

## 4.1 IMPORTANCE of WEAK NON-COVALENT INTERACTIONS in BIOLOGY

*“Apart from consideration of the hydrogen bond, we organic chemists have really paid little attention to linkages other than the purely covalent. I believe that it will be the duty of organic chemists in the future to study the weak non-bonding interactions which are of enormous importance in the large natural macromolecules. Such studies will lead to a new blossoming of organic chemistry in the future.”*

-----Lord Alexander R. Todd

**Weak Non-Covalent Interactions**

- 1) the power of supramolecular chemistry,
- 2) helps in conformational adaptation and showing biological activities by biopolymers or small peptides
- 3) thus allowing strong and selective interactions and recognition of specific guests or within themselves to be achieved.

↓

- 1) The Electrostatic Interactions
- 2) The Aromatic Ring p-p Stacking Forces
- 3) The Hydrophobic Interactions
- 4) The Van der Waals interactions
- 5) The Hydrogen Bonding Interactions

The diagram also shows a chromosome, DNA double helix, a gene on a single strand (GGATATCCAAGC), and the chemical structure of one nucleotide.

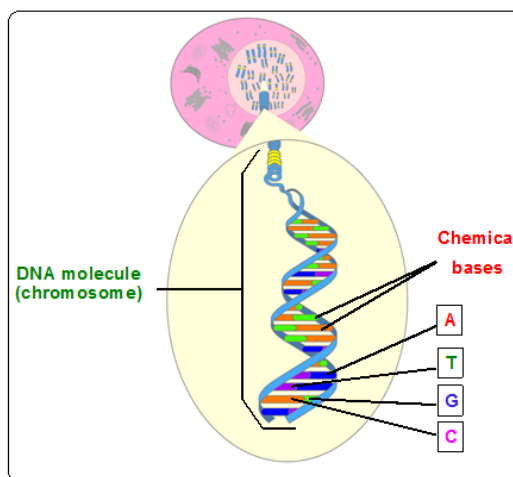
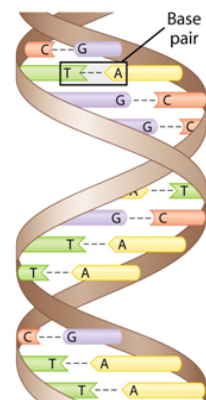
## 4.2. WEAK INTERACTIONS AND THE DNA

Weak non-covalent interactions play important role in many chemical and biological processes. As a matter of fact, the phenomenon of “Molecular Recognition” is dominated by weak forces, which include H-bonding, electrostatic, stereo-electronic, p-stacking and related hydrophobic interactions.

In the biological world, maintenance of protein’s secondary structure is governed by such weak forces.

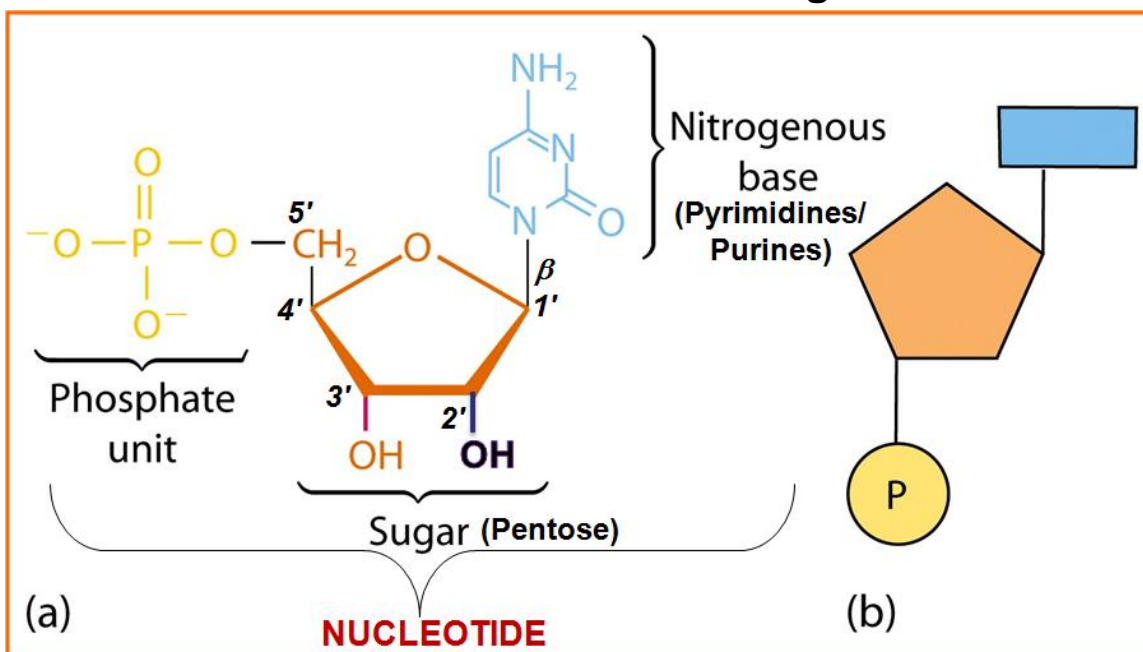
The, nucleic acids are no exceptions.

### WHAT IS DNA?



Deoxyribonucleic acid, or DNA, is a **nucleic acid** molecule that contains the **genetic instructions** used in the **development** and functioning of all known **living organisms**.

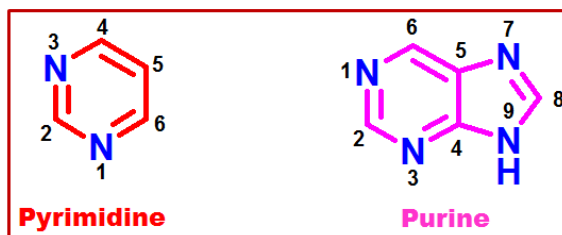
### 4.3. Nucleic Acids: The Ultimate Building Blocks



### 4.4 THE DNA and RNA BASES

In order to understand the structure and properties of DNA and RNA, we need to look at their structural components.

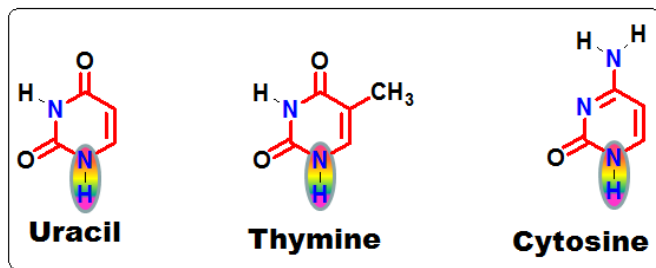
We begin with certain heterocyclic aromatic compounds called pyrimidines and purines.



- Pyrimidine and purine are the names of the parent compounds of two types of nitrogen-containing heterocyclic aromatic compounds.
- Nitrogenous bases are heterocyclic amines
- Cyclic compounds with at least 1 N atom in the ring structure
- Purines are a double ring structure
- A 6-member ring fused to a 5-member ring
- Pyrimidines consist of a single 6-membered ring

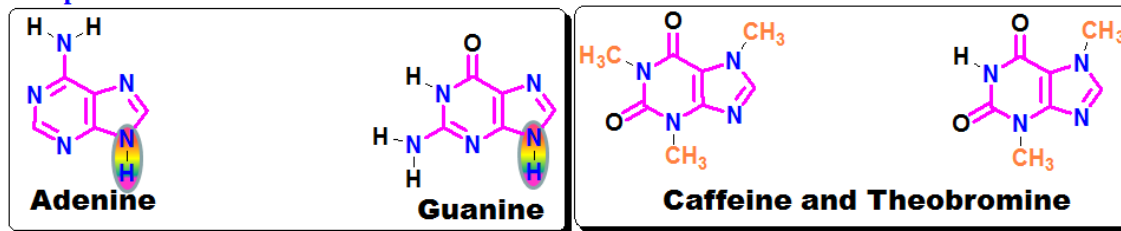
### Important Pyrimidine Bases

- The bases pyrimidines that occur in DNA are cytosine and thymine. Cytosine and uracil are the pyrimidines in RNA.



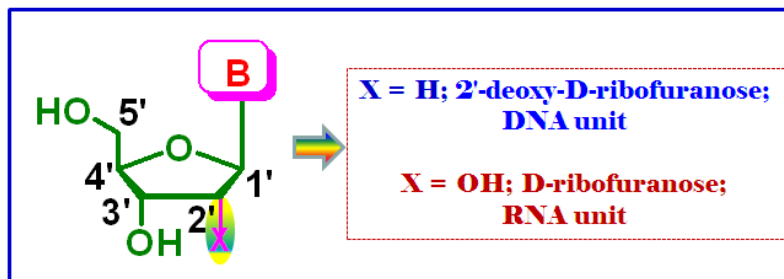
### Important Purine Bases

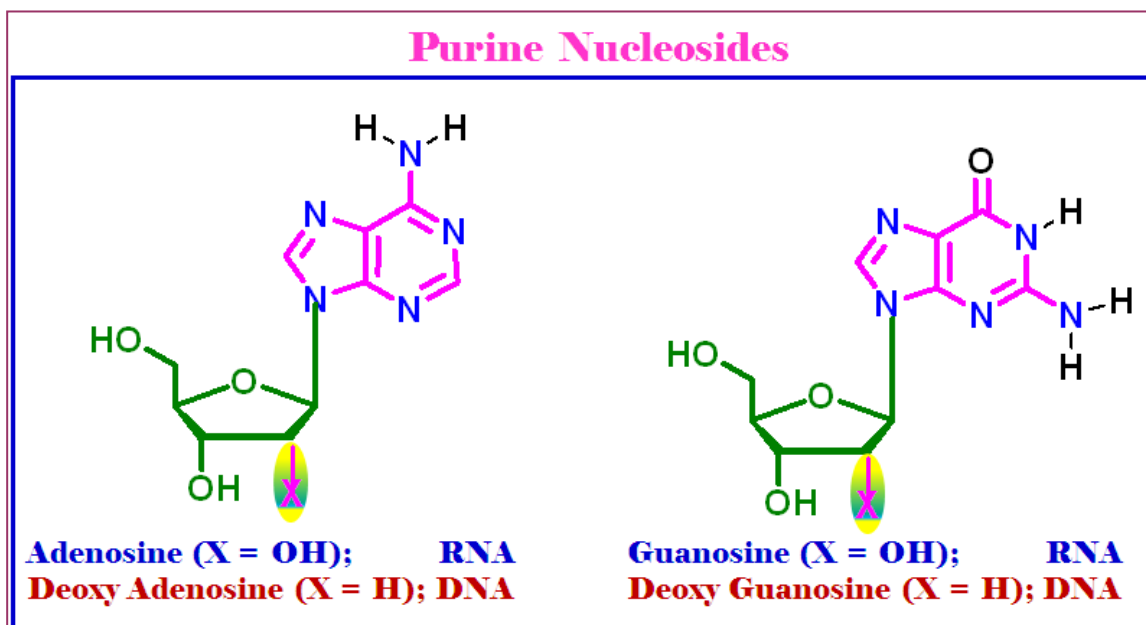
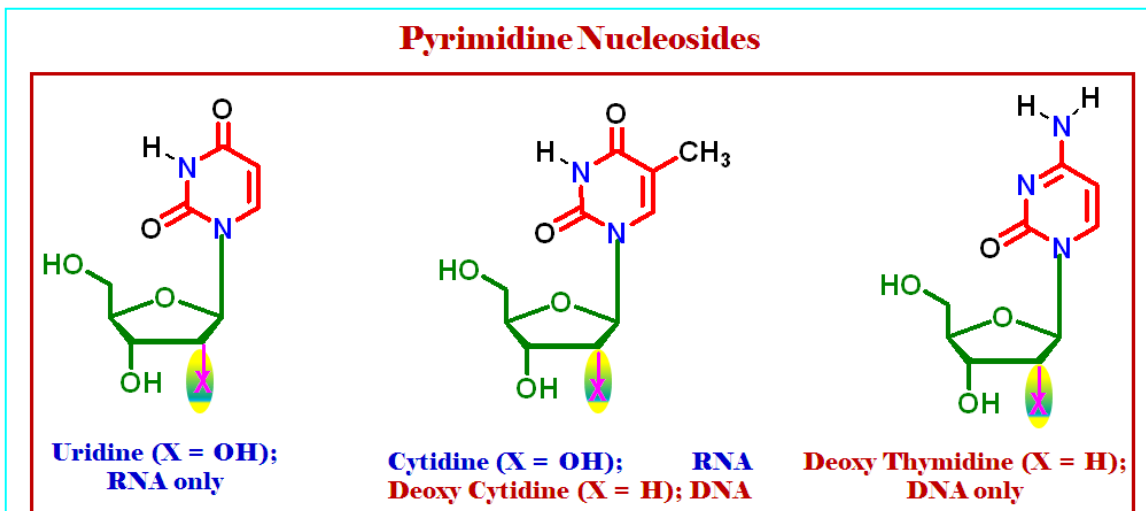
- Adenine and guanine are the principal purine bases of both DNA and RNA
- Caffeine (coffee) and theobromine (coffee and tea) are naturally occurring purines.



## 4.5. THE NUCLEOSIDES

- The classical structural definition is that a nucleoside is a pyrimidine or purine N-glycoside of D-ribofuranose or 2'-deoxy-D-ribofuranose
- Informal use has extended this definition to apply to purine or pyrimidine N-glycosides of almost any carbohydrate.
- The purine or pyrimidine part of a nucleoside is referred to as a *purine or pyrimidine base*.



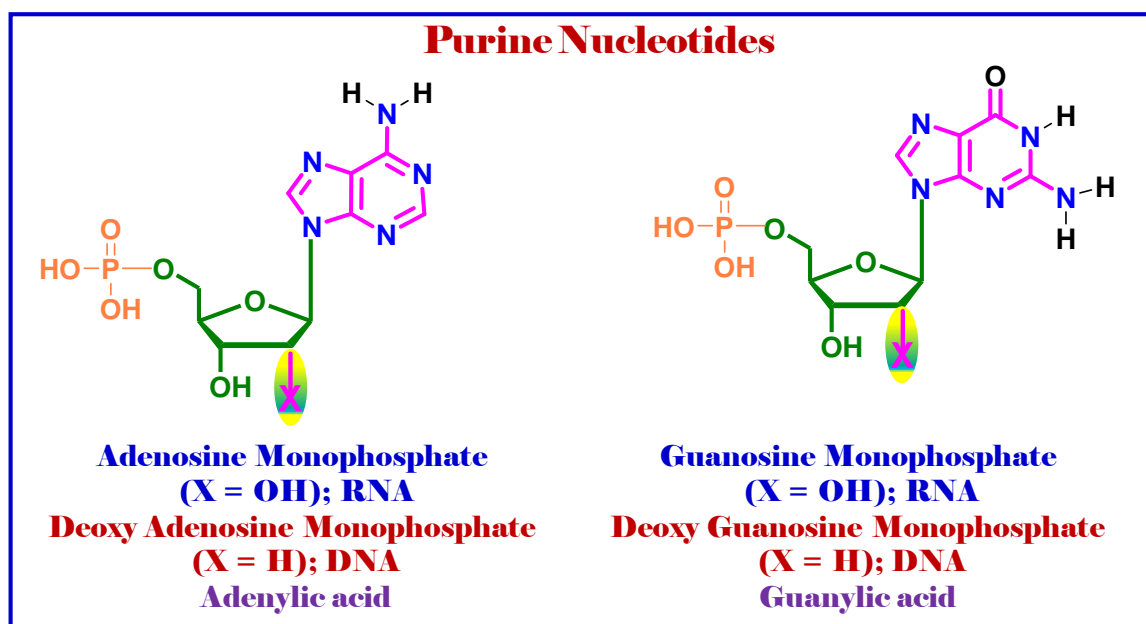
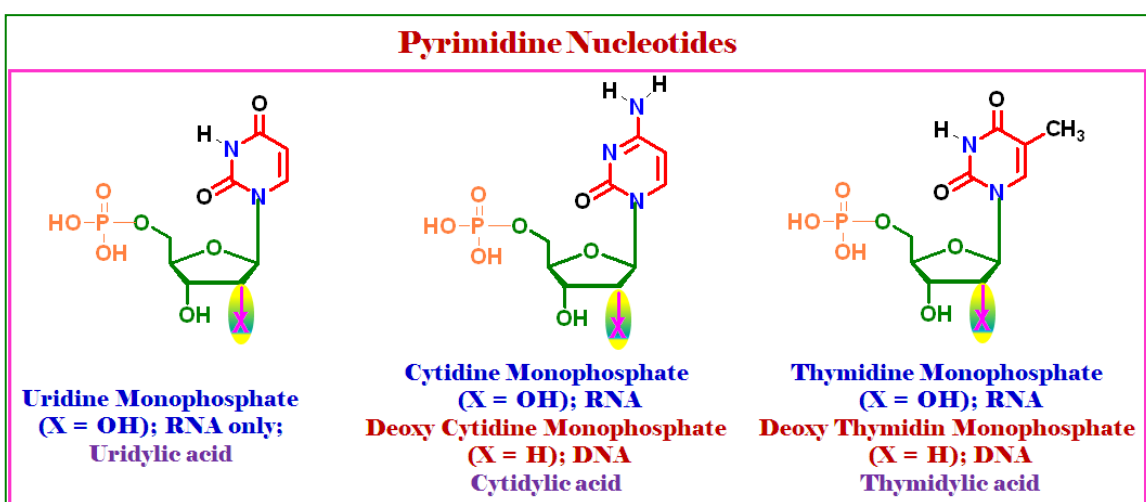


## 4.5. THE NUCLEOTIDES

- DNA and RNA are long polymers whose monomer units are called nucleotides
- A nucleotide consists of:
  1. Nitrogen containing heterocyclic base
    - Purine
    - Pyrimidine
  2. Five-carbon sugar ring
    - Ribose
    - Deoxyribose
  3. Phosphoryl group



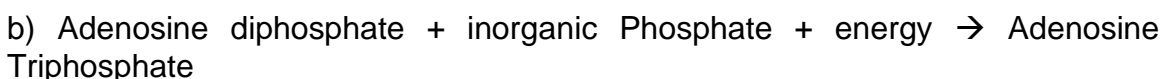
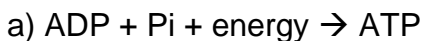
- Ring structures are found in both the base and the sugar
  - Base rings are numbered as usual
  - Sugar ring numbers are given the designation ' or prime
- Covalent bond between the sugar and the phosphoryl group is a phosphoester bond
- Bond between the base and the sugar is a β-N-glycosidic linkage joining the 1'-carbon of the sugar and a nitrogen atom of the base
- A nucleotide is the repeating unit of the DNA or RNA polymer
- The nitrogen base is attached β to
  - ribose (RNA)
  - deoxyribose (DNA)
- The sugar is phosphorylated at carbon 5'



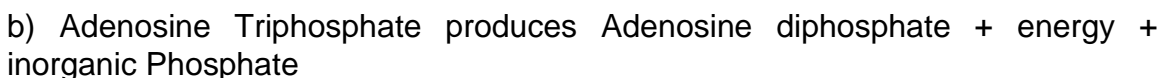
## 4.6. ATP, THE ENERGY SOURCE: ADENOSINE BASED NUCLEOTIDES

Adenosine triphosphate (ATP), the energy currency or coin of the cell pictured in Figures 1 and 2, transfers energy from chemical bonds to endergonic (energy absorbing) reactions within the cell. Structurally, ATP consists of the adenine nucleotide (ribose sugar, adenine base, and phosphate group,  $\text{PO}_4^{-2}$ ) plus two other phosphate groups. Energy is stored in the covalent bonds between phosphates, with the greatest amount of energy ( $\sim 7$  kcal/mole) in the bond between the second and third phosphate groups. This covalent bond is known as a pyrophosphate bond.

