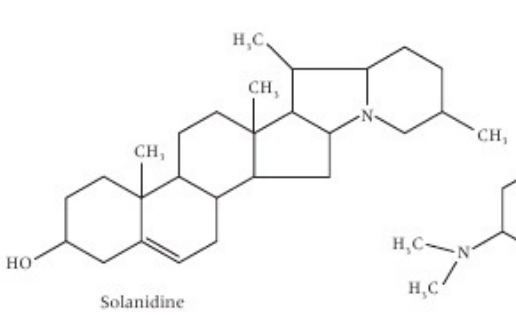
Theobromine

10. Terpenoid alkaloids

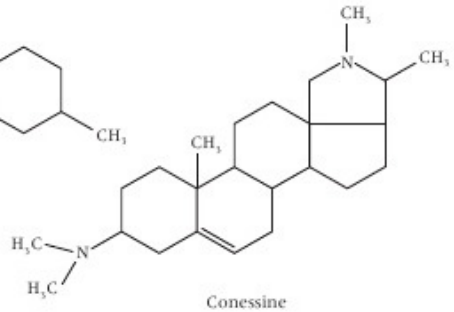


Aconitine

11. Steroidal alkaloids

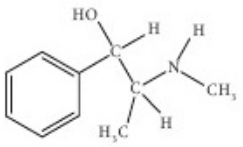


Solanidine

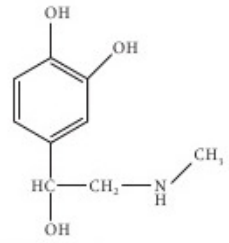


Conessine

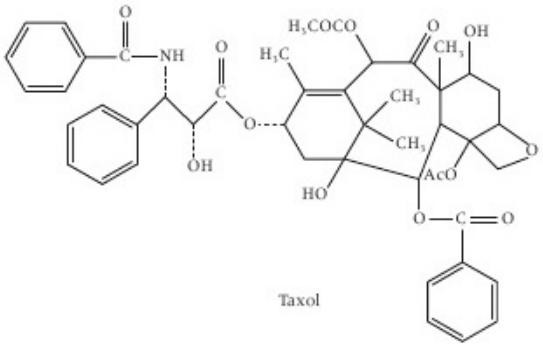
Nonheterocyclic Alkaloids



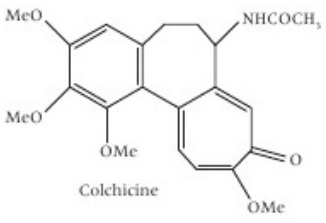
Ephedrine



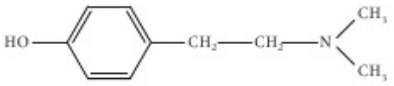
(±)-Adrenaline



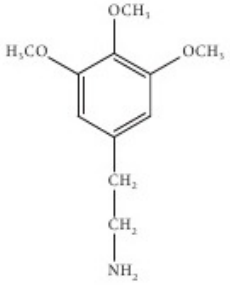
Taxol



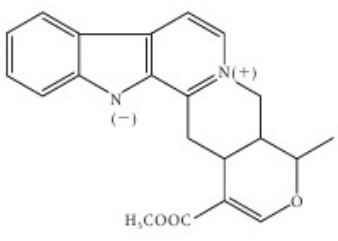
Colchicine



Hordenine



Mescaline



Serpentine

13.7 Qualitative chemical tests for alkaloids

1. Dragendroff's test: To-2-3 ml of the alkaloid solution add few drops of Dragendroff's reagent (potassium bismuth iodide solution). An orange brown precipitate is formed.
2. Mayer's test: To a 2-3 ml of the alkaloid solution add few drops of Mayer's reagent (potassium mercuric iodide solution). White brown precipitate is formed.
3. Hager's test: To a 2-3 ml of the alkaloid solution add few drops of Hager's reagent (saturated solution of piric acid) yellow precipitate is formed.
4. Wagner's test: To a 2-3 ml of the alkaloid solution add few drops of Wagner's reagent (iodine-potassiumiodide solution) redish brown precipitate is formed.
5. For opium alkaloids: These alkaloids are present as salts of meconic acid. Opium is dissolved in water, filtered and to the filtrate, ferric chloride solution is added by which deep reddish purple colour is obtained. the colour persists even upon adding hydrochloric acid.
6. for tropane alkaloids (Vitalis-Morin reaction): Tropane alkaloids is treated with fuming nitric acid, followed by evaporation to dryness and addition of methanolic potassium hydroxide solution to an acetone of nitrated residue. violet colouration takes place because of the presence of tropane derivative.
7. For purine alkaloids (Murexide colour reaction) caffeine is taken in a petri dish to which hydrochloric acid and potassium chlorate are added and heated to dryess. A purple colour is obtained by exposing the residue to vapour of dilute ammonia. The purple colour is lost upon addition of alkali. Caffeine (and other purine alkaloids) gives murexide colour reaction.

13.8 Structural elucidation of alkaloids

As the molecular structure of alkaloids is complex. The various chemical methods performed to determine the structure of alkaloids are as follows

13.8.1 Molecular formula determination

The first step in structural elucidation is the determination of molecular formula and optical rotatory power, elemental composition and hence the empirical formula is found by combustion analysis.

13.8.2 Hydrolysis

Frequently an alkaloid is cleaved into simple fragments by hydrolysis with water, acid or alkali, and the fragments so obtained are examined separately since the structure of the fragment may easily be established than that of the whole molecule.

For example, piperine on hydrolysis gives piperidine and piperic acid indicating that the two fragments are linked with each other by means of an acid amide linkage.

13.8.3 Determination of unsaturation

The presence of unsaturation in an alkaloid may be ascertained by the addition of bromine or halogen acids or by hydroxylation with alkaline permanganate.

13.8.4 Functional group determination

The next step involved in ascertaining the functional nature of oxygen and nitrogen atoms either in the molecule itself or in its fragments obtained.

By using the usual standard chemical tests or by infrared (IR) spectroscopy, functional nature of the alkaloids is determined

The oxygen atom may be present in the form of alcoholic or phenolic hydroxyl (-OH), methoxyl ($-OCH_3$), acetoxyl ($-OCOCH_3$), benzoxyl ($-OCOC_6H_5$), carboxyl (-COOH) aldehyde(-CHO), ketone (C=O) and methylene dioxide ($-O-CH_2-O-$) groups. These various oxygen functional groups can be characterized according to the following characteristics tests.

13.8.5 Phenolic hydroxyl group (=C-OH)

It is identified by the following tests.

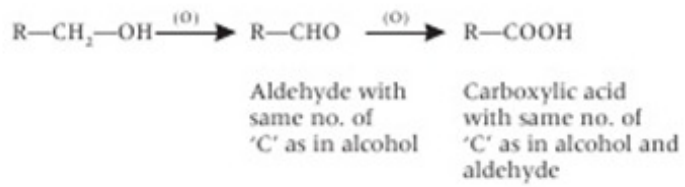
- Soluble in alkali and reprecipitation by CO_2.
- Violet colouration with neutral ferric chloride.
- Yields esters on acetylation. This reaction can be used to determine the number of phenolic –OH.
- Yields ether on reaction with alkyl halide.

13.8.6 Alcoholic hydroxyl group (-C-OH)

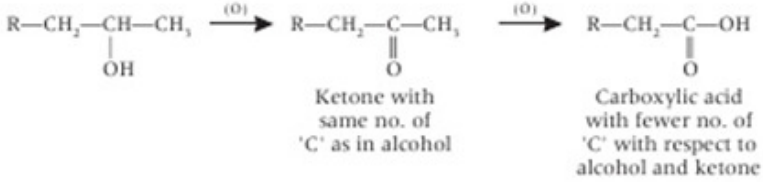
It yields ester on acetylation and benzoylation (but negative answer for phenolic –OH)

This is confirmed by oxidation, dehydration, dehydrogenation and by spectroscopy (IR and NMR). Alcohols are af three different types: 1^0, 2^0 and 3^0 and they are usually distinguished by their oxidation products.

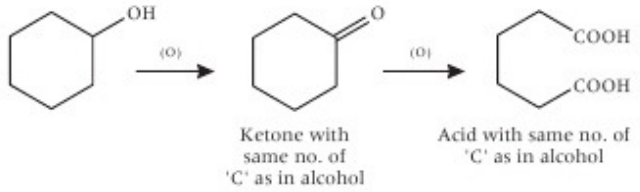
Primary alcohol



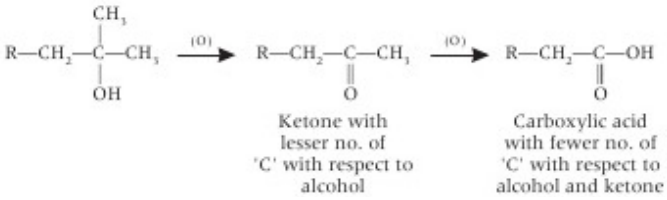
Secondary alcohol



But in cyclic structure, 2° alcohol yields different oxidation products as shown below:

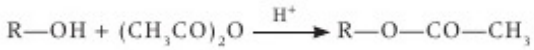


Tertiary alcohol



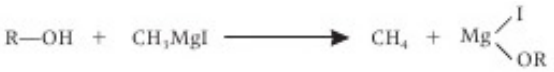
The number of hydroxyl (OH) groups present in the compound is determined by the following methods

Acetylation method:



By determining the amount of acetic anhydride that reacted with alcohol to form an ester, the number of hydroxyl groups is determined.

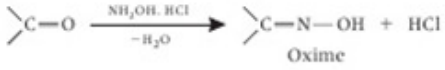
Zerewitinoff active hydrogen determination method: When alcohol is heated with CH_3MgI, methane is obtained, by measuring the methane so formed, the amount of alcohol can be determined.



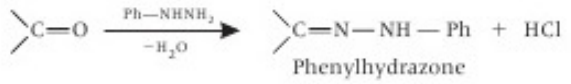
$-OH = CH_4 = 22.4$ L of alcohol at normal temperature and pressure.

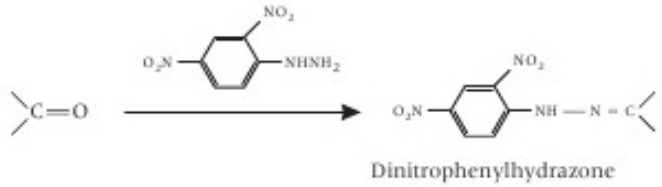
13.8.7 Cabonyl group

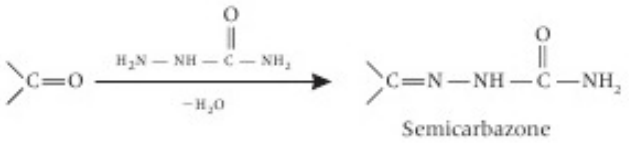
The presence of aldehydes and ketones is determined by their reaction with hydroxylamine, semicarbazide and phenylhydrazine to form the corresponding oxime, semicarbazone and phenylhydrazone, respectively.



By determining the HCl formed, the ketones are estimated quantitatively.







The aldehydes and ketones are distinguished by their oxidation or reduction products.

The carbonyl groups of alsdehydes, ketones and carboxyl are further confirmed by their spectral data such as IR UV and NMR.

(ii) Moderate oxidation is carried out by means of acid or alkaline potassium permanganate, or chromium trioxide in acetic acid.

(iii) Vigorous oxidation is generally carried out by potassium dichromate-sulfuric acid, chromium trioxide – sulfuric acid, concentrated nitric acid, or manganese dioxide – sulfuric acid system.

However, the above method is not rigid and the nature of an oxidizing agent depends upon the nature of the compound to be oxidized.

In some cases, where it is possible the compound should first be converted into unsaturated compound which is then oxidized at the double bond.

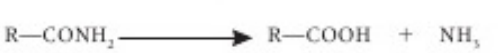
13.8.8 Carboxyl group (-COOH)

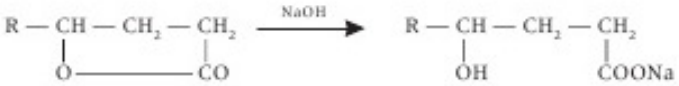
The presence of carboxyl group is determined by the following methods.

- Its solubility in weak bases such as NH_3, $NaHCO_3$, and Na_2CO_3.
- Esterification with alcohols.
- Specific IR and NMR signals.
- Quantitatively by acid-alkali titration: Performed by titrating the carboxylic acid with NaOH using phenolphthalein as an indicator. By knowing the volume of NaOH consumed the number of –COOH groups are determined.

13.8.9 Ester group (-COOR)

Esters and related groups like amides and lactones are detected by their reaction with water, dilute acids or alkali to the hydroxyl and acidic compounds. By elucidating the acid and alcohol, the nature of alkaloids is determined.

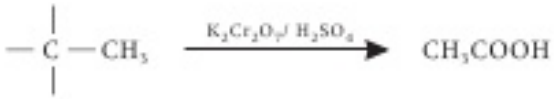




13.8.10 Alkoxy group (-OR)

Determined by Zeisel's method, alkoxy group such as methoxy on reacting with hydroiodic acid followed by silver nitrate yields equal amount of silver iodide. Form the amount of silver iodide formed, the number of alkoxy group is calculated.

Estimation of C-methyl group (Kuhn Roth method): By estimating the acetic acid formed upon oxidation, the C-methyl; groups are quantified.



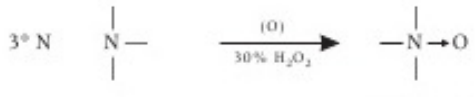
13.9 Functional nature if Nitrogen

Most alkaloids contain N in their ring structure, which may exist as 2^0or 3^0. The 2^0or 3^0 amines are distinguished as follows

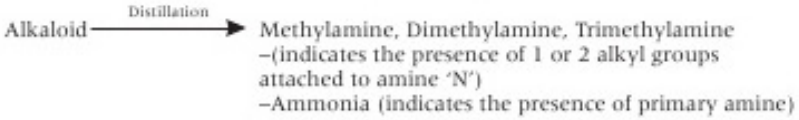
2^0 amines (acetylated or benzoylated) under go Libermann's nitroso reaction.

2^0 amines take up 2 moles of alkyl halides to form quaternary ammonium salt.

3^0 amines take up 1 mole of alkyl halides to form quaternary salt.

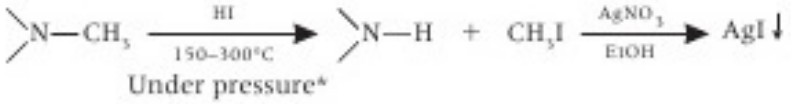


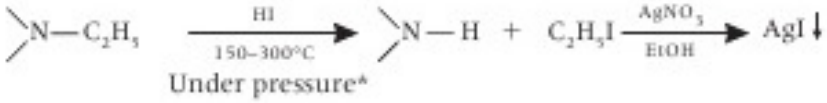
Amine oxide



Further the nature of N is confirmed by degradation methods auch as Hoffmann exhaustive methylation (HEM)

The N-alkyl groups are estimated by Herzig-Meyer method, which is differ form –OR(alkoxygroup) estimation.





From the amount of silver iodide formed, the number of N-alkyl grous is calculated.

13.10 Degradation of alkaloids

Degradation of alkaloids gives some identifiable product of known structure and hence by knowing structure of the degraded products and the changes ordered during the degradation it is convenient to know the structure of the original molecule. Different degradation reactions carried out in elucidating the structure of alkaloids are as follows

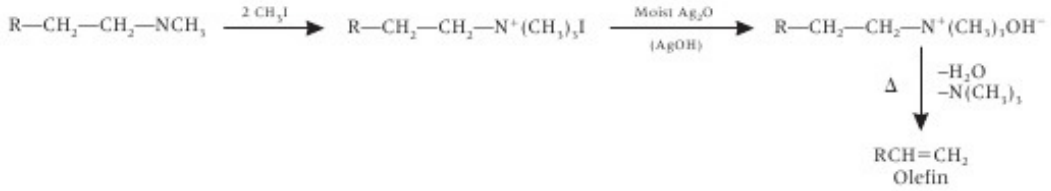
1. HEM method
2. Emde method
3. Von Braun's (VB) method for 3^0 cyclic amines
4. Reductive degradation
5. Oxidation
6. Zinc distillation
7. Alkali fusion
8. Dehydrogenation

13.10.1 HEM method

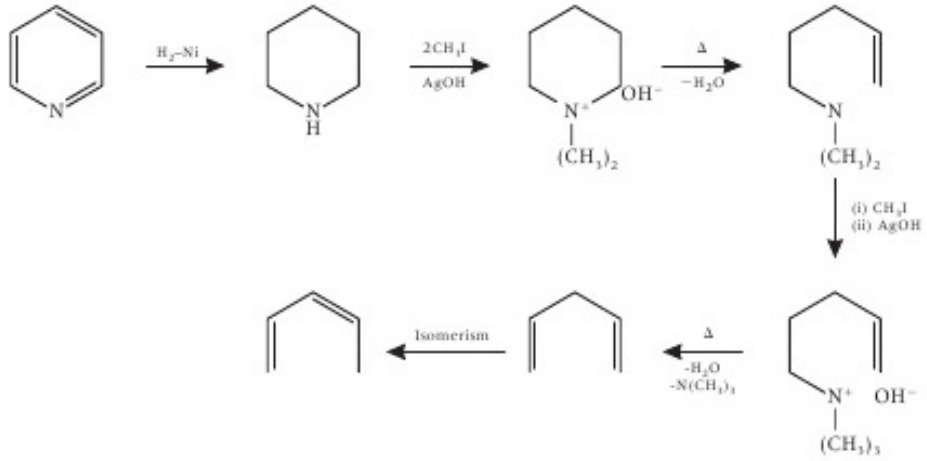
Originally this method was applied by Willstater in 1870 for naturally occurring alkaloids. It was further developed by Hoffmann and hence it is known as HEM.

Principle of this method is that the quaternary ammonium hydroxides yields olefin with the cleavage of carbon-nitrogen linkage upon heating with the loss of water molecule (H from β-carbon atom with respect to N and OH from the 4^0 ammonium hydroxide)

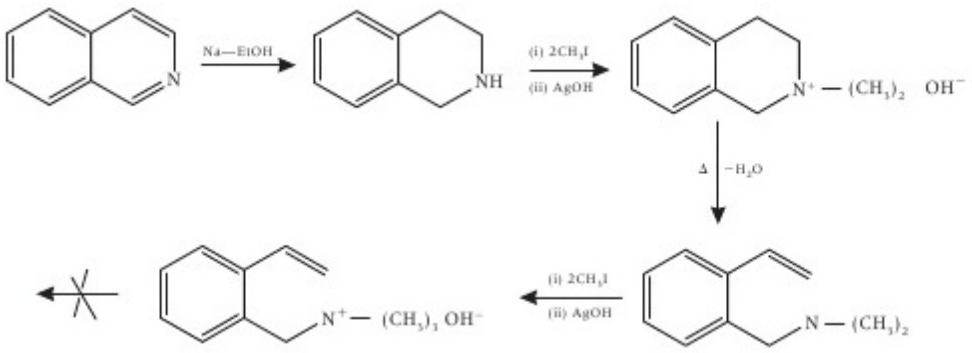
quaternization is done by complete methylation of the amine followed by hydrolysis with moist Ag_2O or KOH



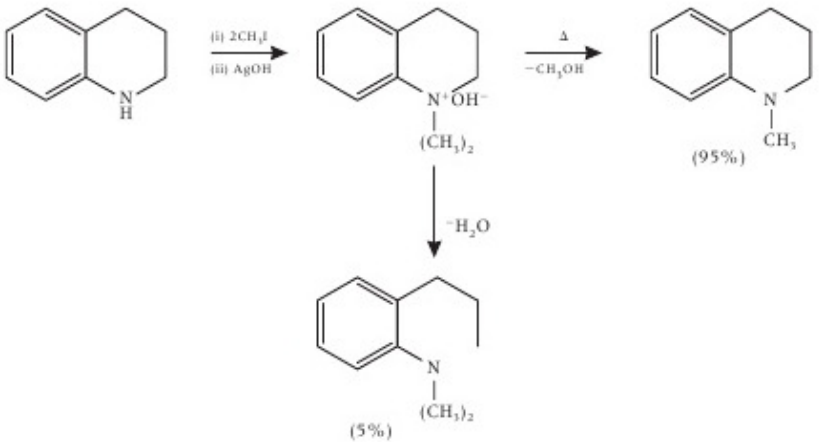
This method can be applied to the reduced ring system but fails with unsaturated analogues and hence the unsaturated rings are first saturated and then HEM is performed.



As β-hydrogen atom is needed to cleave C-N bond and eliminate water molecule, the HEM fails on the ring system that does not have β-hydrogen atom. For example, the degradation of isoquinoline, the cleavage of N does not occur at the final step as there is no β-hydrogen with respect to N.

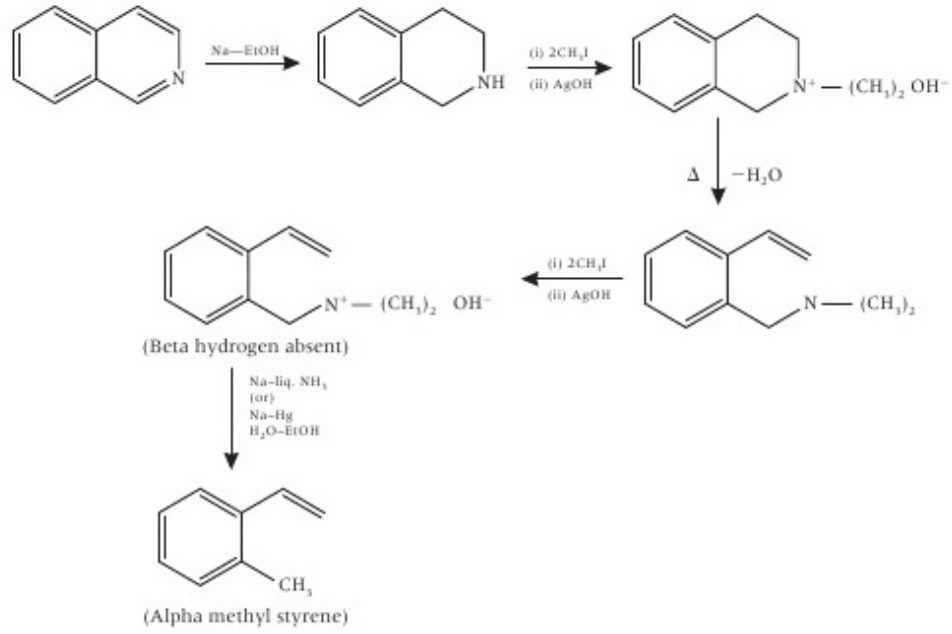


However, there are some cases in which HEM fails even if the β-hydrogen atom is present (the following reaction explain this)

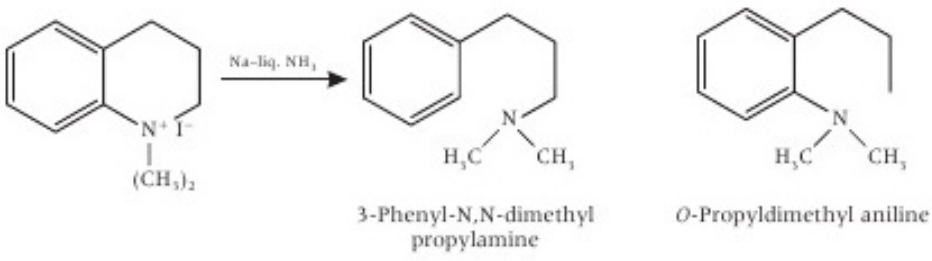


13.10. 2 Emde method

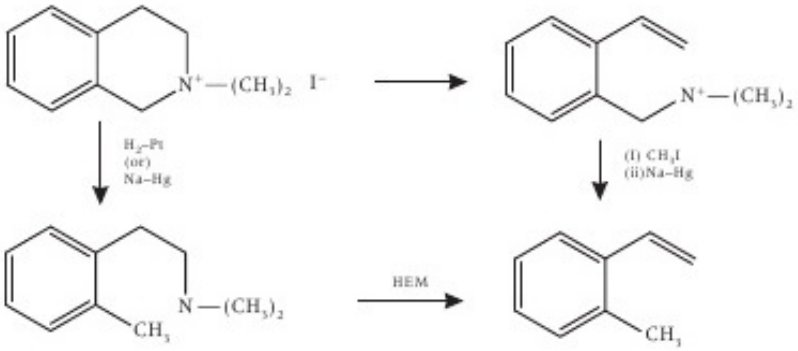
Emde modification may be used in the cases where HEM failed. In this method, 4^0 ammonium halide is reduced with sodium amalgam in aqueous ethanol or Na-liquid NH_3 or catalytically



Tetrahydroquinoline is degraded as follows

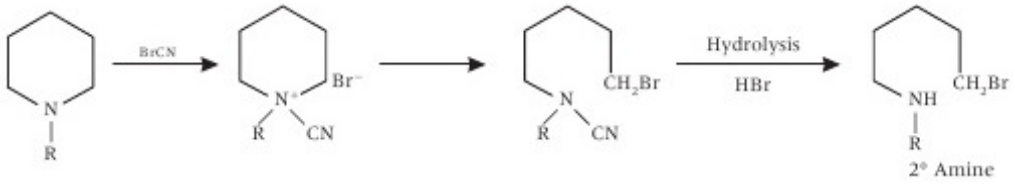


Emde degradation on tetrahydroisoqunoline also proceeds as follows

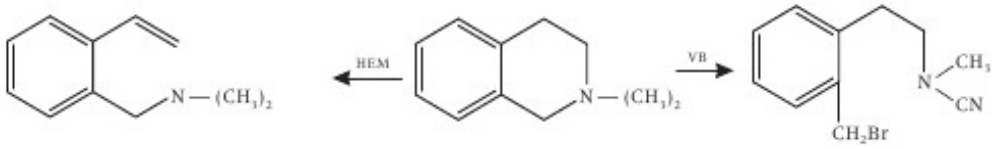


13.10. 3 VB method

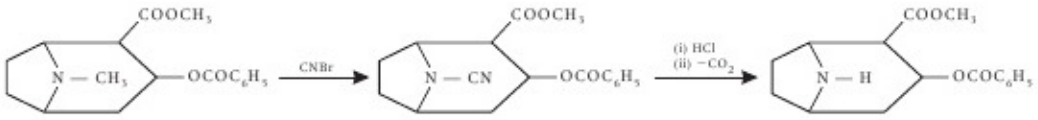
a) For 3^0 cyclic amines: The 3^0 N atom in the ring upon reaction with CNBr followed by hydrolysis yields brominated 2^0 amine.