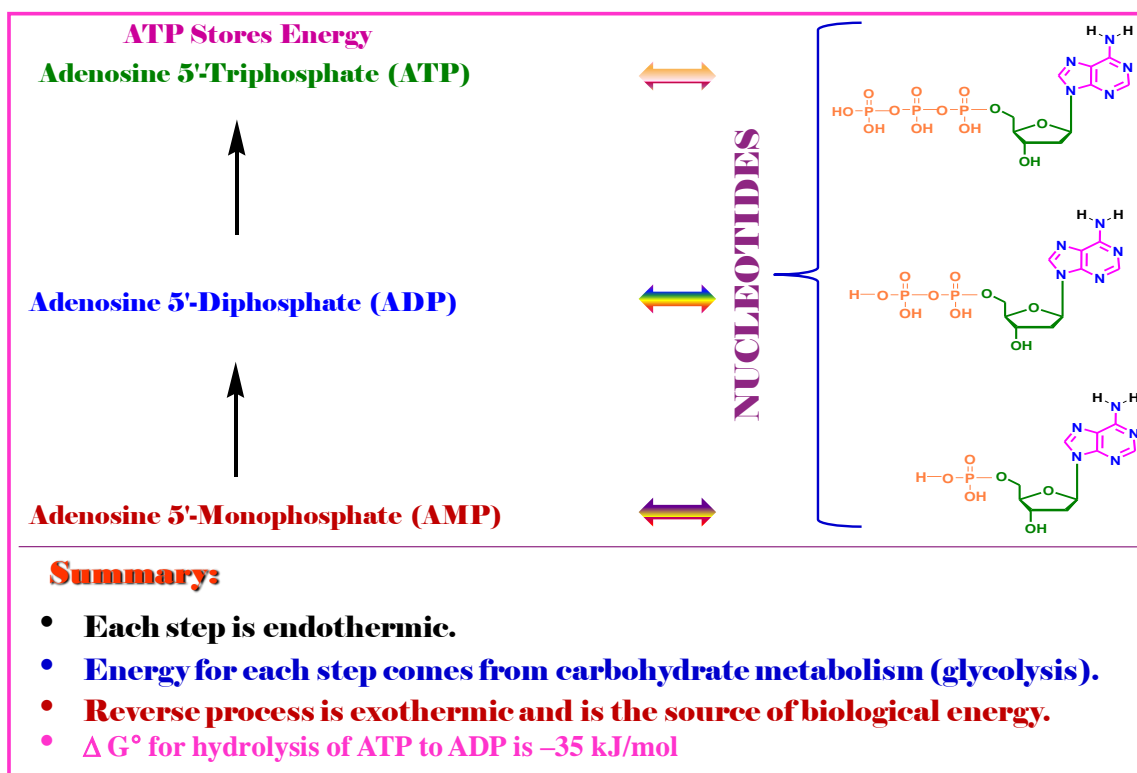
**We can write the chemical reaction for the formation of ATP as:**



**The chemical formula for the expenditure/release of ATP energy can be written as:**



An analogy between ATP and rechargeable batteries is appropriate. The batteries are used, giving up their potential energy until it has all been converted into kinetic energy and heat/unusable energy. Recharged batteries (into which energy has been put) can be used only after the input of additional energy. Thus, ATP is the higher energy form (the recharged battery) while ADP is the lower energy form (the used battery). When the terminal (third) phosphate is cut loose, ATP becomes ADP (Adenosine diphosphate; di = two), and the stored energy is released for some biological process to utilize. The input of additional energy (plus a phosphate group) "recharges" ADP into ATP (as in my analogy the spent batteries are recharged by the input of additional energy).

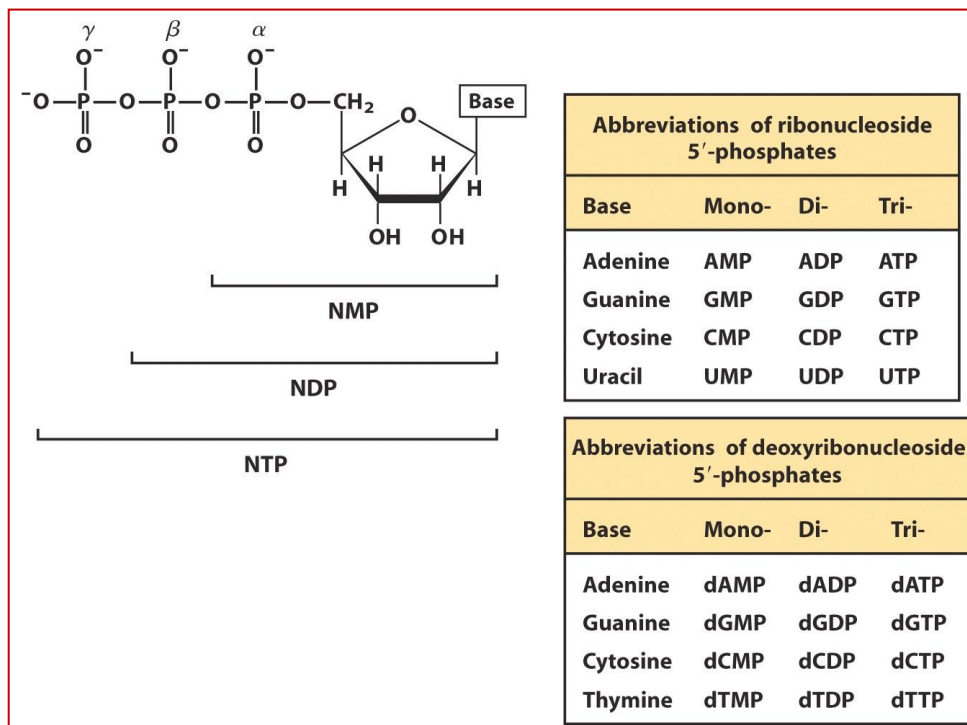


#### 4.7. SUMMARY TABLE OF NUCLEIC ACID UNITS

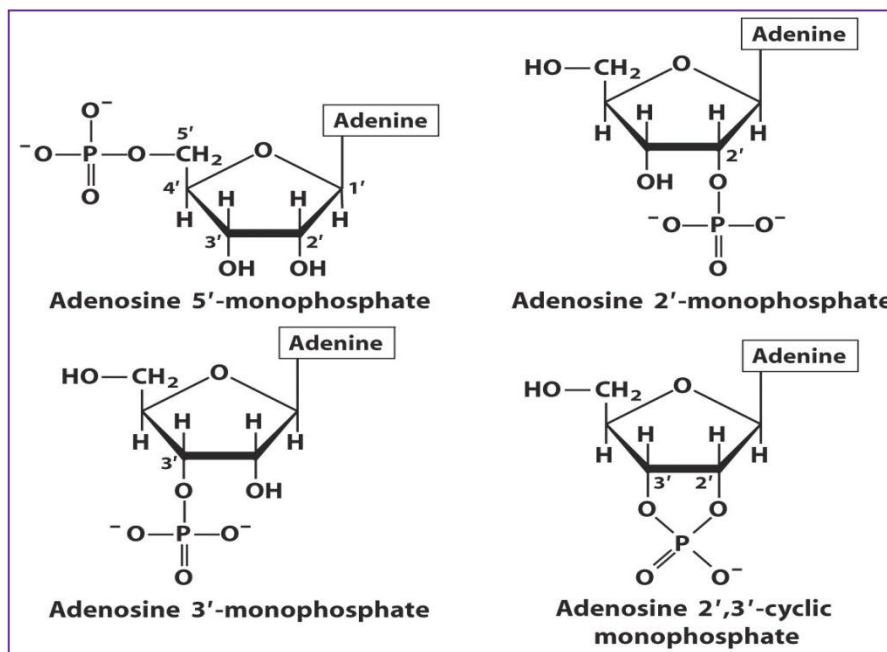
<i>Bases</i>	<i>Nucleosides</i>	<i>Deoxy Nucleosides</i>	<i>Nucleotides</i>	<i>Deoxy Nucleotides</i>
<b>Adenine A</b>	Adenosine <b>A</b>	Deoxyadenosine <b>dA</b>	Adenosine mono, di, triphosphate AMP, ADP, ATP	Deoxyadenosine mono, di, triphosphate dAMP, dADP, dATP
<b>Thymine T</b>	Thymidine <b>T</b>	Deoxythymidine <b>dT</b>	Thymidine mono, di, triphosphate TMP, TDP, TTP	Deoxythymidine mono, di, triphosphate dTMP, dTDP, dTTP
<b>Cytosine C</b>	Cytidine <b>C</b>	Deoxycytidine <b>dC</b>	Cytidine mono, di, triphosphate CMP, CDP,	Deoxycytidine mono, di, triphosphate dCMP, dCDP, dCTP

			CTP	
<b>Guanine</b> <b>G</b>	Guanosine <b>G</b>	Deoxyguanosine <b>dG</b>	Guanosine mono, di, triphosphate GMP, GDP, GTP	Deoxyguanosine mono, di, triphosphate dAMP, dADP, dATP
<b>Uracil</b> <b>U</b>	Uridine <b>U</b>	Deoxyuridine <b>dU</b>	Uridine mono, di, triphosphate UMP, UDP, UTP	Deoxyuridine mono, di, triphosphate dUMP, dUDP, dUTP

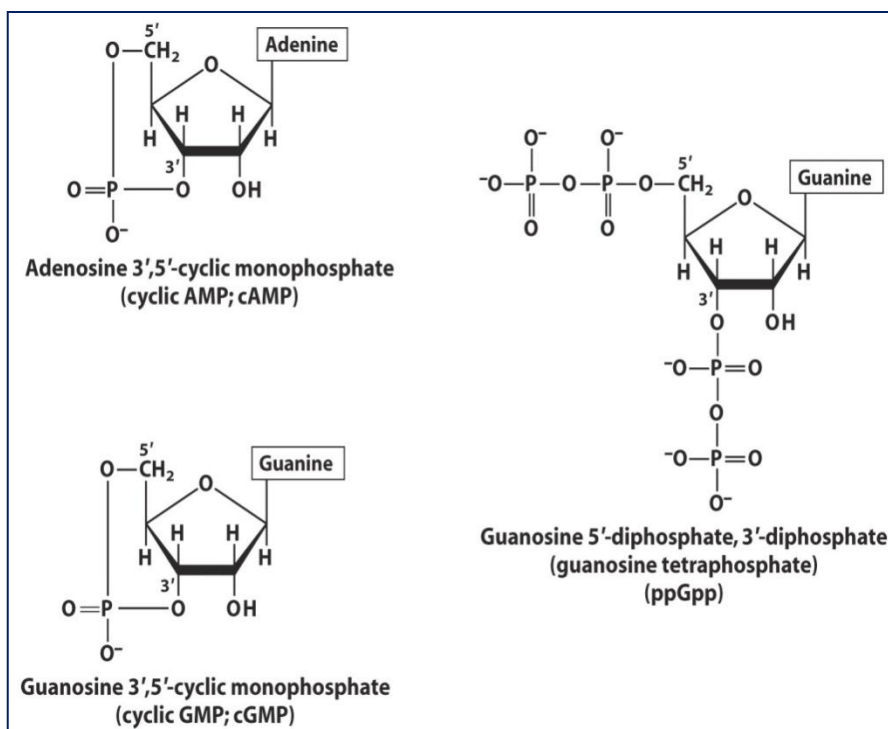
#### 4.8. ABBREVIATIONS OF RIBONUCLEOSIDE 5'-PHOSPHATES



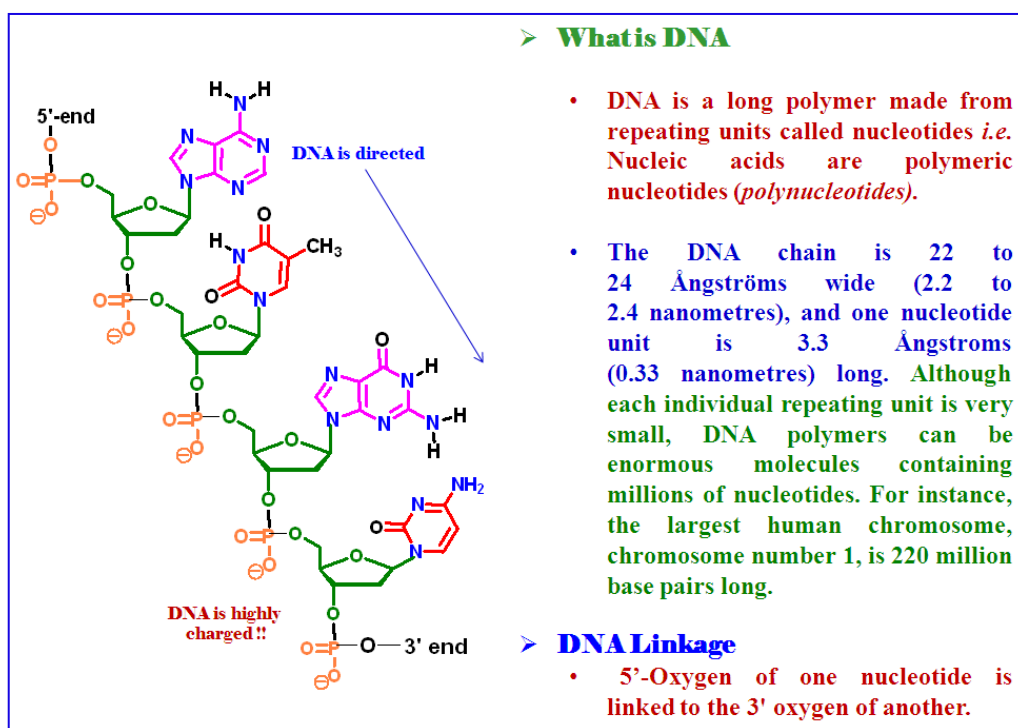
## 4.9. SOME EXAMPLES OF MONOPHOSPHATE NUCLEOTIDES



### The Signaling nucleotides



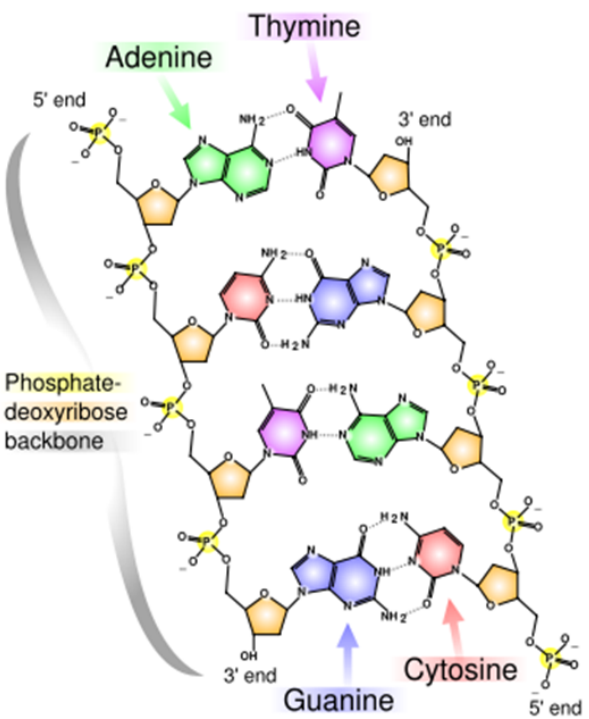
## 4.10. THE DEOXYRIBONUCLEIC ACID (DNA)



### Deoxyribonucleic acid (DNA) is the genetic material that

- Stores genetic information in the form of a code: a linear sequence of nucleotides.
- DNA is a double stranded biopolymer containing repeating units of nitrogen base, deoxyribose sugar, and phosphate.
- DNA can be arranged in 3 types of duplexes which contain major and minor grooves.
- DNA can adopt several topological forms.
- There are enzymes that will cut DNA, ligate DNA, and change the topology of DNA.
- Replicated by copying the strands using each as a template for the production of the complementary strand.
- Human genome contains about 3.2 billion base pairs. Inter-individual differences are observed at about 1 per 1,000 nucleotides.

## So Where DNA Is??...Everywhere In LIFE...In Cell...In Cytoplasm.



The diagram illustrates a segment of a DNA double helix. The backbone consists of alternating phosphate groups (yellow) and deoxyribose sugar molecules (orange). The nitrogenous bases are shown in the center, forming major and minor grooves. The bases are color-coded: Adenine (green), Thymine (purple), Guanine (blue), and Cytosine (red). Hydrogen bonds are shown between the bases: Adenine pairs with Thymine (two bonds), and Guanine pairs with Cytosine (three bonds). The 5' end is on the left and the 3' end is on the right. A label 'Phosphate-deoxyribose backbone' points to the outer structure.

- **What Is DNA (Biology):**  
**Deoxyribonucleic acid, or DNA, is a nucleic acid molecule that contains the genetic instructions used in the development and functioning of all known living organisms.**
- **Where is DNA:**
  - Within cells, DNA is organized into structures called chromosomes and the set of chromosomes within a cell make up a genome.**
  - These chromosomes are duplicated before cells divide, in a process called DNA replication.**
  - Eukaryotic organisms such as animals, plants, and fungi store their DNA inside the cell nucleus, while in prokaryotes such as bacteria it is found in the cell's cytoplasm.**
  - Within the chromosomes, chromatin proteins such as histones compact and organize DNA, which helps control its interactions with other proteins and thereby control which genes are transcribed.**

### 4.10.1. ALL LIVING THINGS CONTAIN DNA

The stringy stuff in the test tube is DNA that has been extracted from an organism. But you can't tell which one of these organisms it came from just by looking at it. That's because DNA looks exactly the same in every organism on Earth.

## 4.10.2. THE CHEMICAL COMPOSITION OF DNA

● **Erwin Chargaff (Columbia Univ.) studied DNAs from various sources and analyzed the distribution of purines and pyrimidines in them.**

● **The distribution of the bases adenine (A), guanine (G), thymine (T), and cytosine (C) varied among species.**

● **But the total purines (A and G) and the total pyrimidines (T and C) were always equal.**

● **Moreover: %A = %T, and %G = %C**

### Composition of Human DNA

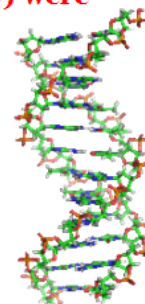
For example:

#### Purine

• **Adenine (A) 30.3%**  
 • **Guanine (G) 19.5%**  
 • **Total purines: 49.8%**

#### Pyrimidine

• **Thymine (T) 30.3%**  
 • **Cytosine (C) 19.9%**  
 • **Total pyrimidines: 50.2%**



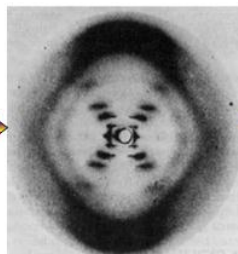
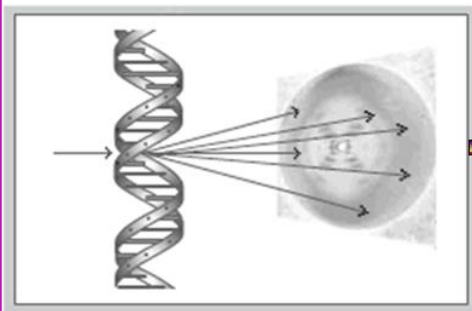


### 4.10.3. DNA STRUCTURE: HISTORY OF INVENTION

#### DNA DOUBLE STRAND STRUCTURE: NOBEL PRIZE in 1962

- **In 1869: DNA Isolation:** [Swiss physician Friedrich Miescher](#): DNA was first isolated from [pus](#) of discarded surgical bandages. As it resided in the nuclei of cells, he called it "nuclein".
- **In 1919: Component Identification:** [Phoebus Levene](#): Identification of the base, sugar and phosphate nucleotide unit. Levene suggested that DNA consisted of a string of nucleotide units linked together through the phosphate groups. bases repeated in a fixed order
- **In 1937: X-ray diffraction:** [William Astbury](#): produced the first X-ray diffraction patterns that showed that DNA had a regular structure.
- **In 1953: DNA's role in heredity:** [Alfred Hershey and Martha Chase](#): Hershey-Chase experiment ----DNA is the genetic material of the T2 phage.
- **In 1953: Rosalind Franklin:** X-ray diffraction images---and DNA bases pair. Finally, based on X-ray images:-----
- **In 1953: DNA Structure:** [James D. Watson and Francis Crick](#): suggested the the DNA structure which is now accepted as the first accurate model of DNA structure.
- Other contribution: [Franklin and Raymond Gosling's ; Maurice Wilkins](#)
- Franklin and Gosling's subsequent paper identified the distinctions between the A and B structures of the double helix in DNA.
- **In 1962: Watson, Crick, and Maurice Wilkins:** Jointly received the Nobel Prize in Physiology or Medicine ([Franklin didn't share the prize with them since she had died earlier](#)).
- In an influential presentation in 1957, Crick laid out the "Central Dogma" of molecular biology-Protein synthesis.
- Later by , [Har Gobind Khorana](#), [Robert W. Holley](#) and [Marshall Warren Nirenberg](#) to decipher the genetic code. These findings represent the birth of molecular biology.

DNA was first isolated by the Swiss physician Friedrich Miescher who, in 1869, discovered a microscopic substance in the pus of discarded surgical bandages. As it resided in the nuclei of cells, he called it "nuclein".



Franklin & Wilkins:  
X-ray diffraction  
on DNA fibers (1950s)



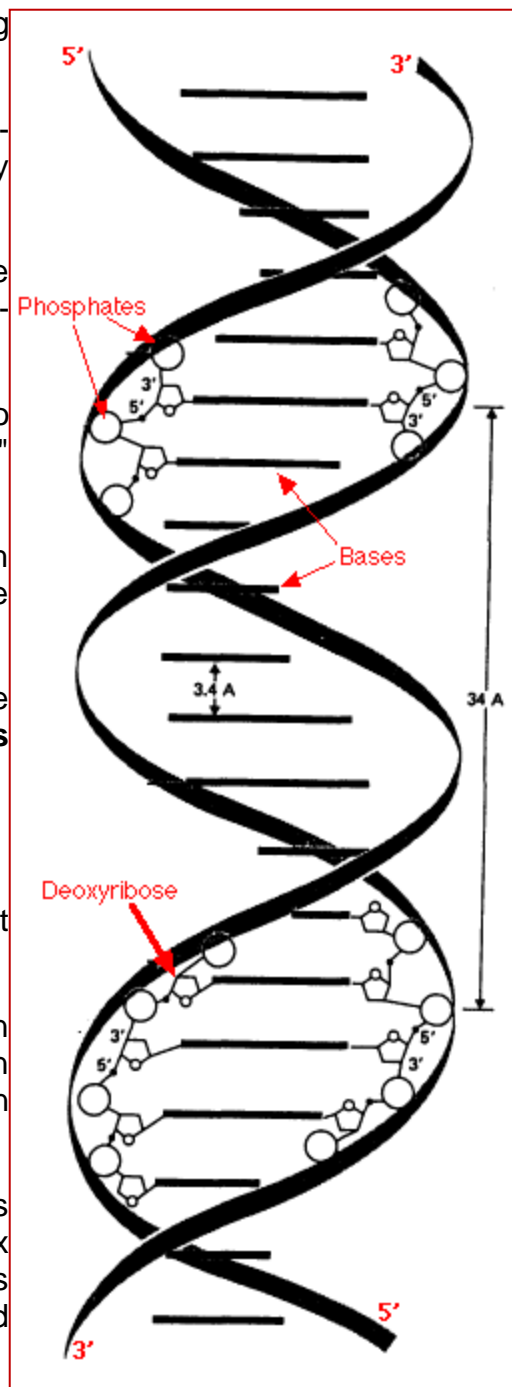
Watson & Crick:  
Interpretation of  
X-ray data (1953)



#### 4.10.4. DNA BASE PAIRES AND DUPLEX STRUCTURE

The double helix of DNA has the following features:

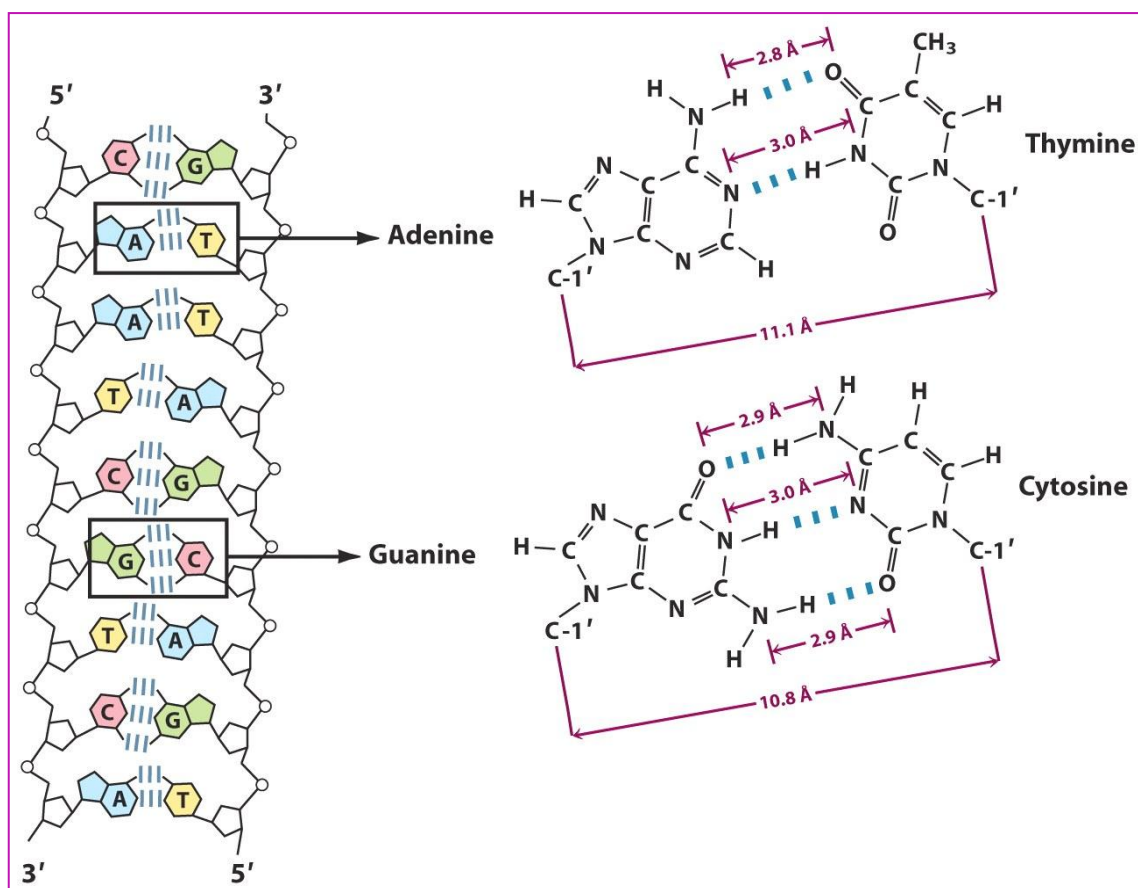
- It contains two polynucleotide strands wound around each other.
- The backbone of each consists of alternating deoxyribose and phosphate groups.
- The phosphate group bonded to the 5'-carbon atom of one deoxyribose is covalently bonded to the 3'-carbon of the next.
- The two strands are "antiparallel"; that is, one strand runs 5'- to 3'- while the other runs 3'- to 5'-.
- The DNA strands are assembled in the 5'- to 3'- direction and, by convention, we "read" them the same way.
- The purine or pyrimidine attached to each deoxyribose projects in toward the axis of the helix.
- Each base forms hydrogen bonds with the one directly opposite it, forming **base pairs** (also called nucleotide pairs).
- 3.4 Å separates the planes in which adjacent base pairs are located.
- 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 shown in the diagram.
- 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.
- The diameter of the helix is 20 Å.



- The helix can be virtually any length; when fully stretched, some DNA molecules are as much as 5 cm (2 inches!) long.
- The path taken by the two backbones forms a major (wider) groove (from "34 Å" to the top of the arrow) and a minor (narrower) groove (the one below).

#### 4.10.4.1. DISCUSSION OF BASE PAIRING in DNA

This structure of DNA was worked out by **Francis Crick** and **James D. Watson** in 1953. It revealed how DNA — the molecule that Avery had shown was the physical substance of the genes — could be replicated and so passed on from generation to generation. For this epochal work, they shared a Nobel Prize in 1962.



The rules of base pairing (or nucleotide pairing) are:

- A with T: the purine adenine (A) always pairs with the pyrimidine thymine (T)
- C with G: the pyrimidine cytosine (C) always pairs with the purine guanine (G)

- This is consistent with there not being 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 why not A with C and G with T?

- The answer: only with A & T and with C & G are there opportunities to establish hydrogen bonds between them (two between A & T; three between C & G). These relationships are often called the rules of Watson-Crick base pairing, named after the two scientists who discovered their structural basis.

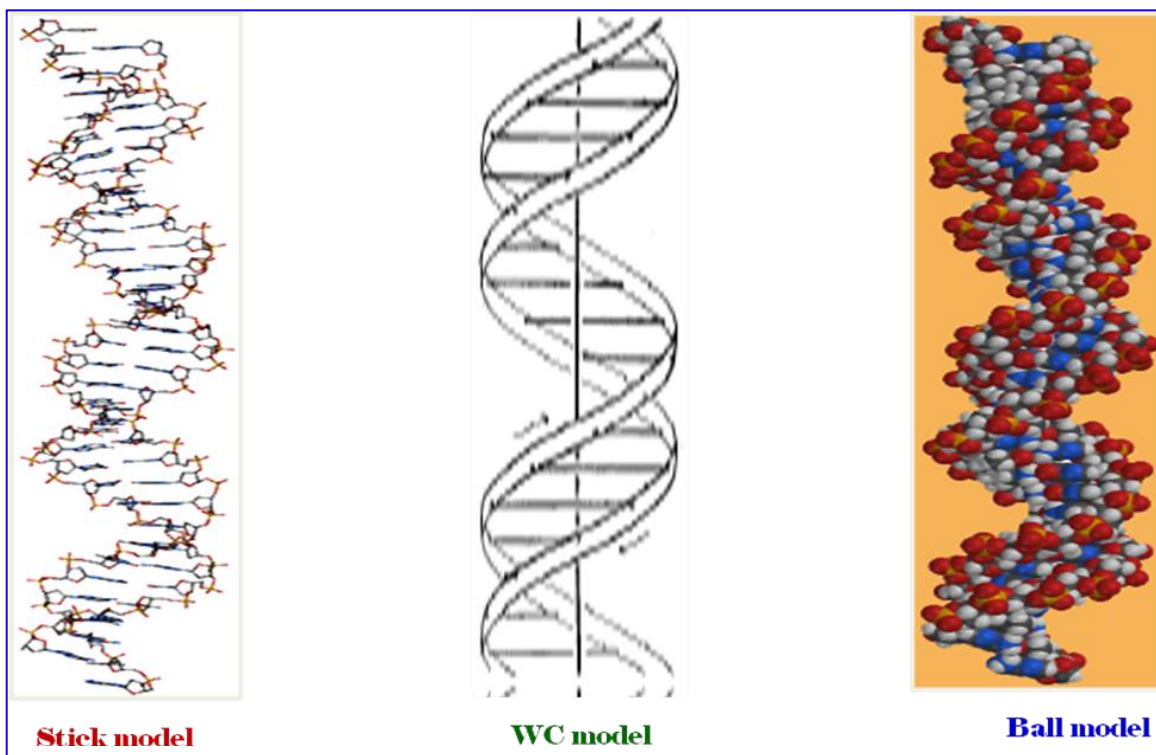
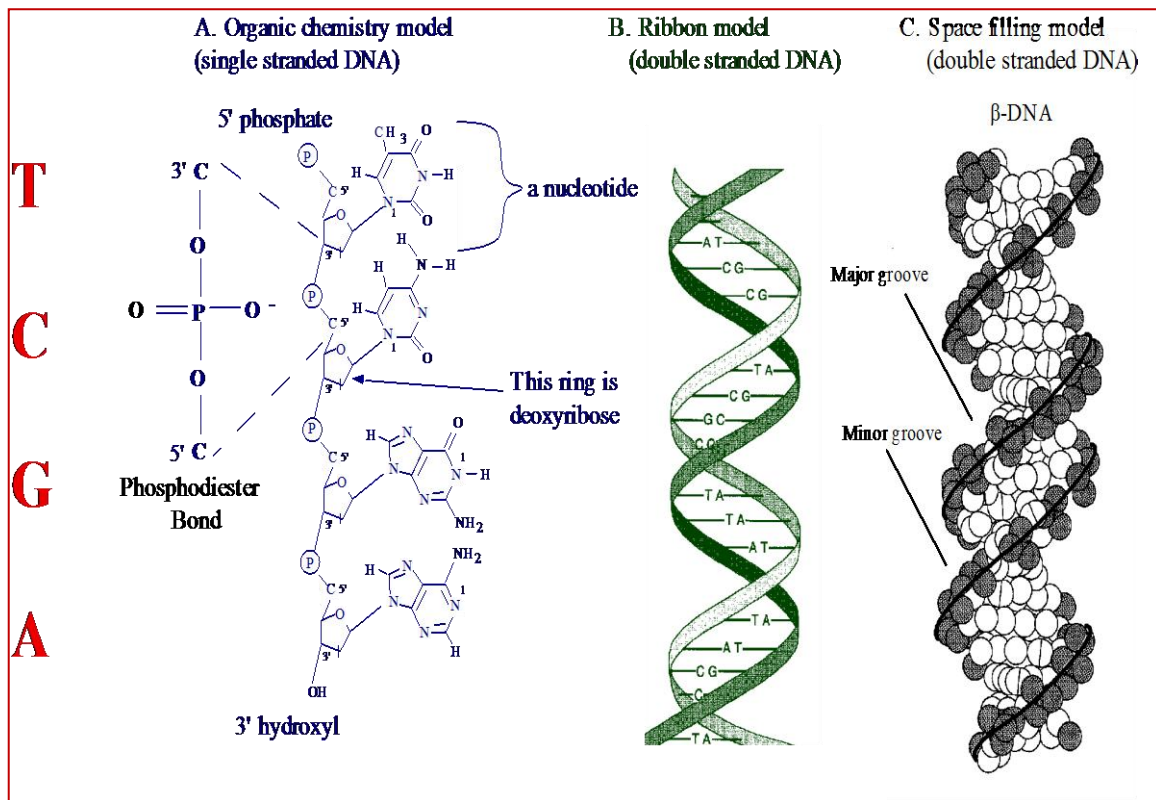
The rules of base pairing tell us that if we can "read" the sequence of nucleotides on one strand of DNA, we can immediately deduce the complementary sequence on the other strand.

The rules of base pairing explain the phenomenon that whatever the amount of adenine (A) in the DNA of an organism, the amount of thymine (T) is the same (called Chargaff's rule). Similarly, whatever the amount of guanine (G), the amount of cytosine (C) is the same.

<b><i>The C+G:A+T ratio varies from organism to organism (particularly among the bacteria), but within the limits of experimental error, A = T and C = G</i></b>				
Relative Proportions (%) of Bases in DNA				
<b>Organism</b>	<b>A</b>	<b>T</b>	<b>G</b>	<b>C</b>
<b>Human</b>	30.9	29.4	19.9	19.8
<b>Chicken</b>	28.8	29.2	20.5	21.5
<b>Grasshopper</b>	29.3	29.3	20.5	20.7
<b>Sea Urchin</b>	32.8	32.1	17.7	17.3
<b>Wheat</b>	27.3	27.1	22.7	22.8
<b>Yeast</b>	31.3	32.9	18.7	17.1
<b>E. coli</b>	24.7	23.6	26.0	25.7

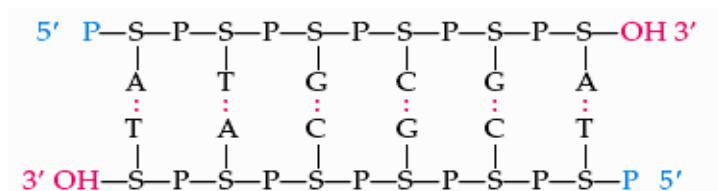


### 4.10.4.2. WAYS of DEPICTING DNA STRUCTURE



### 4.10.4.3. Complementary DNA Strands

- The two DNA strands are complementary strands
  - The sequence of bases on one automatically determines the sequence of bases on the other strand.
- The chains run antiparallel
  - Only when the 2 strands are antiparallel can the base pairs form the H bonds that hold them together.



### 4.10.4.4. DNA MAJOR GROOVES AND MINOR GROOVES

More recent studies used X-ray crystallography and NMR to confirm the original structure by Watson and Crick. Right handed helix- winds about the same direction in which fingers of a right hand curl when the thumb is pointing upward.

Helical axis passes through base pairs and

- Contains a minor and a major groove that wind about the outside of the helix
- Helical structure repeats every 34 Å
- Major and minor groove are the two surfaces that wind about the outside of the helix. They are formed by the edges of the stacked bases. These grooves are distinct because they have different H-bonding patterns and different size.
- Major and minor groove of DNA contain sequence-dependent patterns of H-bond donors and acceptors.
- Sequence-dependent duplex structure (A, B, Z, bent DNA).