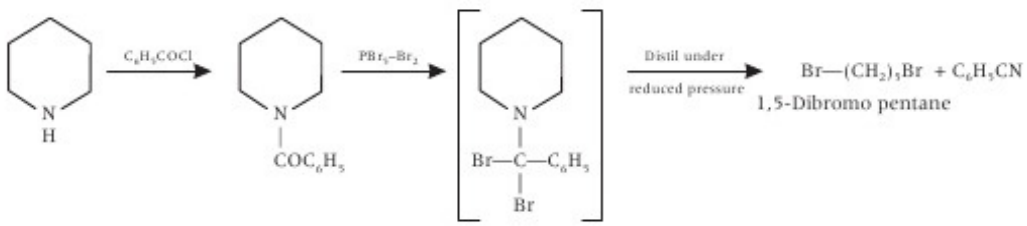
This method is applied on compounds which do not respond to HEM. Ring opening takes place differently in VB and HEM method which is shown in the following degradation.



In general CNBr cleaves the unsymmetrical amines to yields the bromides or shorter bromides. However, in the VB method only dealkylation may occur without ring cleavage in some cases.

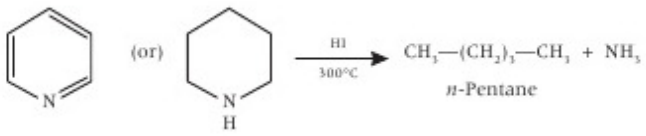


b) For 2^0 cyclic amines



13.10.4 Reductive degradation

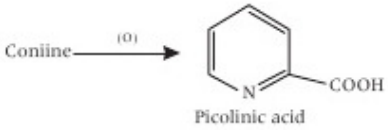
Ring system is opened by treatment with HI in many cases



13.10.5 Oxidation

Oxidation gives valuable information about the fundamental structure of alkaloids like the position and the nature of functional groups, side chains etc…

For example, picolinic acid obtained upon oxidation of coniine indicate that the coniine is an α-substituted pyridine.

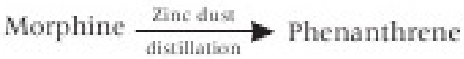


By varying the strength of oxidizing agents, a variety of products may be obtained. Different types of oxidizing agents used are as follows

1. for mild oxidation: H_2O_2, O_3, I_2
2. for moderate oxidation: acid or alkali $KMnO_4$, CrO_3 in CH_3COOH.
3. for vigorous oxidation: $K_2Cr_2O_7$-H_2SO_4, concentrated HNO_3, or MnO_2-H_2SO_4.

13.10. 6 Zinc distillation

Distillation of alkaloids over Zinc dust degrades it in to a stable aromatic derivative.

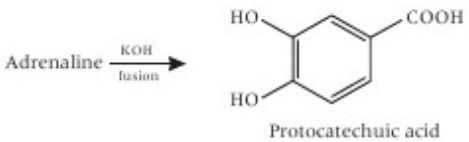


The reaction indicates that morphine is possessing phenanthrene nucleus.

13.10. 7 Alkali fusion

Fusion of alkaloids with KOH gives simple fragments, from them nature of alkaloid can be derived.

The reaction indicates papaverine is containing isoquinoline nucleus.



The reaction indicates adrenaline is a monosubstituted catechol derivative.

13.10.8 Dehydrogenation

Distillation of alkaloids with catalysts such as S, Se and Pd yields simple and recognizable products from which the gross skeleton of the alkaloid may be derived.

Thus with the help of degradation, nature of various fragments obtained, nature of nucleus and type of linkages are established. The fragments obtained are arranged in the possible ways with the possible linkages and the one that will explained all the properties is selected and confirmed by synthesis. Optical activity of an alkaloid helps greatly in establishing the structure of alkaloid.

13.11 Physical methods for elucidation of the structure

The development of spectroscopic methods has revolutionized the elucidation of structure of complex alkaloids. Analysis of spectra replaced the most of the extensive and time-consuming chemical degradation methods. For example, the complete structure of vindoline (an indole alkaloid) starting from the nature of functional group to the complete structure and even configuration has been established by the various spectral studies.

The important physical techniques used for elucidating the structures of alkaloids are

1. Infrared spectroscopy
2. Ultraviolet spectroscopy.
3. NMR spectroscopy
4. Mass spectroscopy.
5. X-ray analysis.
6. Optical rotator dispersion (ORD) and circular dichroism (CD)
7. Conformational analysis

13.11.1 IR spectroscopy

Infrared spectral studies are generally used to identify the presence of various functional groups usch as –OH, -NH_2, -NH, and –C=O. the groups such as –OCH_3, -NCH_3,-OH, -NH_2, and –NH can be detected by IR spectroscopy but quantified by NMR spectroscopy.

13.11.2 NMR spectroscopy

This method helps to detect protons of alkyl, alkenyl, N-methyl, O-methyl, C-methyl, aryl and heteroaryl groups etc. It is also helps in quantitative estimation of these groups. Aromatic and heteroaromatic protons are exactly quantified by using NMR spectroscopy.

13.11.3 UV spectroscopy

UV spectrum of a compound is characteristic of chromophoric system and not the whole compound. Hence, it helps to establish the likely structural type or class of the alkaloid under investigation.

13.11.4 Mass spectroscopy

This method is used to confirm the proposed molecular structure of the alkaloid by determining the molecular weight of compounds and the fragments of the degradation

products. It also helps to confirm the side chain or attached groups by analysis of the fragmentation pattern.

13.11.5 X-ray analysis

This method is used to distinguish the various possible structures of alkaloids

13.11.6 ORD and CD

This method is used to confirm the structure of optically active stereoisomers.

13.11.7 Conformation analysis

It is an experimental technique used to establish the stereochemistry as well as physical properties and chemical reactivity of alkaloids.

13.12 Structural elucidation of atropine

The structural elucidation of alkaloid using chemical fragmentation method is illustrated below taking tropane alkolids as an example.

Atropine is a naturally occurring belladonna alkaloid that is extracted from the belladonna plant. It is the racemic mixture of L-hyoscyamine and hence it can also called as (±) hyoscyamine. It is the tropine ester of racemic tropic acid and occurs mainly in the roots of deadly nightshade (Atropa belladonna), thron apple (Daturs stramonium) of the solanaceae family along with L-hyoscyamine.

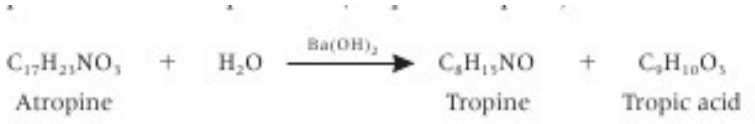
It is a white crystalline powder or colourless crystals. Freely soluble in alcohol and well soluble in water.

It is a competitive antagonist for the muscarinic acetylcoline receptor. Atropine reduces the parasympathetic activity of all muscles and regulatd by the parasympathetic nervous system through muscarinic (M) receptors. M_1 and M_3 receptors function through Gq protein and active membrane-bound phospholipase C, generating inositol triphosphate and diacylglycerol which release calcium ions to produce depolarization in glands and smooth muscles.

The structure of atropine is established as follows

1. *Molecular formula:* The molecular formula of atropine is found to be $C_{17}H_{23}NO_3$.

2. *Action of alkali:* Atropine on hydrolysis yields tropine and (±) tropic acid. Therefore atropine is a tropine ester of tropic acid (Tropinetropate).

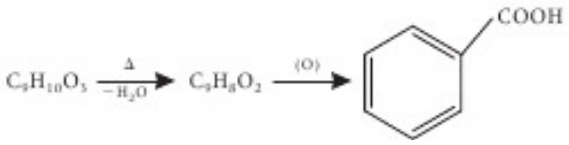


Therefore the structure of atropine will be resolved by solving the structure of tropane and tropic acid

13.12.1 Structure of tropic acid

Results of the usual standard test reveal that tropic acid is found to possess one 1^0 alcohol (-OH) and –COOH group.

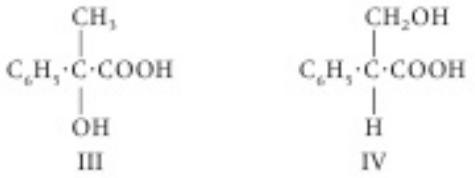
Upon heating tropic acid is converted in to atropicacid, which on further oxidation yields benzoic acid. This indicates atropic acid and tropic acid have benzene ring with a side chain.



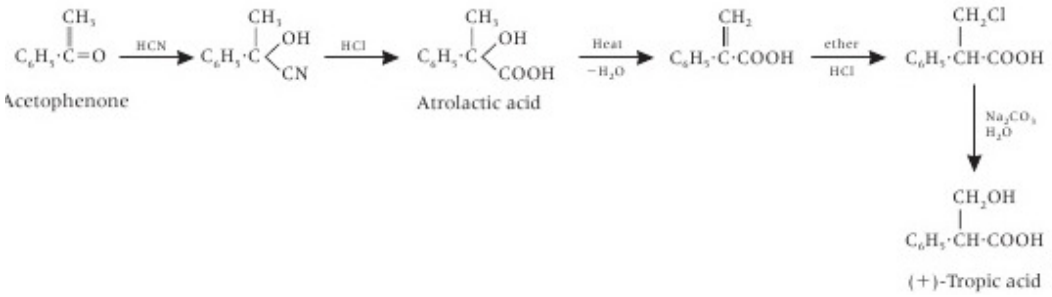
Therefore atropic acid is having one –COOH, one double bond and one benzene nucleus.

Structure II is known to be cinnamic acid hence I is atropic acid.

As astropic acid is obtained by behydration of tropic acid, addition of water molecule to atropic acid yields tropic acid- compound III and IV.



The structure (IV) is found be be correct, as it has 1^0 alcohol. The structure of tropic acid is confirmed by its synthesis



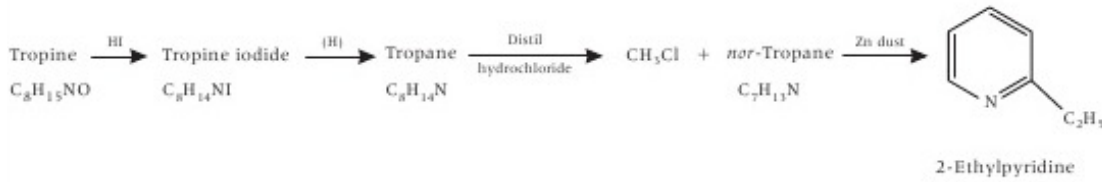
13.12.2 Structure of Tropine

The results of the usual standard tests reveal that N atom is found to be present as 3^0 N.

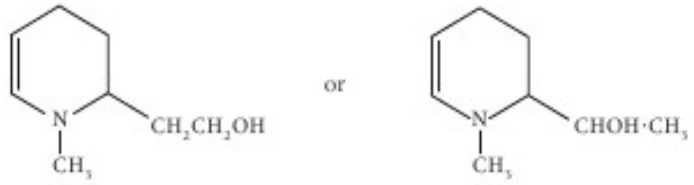
By Herzig-Meyers method, it is found tropine possess one –N-CH_3 grouop.

Results of the usual standard tests reveals that it is found to possess one 2^0 alcohol (-OH) group (benzoylation and oxidation reaction)

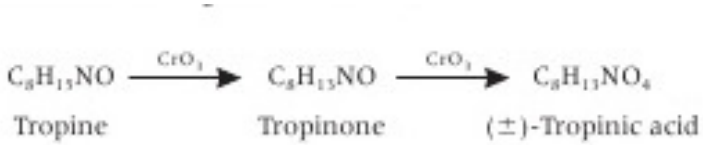
Ladenburg performed the following reactions on tropine.



On this basis, Landenburg proposed that tropine is a reduce pyridine derivative.

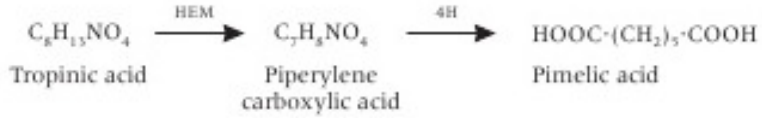


Oxidation of tropine: Tropimic acid is a dicarboxylic acid obtained upon oxidation and possess same number of carbon atom as in alcohol, hence the alcoholic group in the tropine must be present in ring structure and hence Ladenburg structure is discarded.

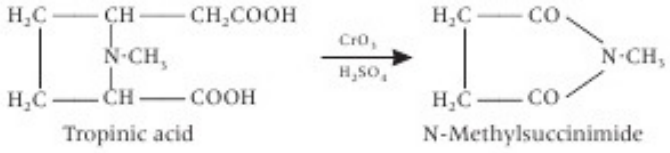


Furthermore, it found that tropinone yields dibenxyidene derivative which is a characteristic of $–CH_2\text{-}CO\text{-}CH_2$-. This confirms that the ketone and alcohol in tropine is linked to a ring structure.

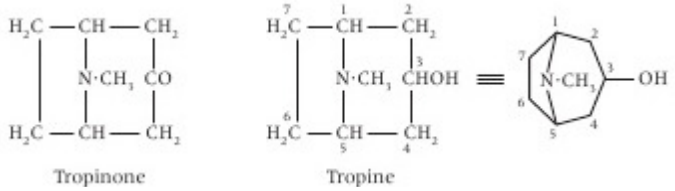
In HEM, tropinic acid yields pimelic acid. Formation of pimelic aicd confirms the presence of seven-membered carbon chain in tropinic acid and tropine.



Presence of five-membered ring is confirmed on the basis of the formation of N-methylsuccinimide from tropic acid on oxidation.



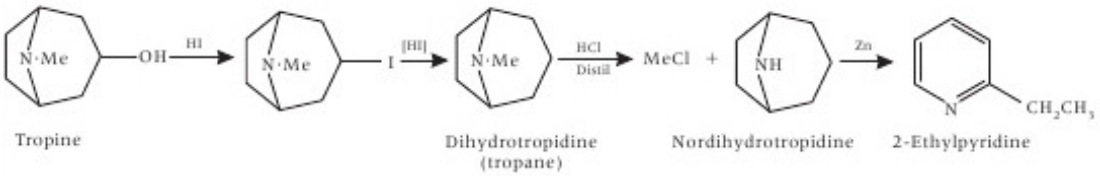
Thus the structure of tropinone and tropine can be proposed as follows



Tropinone Tropine

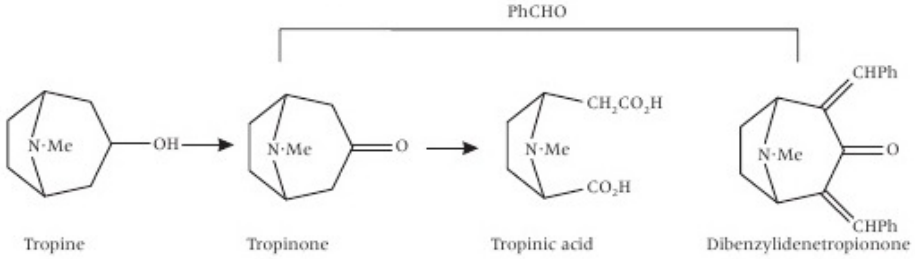
All the foregoing reactions are explained with the above structure as mentioned below

Formation of 2-ethyl pyridine from tropine



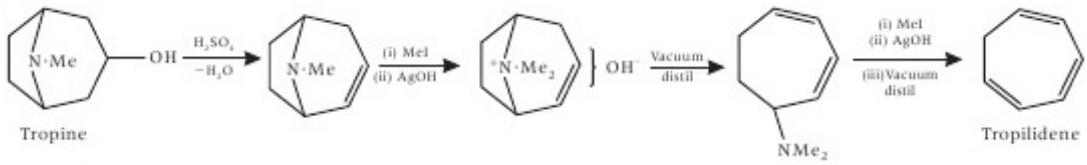
Tropine Dihydrotropidine (tropane) Nordihydrotropidine 2-Ethylpyridine

Froamtion of tripinone and tropic acid from tropine



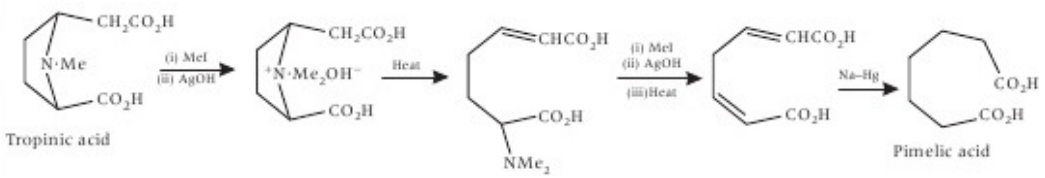
Tropine Tropinone Tropinic acid Dibenzylidenetropionone

Fromation of tropilidene (Cycloheptatriene) from tropine



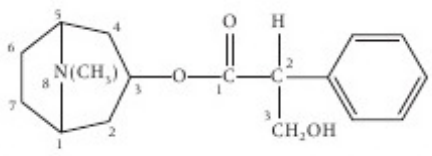
Tropine Tropilidene

Formation of pimelic acid from tropinic acid

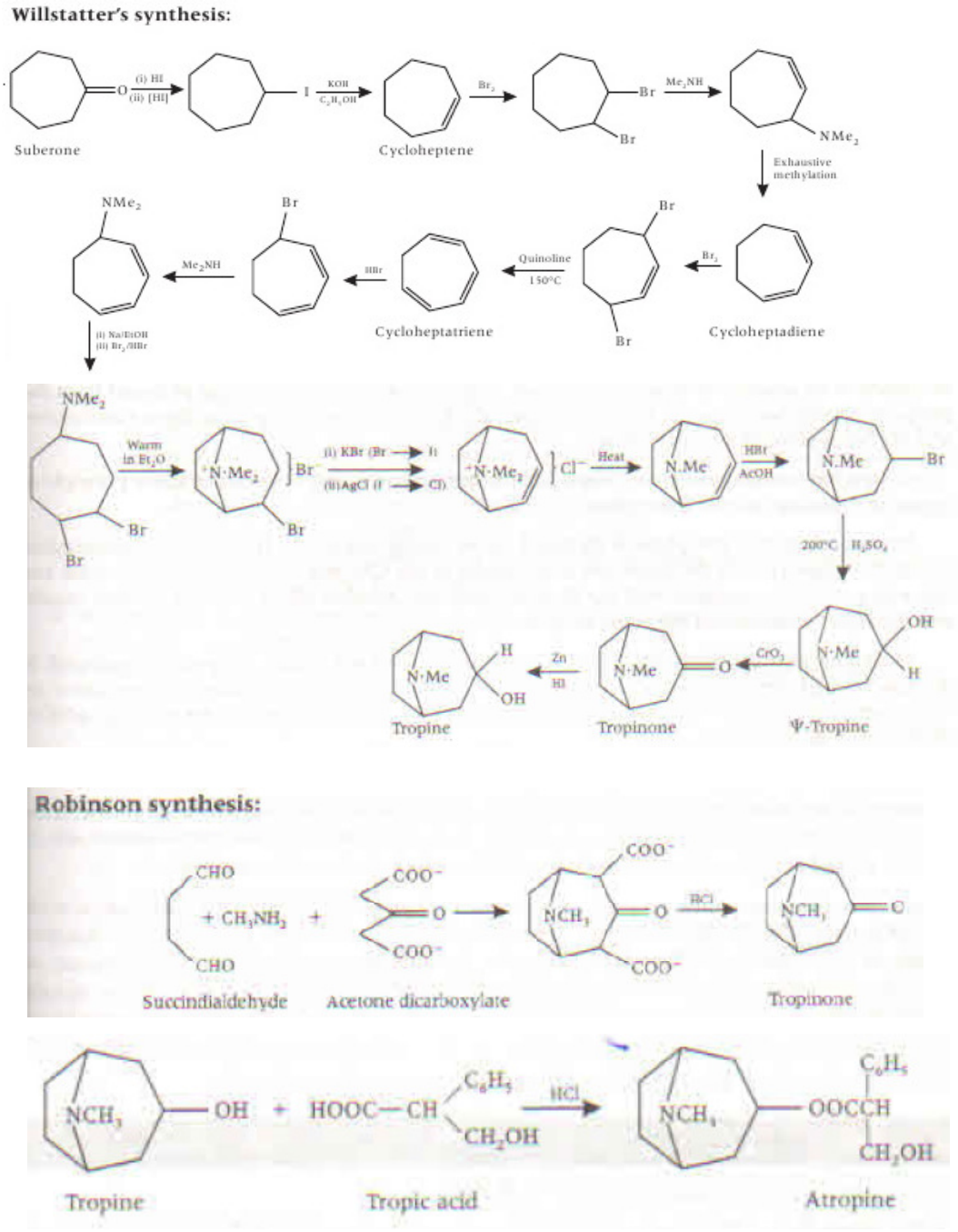


Tropinic acid Pimelic acid

The proposed structure of tropine is shown below and is confirmed by the synthesis



8-Methyl-8-aza-bicyclo[3.2.1]octan-3-yl-3-hydroxy-2-phenylpropanoate



13.13 Summary of the unit

Alkaloids are naturally occurring chemical compounds containing basic nitrogen atoms. The name derives from the word alkaline and was used to describe any nitrogen containing base. Alkaloids are produced by a large variety of organisms, including bacteria, fungi, plants, and animals and are part of the group of natural products (also called secondary metabolites). Many alkaloids are purified from crude extracts by acid-base extraction. Many alkaloids are toxic to other organisms. They often have pharmacological effects and are used as

medications, as recreational drugs, or in entheogenic rituals.Examples is the local anesthetic and stimulant cocaine, the stimulant caffeine, nicotine, the analgesic morphine, or the antimalarial drug quinine. Some alkaloids have a bitter taste. The classification of the alkaloids is complex and may be guided by a set of rules that take into account the structure and other chemical features of the alkaloid molecule, its biological origin, as well as the biogenetic origin where known.

13.14 Key words

Alkaloids; Characteristics of alkaloids; Isolation of alkaloids; Nomenclature of Alkaloids; Classification of alkaloids; Qualitative chemical tests for alkaloids; Structural elucidation of alkaloids; Degradation of alkaloids; HEM method; Emde method; VB method; Atropine; Tropic acid; Tropine

13.15 References for further studies

1) Pharmaceutical Chemistry of Natural Products; V. Alagarsamy; *Elsevier India*, **2012**.

2) Chemistry of Plant Natural Products: Stereochemistry, Conformation, Synthesis, Biology, and Medicine; Sunil Kumar Talapatra, Bani Talapatra; *Springer*, **2015**.

3) Chemistry of Natural Products; Sujata V. Bhat, B.A. Nagasampagi, Meenakshi Sivakumar; *Springer Science & Business Media*, **2005**.

4) Organic Chemistry; R.L. Madhan; *Tata McGraw-Hill Education*, **2013**.

5) Alkaloids; Ane's chemistry active series; Vk Ahluwalia; *Ane Books Pvt Ltd*, **2010**.

6) Chemistry of Alkaloids; P. B. Saxena; *Discovery Publishing House*, **2007**.

13.16 Questions for self understanding

1) What are Alkaloids? Give five examples.

2) Why alkaloids are called secondary metabolites?

3) List out the general characteristics of alkaloids.

4) Describe the general method of isolation of alkaloids from plants with schematic diagram.

5) Discuss the methods followed for nomenclature of Alkaloids

6) Discuss the different methods of classification of alkaloids

7) What are the qualitative chemical tests performed for detection of alkaloids? Explain each test

8) What are the different methods followed in structural elucidation of alkaloids

9) What information can we get form hydrolysis?

10) How presence of unsaturation is determined in alkaloids?

11) Explain how presence of Phenolic hydroxyl functional group (=C-OH) is determined in alkolids?

12) What are the tests performed for identification of presence of alcoholic hydroxyl functional group (-C-OH)?

13) How presence of cabonyl functional group is determined in alkaloids?

14) How carboxyl group (-COOH) and ester group (-COOR) are identified in alkaloids?

15) Explain the methods followed for identification of presence of alkoxy group (-OR) in alkaloids.

16) Discuss the functional nature of Nitrogen atom exhibit in alkaloids.

17) What is meant by degradation of alkaloids?

18) Discuss the HEM method with example.

19) Discuss the Emde method with example.

20) Explain the VB method with example.

21) Discuss the oxidation method used in structure determination of alkaloids.

22) Explain the porpoise of zinc distillation in structure determination of alkaloids.

23) Examine the kind of information obtained by following physical methods in elucidation of the structure of alkaloids

a) IR spectroscopy

b) NMR spectroscopy

c) UV spectroscopy

d) Mass spectroscopy

24) Write the alkali hydrolysis product of Atropine alkalide.

25) Discuss the structural elucidation of tropic acid

25) Discuss the structural elucidation of Tropine

UNIT-14

Structure

14.0 Objectives of the unit

After studying this unit you are able to

- Identify the biological importance of alkaloids
- Write the structure of quinine
- Elucidate the structure of Quininic acid
- write the total Synthesis of quinine
- Elucidate the structure of Morphine
- Write the total synthesis of Morphine
- Elucidate the structure of Reserpine
- Elucidate the structure of reserpic acid

14.1 Introduction

Alkaloids play a very important role in metabolism and functional activity. They are metabolic products in plants, animals and micro-organisms. They occur in both vertebrates and invertebrates as endogenous and exogenous compounds. Many of them have a distributing effect on the nervous systems of animals. Alkaloids are the oldest successfully used drugs throughout the historical treatment of many diseases.