	Width	Depth
Major Groove	11.6Å	8.5Å
Minor Groove	6.0Å	8.2Å

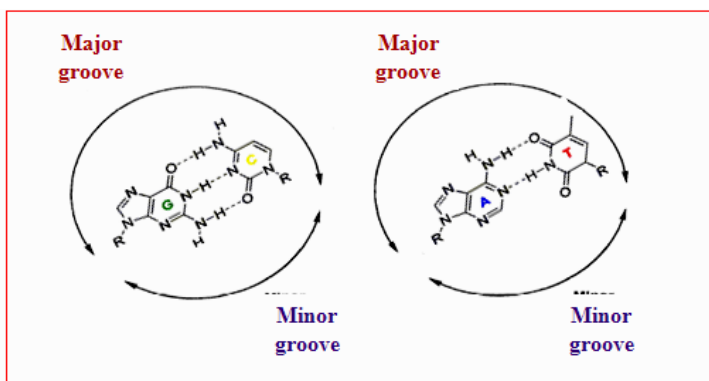
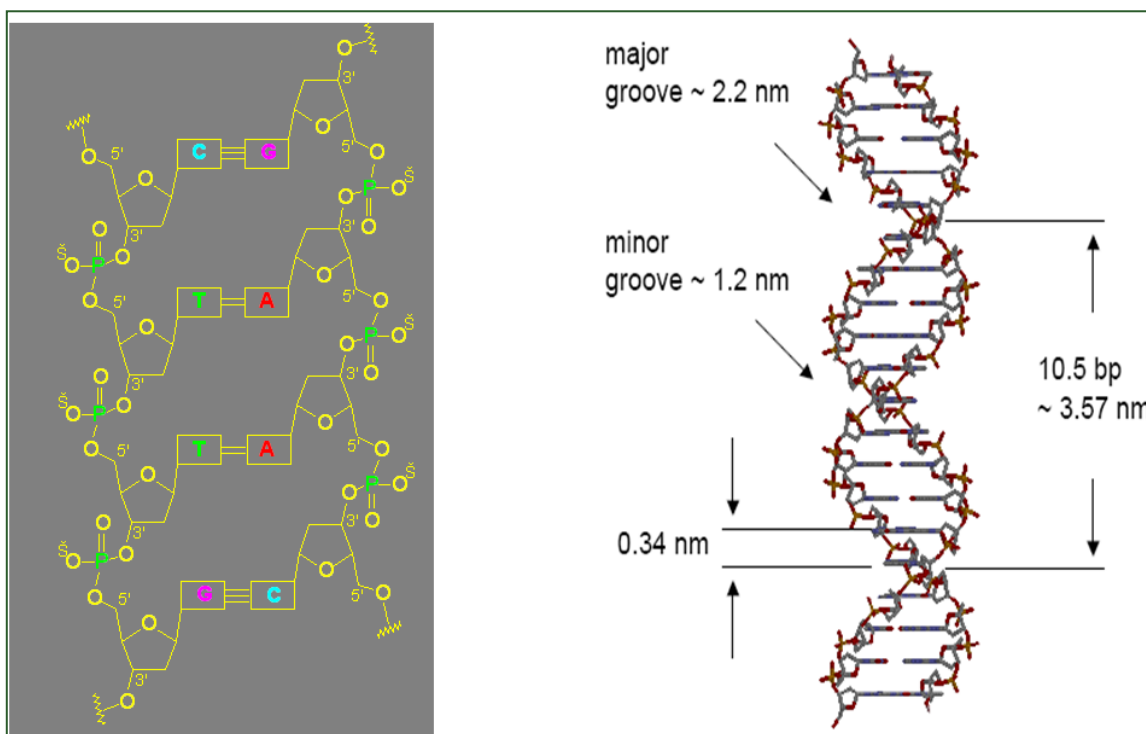


Figure 4.1: Groove dimensions and the WC base pairs.

- The grooves are formed by the edges of stacked bases and have different sizes because deoxyribose residues are attached asymmetrically (not 180 degrees). The minor groove is the one in which C-1' –helix axis'c1' angle is less than 180 degrees (recall that the helix axis passes through the middle of each base pair in B-DNA) and the major groove is on the opposite side of each base pair.
- A/T sequences have a higher negative potential while G/C sequences have higher positive potential
- H-bonding to base edges in the minor groove can be used to distinguish between A:T and G:C base pairs
- A:T is distinguished from a T:A base pair through indirect read out
- DNA binding drugs are generally flat and small, which makes them fit well into the minor groove.
- Groove width varies with sequence; A/T rich tracks tend to make the groove width narrower.
- The convex shape of the minor groove floor complements the typical shapes of minor groove binding drugs
- A/T sequences result in a smooth convex curve whereas G/C sequences have “little” bumps due to the 2-amino groups of guanine.





#### 4.10.4. 5. Sugar-Phosphate Chain Conformations

- Double-stranded DNA has limited structural complexity compared to proteins (only 4 nucleotides vs. 20 amino acids)
- Limited secondary structures, no tertiary or quaternary structures.
- RNA has some well-defined tertiary structure.
- Conformation of a nucleotide is specified by 6 torsion angles of the sugar phosphate backbone and the torsion angle that describes the orientation of the base about the glycosidic bond (7 totals).
- Despite 7 degrees of freedom per nucleotide, they have restricted conformational freedom.

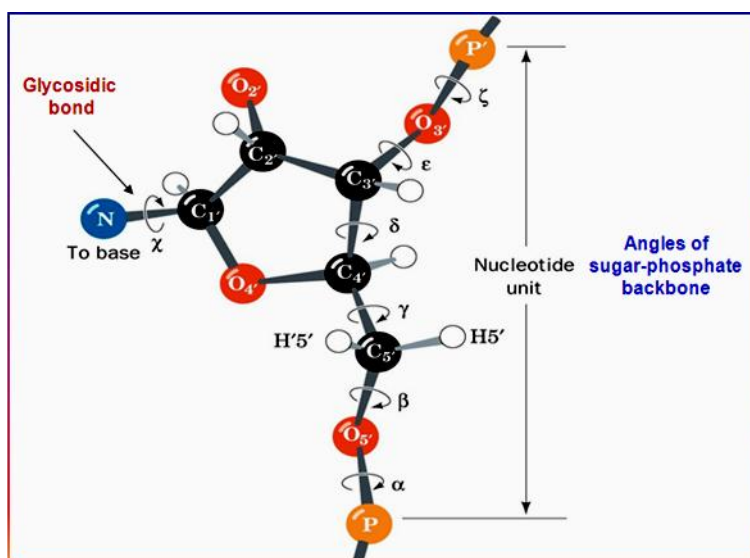


Figure 4.2: The conformation of a nucleotide unit defined by the seven torsion angles.

#### Torsion Angles about Glycosidic Bonds Have only 1 or 2 Stable Positions

- Purine residues have 2 sterically allowed orientations relative to the ribose group, *syn*- and *anti*-
- For pyrimidines only the *anti* conformation is allowed due to steric hindrance between the sugar and the C2 of the pyrimidine.
- Most double helical nucleic acids are in the *anti*- conformation
- Exception is Z-DNA which has alternating *anti* and *syn*- pyrimidine and purine residues.

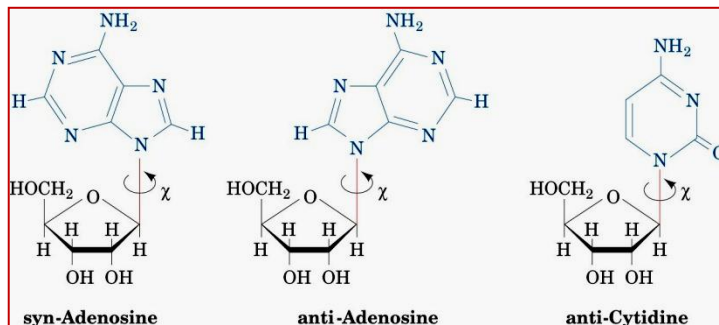


Figure 4.4: The sterically allowed orientations (*syn*- vs. *anti*-)of purine and pyrimidine bases with respect to their attached ribose units.

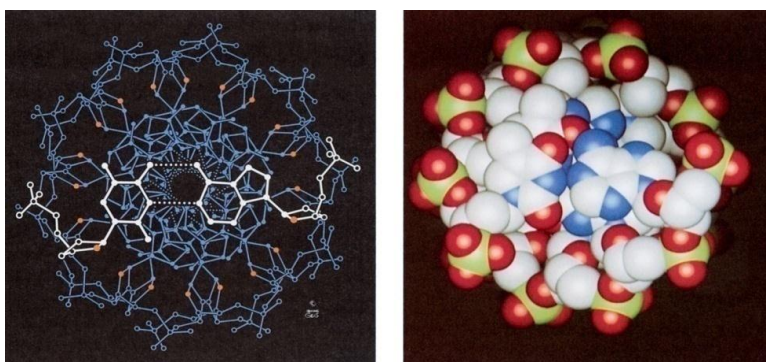


Figure 4.5: The Structure of B-DN (Ball and stick) and corresponding space-filling model viewed down the helix axis.

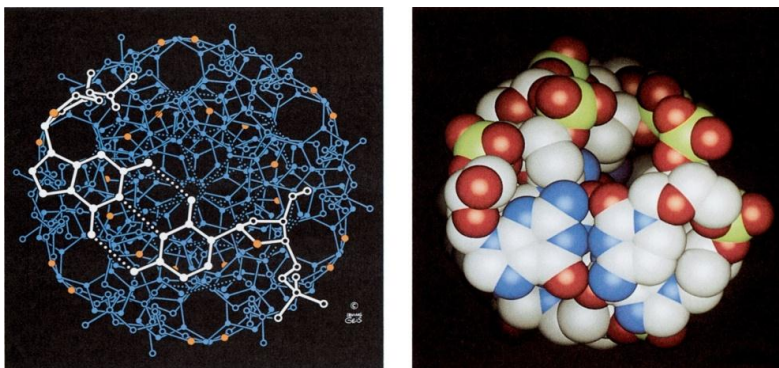


Figure 4.6: The Structure of Z-DNA (Ball and stick) and corresponding space-filling model viewed down the helix axis.

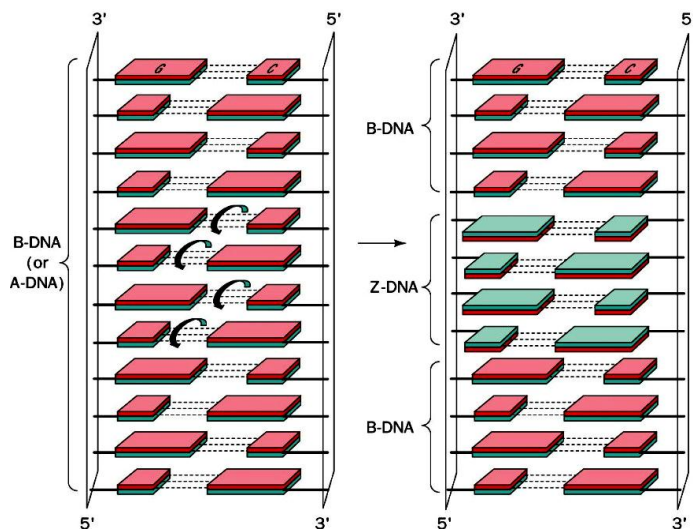


Figure 4.7: The Conversion of B-DNA to Z-DNA.

- Sugar Pucker
  - Ribose rings undergoes sugar pucker and become slightly nonplanar.
  - Most structures show that 4 of the five ring atoms are coplanar, 5th atom is out of the plane in a half-chair conformation.
  - **Endo-** conformation → if out of plane atom is the same side of the ring as the C5'
  - **Exo-** conformation → if out of plane atom is on the opposite side of the ring as the C5'
  - Most nucleoside and nucleotide structures (not in double helices) out of plane atom is either C2' or C3'.
  - C2'-endo most common; C3'-endo and C3'-exo are also common.

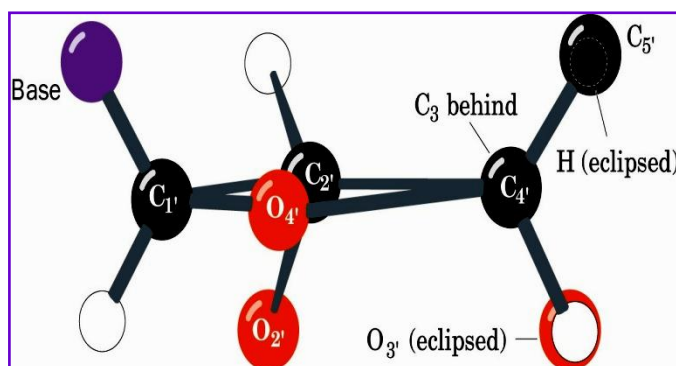


Figure 4.8a: The Sugar ring pucker-A planar ribose ring, viewed down the C3'—C4' bond, are all eclipsed.

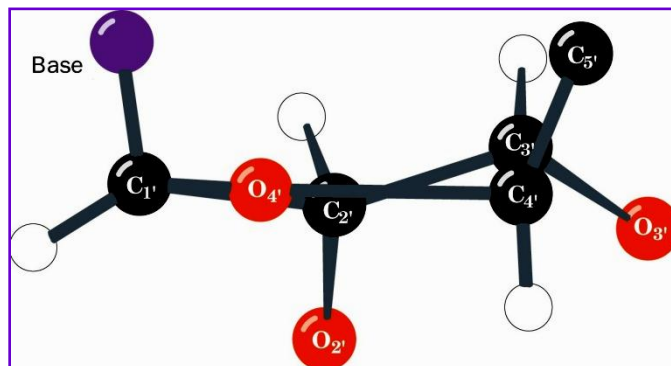


Figure 4.8b: The Sugar ring pucker—The steric strain resulting in planar sugar ring (in Figure 4.8a) is partially relieved by ring puckering in a half-chair conformation in which C3' is the out-of-plane atom (same as C5' so C3'-endo).

- Endo conformation → if out of plane atom is the same side of the ring as the C5'
- Exo- conformation → if out of plane atom is on the opposite side of the ring as the C5'
- Most nucleoside and nucleotide structures (not in double helices) out of plane atom is either C2' or C3'.
- C2'-endo most common; C3'-endo and C3'-exo are also common.
- This ribose pucker determines the relative orientation of the phosphates to the sugar.
- To have a regularly repeating model—need C2'-endo (B-DNA) or C3'-endo (A-DNA, RNA-11)
- Z-DNA purines are all C3'-endo, pyrimidines C2'-endo (dinucleotide repeating unit).
- B-DNA has some flexibility (can be observed as C4'-exo, O4'-endo, C1'-exo and C3'-exo).

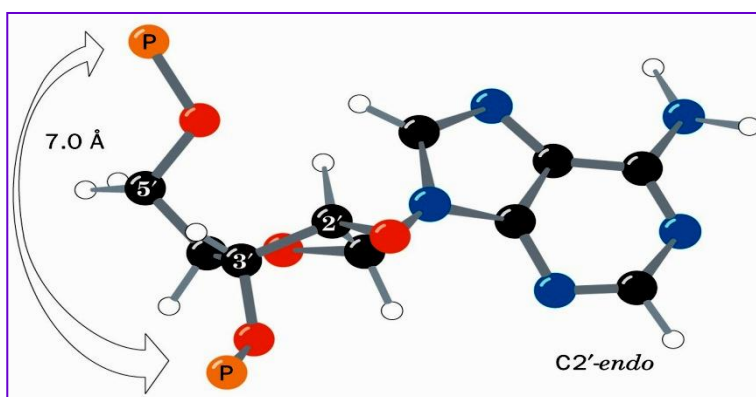


Figure 4.9a: Nucleotide sugar conformations—The C3'-endo conformation (on the same side of the sugar ring as C5'), occurs in A-RNA and RNA-11.

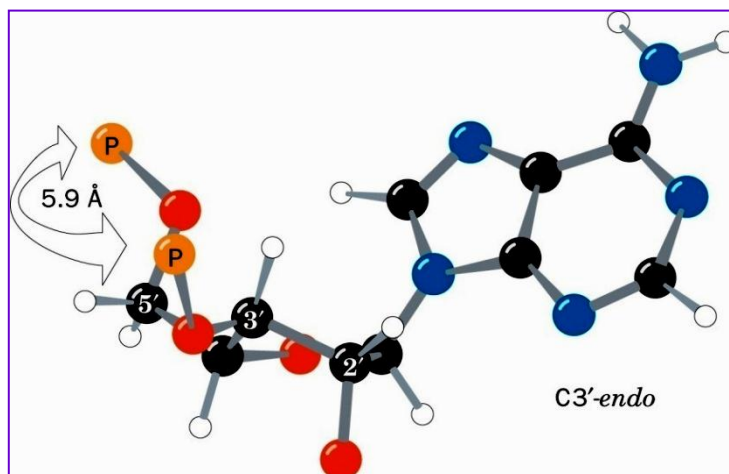


Figure 4.9b: The Nucleotide sugar conformations-The *C2'-endo* conformation, occurs in B-DNA.

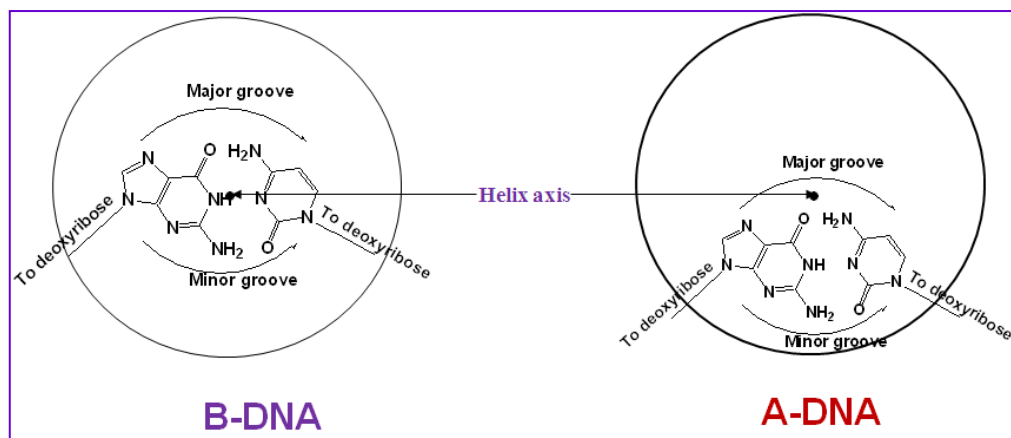
#### 4.10.4.6. DIFFERENT CONFORMATIONS OF DNA

	A form	B form	Z form
Helical Sense	right handed	right handed	left handed
Diameter	2.6 nm	2.0 nm	1.8 nm
bp/turn	11	10.5	12
base tilt	20°	6°	7°

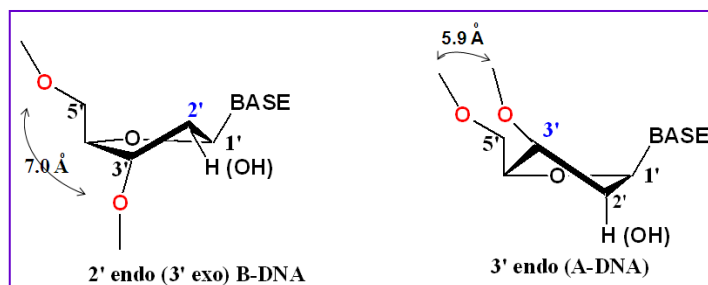
**A-DNA**      **B-DNA**      **Z-DNA**

- The A form is a wider right-handed spiral, with a shallow and wide minor groove and a narrower and deeper major groove.
- In more common B form the strands turn about the helical axis in a right-handed spiral.
- Segments of DNA where the bases have been chemically-modified may undergo a larger change in conformation and adopt the Z form from the natural B-form.
- Here, the strands turn about the helical axis in a left-handed spiral, the opposite of the more common B form.