Alkaloids are compounds needed for cell activity and gene code realization in the genotype. They are biologically significant as active stimulators, inhibitors and terminators of growth, a part of an endogenous security and regulation mechanism. Some alkaloids have significance as haemoglobinizators of leukaemia cells and they can be biologically determined to be estrogenically active molecules. They display antimicrobial and anti-parasitic properties. Recent research has proved that they are not toxic to the organisms that produce them. Biotoxicity is directed only towards foreign organisms or cells and it is selective.

Alkaloids can alter DNA, selectively deform cells and cause locoism. Some alkaloid molecules, both natural and synthetic, can act as narcotics. Moreover, they play a very important role in the immune systems of animals and plants. Alkaloid metabolism is genetically coded, and to date more than 30 genes coding for the enzymes involved in alkaloid synthesis have been isolated. Alkaloid molecules are active agents in evolutionary interactions

14.2 Alkaloids in biology

For many years, the nature of alkaloids in biology was a mystery. It has been difficult to understand the function of these compounds in plant metabolism. There are many explanations for why plants, animals and micro-organisms produce alkaloids. Despite the advanced research in the field, a final comprehensive biological explanation of the nature of

alkaloids is still on the way. In this sense the alkaloid mystery continues to exist. New compounds are being discovered all the time, however, their biological significance remains unexplained. Alkaloids are non-toxic in vacuoles where they are stored but toxic when they escape from the vacuoles. They have to change their chemical configurations and biological activity in different cells and tissues according to pH changes. This means that some alkaloids can have different biological activity in different cell

14.3 Uses of alkaloids

Alkaloids are mainly used

- As medicines.
- As Euphoric & addicting drugs
- As pesticides or insect repellents.
- For research and scientific study.
- To catch animals.

In medicine

Alkaloids are mainly used in medicine as health care agents. They act as life saving drugs in some serious disorders like heart-failure, cancer, blood pressure etc. for example

Digoxin is used for heart-failure. This helps to improve the heart function and maintain life. Vincristine from vinca roseus is used as anticancer drug. Ephedrine from ephedra is used in blood pressure. Codeine is used as cough suppressant. Ergot alkaloids are used as drugs to relieve migraine,

Euphoric & addicting drugs

Many psychotropic substances like marijuna, cannabis, opium are alkaloids. They have been used since ancient period as instruments for mental excitement and euphoria, they are considered harmful as per modern medicine.

Insect repellents and pesticides

The alkaloids are toxic in nature. They are toxic in varying concentrations. Alkaloid pyrethrin is used as insect repellent in mosquito coils, mosquito repellents and even in agriculture.

For research and scientific study

Due to their specific effects on body, they are used in research and scientific study. For example, atropine an alkaloid can cause dilation of pupil. To test if a new substance is having similar effects or opposite effects, it is compared with atropine. So here atropine is used as standard for comparison in research.

To catch animals

Heavy animals like elephants, were captured by making them immovable. For this the hunters use arrow poison (D-tubocurarine) which is applied to the point of arrow and shot at animal. When the arrow gets stuck to the animal, the alkaloid (D-TC) will enter the muscle and cause paralysis. This is a short term action and is reversible. Since the animals are immovable, they can be trapped and transported. The same can be used for deer, wild animals etc.

Besides the above, alkaloids are produced by plants and large trees. They are waste products of body metabolism. They are excreted through leaves, fruits etc. They are toxic in nature, so they help the plant by preventing animals from grazing on them. Thus these alkaloids save the plants from extensive grazing by cattle. Due to bitter taste and other poisonous effects in the body, animals do not prefer to feed on plants containing alkaloids

14.4 Biological importance of alkaloids

Alkaloids in plants play significant role as growth stimulators, inhibitors, protective agents and reservoirs of nitrogen. Some alkaloids are known as neurotransmitters in animals and can also be considered part of the signalling system.

Alkaloids have their own signalling system. Receptors and membranes play an active role in this system. The role of biological membranes in alkaloid signalling is connected with the action of the specific ion channels of Ca^{2+}, Na^{+} and K^{+} and their active pumps. Only alkaloids can promote or inhibit activity of ion channels and their active pumps. Therefore, these channels are important in an alkaloid signalling system.

Alkaloids have strong antimicrobial, antibacterial and antifungal biological properties.

Anti-parasitic activity

A parasite is an organism living in or on, and metabolically depending on, another organism. Endoparasites live inside an organism, and ectoparasites live on the surface of the host. Parasites can be carnivorous if living with animals or herbivorous if living with plants.

Biotoxicity

Many alkaloids are toxic to foreign organisms. Toxicity is a secondary function of the alkaloids, because they are generally non-toxic to the organisms producing them. This is very important for understanding alkaloid nature. biotoxicity of alkaloids is selective and dependent on different organisms and the chemical structure of alkaloids themselves.

Alkaloids generally exert pharmacological activity particularly in mammals as humans. Even today many of our commonly used drugs are alkaloids from natural sources and new alkaloidal drugs are still developed for clinical use. Most alkaloid with biological activity in humans effect the nervous system, particularly the action of the chemical transmitters,

e.g, acetylcholine, epinephrine, norepinephrine, γ-aminobutric acid (GABA), dopamine and serotonin.

Many alkaloids serve as models for the chemical synthesis of analogues with better properties. For examples, hyoscyamine and scopolamine as models for synthetic parasympatholytic agents.

Antibiotic activities are common for alkaloids and some are even used as antiseptics in medicine. For example, berberine in ophthalmics and sanguinarine in toothpastes.

Many medically useful alkaloids act by way of the peripheral nervous system, others work directly on the brain. Prominent among the latter are the pain relievers morphine and codeine, derived from the opium poppy (Papaver somniferum). Morphine is the stronger of the two, but codeine is often prescribed for moderate pain. Codeine is also an effective cough suppressant. Both morphine and codeine are addictive drugs that produce a state of relaxed, dreamy euphoria—an exaggerated state of "feeling good," referred to by drug addicts as a "high."

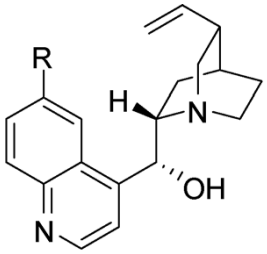
The effects of cocaine are almost the opposite of those of morphine. Cocaine's are classified as a narcotic drug that produces stupor. This produces a state of euphoric hyperarousal. Upon intake of this, user feels excited, elated, and intensely aware of his or her surroundings. These feelings are accompanied by the physical signs of arousal, elevated heart rate and blood pressure. The increased heart rate caused by a high dose may lead to fibrillation and death.

Another pleasurable yet addictive drug is nicotine, usually obtained by either smoking or chewing leaves of the tobacco plant, Nicotiana tabacum.

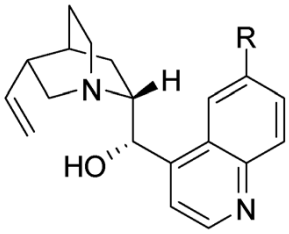
Nearly all of the alkaloids are poisonous in large amounts. Some alkaloids are almost solely known as poisons. One of these is strychnine, derived from the small Hawaiian tree Strychnos nux-vomica. Symptoms of strychnine poisoning begin with feelings of restlessness and anxiety, proceeding to muscle twitching and exaggerated reflexes.

14.5 Quinine

Cinchona alkaloids isolated from the bark of several species of cinchona trees, are the organic molecules with the most colourful biography. They were first introduced into the European market as antimalarial compounds. The active compound, quinine was isolated in 1820 by Pierre-Joseph Pelletier and Joseph Bienaime Caventou. Since then, cinchona alkaloids especially, quinine have played a pivotal medicinal role in human society.



R = H
Cinchonidine
R = OMe
Quinine



R = H
Cinchonine
R = OMe
Quinidine

Quinine and other cinchona alkaloids are extracted from the bark of the cinchona tree

14.5.1 Isolation of quinine

It is isolated from finely ground cinchona bark mixed with lime. The mixture is extracted with hot, high boiling paraffin oil. The solution is filtered shaken with dilute H_2SO_4 and then neutralizes with hot sodium carbonate. Upon cooling, quinine sulphate crystallizes out. Pure quinine is obtained by treating the sulphate with ammonia.

14.5.2 Physiological uses of quinine

Quinine is used in medicines as an anti-malarial drug. In addition quinine and its salts are used in soft drinks and other beverages.

14.5.3 Structure of quinine

The structure of quinine has been established as follows

The molecular formula of quinine $C_{20}H_{24}O_2N_2$

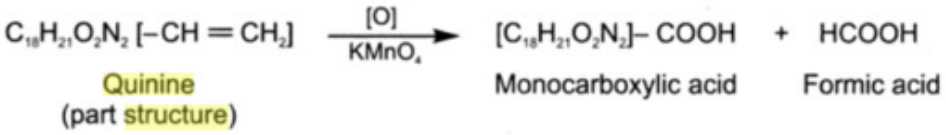
Quinine has one hydroxyl group since it gives a monoacetate and a monobenzoate.

Quinine on oxidation with chromium trioxide gives a ketone, quinine, $C_{20}H_{22}N_2O_2$ (i.e CHOH→CO) therefore the hydroxyl group is secondary.

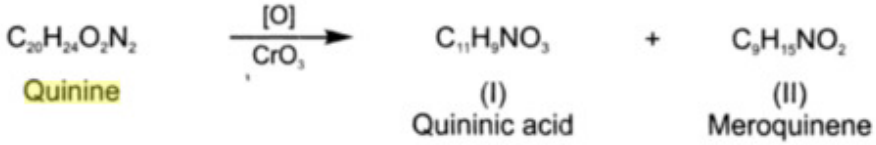
It adds on two molecules of methyl iodide to form a diquaternary salt $C_{20}H_{24}O_2N_2.2CH_3I$. Therefore both the nitrogen in quinine are tertiary.

Quinine also has one ethylenic double bond since it adds on one molecules of hydrogen bromine or halogen acid.

Quinine on oxidation with $KMNO_4$ gives a mono carboxylic acid and formic acid. Hence the ethylenic double bond is present as vinyl group.



Controlled oxidation of quinine with chromic acid gives quininic acid $C_{11}H_9NO_3$(I) and a compound designated as the second half and called meroquinine $C_9H_{15}NO_2$(II)

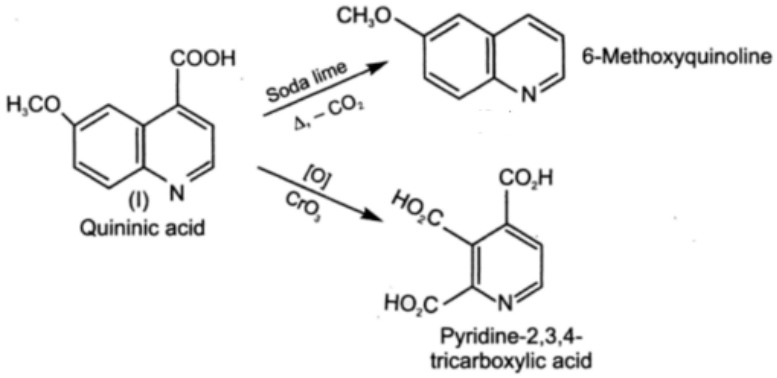


Solving the structure of quininic acid (I) and meroquinene (II) is important to elucidate the structure of quinine.

14.5.4 Quininic acid

Molecular formul of quininic acid is $C_{11}H_9NO_3$

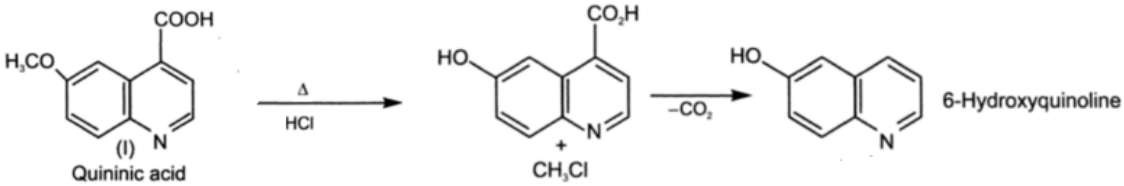
On heating with soda lime quininic acid undergoes decarboxylation yield a methoxyquinoline identified as 6-methoxyquinoline.



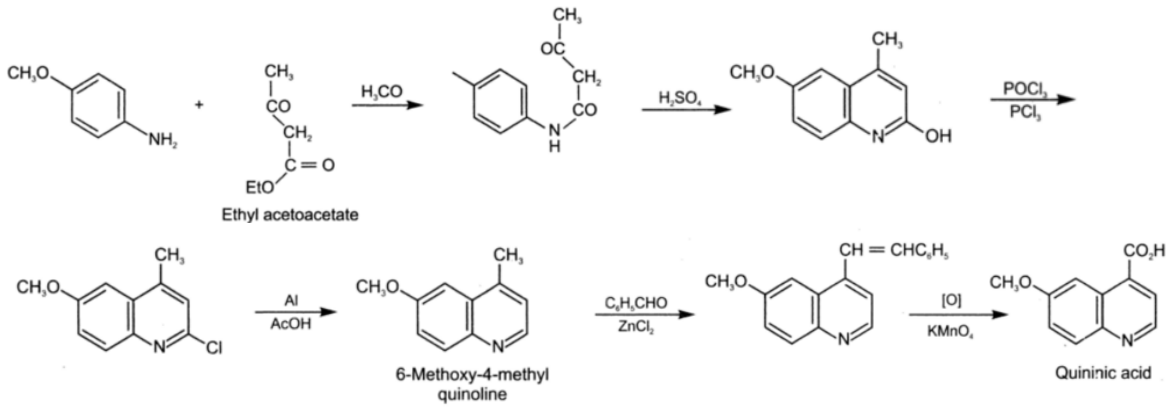
This confirms that quinine acid has a quinoline nucleus.

Further quininic acid on oxidation with chromic acid gives pyridine-2,3,4-tricarboxylic acid. This proves the presence of methoxyl group in benzene ring (of quinoline) and carboxyl group in position 4.

Quininic acid on heating with hydrochloric acid undergoes demethylation and decarboxylation to give 6-hydroxy quinoline a known product. Thus quininic acid is 6-ethoxycinchonic acid



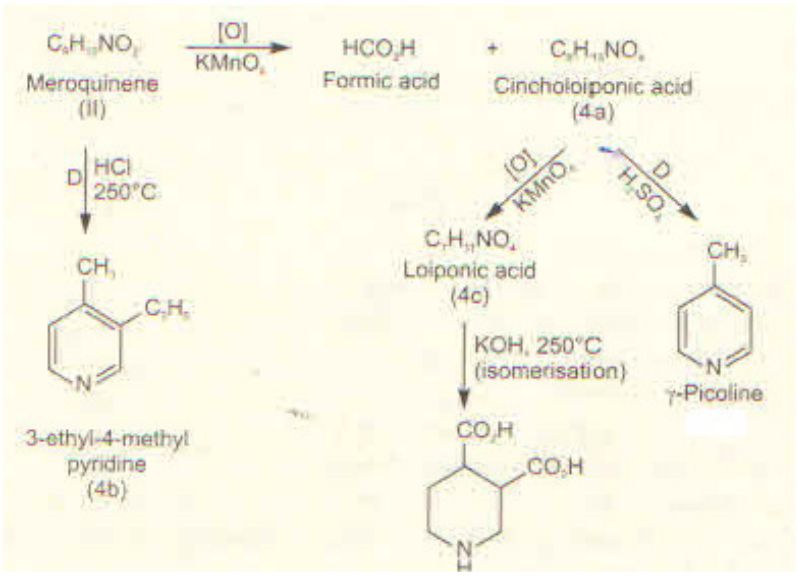
The structure (I) of quininic acid has been confirmed by its synthesis as shown below



The direct oxidation of methyl group in 6-methoxy-4-methyl quinoline to quininic acid is extremely difficult. This is because the oxidation reaction is accompanied by the oxidation of the benzene ring to give pyridine-2,3,4-tricarboxylic acid.

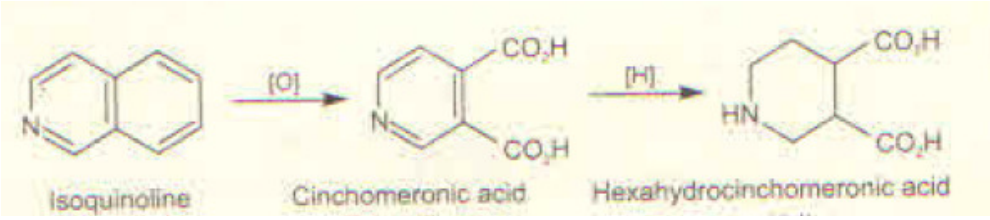
14.5.5 Structure of meroquinene(II)

- Molecular formulas of meroquinene (II) is $C_9H_{15}NO_2$
- It contains one carboxyl group and one double bond as shown by routine tests.
- Oxidation of meroquinene with cold potassium permanganate in acid media gives formic acid and a dicarboxylic acid $C_8H_{13}NO_4$ cincholoiponic acid.

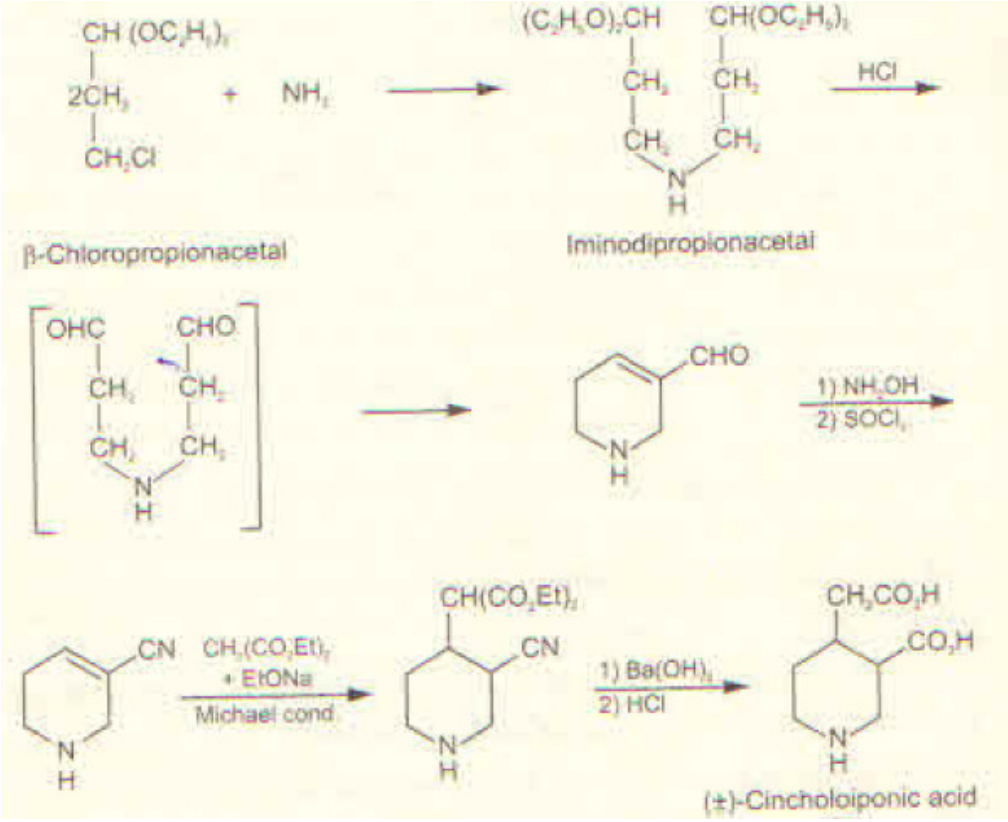


- The formation of formic acid indicates the presence of a vinyl side chain (-CH=CH_2) in meroquinene. The presence of this group is also demonstrated by ozonolysis of meroquinene which gives formaldehyde

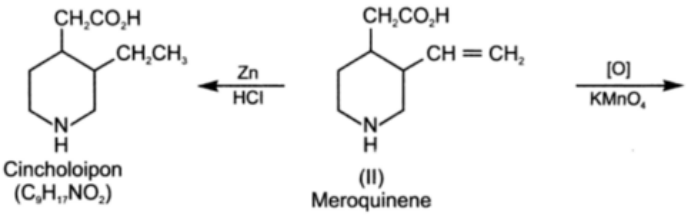
- Meroquinene on heating with HCl at 240°C gave 3-methyl pyridine. Cincholoipoinic acid on further oxidation with cold acid potassium permanganate results in the formation of loiponic acid, $C_7H_{11}NO_4$.
- On heating with KOH it isomerizes to more stable form hexahydrocinchomeronic acid. Lipopnic acid or its isomerised product contain a methylene less than its precursor, cincholoiponic acid, this suggests that the latter contains a sdie chain of the type-CH_2COOH Cincholoponic acid on treatment with concentrated H_2SO_4 gives γ-picoline. This suggests that the additional $–CH_2$ group is present in position 4.

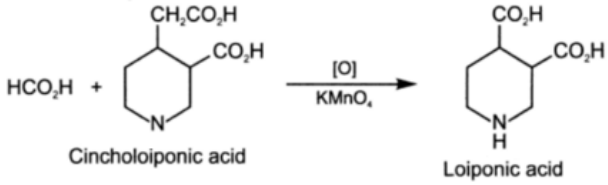


The structure of hexahydrocinchomeronic acid as piperidine-3,4 dicarboxylic acid has been confirmed by its synthesis.



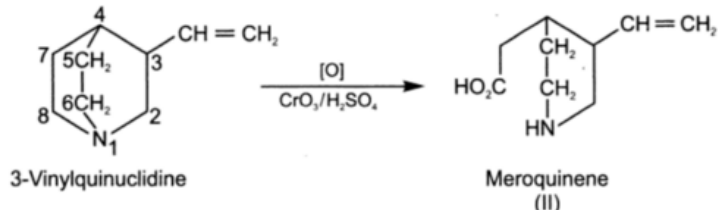
A careful consideration of the above result shows that the structure of meroquinene (II) is as shows on the basis this structure all the reactions of meroquinene can be explained as follows.





The structure (II) of meroquinine also explains the formation of cinholoipon $C_9H_{17}NO_2$ on reduction with zinc and HCl

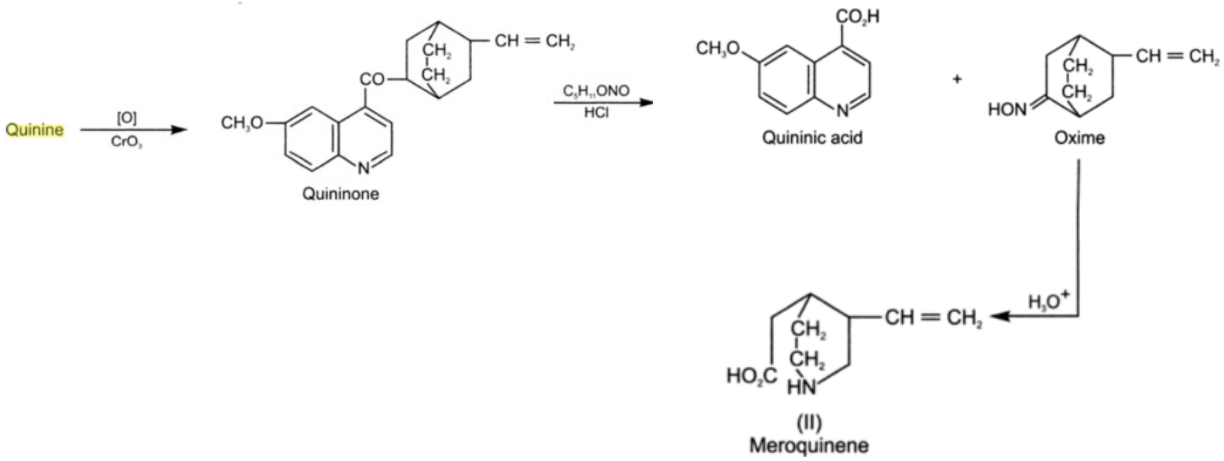
It has already been stated that quinine is ditertiary base and on oxidative degradation gives meroquinene which is a secondary base. Therefore in its formation a tertiary nitrogen atom is converted into a secondary nitrogen atom and a carboxyl group is produced at the same time. A reasonable explanation for this observation is that the tertiary nitrogen atom is part of a bridged ring, one C-N bond is broken when quinine is oxidized.



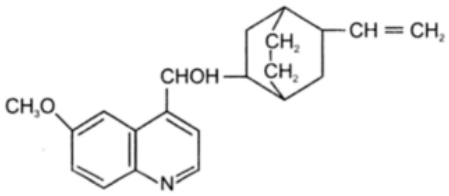
Having established the structures of quininic acid and meroquinene the structure of quinine is established as follows.

The point of linkage of the quinuclidine group to the rest of the molecule was settled by Rabe in 1980

He converted the quinine into a ketone (quinione) by mild oxidation with chromic acid. Both the nitrogen atoms in this ketone are still tertiary and on treatment with amyl nitrite and hydrogen chloride gives quininic acid and an oxime, the formation of an oxime and an acid indicates the presence of a methane group adjacent to a carboxyl group. Hence the structure of the oxime as 8-oximino-3-vinylquinuclidine is assigned by its hydrolysis to the hydroxylamine and meroquinene.



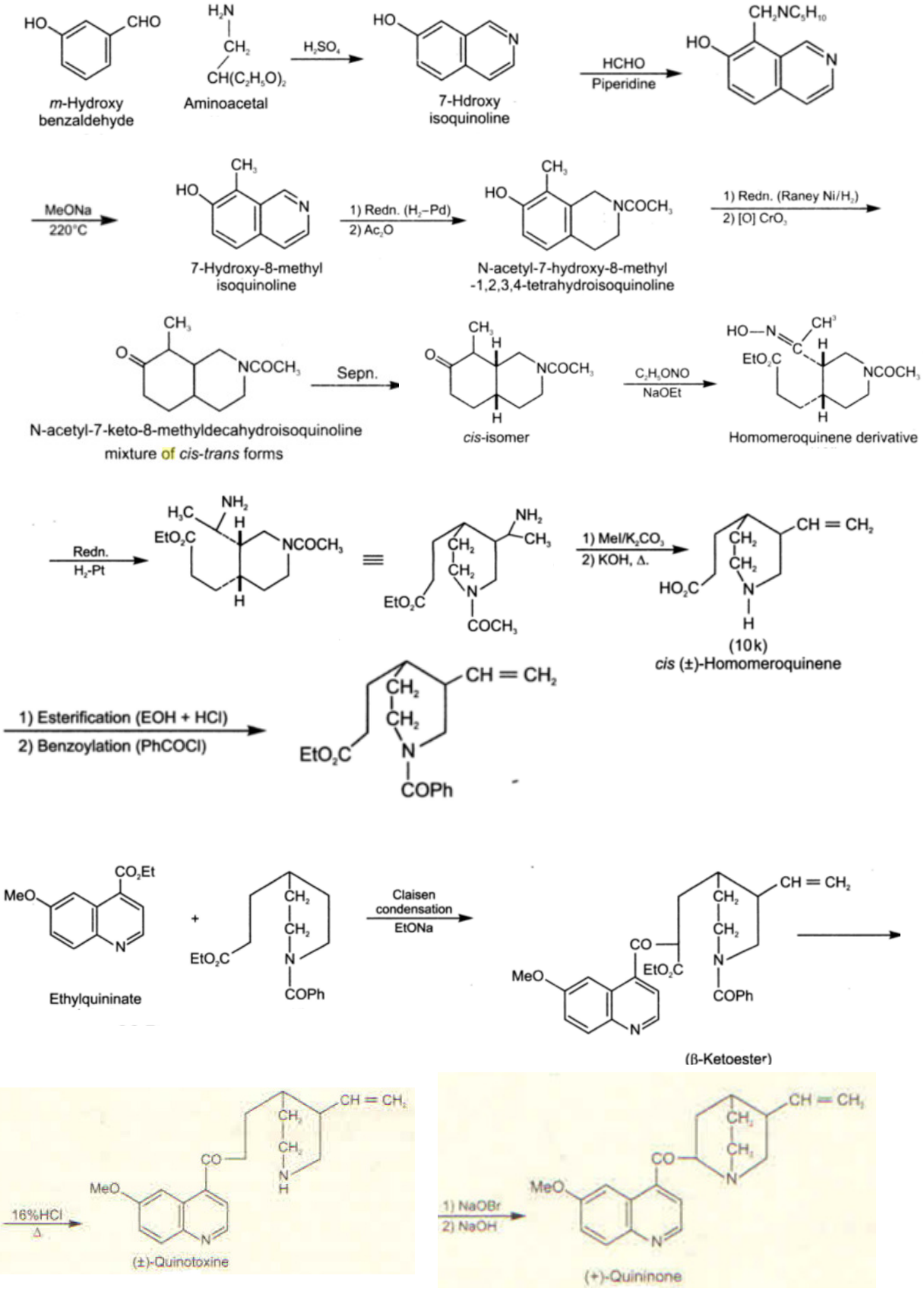
On the basis of the above the following structure is assigned for quinine and this structure explains all the above reaction

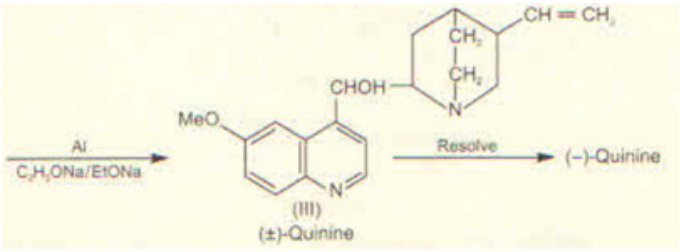


14.6 Total Synthesis of quinine

This structure of quinine has finally established by its synthesis. A partial synthesis of quinine was achieved starts from quinotoxine which was obtained by heating quinine in acetic acid.

Quinotoxine was synthesized by woodword and Doering as follows. Condensation of m-hydroxy benzadehyde with aminoacetal and cyclization of the product with H_2SO_4 gave 7-hydroxisoquinolin. Its treatment with formaldehyde in methanol containing piperidine afforded the complex which on heating with sodium methoxide gave 7-hydroxy-8-methylisoquinoline. Catalytic reduction of this followed by acetylation afforded N-acetyl-7-hydroxy-8-methyl -1,2,3,4-tetrahydroisoquinoline. Further catalytic reduction with Raney nickel under pressure followed by oxidation with chromic acid gave N-acetyl-7keto,8-methyldecahydroquinoline which was a mixture of cis- and trans isomers. The cis from was separated and treated with ethyl nitrite in presence of sodium ethoxide to give the homomeroquinene derivative which on reduction gave amine derivative. Its exhaustive methykation followed by hydrolysis afforded cis(±)- homomeroquinene, which on esterfication and benzoylation afforded N-bezoyl derivative. Claisen condensation of this with ethyl quininate gave the β-keto ester, which on heating with hydrochloric and afforded (±) – quinotoxine.


m-Hydroxy benzaldehyde
Aminoacetal
7-Hdroxy isoquinoline
HCHO Piperidine
MeONa 220°C
7-Hydroxy-8-methyl isoquinoline
1) Redn. (H2–Pd) 2) Ac2O
N-acetyl-7-hydroxy-8-methyl -1,2,3,4-tetrahydroisoquinoline
1) Redn. (Raney Ni/H2) 2) [O] CrO3
N-acetyl-7-keto-8-methyldecahydroisoquinoline mixture of cis-trans forms
Sepn.
cis-isomer
C2H5ONO NaOEt
Homomeroquinene derivative
Redn. H2-Pt
1) MeI/K2CO3 2) KOH, Δ.
(10k)
cis (±)-Homomeroquinene
1) Esterification (EOH + HCl) 2) Benzoylation (PhCOCl)
Ethylquininate
Claisen condensation EtONa
(ß-Ketoester)
16%HCl Δ
(±)-Quinotoxine
1) NaOBr 2) NaOH
(+)-Quininone



The (±) quinotoxine obtained is resolved and the (+)- isomer reacted with sodium hypobromite followed by alkali treatment to give (+)- quininone its reduction afforded(±)-quinine (III) it could be resolved to give (-)-quinine identical with the natural product.

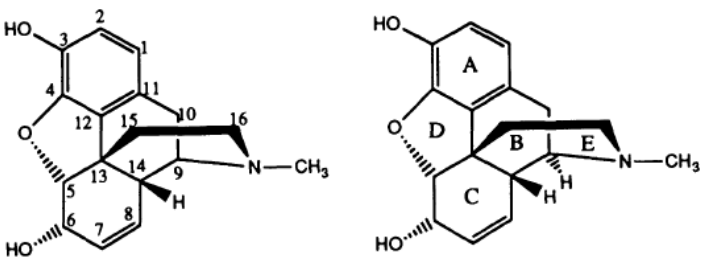
14.7 Morphine

The term 'opium alkaloids' has been used for narcotic analgesics. The opiates are the compounds which are used in the relief of pain. The opiates have proved ideal for the treatment of 'deep' chronic pain and work in the central nervous system (CNS).

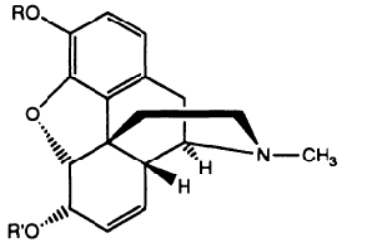
Opium contains a complex mixture of almost twenty-five alkaloids. The principle alkaloid in the mixture, and the one responsible for analgesic activity, is morphine, named after the Roman god of sleep—Morpheus.

Morphine is the most abundant of opium's 24 alkaloids, accounting for 9 to 14% of opium-extract by mass. The chemical formula for morphine is $C_{17}H_{19}NO_3$. It is a benzylisoquinoline alkaloid.

The three dimensional structure of morphine is fascinating. It consists of five rings, of these, three lie approximately in the same plane. One nitrogen-containing ring and another one oxygen containing ring are at right angles to the other three. Morphine also has a methyl group attached to a nitrogen atom. If this methyl group is replaced by a propenyl group, an antagonist of morphine called nalorphine is formed.



Structure of morphine



R = R' = H Morphine
R = Me R' = H Codeine
R = R' = Ac Diamorphine (Heroin)

14.7.1 Biological activity

Morphine is still one of the most effective painkillers available to medicine. It is especially good for treating dull, constant pain rather than sharp, periodic pain. It acts in the brain and

appears to work by elevating the pain threshold, thus decreasing the brain's awareness of pain. Unfortunately, it has a large number of side-effects which include the following

• Depression of the respiratory centre • Constipation • Excitation • Euphoria • Nausea

• Pupil constriction • Tolerance • Dependence

14.7.2 Mode of action

Morphine acts on a specific receptor of nerve cells. More specifically many such receptors are found in the spinal cord's substantia gelatinosa, a region where pain signals are first processed. The architecture of the morphine receptor is what dictates the morphine rule. There is a flat part that binds to the aromatic ring, a cavity that attracts the two carbon atoms and an anionic site that accommodates the tertiary nitrogen atom. When morphine or another agonist binds to the receptor, the cell membrane's affinity for sodium ion changes. This eventually reduces the release of neurotransmitters from the affected neurons. Investigators learned about morphine's mode of action by applying it and other opiates (including enkephalin) to guinea-pig intestines. In the presence of antagonists, Na^+ affinity was restored and intestinal contractions which had dropped precipitously shot up again.

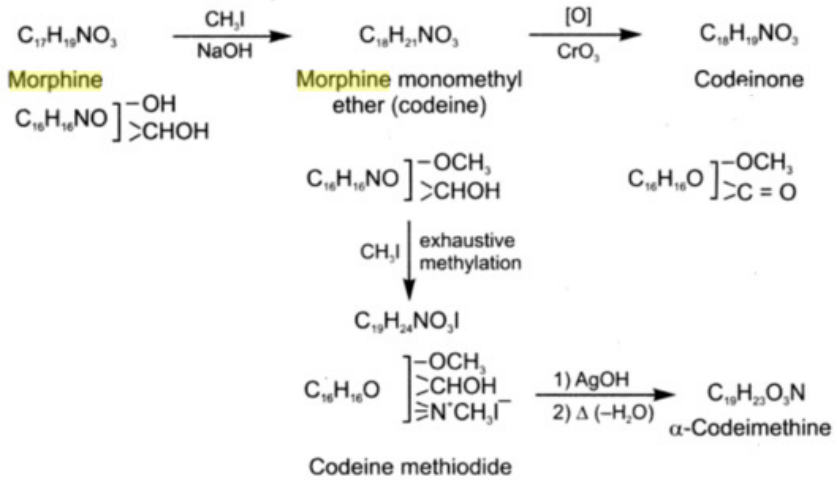
14.7.3 Structure elucidation of Morphine

Morphine was first isolated from opium by Serturner in 1816. It has hydrophenanthrene skeleton with fused N-methyl-piperidine ring and contains a phenolic and an alcoholic hydroxyl groups.

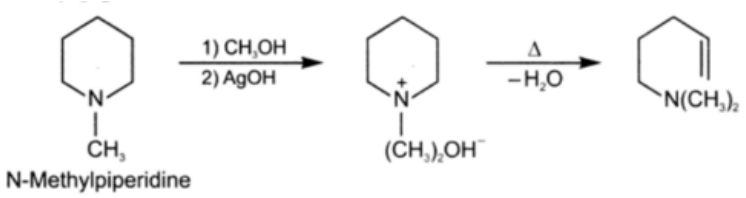
- Morphine has the molecular formula $C_{17}H_{19}NO_3$
- It contains two hydroxyl groups since it forms a diacetate $C_{17}H_{17}ON(OAc)_2$ and a dibenzoate. The diacetate has a bitter taste and is highly toxic and is a habit forming narcotic.
- Of the two hydroxyl groups one hydroxyl is phenolic as indicated by its positive ferric chloride reaction and solubility in sodium hydroxide solution to form a mano sodium salt. The other hydroxyl group in morphine is secondary alcoholic, since on treatment with halogen acid it is replaced by halogen atom.
- Morphine on heating with methyl iodide in queous sodium hydroxide gives a monomethyl ether $C_{18}H_{21}NO_3[C_{17}H_{18}O_2(OCH_3)]$ also known as codeine. Since codeine is insoluble in alkali it follows that the only phenolic hydroxyl group in morphine is methylated.
- Most of the reactions leading to structure elucidation of morphine have been carried out using its methyl ether. Oxidation of dodeine with chromic acid gives a ketone codeinone $C_{18}H_{19}NO_3$, which has two hydrogen less than codeine thus the second hydroxyl group in

morphine is secondary alcoholic group. The third oxygen in morphine or codeine appears to be part of the cyclic ether system, since it is remarkably inactive under a variety of conditions.

- Codeine has one ethylene bond since its catalytic hydrogenation gives dihydrocodeine.
- The nitrogen in codeine is tertiary since it adds on a molecule of methyl iodide to form a quaternary salt codeine methiodide $C_{19}H_{24}NO_3I$ or $C_{18}H_{24}NO_3CH_3I$.
- The nitrogen is present as a cyclic tertiary base, this is supported by the exchaustive methylation of codeine to give α- codeimethine, which contains one CH_2 more than codeine and the nitrogen is not lost



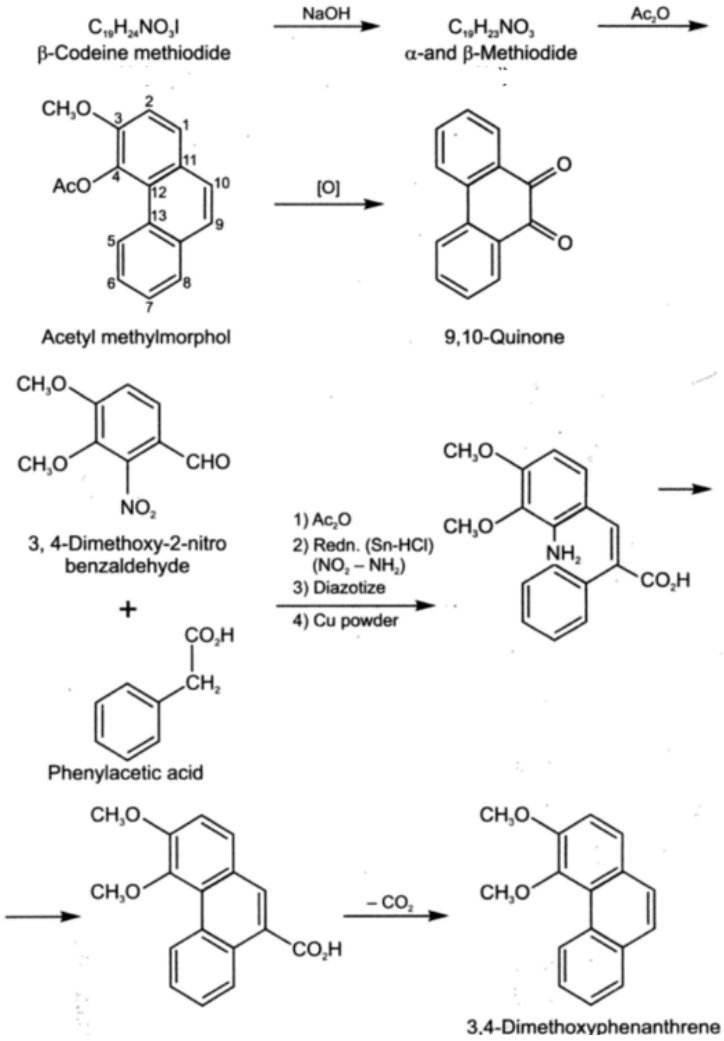
The exhaustive methylation of codeine methiodide to α- codeimethine is expected of a cyclic tertiary amine and is represented as shown in using N-methylpiperidine.



Morphine has a phenanthrene nucleus since on distillation with zinc dust it gives phenanthrene.

Codeine methoiodide on heating with sodium hydroxide solution gives α- codeimethine, under basic conditions it isomenises into β-codeimethine. The α-β-codeimethine on heating with acetic anhydride give acetyl methylmorphol. It is a phenanthrene derivative and on oxidation gives the 9,10 quinone without losing any group this shows that no substituents are present in position 9 and 10.

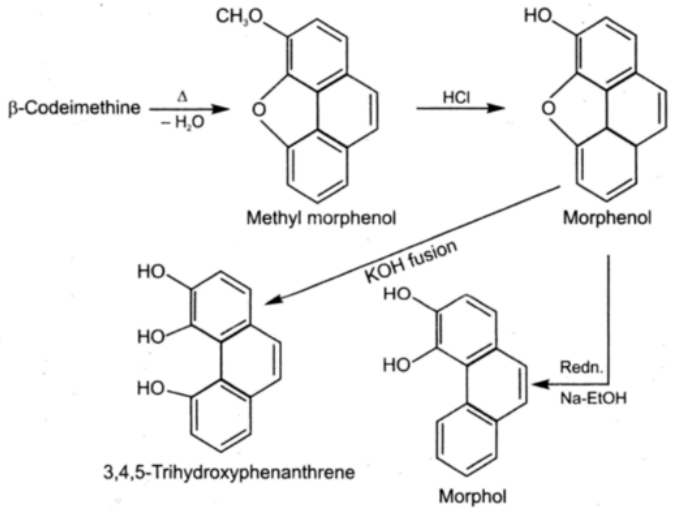
The structure of acetylmethylmorphol is established by its methylation to 3,4-dimethylamorphol identical with a synthetic sample synthesized from 3,4-dimethoxy-2-nitrobenzaldehyde and phenylacetic acid. The formula for acetyl methylmorphol is its correct representation. Thus the exact position of OCH_3 and OAc have been determined.



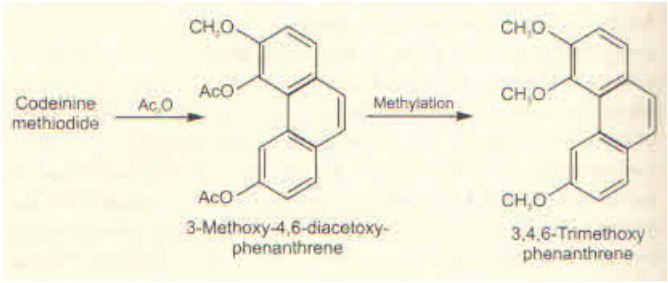
- Deacetylation of acetyl methylmorphol obtained above gives methylmorphol it is shown to be 4-hydroxy-3-methoxyphenathrene by its synthesis using Pschorr synthesis starting from 3-acetoxy-4-methoxy-2-nitro-benzaldehyde.
- β-dedeimethine on heating with water gives methylamine ethylene and methylmorphenol its demethyalation with HCl gives morphenol which contains one phenolic hydroxyl group and inert oxygen. Morphenol on fusion with potassium hydroxide gives the 3,4,5-trimethoxy phenanthrene identical with the product obtained by methylation of 3,4,5-

trihydroxy phenanthrene. Further reduction of morphenol with sodium and alcohol gives morphol.

On the basis of above it is assumed that morphenol has structure containing an ether linkage in position 4,5 of the phenanthrene nucleus.



- The position of the alcoholic group is determined by the degradation of methiodide or codeinine with acetic anhydride to give 3-methoxy-4,6-diacetoxy phenanthrene which was confirmed by its conversion to 3,4,6-trimethoxyphenanthrene identical with a synthetic sample.



- The postion 3 of the methoxy group and position 4 of the hydroxyl group have already been accounted for in morphine. The above results with codeine suggest that the alcoholic group is at position 6.
- The presence of N-methyl group in morphine or codeine is deduced as follows. Codeine methiodide on heating with sodium hydroxide gives a mixture of α-β-codeimethine along with ethanoldimethylamine $(CH_3)_2NCH_2CH_2OH$ its formation indicates that there is $>N\text{-}CH_3$ group in codeine it has been shown that the N atom in codeine is in a heterocylic ring and adds only one molecule of methyl iodide to from codeine methiodide. This has been confirmed by subjecting codeine to Von Braun degradation,

three hydrogen atoms are lost and one nitrogen atom is added. This can be explained by the conversion of $>NCH_3$ into >NCN thus it follows that codeine has an N-methyl group.

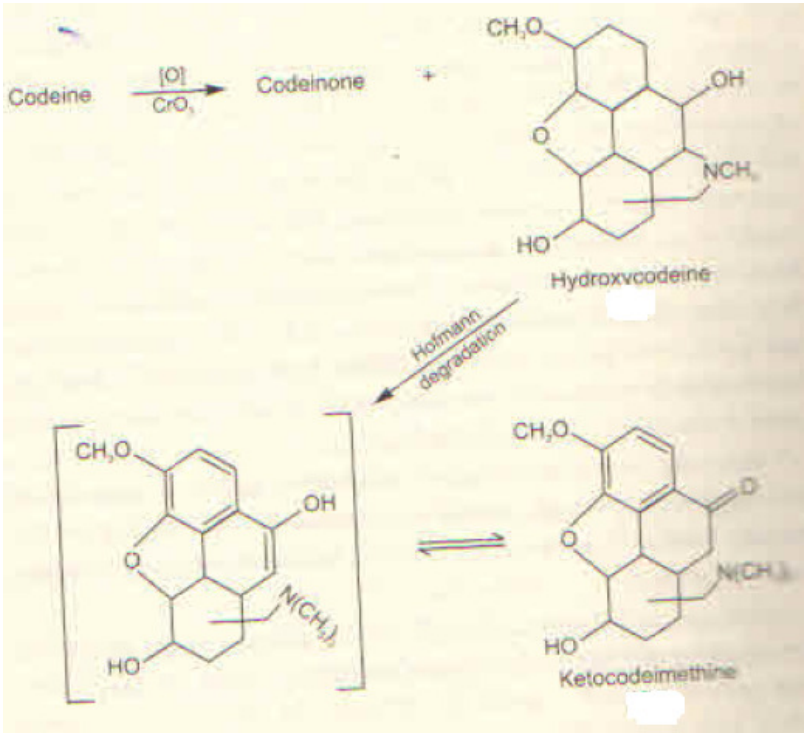
The N atom has been shown to be attached to position to 9 or 10 on the basis of the following evidence.

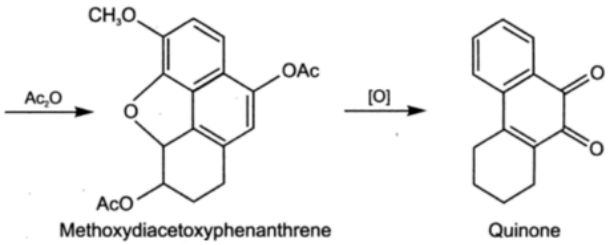
It is known that codeine on oxidation with chromic acid gives codeinone. Small amount of hydroxycodeine is also obtained in this oxidation.

The hydroxycodeine on subjected to Hofman degradation gives a ketocodeimethine in which the oxygen appears as a carbonyl group.

The ketocodeimethine on heating with acetic anhydride gives a methoxy deacetoxyphenanthrene, which on oxidation loses an acetoxy group to give quinine. The new hydroxyl group in the phenanthrene and thus the new hydroxyl group in hydroxycodeine must therefore be in position 9 or 10.

Since this hydroxyl group is converted into a carbonyl group during Hofmann degradation a double bond may be introduced at position 9 and 10 during the scission of the nitrogen ring. On this basis the nitrogen must be linked C-9 or C-10. No positive proof for the attachment at C-9 or C-10 was obtained until the morphine skeleton was constructed. However on steric consideration position 9 is not the most likely point of attachment. The point of attachment of the carbon end of the nitrogen-containing side chain could not be settled for a number of years. The above transformations are tentatively represented as shown.





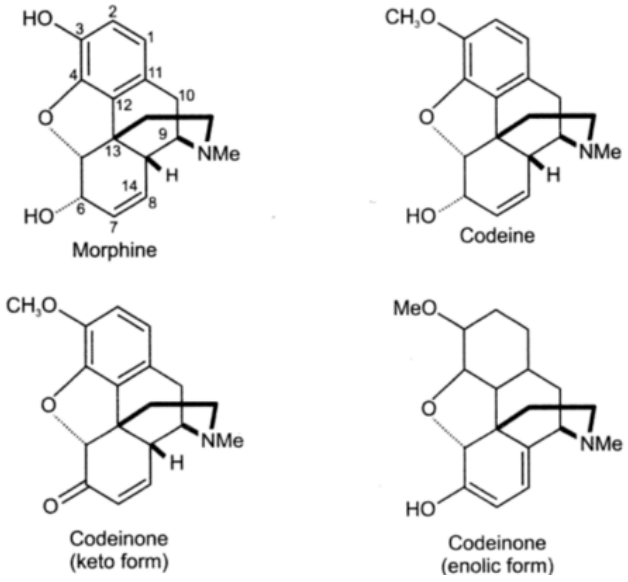
Methoxydiacetoxyphenanthrene Quinone

The attachment of the other end of N-containing side-chain is decided as follows

- Codeine on treatment with phosphorouspentachloride with acetic acid-water mixture gives a mixture of three alcohols, isocodeine, pseudocodeine and allopseudicodeine.
- Codeine and isocodeine differ in the steric arrangement of the –CH(OH) group as they give the same codeinone on oxidation.
- Pseudocodeine and allopseudocodeine also form an epimeric pair giving the same pseudicideinone on oxidation.
- Pseudocodeinone on heating with aceticanhydride gives 3-methoxy-4,8-diacetoxy phenanthrene identified by conversion into 3,4,8-trimethoxyphenanthrene. These can be best explained by fixing the isolated double bond at position 7,8 in codeine and isocodeine and position-6,7 in pseudocodeine and allopseudocodeine. Pseudocodeinone has a C=O group at C-8 and attachment of side chain at this point is not possible.