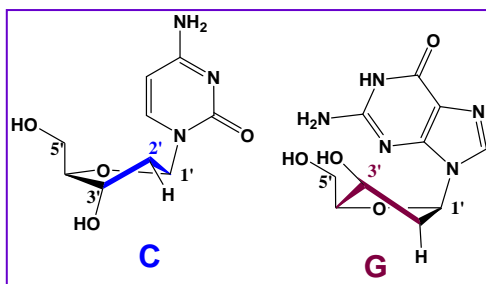
- **A-DNA has a shallow minor groove and a deep major groove:**



- **The sugar pucker in A-DNA is 3'-endo.**



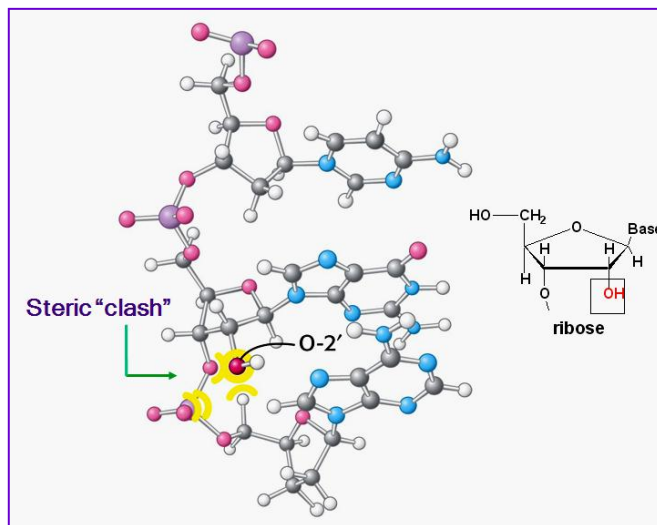
- **The Sugar and base conformations in Z-DNA alternate:**  
In the Z-DNA duplex, d(5'-GCGCGCGCGCGC-3')/d(3'-CGCGCGCGCGCGC-5'), C: sugar is 2'-endo, base is *anti* and G: sugar is 3'-endo, base is *syn*.



#### 4.10.4.7. B-type Duplex is not Possible for RNA

Although B-DNA solved by Watson and Crick is the major form of DNA in chromosomes, other types of secondary structure are possible. For example, duplexes containing RNA cannot form B type duplex because of the steric clash between its 2'-OH and the phosphodiester. Instead, it forms an A-type duplex.

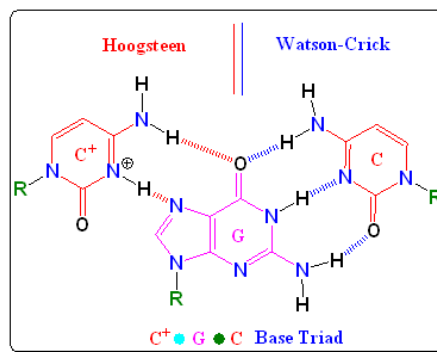
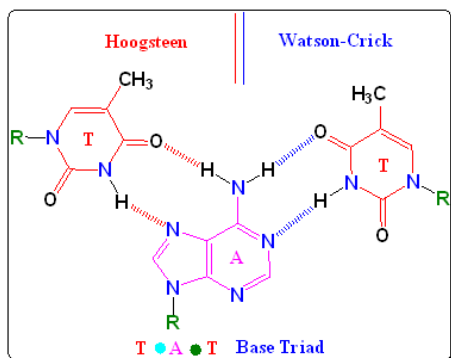




#### 4.10.4.8. SOME OTHER TYPE OF BASE PAIRING AND DNA SECONDARY STRUCTURE

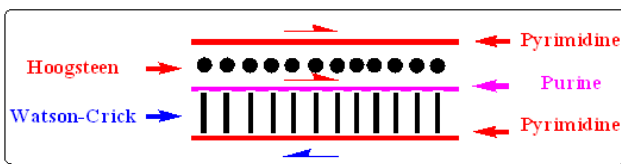
##### HOOGSTEEN BASE PAIR AND DNA TRIPLEX:

In Hoogsteen base pair, two nucleobases on each strand can be held together by hydrogen bonds in the major groove.



Hoogsteen pairs have quite different properties from Watson-Crick base pairs. Hoogsteen base pairs, allows a third strands to wind around the duplexes, which are assembled in the Watson-Crick pattern, and form triple-stranded helices .

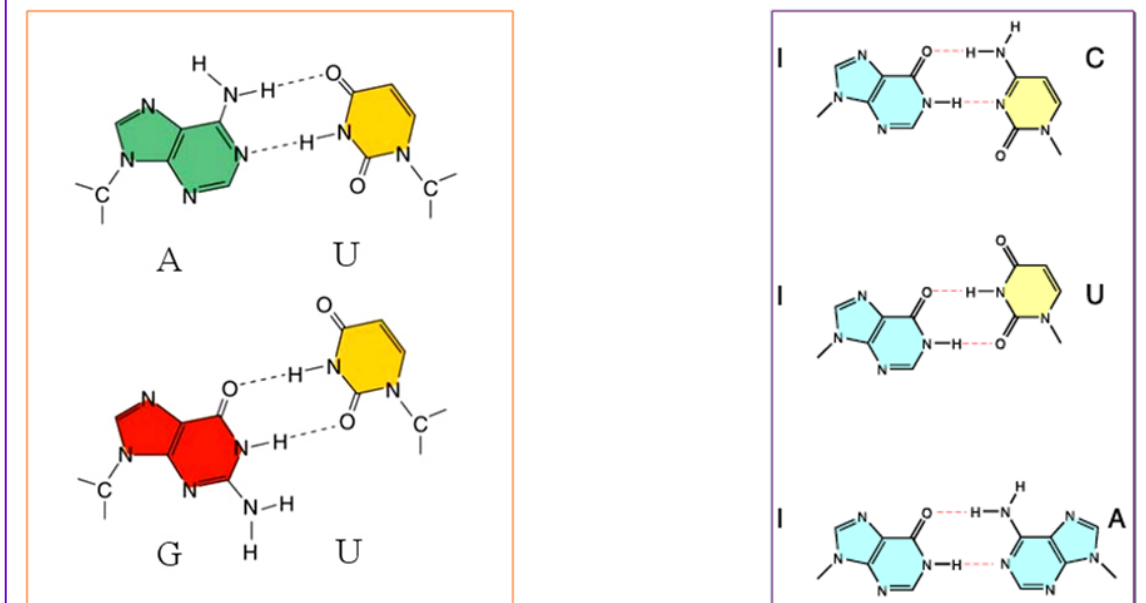
##### A DNA Triple Helix



### 4.10.4.9. WOBBLE BASE PAIRS AND RNA SECONDARY STRUCTURE

A wobble base pair is a G-U and I-U / I-A / I-C pair fundamental in RNA secondary structure. Its thermodynamic stability is comparable to that of the Watson-Crick base pair. Wobble base pairs are critical for the proper translation of the genetic code. The genetic code makes up for disparities in the number of amino acids (20) for codons (64), by using modified base pairs in the first base of the anti-codon. One important modified base is **inosine** which can pair with three bases: **uracil**, **adenine**, and **cytosine**.

Another critical base pair is the **G-U base pair**, which allows **uracil** to pair with two bases: **guanine** and **adenine**.



### 4.10.4.10. G-QADRUPLEX

- Structure of a DNA quadruplex formed by telomere repeats. The conformation of the DNA backbone diverges significantly from the typical helical structure.
- These guanine-rich sequences may stabilize chromosome ends by forming very unusual structures of stacked sets of four-base units, rather than the usual base pairs found in other DNA molecules. Here, four guanine bases form a flat plate and these flat four-base units then stack on top of each other, to form a stable G-quadruplex structure.
- Telomeres and telomerase have recently received great attention because of their potential links to cancer, HIV and other diseases. A unique G-rich DNA sequence in the telomeres was found to protect the chromosomes from recombination, end to end fusion, and degradation through forming G-quadruplexes with highly polymorphic structures in the presence of alkali metal cations. This unusual structure and extensive cellular



functions make G-quadruplex a very attractive target for drug design, which made it important for determination of G-quadruplex.

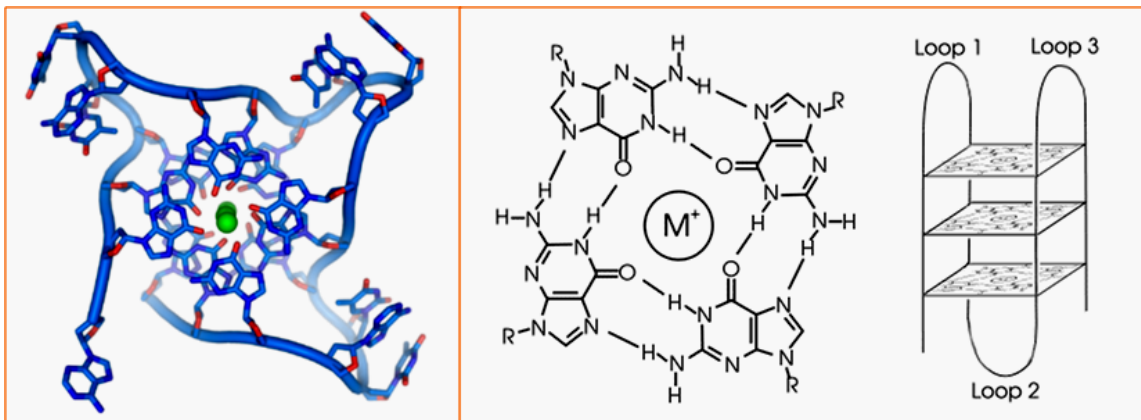
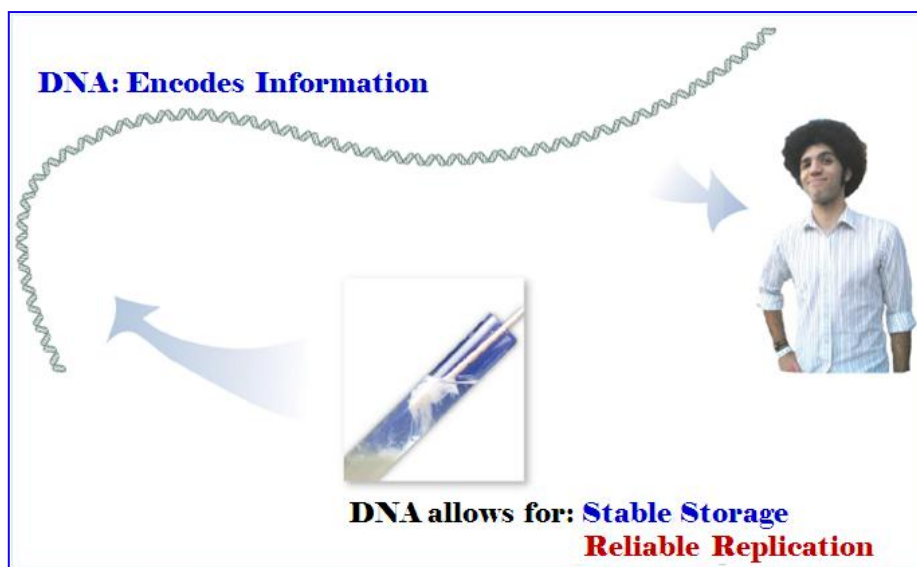


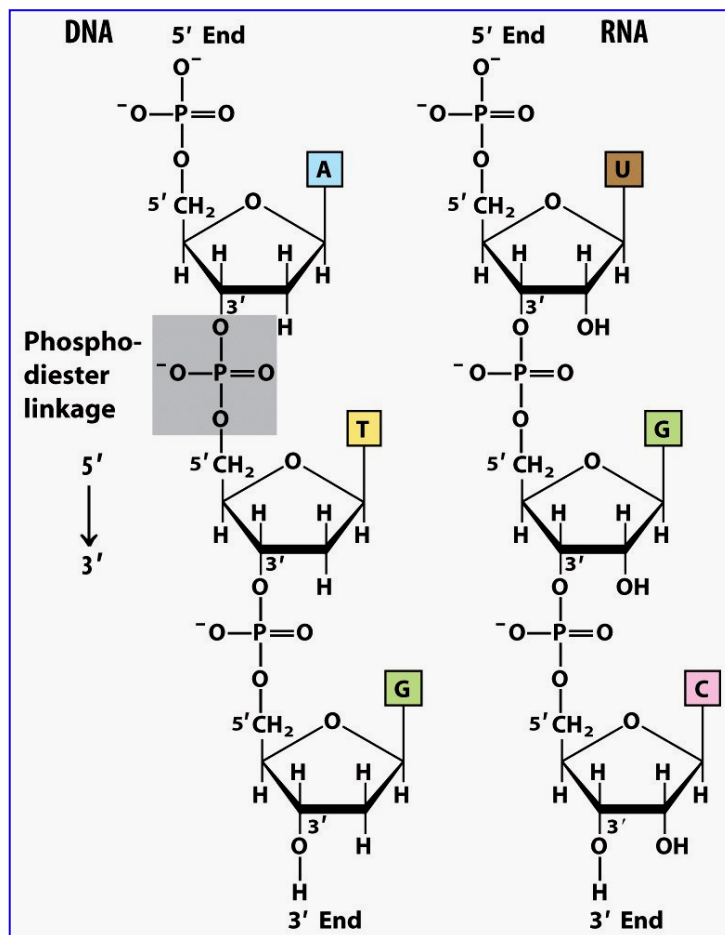
Figure 4.10: Structure of G-quadruplex.

#### 4.10.4.11. DNA IS A STORE OF GETIC INFORMATIONS

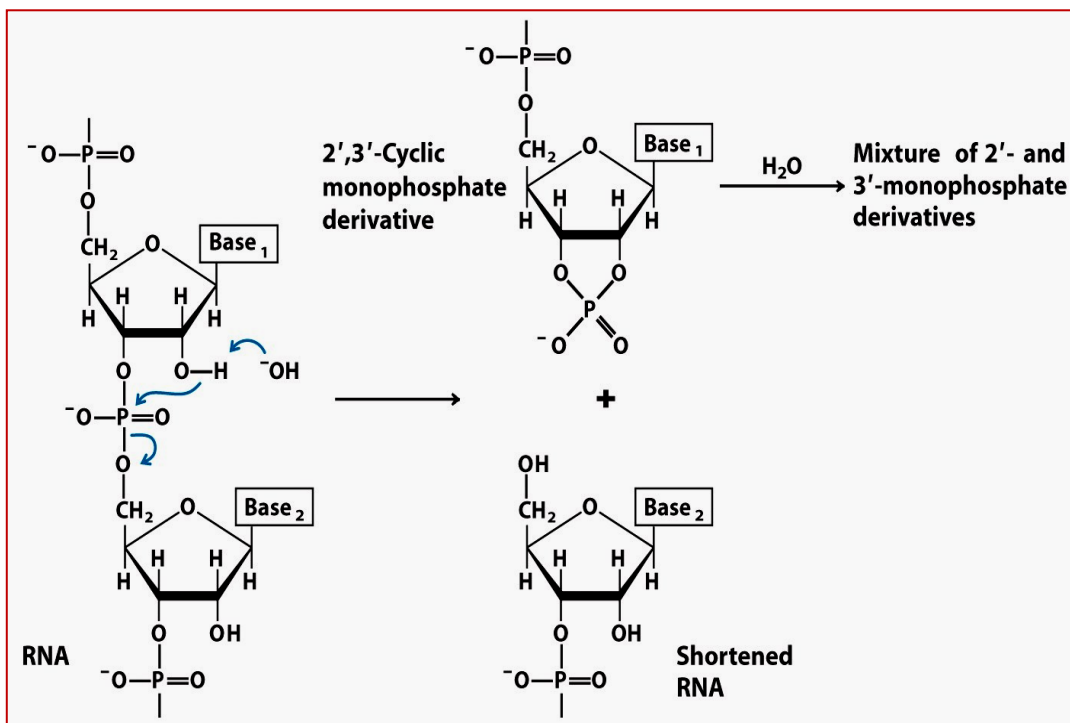
DNA contains all of the information necessary to make a complete organism, like the human shown here. DNA is a very useful molecule for storing this information because it is extremely stable - you can't break it apart easily. It's also easy to replicate. That is, machinery in the cell can make many copies of it with few mistakes.



## 4.11. THE COMPARATIVE STRUCTURE OF DNA AND RNA



## 4.12. The Stability of RNA v. DNA



## 4.13. PHYSICAL PROPERTIES OF DNA

1. DNA base pairs--called complementary base pairing ---- hydrogen bonds between bases.
2. This arrangement of two nucleotides binding together across the double helix is called a base pair.
3. In a double helix, the two strands are also held together via forces generated by the hydrophobic effect and pi-stacking, which are not influenced by the sequence of the DNA.
4. As hydrogen bonds are not covalent, they can be broken and rejoined relatively easily. The two strands of DNA in a double helix can therefore be pulled apart like a zipper, either by a mechanical force or high temperature.
5. In the laboratory, the duplex stability can be measured by finding the temperature required to break the hydrogen bonds, their melting temperature (also called *T<sub>m</sub>* value). When all the base pairs in a DNA double helix melt, the strands separate and exist in solution as two entirely independent molecules. These single-stranded DNA molecules have no single common shape, but some conformations are more stable than others.

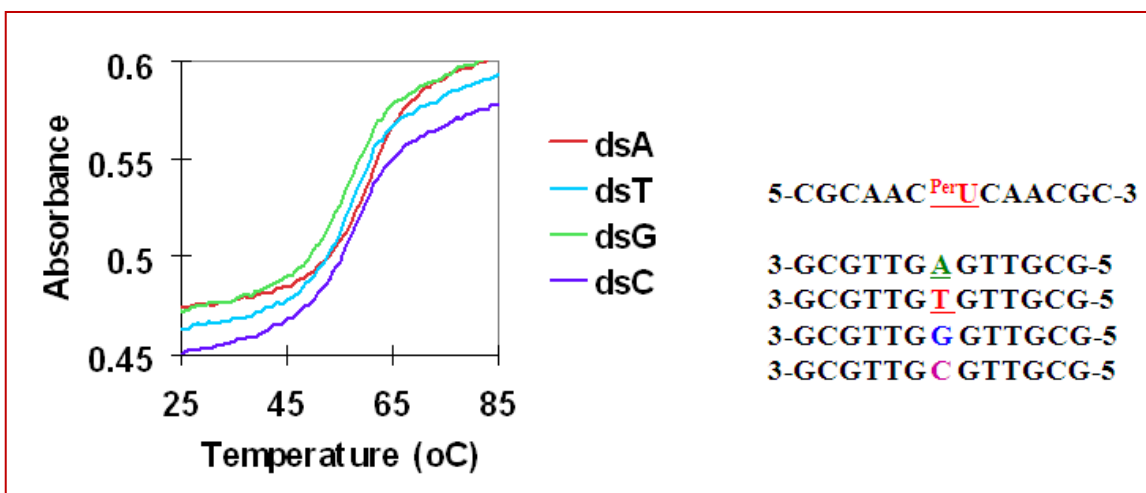


Figure 4.11: Melting curve of various duplexes shown.

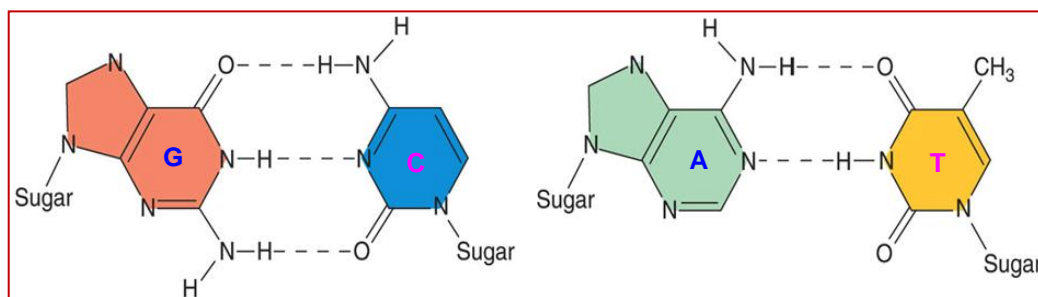
#### 4.14. Stability of Double Helix: Forces in duplex DNA

Since biological processes are regulated, fundamentally, through specific DNA-protein recognition, and DNA protein recognition is controlled by DNA conformation, the study of duplex DNA structure and stability is highly significant. Various Weak Forces come together to stabilize the DNA structure.

1. Hydrogen bonding (2-3 kcal/mol per base pair)
2. Stacking interactions (4-15 kcal/mol per base pair)
  - a. Hydrophobic interactions
  - b. van der Waals interactions
3. **Charge-Charge Interactions**
4. Solvation

These are cooperative forces; each contributes a little, but adds up because DNA chains can be millions of nucleotides long.

- **Hydrogen bonds**, linkage between bases, although weak energy-wise, is able to stabilize the helix because of the large number present in DNA molecule. Also important that the purine-pyrimidine base pairs are of similar size.



- **Stacking interactions**, or also known as Van der Waals interactions between bases are weak, but the large amounts of these interactions help to stabilize the overall structure of the helix.
  - Double helix is stabilized by **hydrophobic effects** by burying the bases in the interior of the helix increases its stability; having the hydrophobic bases clustered in the interior of the helix keeps it away from the surrounding water, whereas the more polar surfaces, hence hydrophilic heads are exposed and interaction with the exterior water
  - Stacked base pairs also attract to one another through **Van der Waals forces** the energy associated with a single van der Waals interaction has small significant to the overall DNA structure however, the net effect summed over the numerous atom pairs, results in substantial stability.
  - Stacking also favors the conformations of rigid five-membered rings of the sugars of backbone.

**Evidence of Stacking interactions:** Compounds that interfere with Hydrogen bonds (urea, formamide) don't separate strands by themselves, still requires heat.

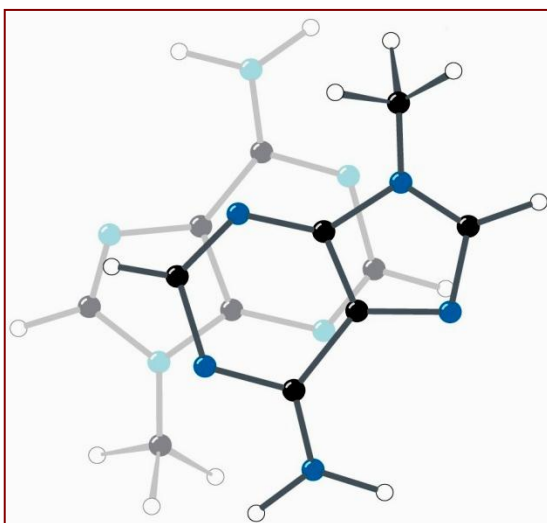


Figure 4.12: The Stacking of adenine rings in the crystal structure of 9-methyladenine.

Table 4.1: Association Constants for Base Pair Formation.

Base Pair	K (M <sup>-1</sup> ) in CDCl <sub>3</sub>
Self-association	
A·A	3.1
U·U	6.1
C·C	28.0
G·G	10 <sup>3</sup> -10 <sup>4</sup>
Watson-Crick Base Pairs	
A·U	100
G·C	10 <sup>4</sup> -10 <sup>5</sup>
Kyogoku, Y.; Lord, R. C.; Rich, A. <i>Biochim. Biophys. Acta.</i> <b>1969</b> , 179, 10.	

- **Charge-Charge Interactions-** refers to the electrostatic (ion-ion) repulsion of the negatively charged phosphate is potentially unstable, however the presence of Mg<sup>2+</sup> and cationic proteins with abundant Arginine and Lysine residues that stabilizes the double helix. Double-stranded helix structure thus, promoted by having phosphates on outside, interact with H<sub>2</sub>O and counter ions (K<sup>+</sup>, Mg<sup>2+</sup>, etc.).
- **Solvation** also plays a role in stabilizing the double helix that affects base pairing to mediating binding events.

Thus, the DNA strands in a double helix are held together by the H-bonds between the bases. The H-bonds already provide specificity but they also confer stability to the structure. The phosphate groups must be neutralized (by Na<sup>+</sup> or Mg<sup>2+</sup> ions) to allow the negatively charged phosphates to be in close proximity. The hydrophobic interactions between the planar base pairs stabilize the bases on the inside of the helix, so these provide stability to the structure but do not contribute to the specificity.

State-of-the-art experimental techniques, including denaturation study, study of salt effect are used to explore the binding affinities of DNA double-stranded oligomers in the solution. Besides these, electrospray ionization, liquid chromatography mass spectrometry (LCMS) and Fourier transform mass spectrometry (FTMS), are used to explore the binding affinities of DNA double-stranded oligomers in the gas phase, in the absence of solvent.

### DNA Denaturation

DNA "Melting" is the term given to the separation of the two DNA strands. There are a number of ways to do this experimentally: Helical formation can be monitored by observing the optical density of a solution. The disruption of base stacking alters the electronic interaction between the bases. As the electronic interaction decreases, it becomes easier for an electron to absorb a photon. Hence, denaturation of DNA leads to the "hyperchromic" effect, i.e., the increased absorption of light. Thus, an increase in Temperature - when temperature of a DNA solution increases to the melting point (T<sub>m</sub>), the strands

separate. The absorbance of the solution changes as shown in the graph below. The increase in absorbance as the strands separate is due to the irregular orientation of the bases in the SS-DNA compared to the regular planar orientation in the helix.

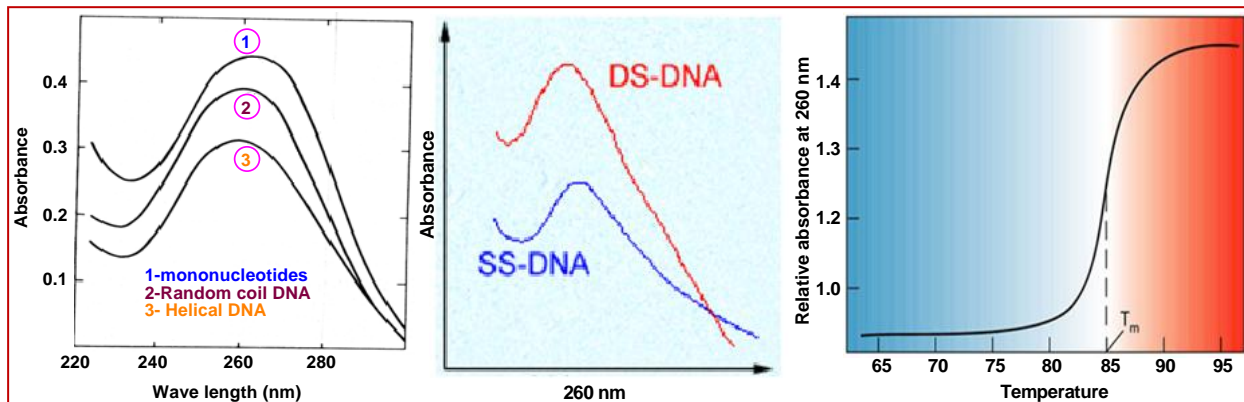


Figure 4.13: Temperature dependent absorbance and the melting temperature of DNA.

### Double Helix Stability and Base Composition:

In general, helical stability is linearly related to fractional G+C base pair content in DNA. As G+C increases so does stability (Figure 21). An empirical formula for calculating the melting temperature of a particular helix is given as  $T_m$  ( $^{\circ}\text{C}$ ) =  $69.3 + 41 \cdot f_{\text{G/C}}$ . This expression quantifies the observed result that there is a linear relation between  $T_m$  and G+C content. This observation argues that the energetic contributions of the bases in the helix to its stability are independent and therefore additive--->this implies that stabilization energies are sequence independent. That is the base pairs are all contributing equally and independently a constant amount of stacking energy, independent of the neighbors.

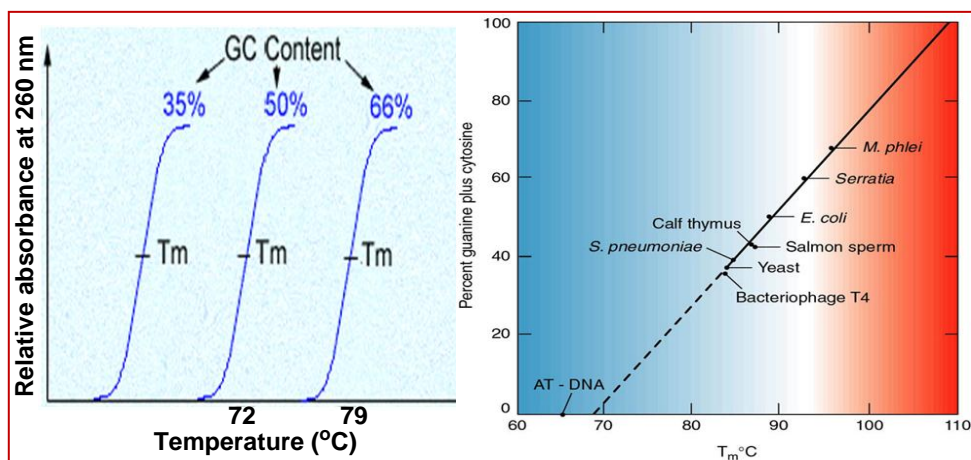


Figure 4.14: Double helix stability depends on base composition.

## Helical Stability and Salt

It has been long observed that multiple stranded polynucleotide helices are stabilized by increasing monovalent cation concentration (**Figure 4.15**). In fact the  $T_m$  of a given DNA is linearly dependent on the log of the monovalent cation concentration. We will not spend a lot of time on the polyelectrolyte behavior of nucleic acids, but instead we will simplify the treatments and take an empirical and thermodynamic approach.

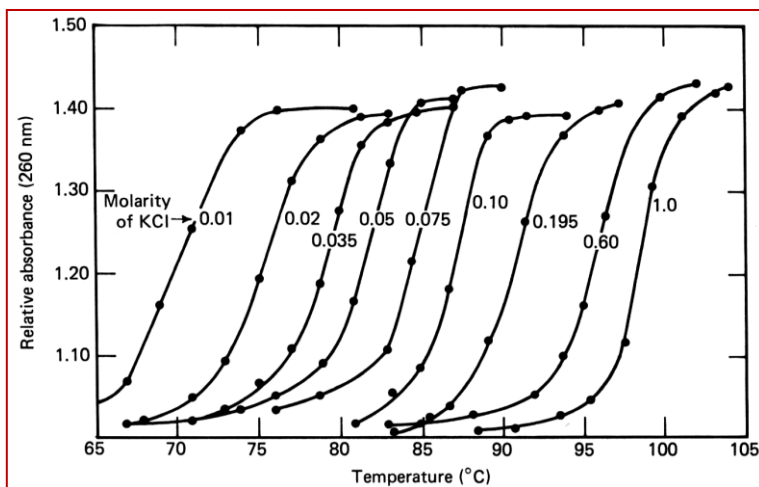
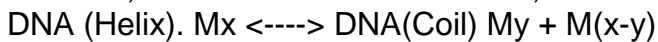


Figure 4.15: Double helix stability depends on salt concentration.

The DNA phosphate backbone is negatively charged. In salt solutions, cations are associated with it. When DNA is denatured fewer total cations are associated with the separated strands than with the nucleic acid helix in its native state. This is because the charge density on double stranded DNA is higher than single strand nucleic acids. This creates a larger electrostatic potential, which more effectively attracts counter ions.

Thus, in the denaturation reaction, the mass action equation can be written.



$x$  = no. of ions bound/base pair in a helix

$y$  = no. of ions bound/base in a coil

$x - y$  = net gain in free cations due to denaturation

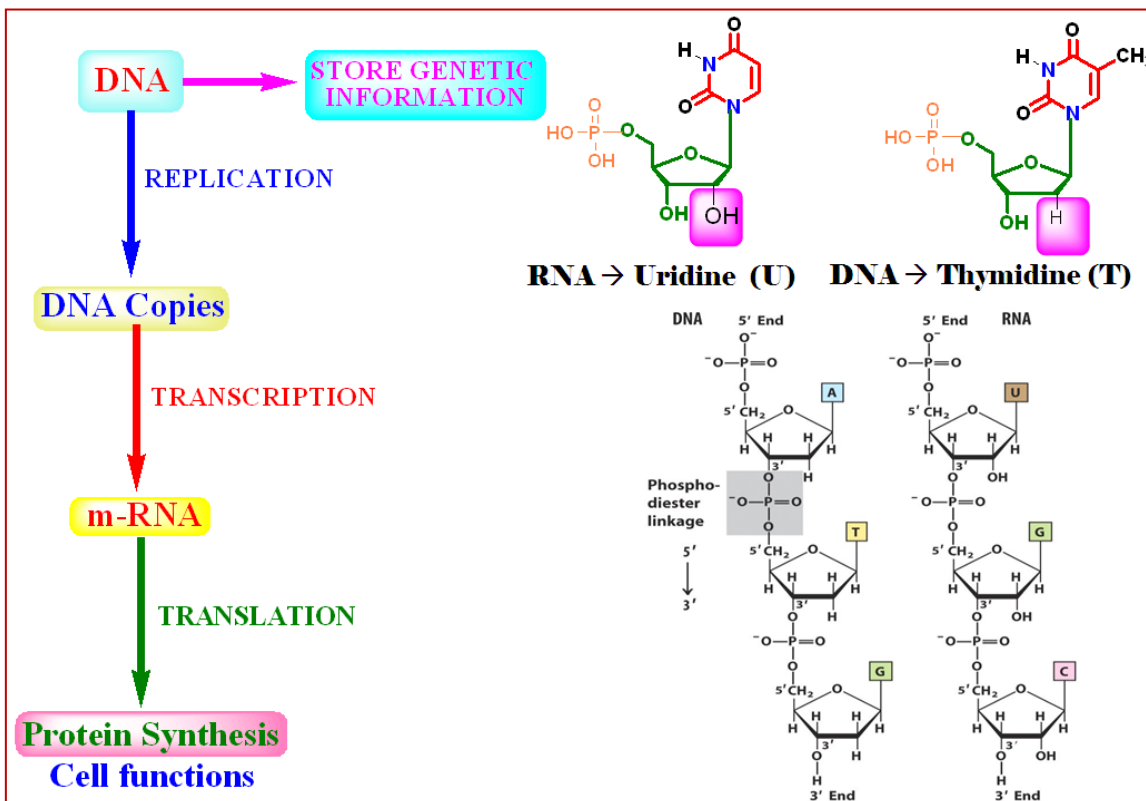
Therefore, the denaturation reaction equilibrium can be shifted by adjusting the cation concentration. We have already discussed that effect of temperature on helix  $\rightarrow$  coil transition. The two effects can be balanced at particular conditions. That is if the salt is raised, increases helix potential, can increase temperature to denature.

We have only discussed here the effect of monovalent cations on structure. The effects of divalent cations are much more complex due to their multiple interactions with the DNA phosphate backbone—each  $M^{2+}$  can potentially bind one



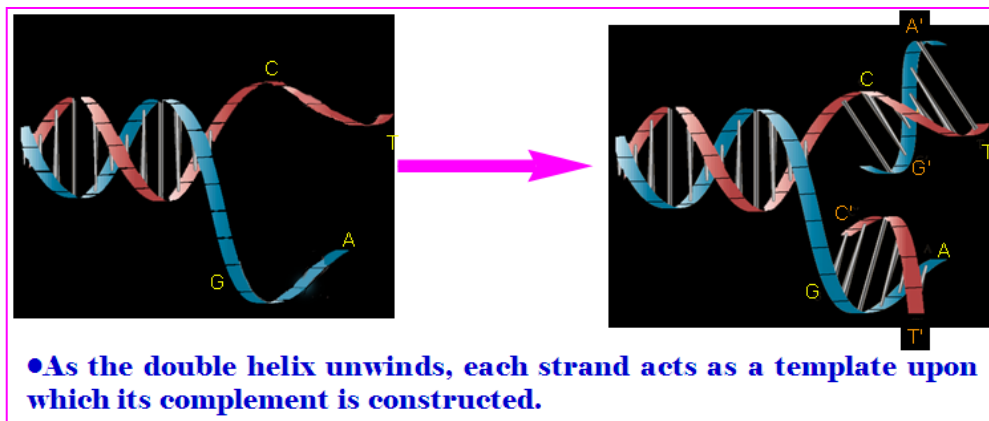
or two DNA phosphates and the binding is likely to be cooperative. Hence, the  $T_m$  dependence on divalent cation concentration is decidedly non-linear.

## 4.15. BIOLOGICAL FUNCTIONS OF DNA



### 4.15.1. DNA REPLICATION

1. Cell division is essential for an organism to grow
2. During this DNA make copy → So, two daughter cells have the same genetic information as their parent.
3. Called DNA replication.
4. Here, the two strands are separated and then each strand's complementary DNA sequence is recreated by an enzyme called DNA polymerase.
5. This enzyme makes the complementary strand by finding the correct base through complementary base pairing, and bonding it onto the original strand.
6. As DNA polymerases → extend a DNA strand in a 5' to 3' direction → So, different mechanisms are used to copy the antiparallel strands of the double helix.
7. Thus, base on the old strand dictates which base appears on the new strand, and the cell ends up with a perfect copy of its DNA.