However based on the above evidence and large amount of other experiment work it was not possible to get a rigid proof for the attachment of nitrogen residual with certainty till morphine was synthesized.

Gulland and Robinson proposed the following structure for morphine codeine and codeinone these structures have been proved by their respective synthesis.

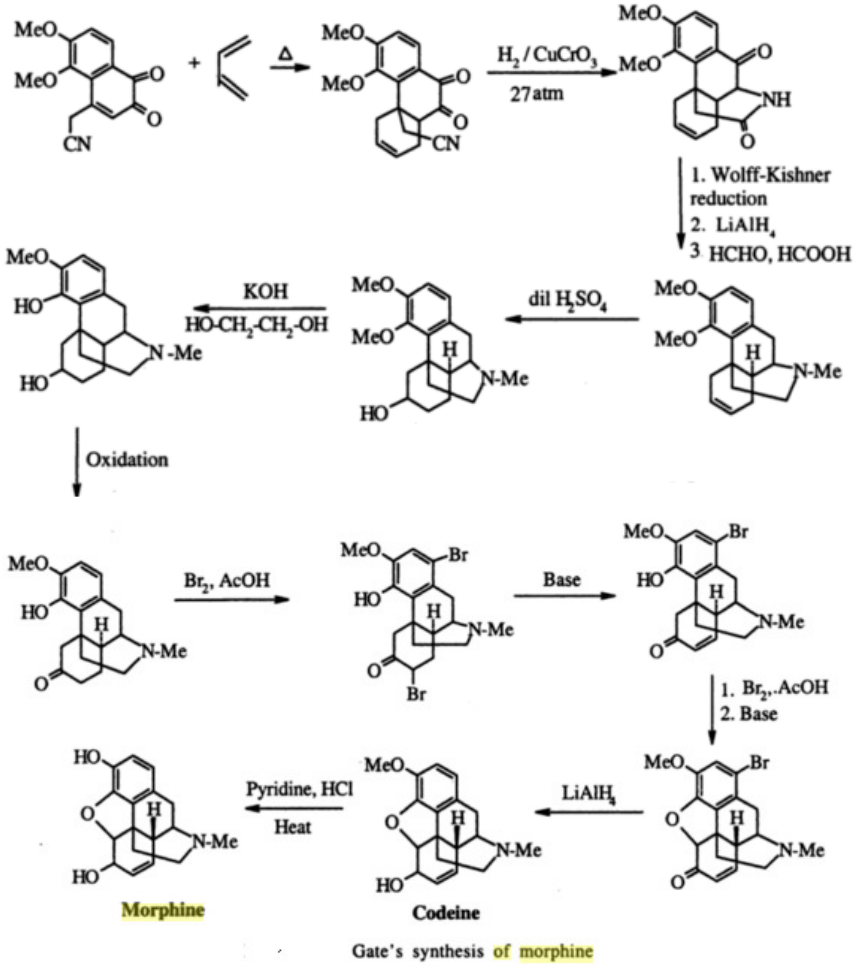


Morphine Codeine

Codeinone (keto form) Codeinone (enolic form)

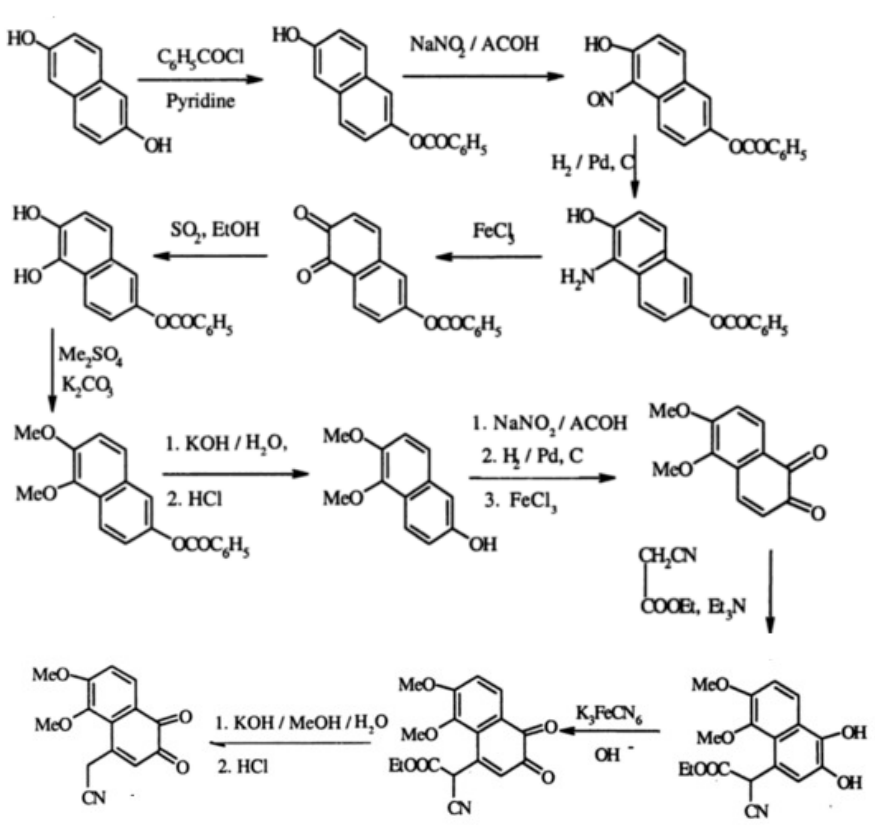
14.7.4 Total synthesis of Morphine

Gates achieved total synthesis of morphine in 1952. The first step in his synthesis is the Diels-Alder addition of butadiene to 4-cyanomethyl-5,6-dimethoxy-1,2-napthoquinone. The second step is the catalytic hydrogenation, in this step the adduct underwent ring closer to keto-lactum. In third step the Wolff-Kishner reduction of carbonyl group followed by reduction of amide carbonyl by treatment with lithium aluminium hydride and N-methylation by treatment with formaldehyde, formic acid yields product. The double bond was hydrated and methyl ester group was selectively demethylated and the product was oxidized to ketone and brominated to yield the α-bromo keto product. In the course of dehydrobromination followed by bromination and dehydrobromination with a base to produce α,β-unsturated ketone, inversion occurs at the adjacent centre at C14 to produce the natural orientation. The reductive removal of the 1-bromine atom gave codeine. Which was concvert to morphine by the cleavage of the ether group by the treatment with pyridine hydrochloride. The entire synthesis scheme is shown below



Gate's synthesis of morphine

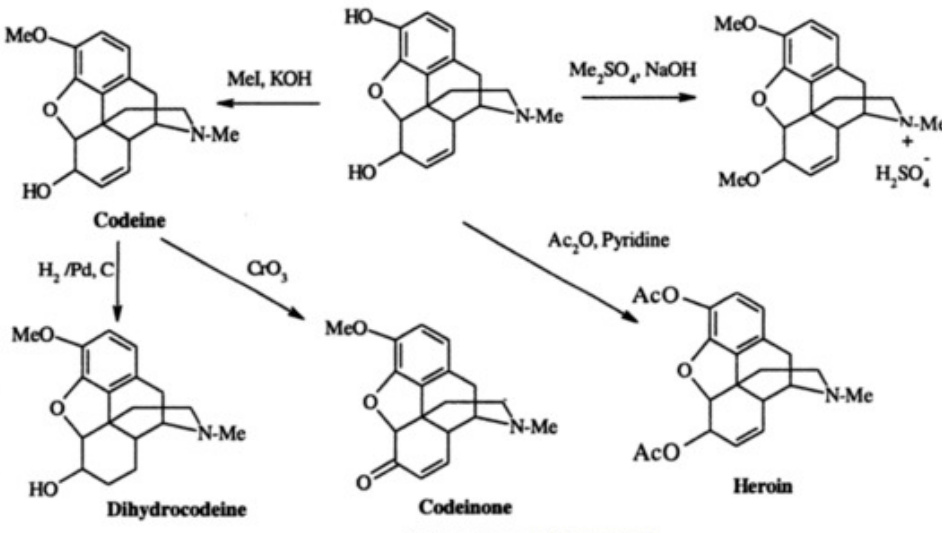
Starting compound was obtained from 2,6-dihydroxynaphthalene as shown below.



Synthesis of O-quinone

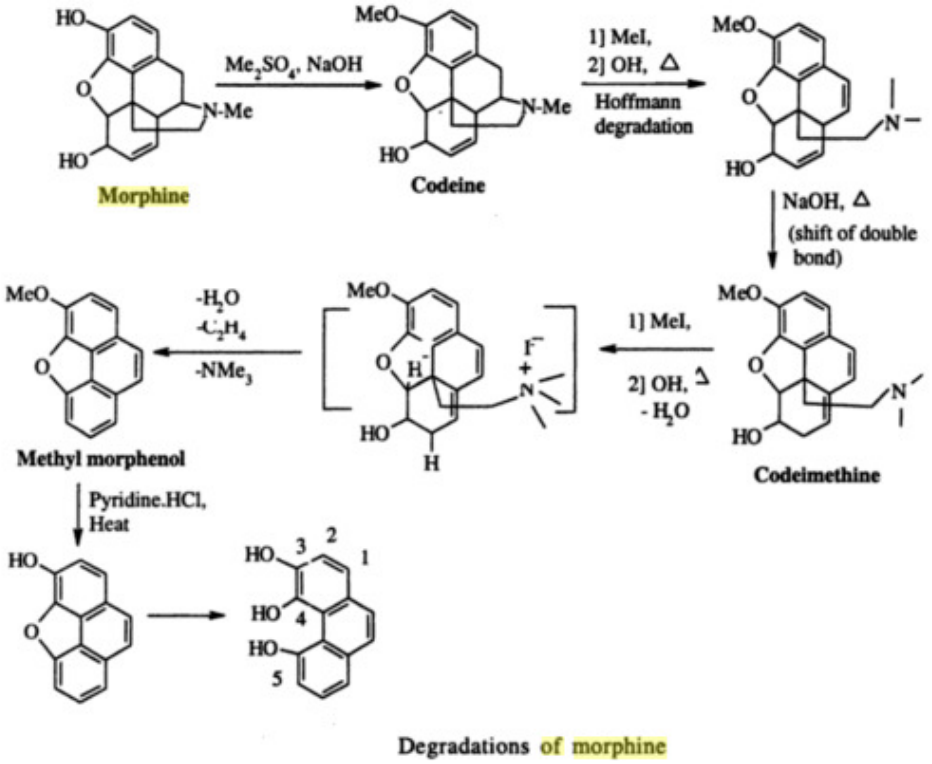
14.7.5 Reactions of morphine

Herion(diacetylmorphine) was prepared by the acetylation of morphine. Heroin has similar structure but more pronounced action than morphine.



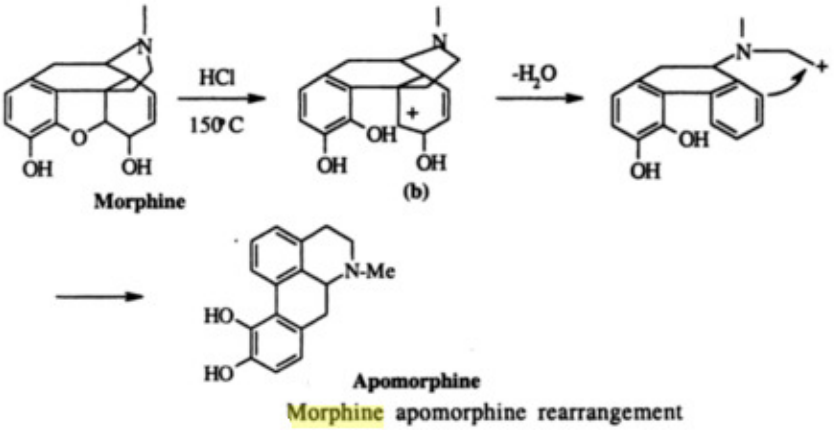
Reactions of morphine

Zinc distillation of morphine yields hydrophenanthrene and the location of three oxygen functions was established by degradation leading to phenanthrene derivatives oxygenated at 3, 4, 5 positions



Degradations of morphine

Morphine on heating with hydrochloric acid gave apomprphine. It can be envisioned as an attack by a proton on ethereal oxygen followed by opening of oxide bridge to give carbonium ion, which further undergoes detachment of the ethanamine bridge at C13 and dehydration leading to aromatization. The reattachment of ethanamine carbonium ion at C6 gives apomorphine is of interest because it estabilished a relationship between the morphine alkaloids and the benzylisoquinoline alkaloids.



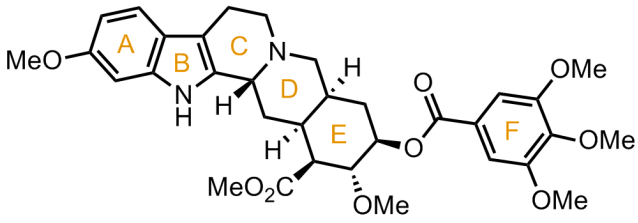
Morphine apomorphine rearrangement

14.8 Rauwolfia Alkaloids

The roots of Rauwolfia plant have been used by the natives Himalayan region in india for the cure of afflictions ranging from snake bite to insanity. The drug obtained from the roots of Rauwolfia was reported to control hypertension, epilepsy, insomnia, fever, dizziness and headache. The structural investigation on the drug was initiated first in India and later in Europe and USA and discovered the Reserpine, which was the main active compound of the plant. The chemical investigation of over sixty species of Rauwolfia led to the isolation of nearly sixty alkaloids.

14.8.1 Reserpine

Reserpine is an indole alkaloid isolated in 1952 from the extract of Rauwolfia sepentina or 'Indian snake root', a popular plant in traditional Indian medicine used as a sedative and antipyretic. It's also used as an antihypertensive and antipsychotic, notably being the first ever drug to successfully demonstrate antidepressant properties. Its structure was solved in just 3 years and was finally reported in 1955.

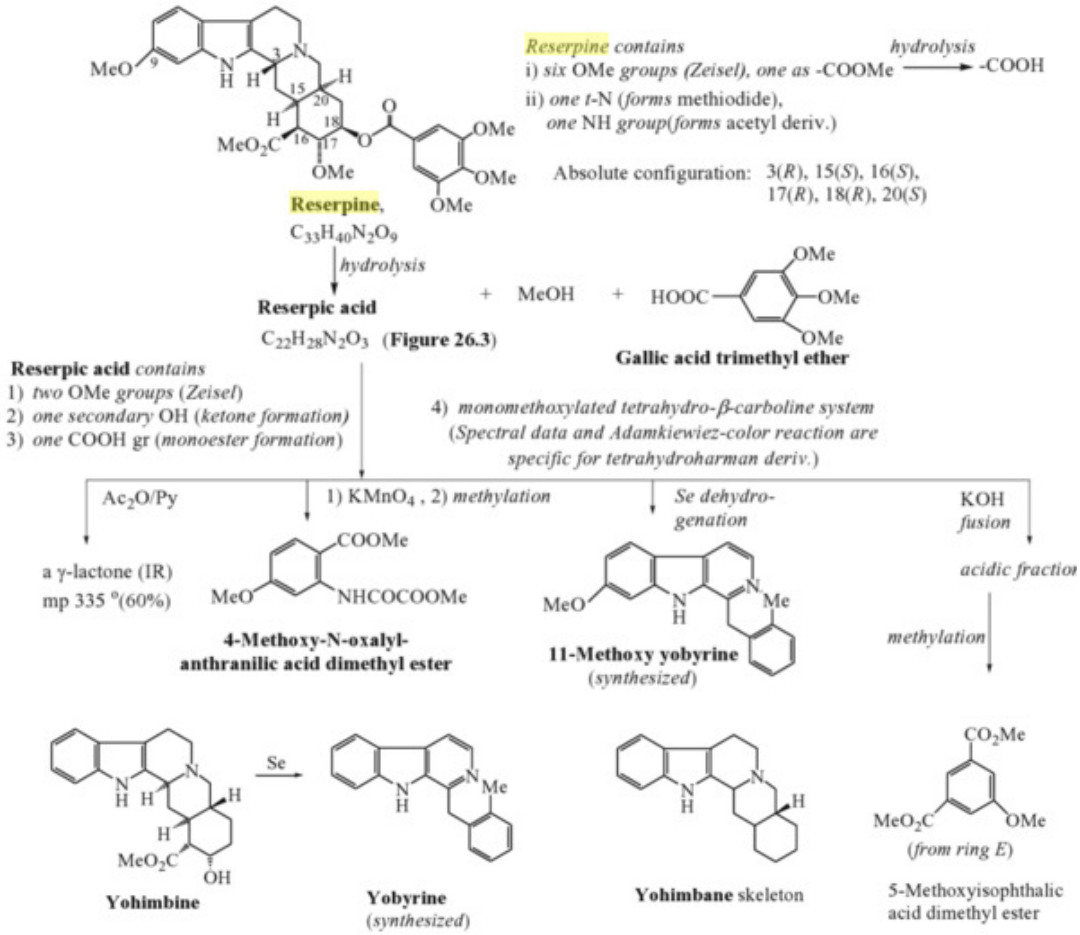


(–)-reserpine

- ✓ The molecular formula of Reserpine is $C_{33}H_{40}N_2O_9$
- ✓ Melting point is 286-288^{0}C and $[\alpha]_D^{24}$ is -116.9 ($CHCl_3$)
- ✓ Reserpine is a weak base and both nitrogens are involved in ring; one nitrogen is present in indole group as revealed by infrared absorption(7.1 cm^{-1}) the other nitrogen is tertiary since reserpine cannot be alkylated.
- ✓ Reserpine contains nine oxygens. Out of these five oxygens are present as methoxyl group. Reserpine is a double ester, one ester group is shown by the action of hydriodic acid and the other is present as O-C=O viz methyl carboxylate (-$COOCH_3$). This accounted for other four oxygens.

Alkaline hydrolysis of Reserpine furnished reserpic acid, trimethoxybenzoic acid and methanol. Reserpine could be regenerated from methyl reserpate and trimethoxylbenzoyl chloride in presence of pyridine suggesting that reserpine is an ester of reserpic acid having a hydroxyl group with trimethoxybenzoic acid and that during hydrolysis reserpic acid part does not undergo structural change. Hence the structure determination of reserpic acid

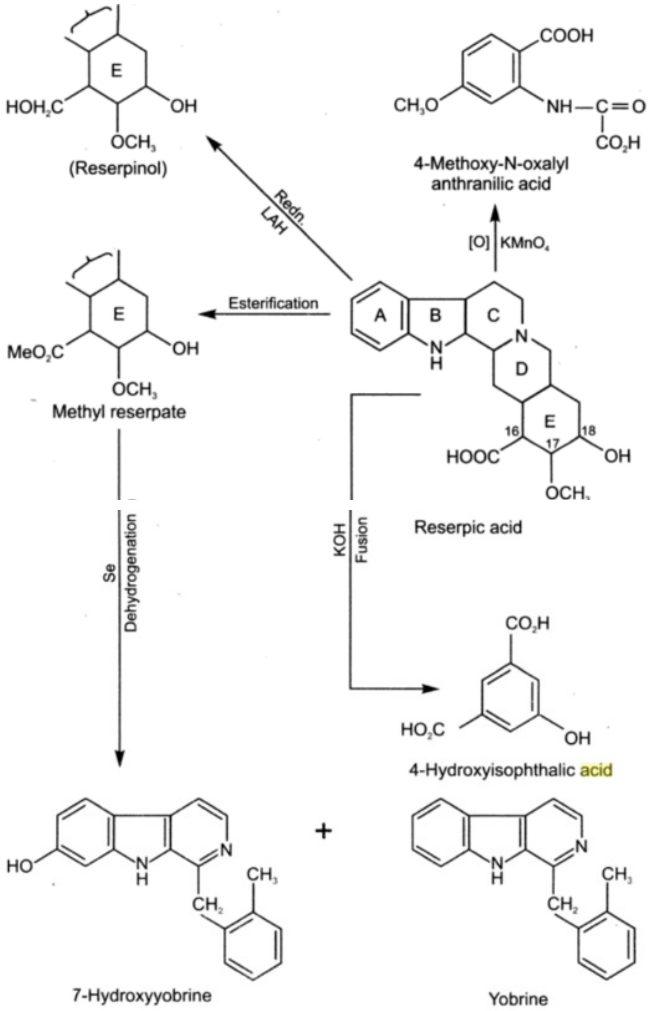
established the structure of reserpine. Degradation of reserpine and reserpic acid are delineated in below scheme. This suggests that reserpic acid have the yohimbine skeleton.



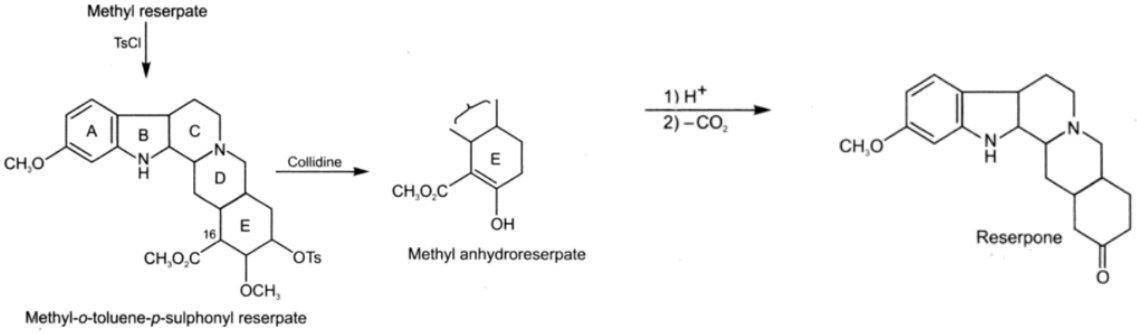
14.8.2 The structure of reserpic acid

- ✓ Usual tests showed the presence of two methoxy, one carboxyl and one hydroxyl groups and that both the nitrogen atoms are present in heterocyclic ring.
- ✓ Reduction of reserpicacid with lithium aluminium hydride gives reserpic alcohol in which the –COOH is converted into $–CH_2OH$.
- ✓ Oxidation of reserpicacid with $KMnO_4$ gives 4-methoxy-N-oxalylanthanilic acid showing that one N atoms is in m-position to one of the methoxy groups.
- ✓ Potash fusion of reserpicacid gives 4-hydroxyisophthalic acid. Treatment of reserpicacid with acetic anhydride gives a γ-lactone the formation of which is possible only when the hydroxyl and carboxyl groups are in 1 : 3 positions.
- ✓ Dehydrogenation of methylreserpate with selenium gives yobrine and 7-hydroxyyobrine the structure of both these compounds were known.

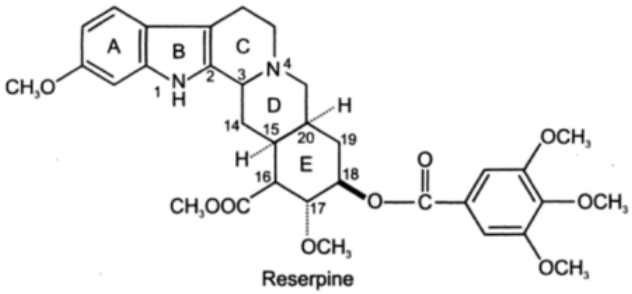
On the basis of the above facts reserpicacid may be formulated as.