#### 4.15.2. TRANSCRIPTION & TRANSLATION = PROTEIN BIOSYNTHESIS

●According to Crick, the "central dogma" of molecular biology is:  
"DNA makes RNA makes protein."

●Three kinds of RNA are involved.

**messenger RNA (mRNA)**

**transfer RNA (tRNA)**

**ribosomal RNA (rRNA)**

●There are two main stages.

**Transcription**

**Translation**

### 4.15.2.1. TRANSCRIPTION

- **Transcription is the formation of a strand of mRNA using one of the DNA strands as a template.**
- **The nucleotide sequence of the mRNA is complementary to the nucleotide sequence of the DNA template.**
- **Transcription begins at the 5' end of DNA and is catalyzed by the enzyme RNA polymerase.**

- **As double-stranded DNA unwinds, a complementary strand of mRNA forms at the 5' end.**
- **Uracil is incorporated in RNA instead of thymine.**

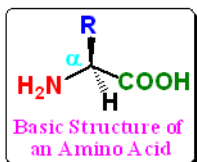
### 4.15.2.2. TRANSLATION

The message on DNA that has been translated to mRNA:

1. Degenerate: more than one three base codon can code for the same amino acid
2. Specific: each codon specifies a particular amino acid
3. Non-overlapping and comma less:
  - None of the bases are shared between consecutive codons
  - No noncoding bases appear in the base sequence
4. Universal: all organisms use the same code
5. All 64 codons have meaning
  - 61 code for amino acids
  - Three code for the “stop” signal
6. Multiple codes for an amino acid tend to have two bases in common
  - CUU, CUC, CUA, CUG code for leucine
  - Makes the code mutation resistant
7. Codons are written in a 5' 3' sequence

### 4.15.2.2.1. The Genetic Code

- The nucleotide sequence of mRNA codes for the different amino acids found in proteins:



- There are three nucleotides per codon:

<b>Alanine</b> GCU GCA GCC GCG	<b>Arginine</b> CGU CGA AGA CGC CGG AGG	<b>Asparagine</b> AAU AAC	<b>Aspartic Acid</b> GAU GAC	<b>Cysteine</b> UGU UGC
<b>Glutamic acid</b> GAA GAG	<b>Glutamine</b> CAA CAG	<b>Glycine</b> GGU GGA GGC GGG	<b>Histidine</b> CAU CAC	<b>Isoleucine</b> AUU AUA AUC
<b>Leucine</b> UUA CUU CUA UUG CUC CUG	<b>Lysine</b> AAA AAG	<b>Methionine</b> AUG	<b>Phenylalanine</b> UUU UUC	<b>Proline</b> CCU CCA CCC CG
<b>Serine</b> UCU UCA AGU UCC UCG AGC	<b>Threonine</b> ACU ACA ACC ACG	<b>Tryptophan</b> UGG	<b>Tyrosine</b> UAU UAC	<b>Valine</b> GUU GUA GUC GUG

- There are 64 possible combinations of A, U, G, and C.
- The genetic code is redundant. Some proteins are coded for by more than one codon.

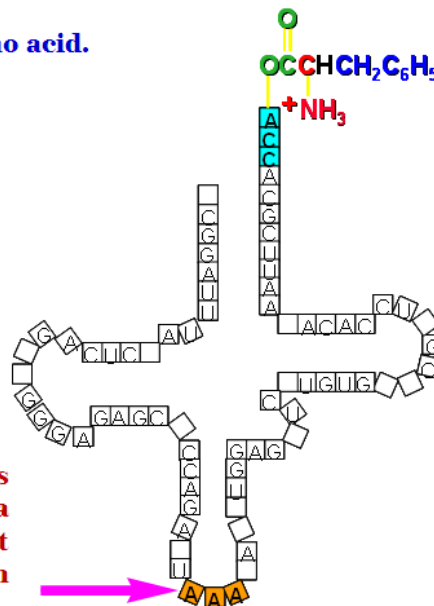
Table 4.2: CORRELATION BETWEEN CODONS AND AMINO ACIDS

		SECOND BASE					
		U	C	A	G		
FIRST BASE	U	UUU	UCU	UAU	UGU	U C A G	
		UUC	UCC	UAC	UGC		
		UUA	UCA	UAA Stop	UGA Stop		
		UUG	UCG	UAG Stop	UGG Trp		
	C	CUU	CCU	CAU	CGU	U C A G	
		CUC	CCC	CAC	CGC		
		CUA	CCA	CAA	CGA		
		CUG	CCG	CAG	CGG		
	A	AUU	ACU	AAU	AGU	U C A G	
		AUC	ACC	AAC	AGC		
		AUA	ACA	AAA	AGA		
		AUG Met or start	ACG	AAG	AGG		
	G	GUU	GCU	GAU	GGU	U C A G	
		GUC	GCC	GAC	GGC		
		GUA	GCA	GAA	GGA		
		GUG	GCG	GAG	GGG		

#### 4.15.2.2.2. TRANSFER RNA (tRNA)

• There are 20 different tRNAs, one for each amino acid.

• Phenylalanine tRNA



This AAA triplet is complementary to a UUU triplet of mRNA; it is an anticodon.

• Each tRNA is single stranded with a CCA triplet at its 3' end.

• A particular amino acid is attached to the tRNA by an ester linkage involving the carboxyl group of the amino acid and the 3' oxygen of the tRNA.

#### 4.15.2.2.3. PROTEIN BIO-SYNTHESIS

- Protein synthesis is called *translation*
  - Carried out on ribosomes, complexes of
    - rRNA
    - Proteins
- Protein synthesis occurs in multiple places on one mRNA at a time
  - mRNA plus the multiple ribosomes are called a polysome
- tRNA
  - Binds a specific amino acid aided by aminoacyl tRNA synthetase
  - Recognizes the appropriate codon on the mRNA
- Proteins built in a ribosome--a protein factory.
- First the mRNA attaches itself to the ribosome.
- Then specific adapter molecules, called transfer RNAs (tRNAs), match themselves to their appropriate codons. Each codon has a corresponding tRNA that carries a specific amino acid at one end.
- Here is the beginning of a new protein. The first tRNA carrying methionine matched itself to the first codon. Then a second tRNA carrying arginine matched itself with the second codon. A chemical reaction joined one free

- end (the hook) of the methionine molecule to a free end (the eye) of the arginine molecule, and methionine's tRNA was released.
- Now a third tRNA carrying tyrosine has matched the third codon, and a chemical reaction has joined the free end of the arginine to tyrosine, and arginine's tRNA has been released. The protein now has three amino acids joined together. This process will continue until a "stop" codon is reached.

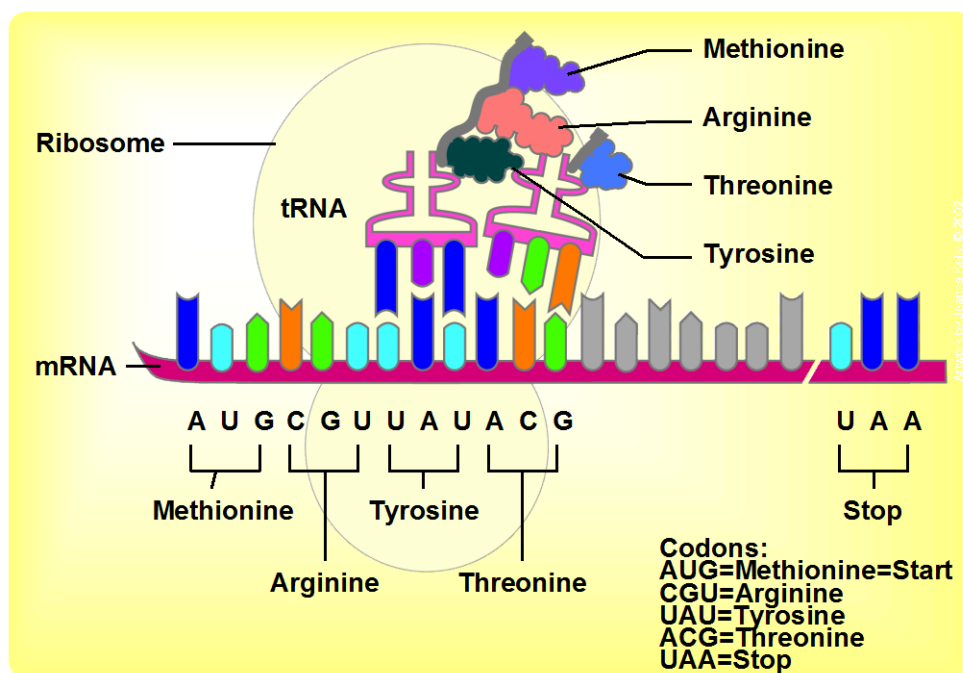


Figure 4.16: The protein synthesis is (*translation*)

## 4.16. DNA Intercalation

**Definition:** There are several ways molecules (in this case, also known as *ligands*) can interact with DNA. Ligands may interact with DNA by covalently binding, electrostatically binding, or intercalating. Intercalation occurs when ligands of an appropriate size and chemical nature fit themselves in between base pairs of DNA. These ligands are mostly polycyclic, aromatic, and planar, and therefore often make good nucleic acid stains. Intensively studied DNA intercalators include berberine, ethidium bromide, proflavine, daunomycin, doxorubicin, and thalidomide. DNA intercalators are used in chemotherapeutic treatment to inhibit DNA replication in rapidly growing cancer cells. Examples include doxorubicin (adriamycin) and daunorubicin (both of which are used in treatment of Hodgkin's lymphoma), and dactinomycin (used in Wilm's tumour, Ewing's Sarcoma, rhabdomyosarcoma).

In order for an intercalator to fit between base pairs, the DNA must dynamically open a space between its base pairs by unwinding.



### 4.16.1. Types of Intercallators

DNA-binding agents tend to interact noncovalently with the host DNA molecule through two general modes: (i) Threading Intercalation: in a groove-bound fashion stabilized by a mixture of hydrophobic, electrostatic, and hydrogen-bonding interactions and (ii) Classical Intercalation: through an intercalative association in which a planar, heteroaromatic moiety slides between the DNA base pairs.

Intercalative binding, the most commonly studied, is the noncovalent stacking interaction resulting from the insertion of a planar heterocyclic aromatic ring between the base pairs of the DNA double helix.

### 4.16.2. The Mechanism of Intercalation

Intercalation as a mechanism of interaction between cationic, planar, polycyclic aromatic systems of the correct size (on the order of a base pair) was first proposed by Leonard Lerman in 1961. One proposed mechanism of intercalation is as follows: In aqueous isotonic solution, the cationic intercalator is attracted electrostatically to the polyanionic DNA. The ligand displaces a sodium and/or magnesium cation that always surrounds DNA (to balance its charge), forming a weak electrostatic bond with the outer surface of DNA. From this position, the ligand may then slide into the hydrophobic environment found between the base pairs and away from the hydrophilic outer environment surrounding the DNA. The base pairs transiently form such openings due to energy absorbed during collisions with solvent molecules.

### 4.16.3. Effect of Intercalation

Intercalation stabilizes, lengthens, stiffens, and unwinds the DNA double helix. The degree of unwinding varies depending on the intercalator. As for example, the ethidium cation unwinds DNA by about 26° and proflavine by about 17°.

These structural modifications can lead to functional changes, often to inhibition of transcription and replication and DNA repair processes, which make intercalators potent mutagens. There is much interest in the ability of intercalators to inhibit nucleic acid synthesis *in vivo*, leading to activity as mutagens, antibiotics, antibacterials, trypanocides, schistosomicides, and antitumor agents. Intercalative interactions between DNA duplexes and planar polycyclic aromatic organic intercalators, such as ethidium bromide, acridine and its derivatives, and benzo[*a*]pyrene (BP), have been thoroughly studied. Studies of bulky intercalators are rare. Bis-intercalators have also been reported, such as bisnaphthalimide, which exhibits antitumor activity. A tetraintercalator has been reported by Iverson, *et al.* Some of the potent polycyclic aromatic intercalators are listed in Figure 4.17 as a representative example.

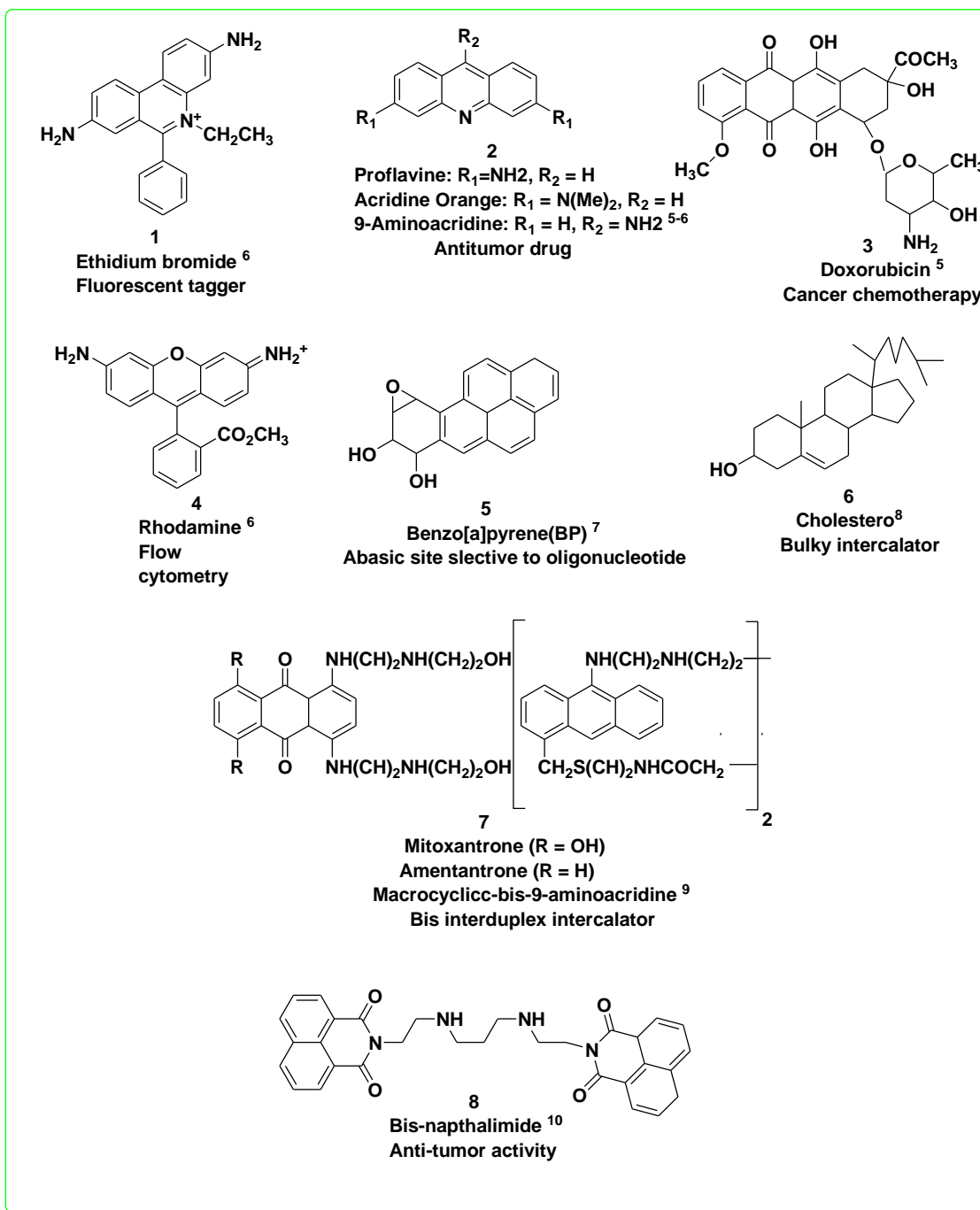


Figure 4.17: Example of some DNA intercalators with their structures and applications.

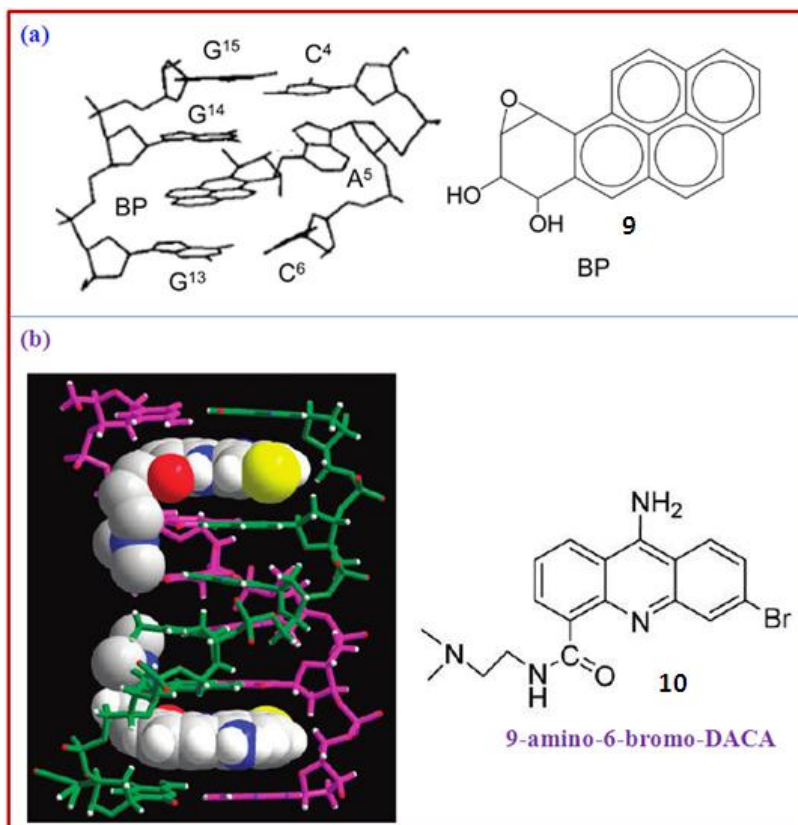


Figure 4.18: (a) NMR solution structure of the DNA duplex and BP showing classical intercalation between the G<sup>13</sup>G<sup>14</sup> base step of the opposite strand. (b) X-ray structure showing threading intercalation of 9-amino-6-bromo-DACA (space-filling model) into the DNA duplex.

Metal complexes that can have intercalative interactions with DNA form two classes. For relatively inert square-planar Pt(II) complexes and octahedral complexes with aromatic ligands, intercalation into DNA mainly involves the aromatic ligands. Metal complexes containing  $\sigma$ -bonded ligands with aromatic side arms as intercalators or organometallic complexes with  $\pi$ -bonded arenes as intercalators can be dual-function complexes: the aromatic side arms or arene ligands can intercalate between DNA bases while the metal coordinates directly to a DNA base. Intercalative interactions between metal complexes and DNA have novel features that can influence biological activity. Square-planar complexes containing aromatic fragments can bind to DNA by intercalation without direct metal coordination to DNA bases. Lippard *et al.* have shown that square-planar platinum(II) complexes containing heterocyclic aromatic ligands, such as terpy, quaterpy, phen, bipy, and phi, bind to DNA duplexes noncovalently, intercalating between the base pairs (**Figure 4.19 a-c**).