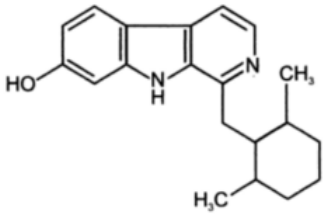
As seen in the structure of reserpic acid the methoxy group is placed at C-17. This assignment is based on biogenetic consideration and is supported by the conversion of methyl-o-toluene-p-sulphonylreserpate with colloidine into methyl anhydroreserpate which is the enol ether of a β-keto acid. Its acid hydrolysis followed by decarboxylation gives reserpone



The position of the keto group in reserpone indicates the position of the methoxyl group in reserpic acid. The position of the CO group and hence the position of other substituents have been proved by degradation of reserpinol. On the basis of the foregoing discussion reserpine is represented as



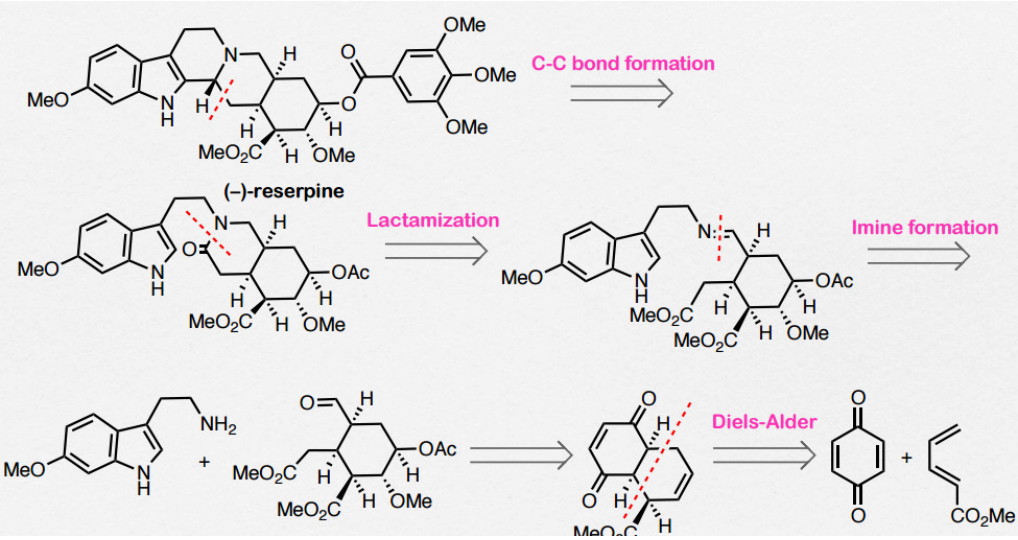
Reserpine



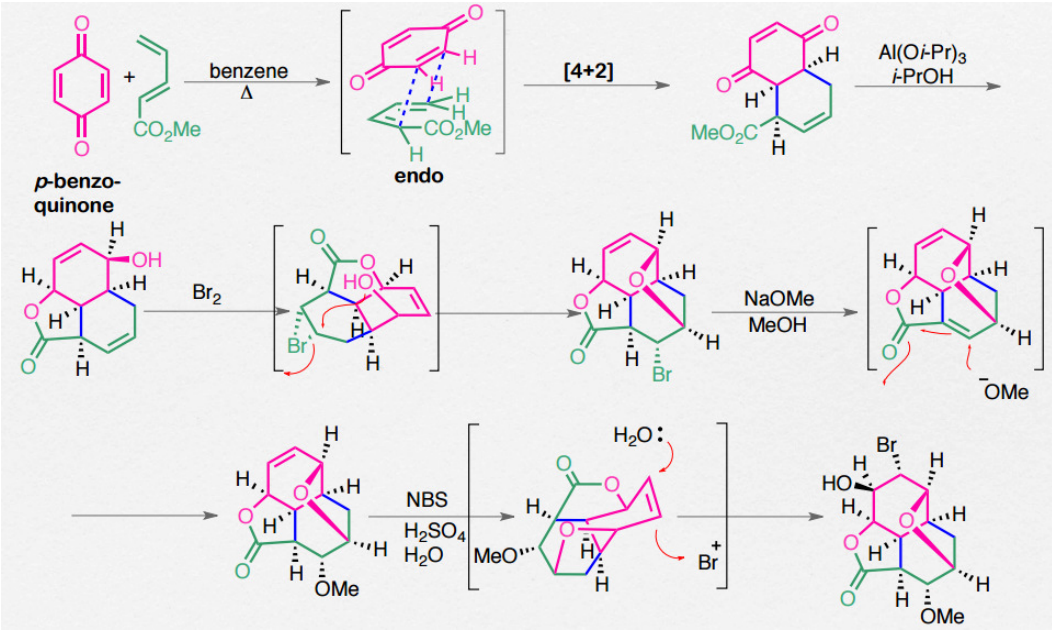
7-Hydroxymethyl yobrine

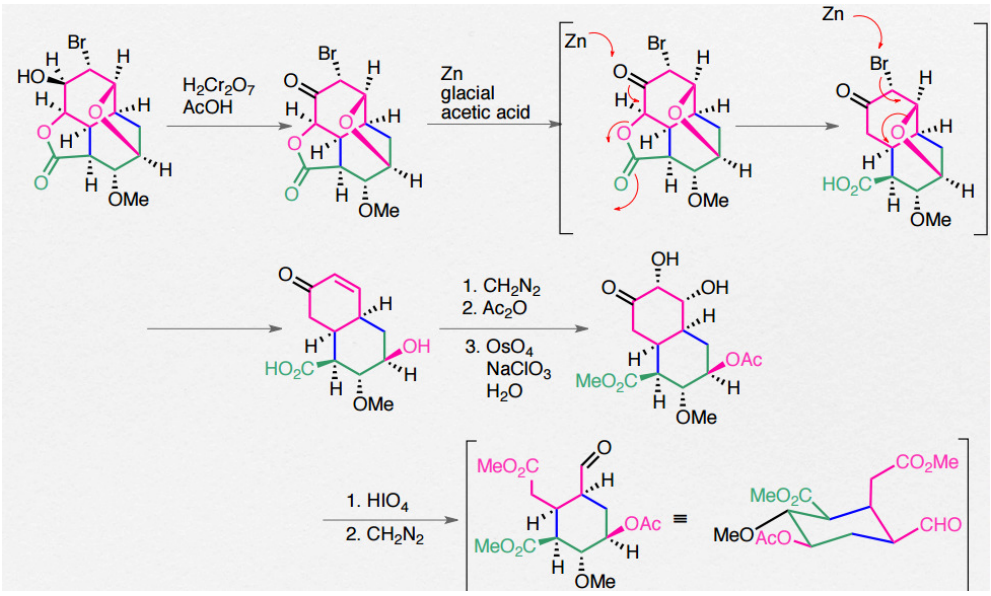
As seen reserpine has six asymmetric centres five of which are in ring E. The D/E ring fusion is cis and is flexible like cis-decalin capable of existing in the two all chair conformation.

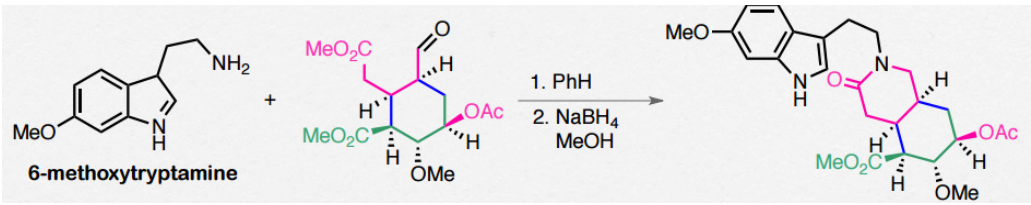
The structure of reserpine has been confirmed by its synthesis. The special features of this synthesis in fixing of the stereochemistry. Synthetic challenge of reserpine is posed by the D/E ring system of the pentacylic nucleus. Synthetic strategy necessitated the preparation of a functionalized hydroisoquinoline derivative that could then be modified to provide the D/E ring system. The retro synthetic pathway for reserpine was shown below

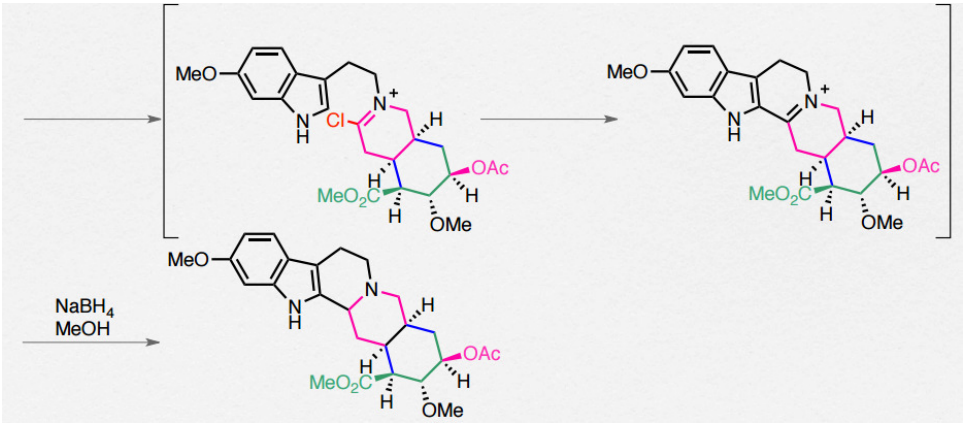


Preparation of a substituted hydroisoquinoline system was made possible by an intramolecular Diels-Alder reaction.









14.9 Summary of the unit

Alkaloids disrupt the integrity of biomembrane, potent inhibitors of ion channel and impair the function of microtubules or microfilaments. Alkaloids are mutagenic and carcinogenic due to their properties to bind DNA and acts at DNA and RNA polymerase level. Alkaloids either activate or inhibit the central process at cellular and organ level in animals. They affect the digestive process by interfering the hydrolytic enzymes. They impaired the kidney and liver functioning. Some alkaloids influence the reproductive system of animals. Alkaloids also possess antimitotic and allergic effects at cellular level. These functions of alkaloids are well known for their toxic and sometimes psychomimmetic, euphoric, and hallucinogenic properties. Alkaloids also have role in chemical ecological perspectives like plant-herbivore interaction, plant-plant interaction and plant-microbial interaction. Many alkaloids, though poisonous, have physiological effect that renders them valuable medicine against various diseases including malaria, diabetics, cancer, cardiac dysfunction etc. These are also used in local anesthesia and relief of pain.

Cinchona alkaloids contain two groups of diastereo-isomers, they are Quinine (levo) /Quinidine (dextro) and Cinchonine (dextro)/Cinchonidine (levo)]. The basic skeleton of Cinchona alkaloids is Ruban-9-Ol. Ruban nucleus is a combined skeleton formed from a quinoline ring attached through a methylene group to a quinuclidine ring (a bicyclic ring contain N). In Rubanol, the methylene group is oxidized to a secondary alcoholic group and the carbon atom becomes Asymmetric. Cinchona alkaloids are Di-acidic bases (due to presence of 2N atoms).

The six opium alkaloids which occur naturally in the largest amounts are morphine, narcotine, codeine, thebaine, papaverine and narceine. Of these, three are phenanthrene alkaloids and are under international control: these are morphine, codeine, and thebaine. They are all used in the drug industry. Morphine is the chief alkaloid of opium both in amount and

in medical importance. The three dimensional structure of morphine is fascinating. It consists of five rings, three of which are approximately in the same plane. The other two rings, including the nitrogen one, are each at right angles to the other trio.

The three medicinally most important alkaloids occurring in Rauwolfia serpertina are: reserpine, deserpidine and rescinnamine. The three alkaloids are levorotatory double esters. They have methyl ester at C-16. C-18 of E-ring having OH esterified with trimethoxy benzoic acid in Reserpine and Deserpidine and with trimethoxy-trans-cinnamic acid in Rescinnamine. Reserpine is 11-methoxy Deserpidine.

Reserpine on alkaline hydrolysis gives trimethoxybenzoic acid, reserpic acid and methanol. Similarly Deserpidine on alkaline hydrolysis gives trimethoxybenzoic acid, deserpidic acid and methanol. Rescinnamine on alkaline hydrolysis gives trimethoxycinnamic acid, reserpic acid and methanol.

14.10 Key words

Biological importance of alkaloids; Quinine; Quininic acid; Morphine; Rauwolfia Alkaloids; Reserpine; Reserpic acid

14.11 References for further studies

1) Pharmaceutical Chemistry of Natural Products; V. Alagarsamy; *Elsevier India*, **2012**.

2) Chemistry of Plant Natural Products: Stereochemistry, Conformation, Synthesis, Biology, and Medicine; Sunil Kumar Talapatra, Bani Talapatra; *Springer*, **2015**.

3) Chemistry of Natural Products; Sujata V. Bhat, B.A. Nagasampagi, Meenakshi Sivakumar; *Springer Science & Business Media*, **2005**.

4) Organic Chemistry; R.L. Madhan; *Tata McGraw-Hill Education*, **2013**.

5) Alkaloids; Ane's chemistry active series; Vk Ahluwalia; *Ane Books Pvt Ltd*, **2010**.

6) Chemistry of Alkaloids; P. B. Saxena; *Discovery Publishing House*, **2007**.

14.12 Questions for self understanding

1) Discuss the biological importance of alkaloids

2) Write the structure of Quinine.

3) How quinine is isolated from its source?

4) Write the various physiological uses of quinine.

5) Discuss the alkali hydrolysis of quinine.

6) Discuss the structural elucidation of quininic acid.

7) Discuss the total Synthesis of quinine.

8) Write the structure of morphine.

9) Discuss the biological activities of morphine.

10) Discus the structural elucidation of morphine.

11) Discuss the total synthesis of morphine.

12) Write the structure of products formed by alkali hydrolysis of Reserpine.

13) Discus the structural elucidation of reserpic acid

UNIT-15

Structure

15.0 Objectives of the unit

After studying this unit you are able to

- Write the basic skeletal structure prostaglandin
- Classify the prostaglandins based on their structure
- identify the biological functions of prostaglandins
- Write the synthesis scheme of PGE1, PGE2, PGF1α and PGE2α

15.1 Introduction

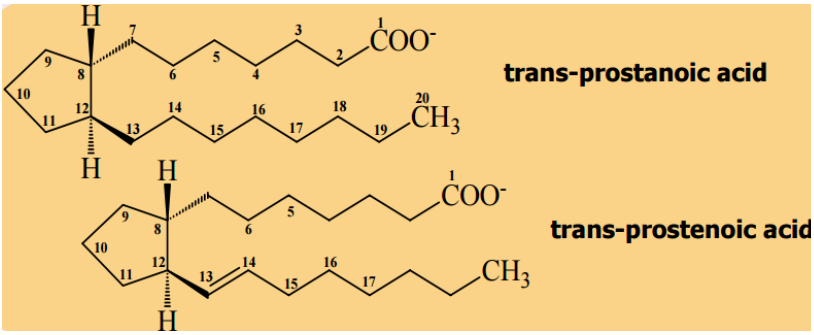
Prostaglandins, thromboxanes, and leukotrienes are enzymatically derived from essential fatty acids and constitute a unique class of polyunsaturated, hydroxylated, 20-carbon fatty acids categorized as eicosanoids. Prostaglandins (PGs) are medicinally interesting due to their broad spectrum of biological activity. PGs act in many parts of the body, including the reproductive system (in both males and females), the nervous system, the cardiovascular system, the immune system and gastrointestinal system. Due to their diverse biological activity, there is potential for prostaglandin analogs (prostanoids) to function as effective therapeutic agents. Indeed, there are already PG analogs used as drugs for the treatment of ulcers, hypertension and other conditions.

In between 1960- 1970, the core structure of the PGs was determined, many different routes for prostaglandin synthesis were explored. However, achieving an asymmetric synthesis of the PGs still presents unique challenges. Development of a robust, asymmetric route to prostaglandins would greatly benefit development of prostanoids by pharmaceutical companies.

15.2 Prostaglandins

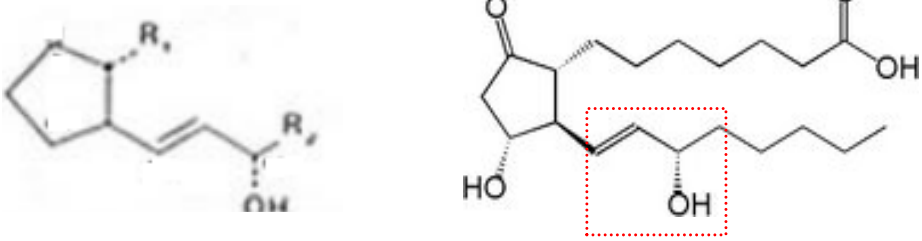
Prostaglandins were first discovered and isolated from human semen in the 1930s by Sweden scientist Ulf von Euler. He thought that they had come from the prostate gland, hence he named them prostaglandins. Later It has been determined that they exist and are synthesized in every cell of the body. Prostaglandins, are like hormones, they act as chemical messengers, but they do not move to other sites and work right within the cells where they are synthesized.

Prostaglandins are unsaturated carboxylic acids, consisting of a 20 carbon skeleton that also contains a five member ring. They are biochemically synthesized from the fatty acid, arachidonic acid.



15.3 Prostaglandin Structure

Prostaglandins are composed of a cyclopentane nucleus with two side chains. One side chain has seven carbon atoms present at carbon-8 on the cyclopentane ring and is indicated by R_7 or R_1. The other side chain consist eight–carbon atoms present at carbon-12 on the cyclopentane ring and is indicated by R_8 or R_2. There may be one, or two, or three double bonds present in the side chains. Generally all prostaglandins contain a hydroxyl group on carbon-15 and a *trans* double bond at carbon13.



The double bonds may also be present in the five member ring in some prostaglandins. The other functional groups like a ketone, or alcohol are also present in the five membered ring.

15.4 Classification of prostaglandins

Mainly three classes of prostaglandins are recognized, and these are categorized on the basis of the number of double bonds present within the prostaglandin molecule and on the fatty acid from which they are derived.

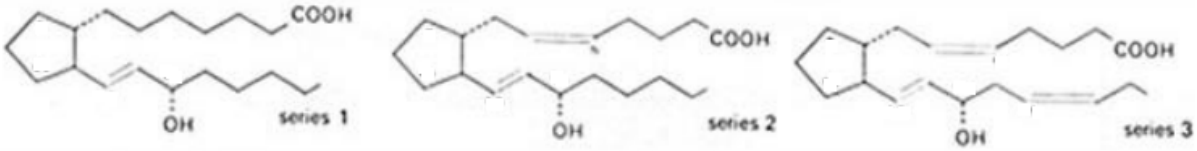
Thus based on the number of double bonds present in the side chain, the prostaglandins are classified as

i) Prostaglandins – 1

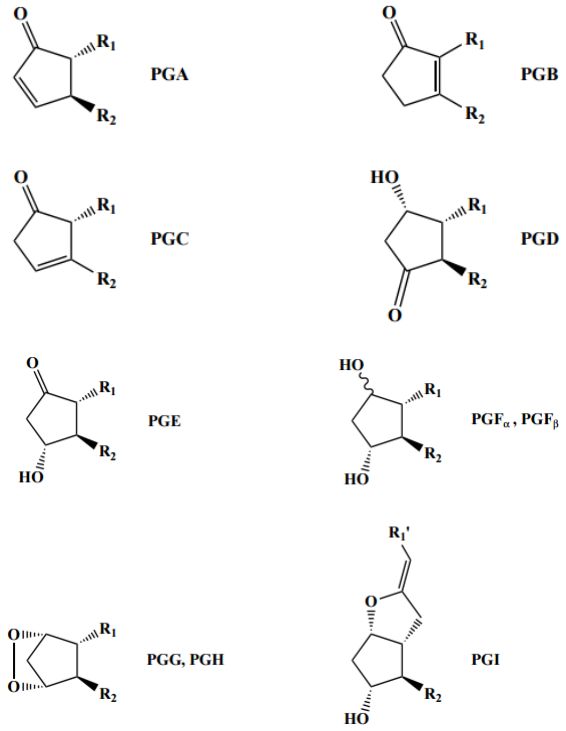
ii) Prostaglandins – 2

iii) Prostaglandins – 3

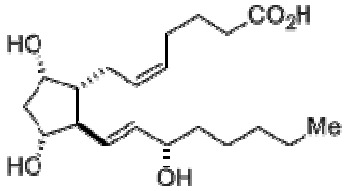
Prostaglandin-1 series have one double bond and are derived from dihomo-γ-linolenic acid, Prostaglandin-2 series have two double bonds and are derived from arachidonic acid, and Prostaglandin -3 series have three double bonds and are derived from eicosapentaenoic acid.



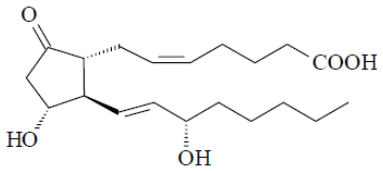
Prostaglandins are also calcified according to the structure and the substituents present on the cylcopentane ring like OH group, keto group, C=C and peroxide function. They are broadly classified as PGA, PGB, PGC, PGD, PGE, PGF, PGG, and PGH based on their cyclopentane/pentene ring substitution patterns. Each general PG class is subclassified based on the degree of unsaturation (i.e. PGE1, PGE2, PGF2). The letters and numbers that follow the initial PG abbreviation indicate the nature of the unsaturation and substitution. For example, the subscript 1 in PGE1 indicates one double bond in the side chains, while the subscript 2 in PGE2 indicates two double bonds in the side chains.



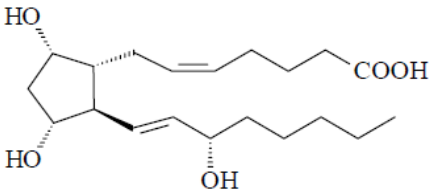
Similarly for example, in prostaglandin F2α (PGF2α), the F indicates that the prostaglandin has two hydroxyl groups in the cyclopentanone ring (F series), the 2 indicates that it has two double bonds, and the α indicates that its hydroxyl grouping at carbon 9 is in the α configuration.



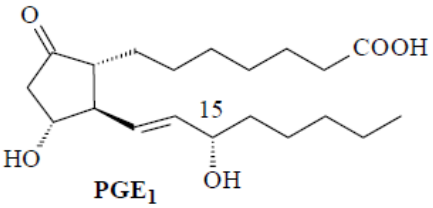
Structure-function relationships and nomenclature of the some prostaglandins are given below



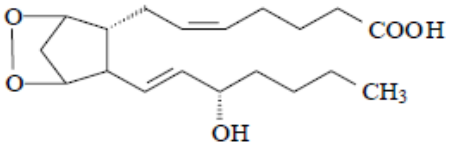
Dinoprostone (PGE_2)



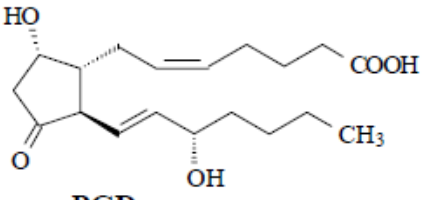
Dinoprost ($PGF_{2\alpha}$)



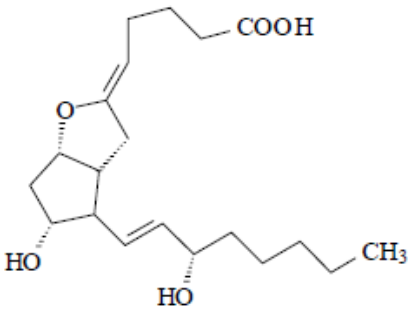
PGE_1



PGH_2



PGD_2



PGI_2 (Prostacyclin)

15.5 Functions of Prostaglandins

Prostaglandins exhibit a wide range of biological effects, and their actions are among the most varied of any naturally occurring compounds. Despite this observation, this group of lipids displays a marked structure-activity specificity, which is determined mainly by cyclopentanone ring substitutions and the degree of unsaturation of the prostanoic acid side chains. The cellular response to prostaglandins is mediated by their interaction with plasma membrane receptors.

There are a variety of physiological effects including:

1. Activation of the inflammatory response, production of pain, and fever. When tissues are damaged, white blood cells flood to the site to try to minimize tissue destruction. Prostaglandins are produced as a result.

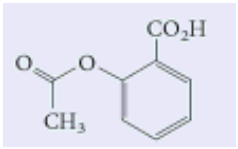
2. Blood clots form when a blood vessel is damaged. A type of prostaglandin called thromboxane stimulates constriction and clotting of platelets. Conversely, PGI2, is produced to have the opposite effect on the walls of blood vessels where clots should not be forming.

3. Certain prostaglandins are involved with the induction of labor and other reproductive processes. PGE2 causes uterine contractions and has been used to induce labor.

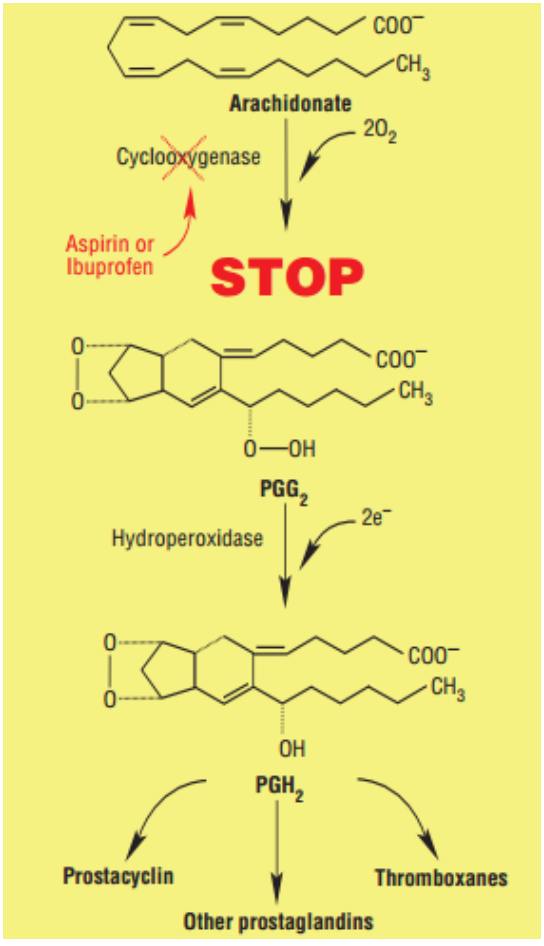
4. Prostaglandins are involved in several other organs such as the gastrointestinal tract (inhibit acid synthesis and increase secretion of protective mucus), increase blood flow in kidneys, and leukotriens promote constriction of bronchi associated with asthma.

15.6 Effects of Aspirin and other Pain Killers

Aspirin is a drug used as pain killer. When prostaglandins induce inflammation, pain, and fever, aspirin blocks an enzyme called cyclooxygenase, COX-1 and COX-2, which is involved with the ring closure and addition of oxygen to arachidonic acid converting to prostaglandins. The acetyl group on aspirin is hydrolyzed and then bonded to the alcohol group of serine as an ester. This has the effect of blocking the channel in the enzyme and arachidonic cannot enter the active site of the enzyme.



Aspirin

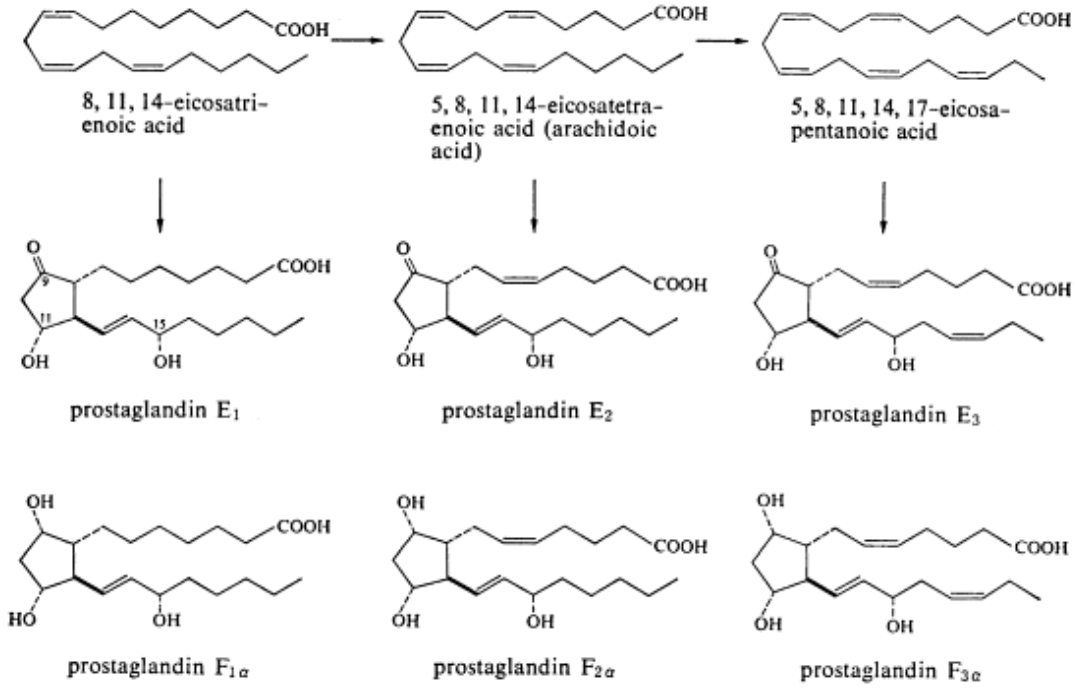


By inhibiting or blocking this enzyme, the synthesis of prostaglandins is blocked, which in turn relives some of the effects of pain and fever.

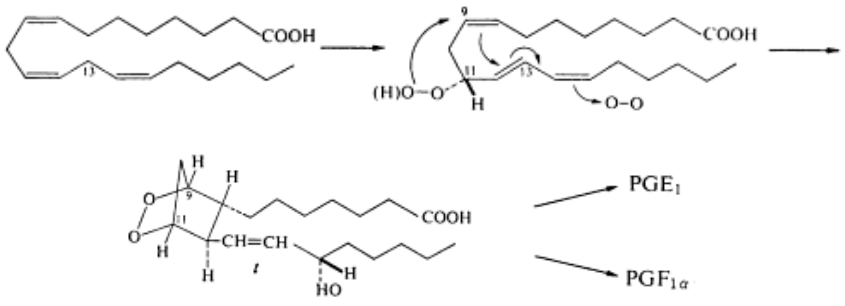
Aspirin is also thought to inhibit the prostaglandin synthesis involved with unwanted blood clotting in coronary heart disease. At the same time an injury while taking aspirin may cause more extensive bleeding.

15.7 Biosynthesis of Prostaglandins

The Prostaglandins are derived biosynthetically from polyunsaturated fatty acids by the enzymatic conversion.



The possible biosynthetic scheme for some Prostaglandins have been proposed on the basis of various experimental results and out lined below

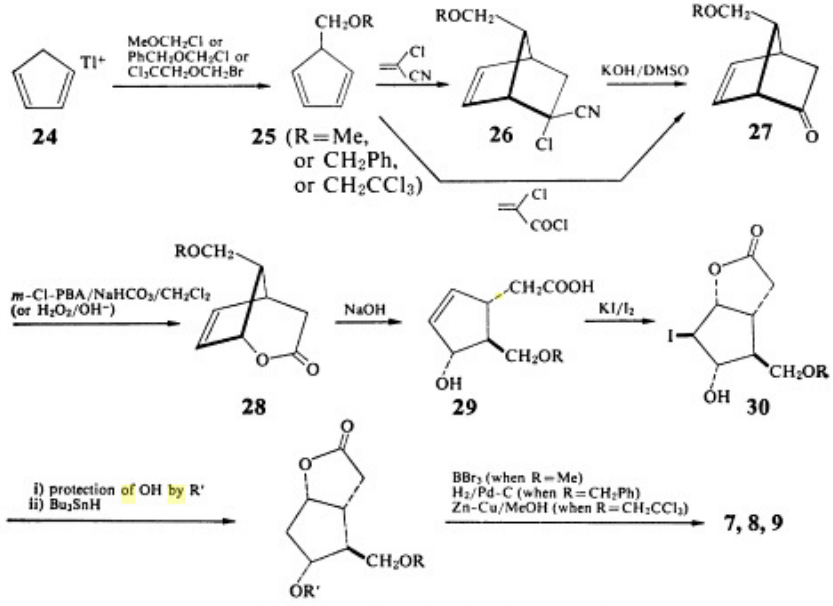


1) The initial step of biosynthesis involves stereospecific elimination of the pro-S hydrogen at C-13 and oxygenation at C-11, giving rise to the peroxide.

2) The peroxide then cyclises into the enedoperoxide by a concentrated oxygenation process

15.8 Synthesis of prostaglandin

Six prostaglandins of the 3 from each E series and the F series are designated as primary prostaglandins. E. J. Corey has been developed the general and srereocontrolled synthesis of all primary prostaglandins (PGE1, PGE2, PGE3, PGF1α, PGE2α and PGF3α) in the naturally occurring forms from common intermediates such as 7, 8 and 9. These key intermediates were synthesized by an efficient route from the bicycle[2,2,2]heptanes system as outlined below.



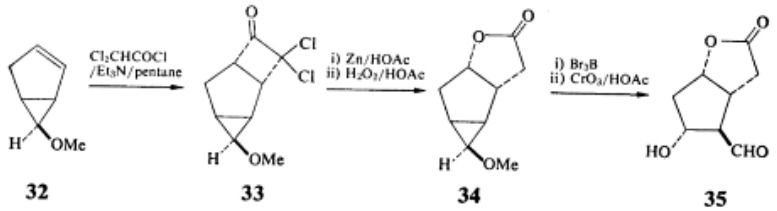
31 (R′ = Ac, or THP or *p*-$C_6H_5C_6H_4CO$)

24→25: alkylation of the thallous cyclopentadienide 24 is performed to the use of sodium or lithium slats.

25→27: Diels-Alder addition of 25 with 2-chloroacrylonitrile followed by hydrolysis afforded the anti-bicyclic ketone 27. Reaction of 25 with 2-chloroacryloylchloride also gave 27 more efficiently.

29: the optically resolution of 29 was achieved via the (+)amphetamine or (+)ephedrine salt. The primary prostaglandins were synthesized as the naturally occurring forms starting with optically active 7-9.

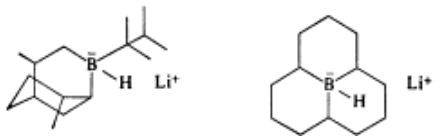
The aldehyde 25 was synthesized by the alternative route shown below

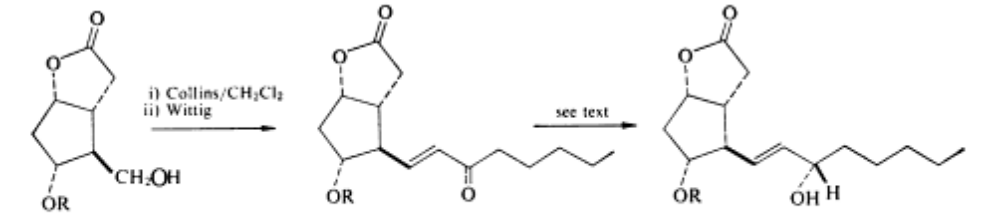


15.9 Synthesis of PGE1, PGE2, PGF1α and PGE2α

7→10: Reaction of the aldehyde derived from 2 with the sodiumsalt of dimethyl 2-oxoheptylphosphonate afforded stereospecifically the *trans* enone lactone 10.

10→13; **11→14**; **12→15**: Reduction of the enone 10 with zinc borohydride gave a mixture of the 15α and 15β hydroxyl derivatives in a ratio of 1:1. When a diisopinocamphenylborane/ methylllithium/hexamethylphosphosphoramide system is used for reduction of p-phenylbenzoate 11, the ratio of the resulting 15α and 15β isomers is improved to 66:31. A trialkylborane derivative 19 prepared from limonene and hexylborane was fond to be more satisfactory reagent for the reduction of 11 to 14 (α:β = 4.5:1). The most outstanding result was obtained when the p-phenylphenylurethane was reduced by below shown reagents (α:β = 92:8)





7 (R = Ac)
8 (R = *p*-$C_6H_5C_6H_4CO$)

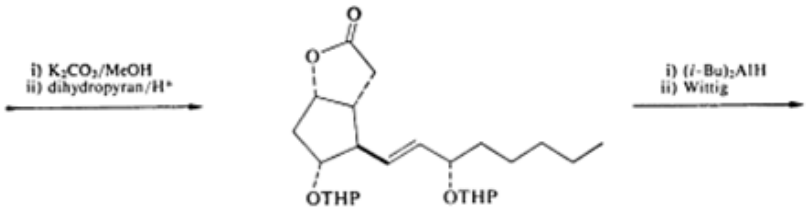
10 (R = Ac)
11 (R = *p*-$C_6H_5C_6H_4CO$)
12 (R = *p*-$C_6H_5C_6H_4NHCO$)

13 (R = Ac)
14 (R = *p*-$C_6H_5C_6H_4CO$)
15 (R = *p*-$C_6H_5C_6H_4NHCO$)

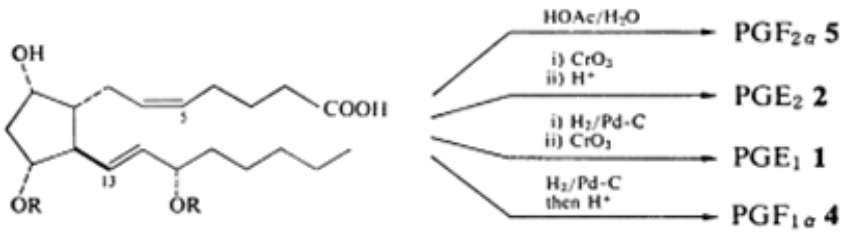
16→17: Condensation of the aldehyde derived from 16 with the wittig reagent prepared from 5-triphenylphosphoniovalerate ion yielded the desired Δ^5-cis product 17 exclusively.

17→4; 17→1: catalytic hydrogenation of 17 afforded a dihydro derivative saturated selectively at the Δ^5 double bond. Selective hydrogenation was also performed more effectively by the use of the bis-dimethylisopropylsilyl derivative 18. This might be due to steric screening of the Δ^{13} double bond by the bulky protecting group.

2→5; 1→4: Sterospecific reduction of PGE's to PGF's was achieved with the bulky trialkyl borohydride reagents mention above



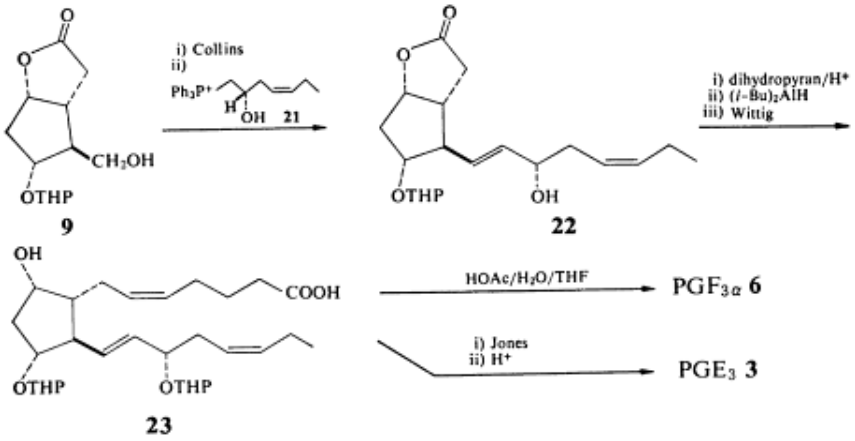
16



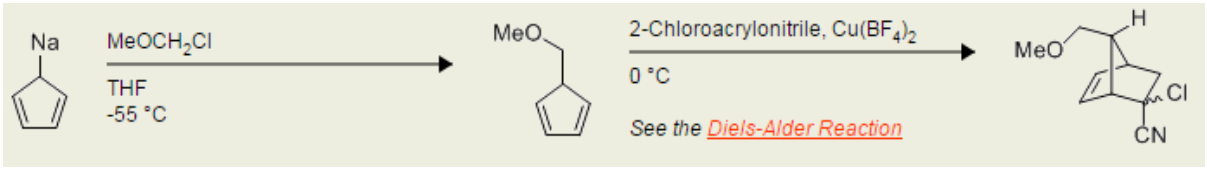
17 (R=THP)
18 (R=dimethylisopropylsilyl)

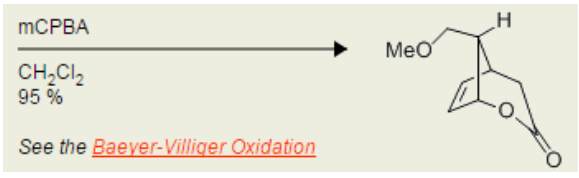
15.10 Synthesis of PGE3 and PGF3α

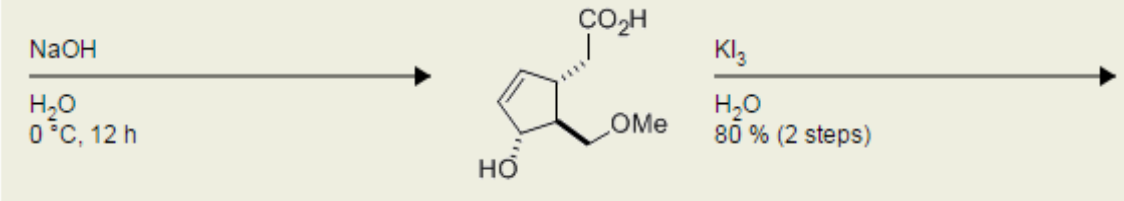
9→22: The optically active aldehyde derived from 9 was converted stereospecifically to the unsaturated alcohol 22 by the β-oxido ylide reagent derived from the hydroxyl phosphonium salt (S)(+)21.

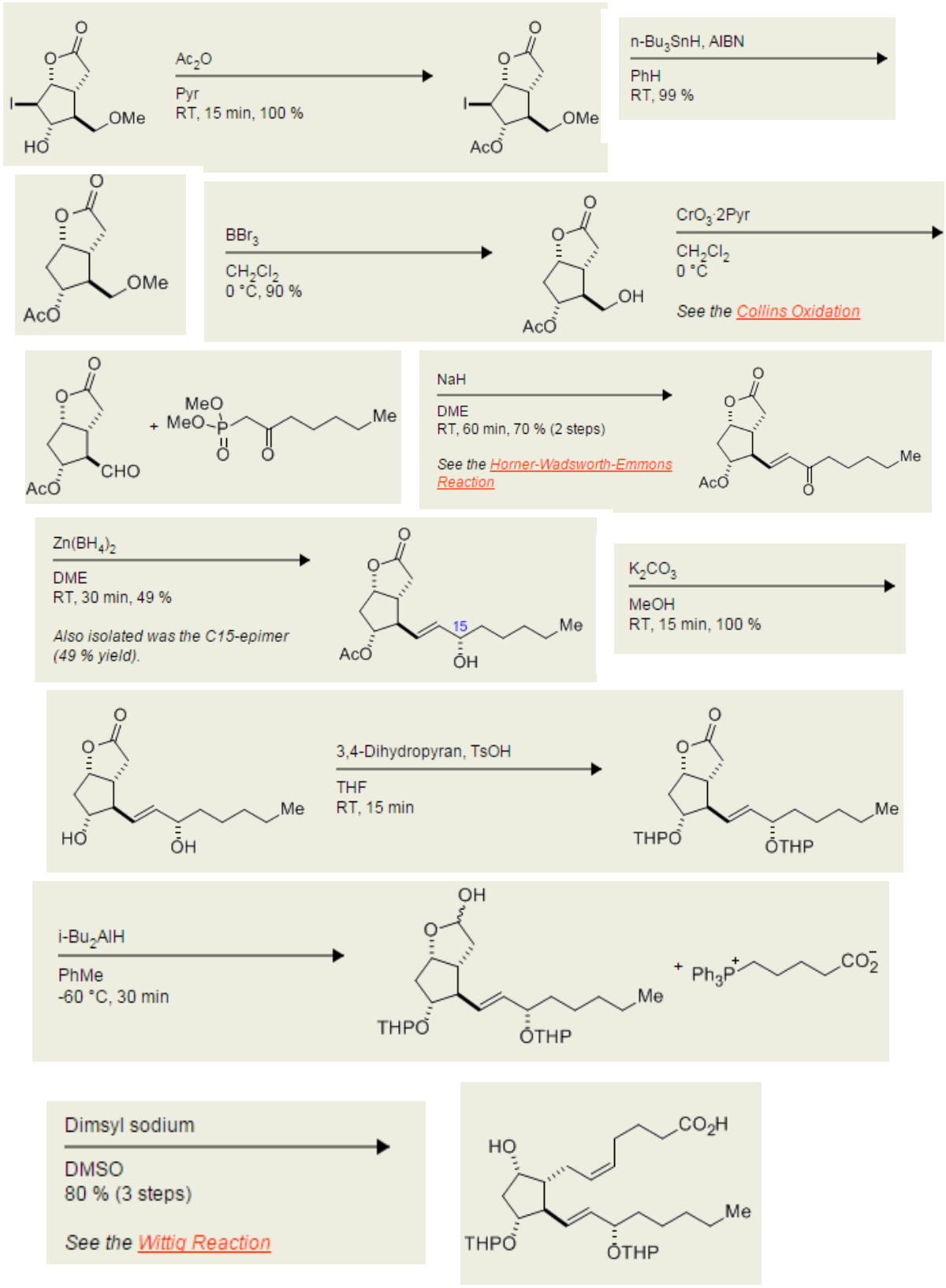


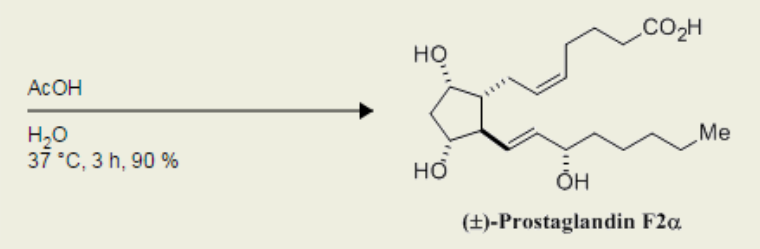
The complete synthesis scheme of prostaglandin with named reaction used in each steps is illustrated below








Ac2O
Pyr
RT, 15 min, 100 %
n-Bu3SnH, AIBN
PhH
RT, 99 %
BBr3
CH2Cl2
0 °C, 90 %
CrO3·2Pyr
CH2Cl2
0 °C
See the Collins Oxidation
NaH
DME
RT, 60 min, 70 % (2 steps)
See the Horner-Wadsworth-Emmons Reaction
Zn(BH4)2
DME
RT, 30 min, 49 %
Also isolated was the C15-epimer (49 % yield).
K2CO3
MeOH
RT, 15 min, 100 %
3,4-Dihydropyran, TsOH
THF
RT, 15 min
i-Bu2AlH
PhMe
-60 °C, 30 min
Dimsyl sodium
DMSO
80 % (3 steps)
See the Wittig Reaction



15.11 Summary of the unit

In the most basic sense, PGs are twenty-carbon molecules made up of a fivemembered (cyclopentane) ring with two aliphatic sidechains. Two other structural features found in all prostaglandins are C-15 is a stereo centre with an attached hydroxyl group, and C-9 is always has an attached to oxygen function (present as a hydroxyl group, ketone, or part of a bridged peroxide moiety). There are a few other distinguishing characteristics of PGs. In most cases, the carbons to which the side chains are attached are stereo centers, and some PGs have second oxygen function attached to either the C-10 or C-11 positions. Also, PGs can have up to three double bonds present in the sidechains. These structural details separate the known prostaglandins into nine families (or "classes"), while the absence or presence of double bonds in the sidechains separates the PGs into three "series".

15.12 Key words

Prostaglandins; Classification of prostaglandins; Functions of Prostaglandins; Effects of Aspirin and other Pain Killers; Biosynthesis of Prostaglandins; PGE1; PGE2; PGF1α; PGE2α; PGE3; PGF3α

15.13 References for further studies

1) The logic of chemical synthesis; E.J. Corey, Xue-Min Cheng; *John-Wiley & Sons*;

2) Natural Products Chemistry, Volume 2; Koji Nakanishi, Toshio Goto, Shô Itô; *Elsevier*, **2013**.

3) Organic Synthesis Via Examination of Selected Natural Products; David J. Hart; *World Scientific,* **2011**.

15.14 Questions for self under standing

1) What are prostaglandins?

2) Write the general skeletal structure prostaglandin

3) Discuss the classification of prostaglandins

4) List the functions of prostaglandins

5) Discus the effects of Aspirin and other Pain Killers on prostaglandins action

6) Explain the biosynthesis of prostaglandins

7) Discuss the synthesis of PGE1, PGE2, PGF1α,PGE2α ,PGE3 and PGF3α

8) Write the structural difference between PGE1, PGE2

9) Write the structural difference between of PGE3 and PGF3α

UNIT-16

Structure

16.0 Objectives of the unit

After studying this unit you are able to

- Identify the different hormones secreted by various glands of body
- List the general characteristics of hormones
- Explain the effect of control of hormonal secretions
- Explain the functions of antibody
- Explain the functions of antioxidant

16.1 Introduction

The general function of the endocrine system is to integrate body systems, in conjunction with the nervous system. The glands are effectors or responsive body parts that are stimulated by motor impulses from the autonomic nervous system. Some of these glands, endocrine glands compose the endocrine system. Endocrine gland is a gland that secretes hormones directly into the bloodstream. Endocrine glands include the following glands

1. Hypothalamus
2. Pituitary
3. Pineal gland
4. Thyroid
5. Parathyroids
6. Thymus
7. Adrenals
8. Pancreas
9. Testes
10. Ovaries

Exocrine gland is a gland that secretes substances into ducts which then leave the body (i.e. sweat/sebaceous glands) or into a internal space or lumen (i.e. digestive glands). Exocrine glands are not part of the endocrine system.

16.2 Hormones

Hormones are very powerful substances secreted by an endocrine gland into the bloodstream that affects the function of another cell or "target cell".

Hormones are mainly classified in to two types they are

a) Steroid hormones and

b) Non-steroid hormones

A steroid hormone passes easily through the target cell membrane whereas a non-steroid hormone requires a receptor on the target cell membrane to allow the hormone to enter the target cell.

16.3 General Characteristics of hormones

1. Needed in very small amounts (potent)
2. Produce long-lasting effects in the cells they target
3. Regulate metabolic processes (maintain homeostasis)
4. Are regulated by negative-feedback mechanisms
5. May be steroid (produced from cholesterol = fat-soluble) or non-steroid (water-soluble).

16.4 Control of hormonal secretions

The Hypothalamus secretes "releasing hormones" that target the anterior pituitary gland. The Anterior pituitary gland (which hangs from the base of the brain) may then secrete 6 different hormones, they are

1. *Human Growth Hormone (HGH)*

Human growth hormone (hGH) plays a vital role in growth and development. This hormone controls the growth of the body and targets the bone, muscle and adipose tissue. It is naturally produced by somatotropic cells in the anterior pituitary gland. The hormone is produced as a 217 amino acid precursor protein. The 26 N-terminal amino acids correspond to a signal peptide, which is essential for hormone secretion. This signal peptide is cleaved during the secretion process to yield the mature, 191 amino acid form of hGH.

Mature hGH travels through the bloodstream and interacts with a specific hGH-receptor on the surface of various cells, including muscle, bone, and cartilage. Binding of hGH to its receptor causes dimerization and signal transduction, which ultimately stimulates cellular division. HGH also indirectly influences growth by stimulating the liver to produce additional growth factors, such as insulin-like growth factor-1

2. *Thyroid stimulating hormone (TSH)*

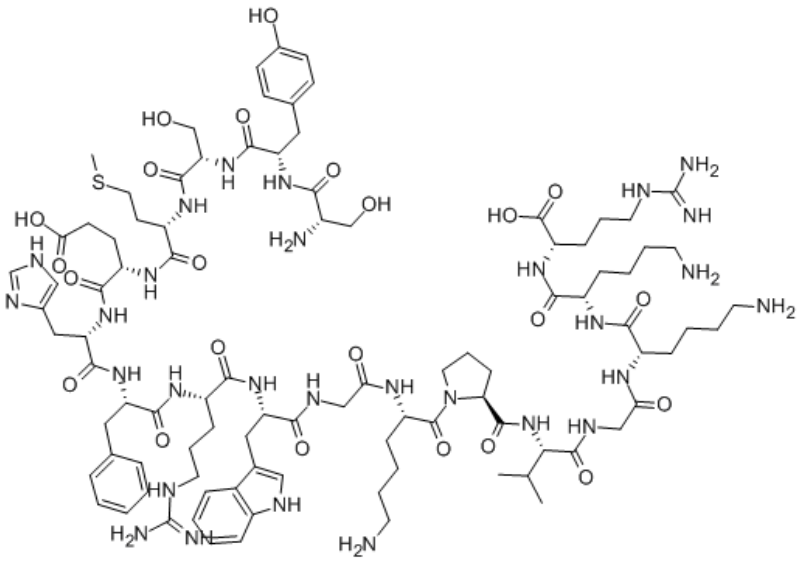
Thyroid-stimulating hormone (TSH) is also called thyrotropin. TSH is a key protein in the control of thyroid function. TSH is synthesized in the anterior pituitary stimulated by thyrotropin-releasing hormone (TRH) and inhibited by thyroid hormone in a classical endocrine negative-feedback loop. TSH controls thyroid function upon its interaction with the G protein-coupled TSH. TSH binding to its receptor on thyroid cells leads to the stimulation of second messenger pathways involving predominantly cAMP and, in high concentrations, inositol 1,4,5-trisphosphate (IP3) and diacylglycerol (DAG), ultimately resulting in the

modulation of thyroidal gene expression. This hormone controls the secretion of hormones by the thyroid gland and targets thyroid gland.

3. *Adrenocorticotropic hormone (ACTH*)

Adrenocorticotropic hormone (ACTH) is also called corticotropin or adrenocorticotropin, a polypeptide hormone formed in the pituitary gland that regulates the activity of the outer region (cortex) of the adrenal glands. This hormone controls the secretion of hormones by the adrenal cortex and targets the outer portion of the adrenal gland (cortex).

ACTH is a 39-amino acid polypeptide of which the N-terminal 24-amino acid segment is identical in all species and contains the adrenocorticotrophic activity. Upon further tissue-specific processing, ACTH can yield ALPHA-MSH and corticotrophin-like intermediate lobe peptide (CLIP).



4. *Prolactin (PRL)*

Prolactin (PRL), also known as luteotropic hormone or luteotropin. Prolactin is a polypeptide hormone that is synthesized in and secreted from specialized cells of the anterior pituitary gland, the lactotrophs. The hormone was given its name based on the fact that an extract of bovine pituitary gland would cause growth of the crop sac and stimulate the elaboration of crop milk in pigeons or promote lactation in rabbits. This hormone stimulates the production of milk by the mammary glands and targets the mammary glands.

5. *Follicle stimulating hormone (FSH)*

Follicle-stimulating hormone (FSH) is a gonadotropin, a glycoprotein polypeptide hormone. FSH is synthesized and secreted by the gonadotropic cells of the anterior pituitary gland and regulates the development, growth, pubertal maturation, and reproductive processes of the

body. This hormone response depends upon sex. In females, FSH stimulates maturation of an ovarian follicle and ovum. In males, FSH stimulates the maturation of sperm in the testes. A gonadotropin targets the primary sex organs (ovary & testis).

6. *Luteinizing hormone (LH)*

Luteinizing hormone (LH) also known as lutropin. It is a hormone produced by gonadotropic cells in the anterior pituitary gland. In females, an acute rise of LH ("LH surge") triggers ovulation and development of the corpus luteum. In males, where LH had also been called interstitial cell–stimulating hormone (ICSH), it stimulates Leydig cell production of testosterone. It acts synergistically with FSH This hormone response also depends upon sex. In females, LH causes ovulation and in males, LH causes secretion of testosterone. A gonadotropin targets ovaries & testes.

16.5 Posterior pituitary gland

The posterior pituitary gland is located behind the anterior pituitary gland. It is continuous with nerve fibers (supraopticohypophyseal tract) of the hypothalamus. It does not actually produce hormones (they are produced by the hypothalamus) but stores them until it is stimulated to release them.

Pituitary gland secretes two hormones they are

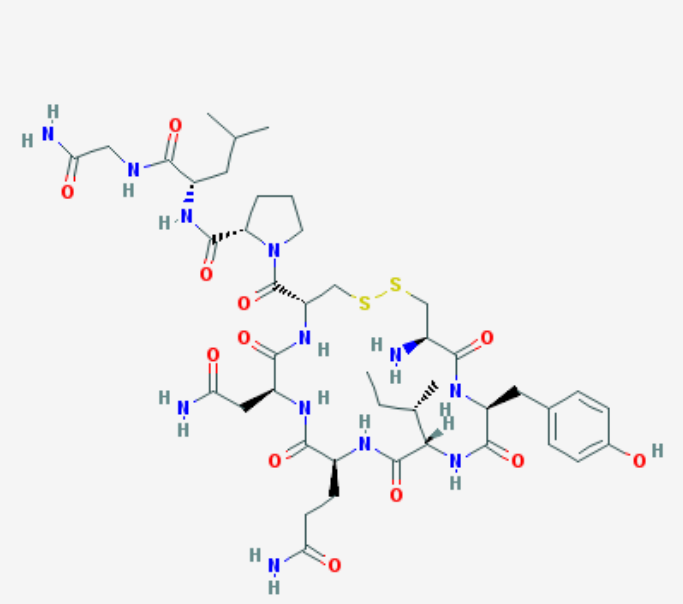
a) Anti-Diuretic hormone (ADH)

Antidiuretic hormone (ADH), is also known as vasopressin. It is a polypeptide hormone secreted by the posterior pituitary gland. Its principal action is to regulate the amount of water excreted by the kidneys. They target the kidney tubules (DCT) and cause the kidney tubules to reabsorb water back into the bloodstream, and therefore control water balance and blood pressure. Thus concentrating the salts and waste products in the liquid, this will eventually become urine.

b) Oxytocin (OT),

Ocytocin, which means "quick birth" in Greek, Oxytocin is a relatively small polypeptide hormone in mammals that plays an important role in birth and ejection of milk from the female breast. It also acts as a neurotransmitter in the brain. It differs from vasopressin by two amino acids at residues 3 and 8. Oxytocin acts on smooth muscle cells, such as causing uterine contractions and milk ejection.

Oxytocin has the chemical formula $C_{43}H_{66}N_{12}O_{12}S_2$. It is a relatively short polypeptide, being composed of only nine amino acids (a nonapeptide). The sequence is cysteine - tyrosine - isoleucine - glutamine - asparagine - cysteine - proline - leucine - glycine (CYIQNCPLG). The cysteine residues form a sulfur bridge.



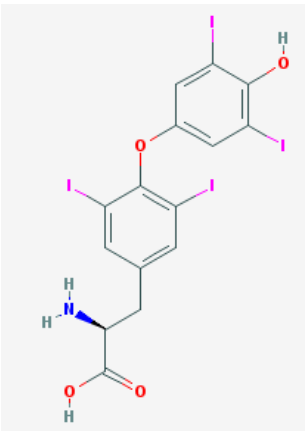
They targets uterine smooth muscle and breasts and causes uterine muscle contraction and milk production.

16.6 Thyroid gland

The thyroid gland is located below larynx and around trachea. This gland is involved in iodine uptake (in order to produce thyroxine (T_4) & triiodothreonine (T_3)). This gland produces 3 hormones when stimulated by TSH they are

a) Thyroxine (T_4) & Triiodothreonine (T_3), they increase basal metabolic rate by stimulating cellular oxygen use to produce ATP thereby regulate metabolism.

L-thyroxine is the major hormone derived from the thyroid gland. Thyroxine is synthesized via the iodination of tyrosines (monoiodotyrosine) and the coupling of iodotyrosines (diiodotyrosine) in the thyroglobulin. Thyroxine is released from thyroglobulin by proteolysis and secreted into the blood. Thyroxine is peripherally deiodinated to form triiodothyronine which exerts a broad spectrum of stimulatory effects on cell metabolism.



b) Calcitonin

Calcitonin also known as thyrocalcitonin is a 32-amino acid linear polypeptide hormone that is produced in humans primarily by the parafollicular cells (also known as C-cells) of the thyroid, and in many other animals in the ultimobranchial body. It acts to reduce blood calcium (Ca^{2+}), opposing the effects of parathyroid hormone. Calcitonin, it release is triggered by an increase in blood calcium levels and targets bone cells (inhibits osteoclast activity) & kidney tubules (causes secretion of calcium into urine). Therefore, causes a decrease in blood calcium and phosphate levels to normal.

16.7 Parathyroid glands

The Parathyroid glands consist of 4 small glands. This gland is located within the thyroid gland and produces a hormone called Parathyroid Hormone (PTH). The release is stimulated by a decrease in blood calcium levels. PTH targets bone cells (activates osteoclasts) and kidney cells (causes kidney tubules to reabsorb more calcium). Therefore, causes an increase in blood calcium and phosphate levels to normal. PTH and calcitonin together maintain the homeostasis of Ca^{++} in the blood.

16.8 Adrenal glands

The Adrenal glands are located atop the kidneys. They are divided into an outer adrenal cortex and an inner renal medulla.

a) The adrenal medulla, it is located in the central portion of the adrenal glands and produces 2 closely related hormones which function in the sympathetic division of the autonomic nervous system

1. Epinephrine
2. Norepinephrine

These hormones target

1. The heart (increased heart rate and blood pressure);
2. Smooth muscle contraction (blood vessels,);
3. The lungs (increased breathing: rate, rhythm, depth).

b) The adrenal cortex, it is located in the outer portion of the adrenal glands and produces many steroid hormones when stimulated by ACTH. Examples are *Aldosterone,* it regulates water and electrolyte levels in the blood (regulates blood pressure). *Cortisol,* it regulates glucose metabolism. *Sex hormones* that supplement those of the ovary and testis.

16.9 Pancreas

The Pancreas is located behind the stomach on the left side of abdomen. It functions as both an exocrine gland (digestion) and endocrine gland. Pancreas contains endocrine organs called *Islets of Langerhans* which produce 2 hormones they are

1. Insulin it is produced by Beta cells Langerhans. It decreases blood glucose levels (i.e. moves glucose from bloodstream into cells and promotes glycogen formation

2. Glucagon, it is produced by Alpha cells Langerhans. It increases blood glucose levels (i.e. causes breakdown of glycogen and release of glucose into bloodstream).

16.10 Thymus gland

The Thymus gland is located in the mediastinum region behind sternum. It produces a hormone called thymosin that affects the maturation of lymphocytes (T-cells). It plays an important role in lymphatic system and immunity also decreases in size as we age.

16.11 Pineal gland

The Pineal gland is attached to the thalamus of the brain stem. It secretes a hormone called melatonin. The production of hormone is stimulated by daylight (circadian rhythm) and affects moods, emotions, etc.

16.12 Ovaries

An ovarian follicle (and ovum) starts to mature each month following puberty under the influence of FSH. The main function is developing follicle secretes estrogen also develops and maintains female secondary sexual characteristics. It targets

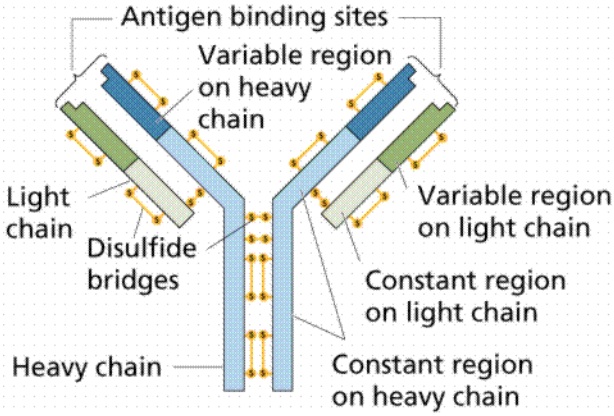
1. Hair follicles
2. Mammary glands/ breasts
3. Adipose tissue.

16.13 Antibody

An antibody is also known as an immunoglobulin, is a large Y-shaped protein produced by B-cells that is used by the immune system to identify and neutralize foreign objects such as bacteria and viruses.

Each antibody consists of four polypeptides in which two are heavy chains and other two are light chains joined to form a "Y" shaped molecule.

Though the general structure of all antibodies is very similar, a small region at the tip of the protein is extremely variable, allowing millions of antibodies with slightly different tip structures, or antigen binding site



The amino acid sequence in the tips of the "Y" varies greatly among different antibodies. This variable region, composed of 110-130 amino acids, give the antibody its specificity for binding antigen. The variable region includes the ends of the light and heavy chains. Treating the antibody with a protease can cleave this region, producing Fab or fragment antigen binding that include the variable ends of an antibody.

The constant region determines the mechanism used to destroy antigen. Antibodies are divided into five major classes, IgM, IgG, IgA, IgD, and IgE, based on their constant region structure and immune function.

The variable region is further subdivided into hypervariable (HV) and framework (FR) regions. Hypervariable regions have a high ratio of different amino acids in a given position, relative to the most common amino acid in that position. Within light and heavy chains, three hypervariable regions exist – HV 1, 2 and 3. Four FR regions which have more stable amino acids sequences separate the HV regions.

The HV regions directly contact a portion of the antigen's surface. For this reason, HV regions are also sometimes referred to as complementarity determining regions, or CDRs. The FR regions form a beta-sheet structure which serves as a scaffold to hold the HV regions in position to contact antigen.

16.14 Isotopes

IgA: Exists as a dimer, and is secreted into mucosal surfaces, such as the gut, respiratory tract, and urogenital tract, and prevents colonization by pathogens.

IgD: Functions mainly as an antigen receptor on B cells that have not been exposed to antigens. It has been shown to activate basophils and mast cells to produce antimicrobial factors.

IgE: Binds to allergens and triggers histamine release from mast cells and basophils, and is involved in allergy. Also protects against parasitic worms.

IgG: In its four forms, provides the majority of antibody-based immunity against invading pathogens. The only antibody capable of crossing the placenta to give passive immunity to fetus.

IgM: Expressed on the surface of B cells (monomer) and in a secreted form (pentamer) with very high avidity. Eliminates pathogens in the early stages of B cell mediated (humoral) immunity before there is sufficient IgG.

16.15 Function

Activated B cells differentiate into either antibody-producing cells called plasma cells, or memory cells that survive in the body years afterward in order for the immune system to remember an antigen and respond faster at future exposures. At the prenatal and neonatal stages of life, the presence of antibodies is provided by passive immunization from the mother. Early endogenous antibody production varies for different kinds of antibodies, usually appearing within the first years of life. Since antibodies exist freely in the bloodstream, they are said to be part of the humoral immune system. Circulating antibodies are produced by clonal B cells that specifically respond to only one antigen. Antibodies contribute to immunity in three ways: preventing pathogens from entering or damaging cells by binding to them; stimulating removal of pathogens by macrophages and other cells by coating the pathogen; and triggering destruction of pathogens by stimulating other immune responses such as the complement pathway.

16.16 Activation of Complement

Antibodies that bind to surface antigens (e.g., on a bacterium) attract the first component of the complement cascade with their Fc region, and initiate activation of the "classical" complement system. This results in the killing of bacteria in two ways. First, the binding of the antibody and complement molecules marks the microbe for ingestion by phagocytes in a process called opsonization; these phagocytes are attracted by certain complement molecules generated in the complement cascade. Secondly, some complement system components form a membrane attack complex to assist antibodies to kill the bacterium directly.

16.17 Activation of effector Cells

To combat pathogens that replicate outside cells, antibodies bind to pathogens to link them together, causing them to agglutinate. Since an antibody has at least two paratopes, it can bind more than one antigen by binding identical epitopes carried on the surfaces of these antigens. By coating the pathogen, antibodies stimulate effector functions against the pathogen in cells that recognize their Fc region. Those cells that recognize coated pathogens have Fc receptors which, as the name suggests, interacts with the Fc region of IgA, IgG, and

IgE antibodies. The engagement of a particular antibody with the Fc receptor on a particular cell triggers the effector function of that cell; phagocytes will phagocytose, mast cells and neutrophils will degranulate, natural killer cells will release cytokines and cytotoxic molecules—resulting ultimately in destruction of the invading microbe. The Fc receptors are isotype-specific, lending greater flexibility to the immune system, and invoking only the appropriate immune mechanisms for distinct pathogens.

16.18 Antioxidant

Antioxidants are compounds found in foods and neutralise or 'mop up' free radicals that can harm our cells. Damage caused to cells by free radicals is linked to certain diseases such as cancer and cardiovascular disease. The best sources of antioxidants are fruits and vegetables.

They act as 'free radical scavengers' preventing and repairing damage done by free radicals. Thus they protect our bodies against certain cancers and diseases. Research shows that antioxidant supplements aren't as effective as natural foods in providing antioxidants.

16.19 Free radical damage

The process of oxidation in the human causes body free radicals, which damage cell membranes and other structures. Our bodies are exposed to free radicals naturally through the process of metabolism. Overload of this free radical is cause for concern. Factors that increase the presence of free radicals in the body include environmental variables such as pollution, pesticides and radiation, as well as lifestyle factors such as smoking, stress, alcohol and poor diet.

Free radicals can cause

- Damage to nerve cells in the brain.
- Arthritis.
- Damage to the lenses of the eye.
- Premature ageing.
- Increased risk of heart disease and certain cancers.

16.20 Mechanism of action

Free radicals react with our cell membranes and other structures such as DNA, lipids and proteins by stealing electrons from these molecules and causing damage. When a free radical attacks these essential molecules they change into free radicals themselves causing a chain reaction that can lead to the destruction of a cell.

Antioxidants are able to neutralise free radicals by using their own electrons to the free radicals and therefore prevent cellular damage. After neutralising a free radical, antioxidants

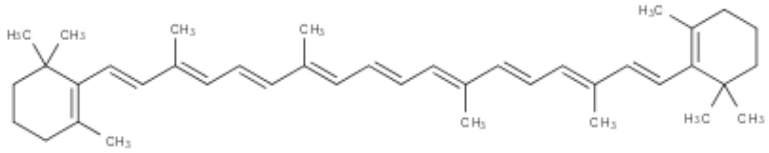
become inactive, which means they need to be constantly re-supplied to our bodies through the right foods.

16.21 Sources of antioxidants

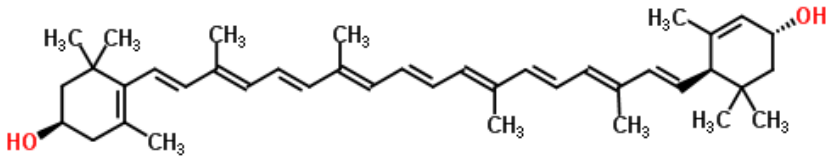
Most of antioxidants come from plants and include vitamins A, C and E and carotenoids such as beta-carotene, minerals, phenolic compounds and other natural chemicals with antioxidant properties. Flavonoids are another powerful antioxidant and are contained in red wine and tea. While vitamin supplements containing antioxidants are also available naturally. It is believed that eating a variety of natural foods like fruits, vegetables, cereals, legumes, nuts, seeds and wholegrains - is the best way to supply your body with antioxidants.

Natural antioxidant sources are

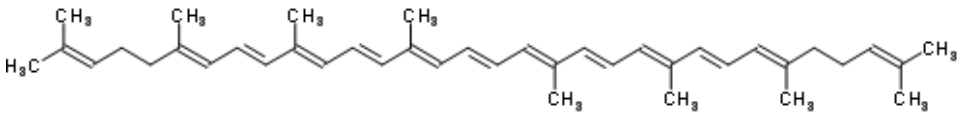
Beta-carotene – orange foods such as carrots, pumpkin, apricots, sweet potatoes and some leafy greens such as kale.



Lutein – green, leafy vegetables such as spinach.

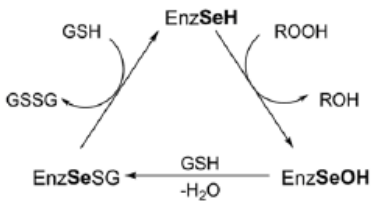


Lycopene – tomatoes, watermelon, papaya.



Selenium – rice and wheat.

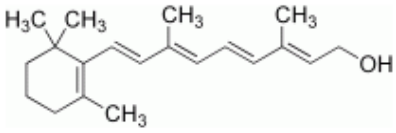
Se



EnzSeH: Glutathione Peroxidase
GSH: Glutathione

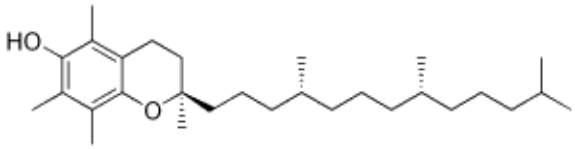
Catalytic cycle of gluthathione peroxidases showing the reduction of an organic hydroperoxide.

Vitamin A – carrots, sweet potato, milk, egg yolks.



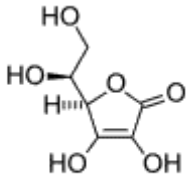
Retinol (Vitamin A)

Vitamin E – almonds, vegetable oils, mangoes, nuts, broccoli.



Alpha-tocopherol (Vitamin E)

Vitamin C – available in many fruits and vegetables such as parsley, broccoli, berries, oranges, cauliflower, kale.



Ascorbic acid (Vitamin C)

16.22 Types of Antioxidants

There are three types of antioxidants found in nature. These are

- Phytochemicals
- Vitamins and
- Enzymes

The most powerful antioxidants are found in plants. This is due to the fact that plants are exposed to UV light all throughout the day. Because plants generate a large number of free radicals, they have a natural built-in protection system that prevents the free radicals from causing the cellular damage that would eventually result in the plant withering and dying. This protection comes from naturally-occurring antioxidants.

a) Antioxidant Enzymes

Enzymes are types of antioxidants that come from the protein and minerals we eat as part of our daily diets. These enzymes are synthesized in the human body, and include superoxide dismutase (SOD), glutathione peroxidase, glutathione reductase, and catalases. In order for antioxidant enzymes to provide optimum antioxidant activity, they require co-factors such as

iron, copper, selenium, magnesium, and zinc. The quality of the protein source does have an impact on the quality of the antioxidant enzymes.

b) Antioxidant Vitamins

The human body does not produce antioxidant vitamins naturally, so it is essential to include dietary sources of them in our daily intake of food, be it through foods or supplements. Common antioxidant vitamins include vitamins A, C, E, folic acid, and beta-carotene.

Vitamin A is particularly important for improving the immune system, eye health, tissue repair, and cholesterol levels. Vitamin C helps to protect the skin from UV damage, promotes better iron absorption, provides greater resistance to infections, and helps to regulate blood cholesterol. Vitamin E is important for maintaining healthy blood vessels, improving skin conditions, and protecting the body's membrane. Meanwhile, folic acid is important to women of childbearing years, particularly in preventing the development of neural tube defects in the fetus.

Beta-carotene is a powerful carotenoid (which is a type of phytochemical) that is considered to offer the best protection against singlet oxygen and free radicals. This vitamin is most commonly found in orange-colored vegetables like carrots, pumpkins, and sweet potatoes, and dark green vegetables like spinach, kale, and collards.

Coenzyme Q10 (or CoQ10), is a vitamin-like substance produced by the body that has been shown to be a necessary component in the basic functioning of cells. The production of this substance does decrease naturally as we age, and its reduction has been linked to the development of various age-related diseases and conditions.

c) Antioxidant Phytochemicals

Phytochemicals are the antioxidants that are naturally used by plants to protect themselves against free radicals. Studies show that humans who eat sources of phytochemicals also benefit from the antioxidant properties of the plant. Phytochemicals are broken down into the following categories:

- Carotenoids
- Flavonoids
- Allyl sulfides
- Polyphenols

Most natural whole foods, such as whole grains, fruits, and vegetables, contain phytochemicals, whereas processed or refined foods contain little to no phytochemicals.

16.23 Summary of the unit

Hormones are chemical messengers that are secreted directly into the blood, which carries them to organs and tissues of the body to exert their functions. There are many types of hormones that act on different aspects of bodily functions and processes. Some of these include, Development and growth, Metabolism of food items, Sexual function and reproductive growth and health, Cognitive function and mood, Maintenance of body temperature and thirst

Hormones are secreted from the endocrine glands in the body. The glands are ductless, so hormones are secreted directly into the blood stream rather than by way of ducts. Some of the major endocrine glands in the body include Pituitary gland, Pineal gland, Thymus, Thyroid, Adrenal glands, Pancreas, Testes, Ovaries

These organs secrete hormone in microscopic amounts and it takes only very small amounts to bring about major changes in the body. Even a very slight excess of hormone secretion can lead to disease.

Antioxidants are a class of molecules that are capable of inhibiting the oxidation of another molecule. Our body naturally circulates various nutrients due to their antioxidant properties. It also manufactures antioxidant enzymes in order to control free radical chain reactions.

16.24 Key words

Hormones; Human Growth Hormone (HGH); Thyroid stimulating hormone (TSH); Adrenocorticotropic hormone (ACTH); Prolactin (PRL); Follicle stimulating hormone (FSH); Luteinizing hormone (LH); Posterior pituitary gland; Anti-Diuretic hormone (ADH); Oxytocin (OT); Thyroid gland; Parathyroid glands; Adrenal glands; Pancreas; Thymus gland; Pineal gland; Ovaries; Antibody; Isotopes ; Antioxidant

16.25 References for further study

1) Hormones; Anthony W. Norman, Helen L. Henry; *Academic Press*, **2014**.
2) Hormones; Book 11 of Basic biology course; *CUP Archive*, **1979**.
3) Hormones; Steve Parker, Carol Ballard; *Heinemann Library*, **2003**.
4) Natural Antioxidants: Chemistry, Health Effects, and Applications; Fereidoon Shahidi; *The American Oil Chemists Society*, **1997**.
5) Oxidants, Antioxidants And Free Radicals; Steven Baskin, Harry Salem; *CRC Press*, **1997**.

16.26 Questions for self understanding

1) What are hormones?
2) Write the general characteristics of hormones.

3) Discuss the effect of control of hormonal secretions on human body.

4) What is Human Growth Hormone (HGH)? Discuss their biological significance.

5) What is Thyroid stimulating hormone (TSH)? Discuss their biological significance.

6) What is Adrenocorticotropic hormone (ACTH)? Discuss their biological significance.

7) What is Prolactin (PRL)? Discuss their biological significance.

8) What is Follicle stimulating hormone (FSH)? Discuss their biological significance.

9) What is Luteinizing hormone (LH)? Discuss their biological significance.

10) What are the hormones serrated by the posterior pituitary gland?

11) What is Anti-Diuretic hormone (ADH)? Discuss their biological significance.

12) What are the hormones serrated by Thyroid gland? Discuss their biological significance.

13) What are the hormones serrated by Parathyroid glands? Discuss their biological significance.

14) What are the hormones serrated by Adrenal glands? Discuss their biological significance.

15) What are the hormones serrated by Pancreas? Discuss their biological significance.

16) What are the hormones serrated by Thymus gland? Discuss their biological significance.

17) What are the hormones serrated by Pineal gland? Discuss their biological significance.

18) What are the hormones serrated by Ovaries? Discuss their biological significance.

19) What is Antibody? Explain their significance for living systems.

20) What are the hormones serrated by Isotopes

21) What are the hormones serrated by Function

22) What is activation of complement?

23) Discuss activation of effectors Cells.

24) What are antioxidants?

25) Discuss the free radical damage in living system.

26) Discuss the mechanism of action of antioxidants

27) Write a note on types of antioxidants. Give two examples for each