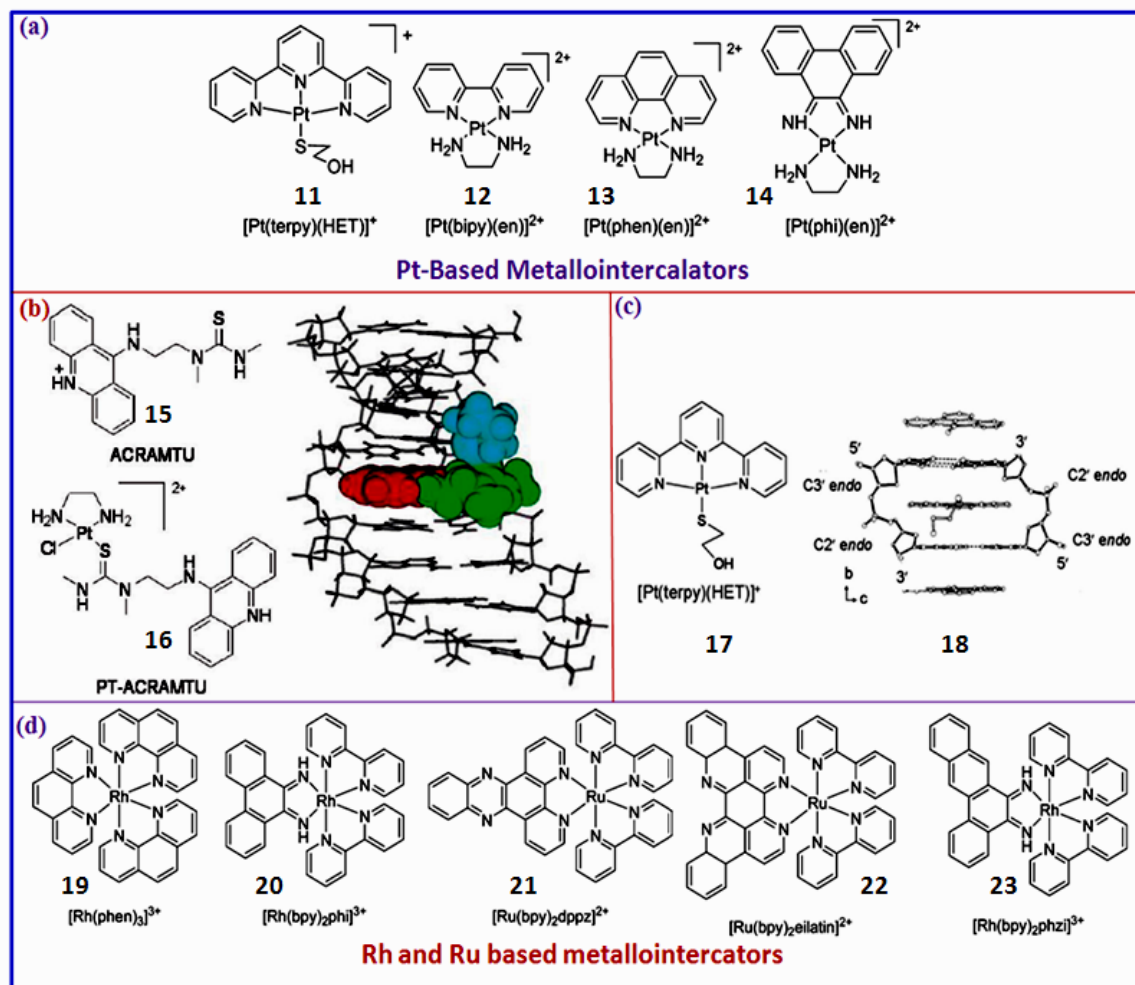


Figure 4.19: (a) Some examples of platinum metallointercalators. (b) Structures of ACRAMTU, Pt-ACRAMTU and model of a DNA duplex containing with the Pt-ACRAMTU adduct (c) Molecular structure of  $[\text{Pt}(\text{terpy})(\text{HET})]^+$  and Crystal structure of  $[\text{Pt}(\text{terpy})(\text{HET})]^{+3} \text{d}(\text{CpG})_2$  showing HET intercalated between two GC base pairs.(d) Some examples of Rh and Ru based metallointercalators.

## 4.17. Chemical Synthesis of DNA

Definition: **Oligonucleotide synthesis** is the chemical synthesis of relatively short fragments of nucleic acids with defined chemical structure/sequence. The technique is extremely useful in current laboratory practice because it provides a rapid and inexpensive access to custom-made oligonucleotides of the desired sequence. Whereas enzymes synthesize DNA and RNA in a 5' to 3' direction, chemical oligonucleotide synthesis is carried out in the opposite, 3' to 5' direction. Currently, the process is implemented as solid-phase synthesis using phosphoramidite method and phosphoramidite building blocks derived from protected 2'-deoxynucleosides (dA, dC, dG, and T), ribonucleosides (A, C, G, and U), or chemically modified nucleosides, e.g. LNA. To obtain the desired oligonucleotide, the building blocks are sequentially coupled to the growing

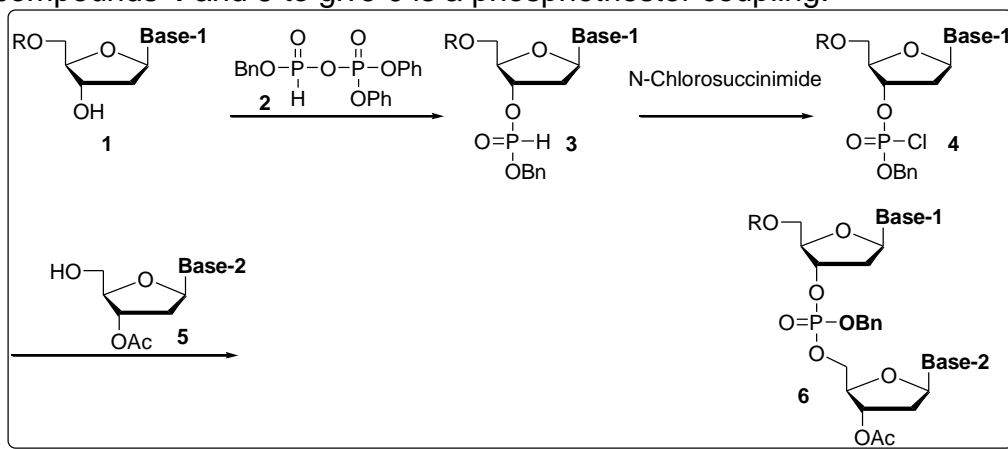
oligonucleotide chain in the order required by the sequence of the product. Upon the completion of the chain assembly, the product is released from the solid phase to solution, deprotected, and collected. The occurrence of side reactions sets practical limits for the length of synthetic oligonucleotides (up to about 200 nucleotide residues) because the number of errors accumulates with the length of the oligonucleotide being synthesized. Products are often isolated by HPLC to obtain the desired oligonucleotides in high purity. Typically, synthetic oligonucleotides are single-stranded DNA or RNA molecules around 15–25 bases in length. They are most commonly used as antisense oligonucleotides, small interfering RNA, primers for DNA sequencing and amplification, probes for detecting complementary DNA or RNA via molecular hybridization, tools for the targeted introduction of mutations and restriction sites, and for the synthesis of artificial genes.

#### 4.17.1. Types of Chemical Synthesis

The evolution of oligonucleotide synthesis comprised of four major methods of the formation of inter-nucleosidic linkages.

##### 4.17.1.1. Early work and contemporary H-phosphonate synthesis

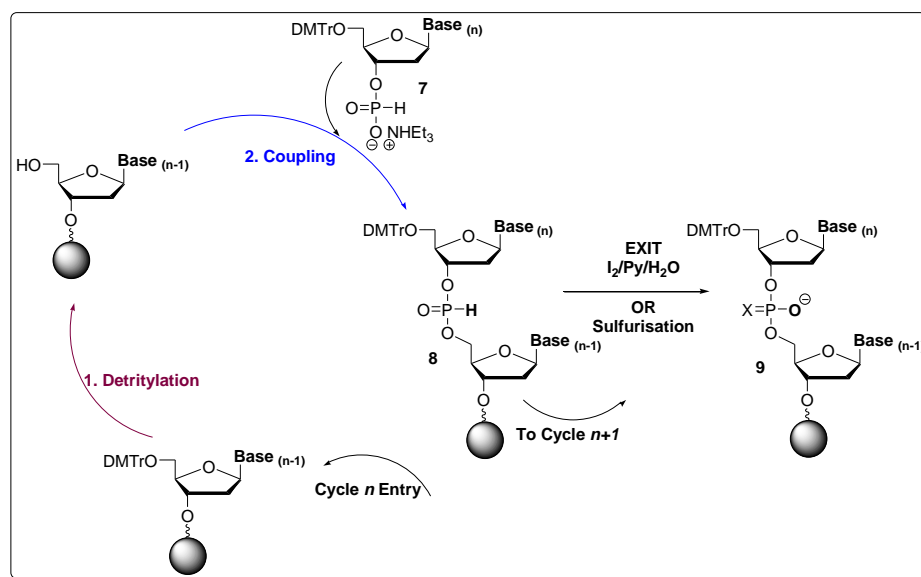
**A. Todd's synthesis:** In the early 1950s, Alexander Todd's group pioneered H-phosphonate and phosphate triester methods of oligonucleotide synthesis. The reaction of compounds **1** and **2** to form H-phosphonate diester **3** is an H-phosphonate coupling in solution while that of compounds **4** and **5** to give **6** is a phosphotriester coupling.



Scheme 4.1: Todd's DNA synthesis.

**B. H-Phosphonate Synthesis:** Thirty years later, this work inspired, independently, two research groups to adopt the H-phosphonate chemistry to the solid-phase synthesis using nucleoside H-phosphonate monoesters **7** as building blocks and pivaloyl chloride, 2,4,6-

triisopropylbenzenesulfonyl chloride (TPS-Cl), and other compounds as activators. The practical implementation of H-phosphonate method resulted in a very short and simple synthetic cycle consisting of only two steps, detritylation and coupling (**Scheme 4.2**). Oxidation of internucleosidic H-phosphonate diester linkages in **8** to phosphodiester linkages in **9** ( $X = O$ ) with a solution of iodine in aqueous pyridine is carried out at the end of the chain assembly rather than as a step in the synthetic cycle. Alternatively, **8** can be converted to phosphorothioate **9** ( $X = S$ ).

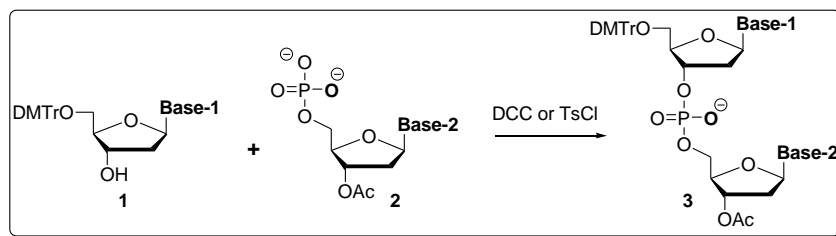


Scheme 4.2: Synthetic cycle in H-phosphonate method of oligonucleotide synthesis.

#### 4.17.1.2. Phosphodiester synthesis

In the 1950s, Khorana and co-workers developed a phosphodiester method where 3'-O-acetylnucleoside-5'-O-phosphate **2** (**Scheme 4.3**) was activated with *N,N*-dicyclohexylcarbodiimide (DCC) or 4-toluenesulfonylchloride (Ts-Cl) and a 5'-O-protected nucleoside **1** was reacted with the activated species to give a protected dinucleoside monophosphate **3**. Upon the removal of 3'-O-acetyl group using base-catalyzed hydrolysis, further chain elongation was carried out. Following this methodology, sets of tri- and tetraoxyribonucleotides were synthesized and enzymatically converted to longer oligonucleotides, which allowed elucidation of the genetic code. The major limitation of the phosphodiester method consisted in the formation of pyrophosphate oligomers and oligonucleotides branched at the internucleosidic phosphate. The method seems to be a step back from the more selective chemistry described earlier; however, at that time, most phosphate-protecting groups available now had not yet been introduced. The lack of the convenient protection strategy necessitated

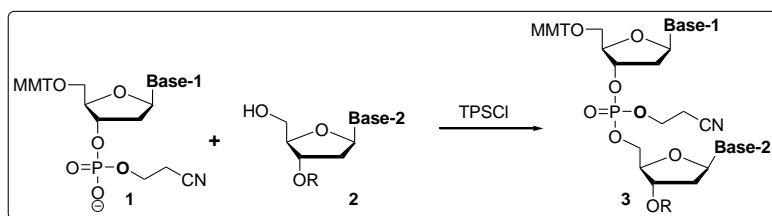
taking a retreat to a slower and less selective chemistry to achieve the ultimate goal of the study.



Scheme 4.3: Oligonucleotide coupling by phosphodiester method.

### 4.17.1. 3. Phosphotriester Synthesis

In the 1960s, groups led by R. Letsinger and C. Reese developed a phosphotriester approach. The defining difference from the phosphodiester approach was the protection of the phosphate moiety in the building block **1** (**Scheme 4.4**) and in the product **3** with 2-cyanoethyl group. This precluded the formation of oligonucleotides branched at the internucleosidic phosphate. The higher selectivity of the method allowed the use of more efficient coupling agents and catalysts, which dramatically reduced the length of the synthesis. The method, led to the automation of the oligonucleotide chain assembly.



Scheme 4.4: Oligonucleotide coupling by phosphotriester method.

### 4.17.1. 4. Phosphite Triester Synthesis

In the 1970s, substantially more reactive P(III) derivatives of nucleosides, 3'-O-chlorophosphites, were successfully used for the formation of inter-nucleosidic linkages. This led to the discovery of the phosphite triester methodology. The group led by M. Caruthers took the advantage of less aggressive and more selective 1*H*-tetrazolidophosphites and implemented the method on solid phase. The use of 2-cyanoethyl phosphite-protecting group in place of a less user-friendly methyl group led to the nucleoside phosphoramidites currently used in oligonucleotide synthesis.



## 4.17.1.5. Synthesis by the Phosphoramidite Method

### 4.17.1.5.1. Building Blocks: Nucleoside phosphoramidites

The naturally occurring nucleotides (nucleoside-3'- or 5'-phosphates) and their phosphodiester analogs are insufficiently reactive to afford expedite synthetic preparation of oligonucleotides in high yields. The selectivity and the rate of the formation of inter-nucleosidic linkages is dramatically improved by using 3'-O-(*N,N*-diisopropyl phosphoramidite) derivatives of nucleosides called nucleoside phosphoramidites that serve as building blocks in phosphite triester methodology. To prevent undesired side reactions, all other functional groups present in nucleosides have to be rendered unreactive by attaching protecting groups. Upon the completion of the oligonucleotide chain assembly, all the protecting groups are removed to yield the desired oligonucleotides.

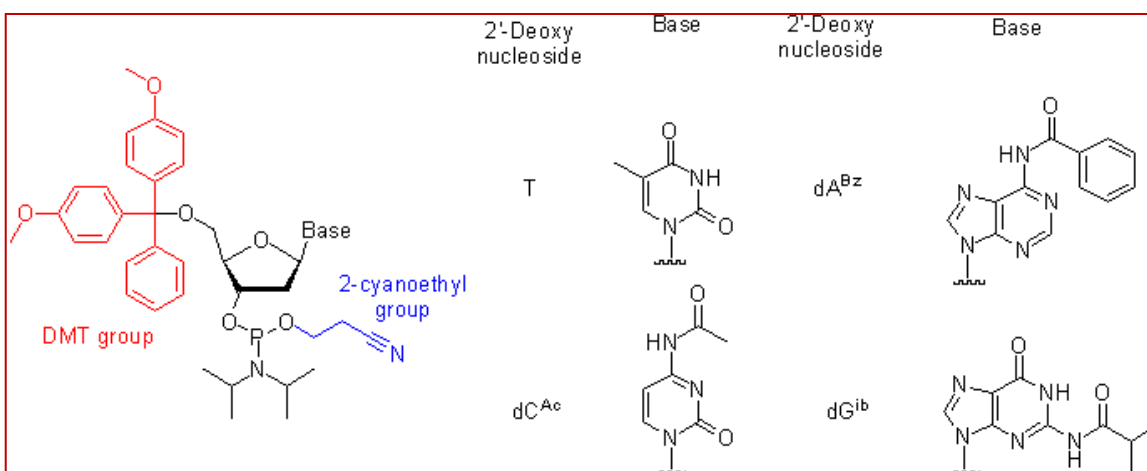


Figure 4.20: Protected 2'-deoxynucleoside phosphoramidites.

Below, the protecting groups currently used in commercially available and most common nucleoside phosphoramidite building blocks are briefly described.

- The 5'-hydroxyl group is protected by an acid-labile DMT (4,4'-dimethoxytrityl) group.
- Thymine and uracil, nucleic bases of thymidine and uridine, respectively, do not have exocyclic amino groups and hence do not require any protection. Although the nucleic base of guanosine and 2'-deoxyguanosine have an exocyclic amino group, its basicity is low to an extent that it does not react with phosphoramidites under the conditions of the coupling reaction. However, a phosphoramidite derived from the N2-unprotected 5'-O-DMT-2'-deoxyguanosine is poorly soluble in acetonitrile, the solvent commonly used in oligonucleotide synthesis. In contrast, the N2-protected versions of the same compound dissolve in acetonitrile well and hence are widely used. Nucleic bases adenine and cytosine bear the exocyclic

amino groups reactive with the activated phosphoramidites under the conditions of the coupling reaction. Although, at the expense of additional steps in the synthetic cycle, the oligonucleotide chain assembly may be carried out using dA and dC phosphoramidites with unprotected amino groups, most often these are kept permanently protected over the entire length of the oligonucleotide chain assembly. The protection of the exocyclic amino groups has to be orthogonal to that of the 5'-hydroxy group because the latter is removed at the end of each synthetic cycle. The simplest to implement and hence the most widely accepted is the strategy where the exocyclic amino groups bear a base-labile protection. Most often, two protection schemes are used.

- In the first, the standard and more robust scheme, Bz (benzoyl) protection is used for A, dA, C, dC, G, and dG are protected with isobutyryl group. More recently, acetyl group is often used to protect C and dC.
- In the second, mild protection scheme, A and dA are protected with isobutyryl or phenoxyacetyl groups (PAC). C and dC bear acetyl protection, and G and dG are protected with 4-isopropylphenoxyacetyl (i-Pr-PAC) or dimethylformamidino (dmf) groups. Mild protecting groups are removed more readily than the standard protecting groups. However, the phosphoramidites bearing these groups are less stable when stored in solution.
- The phosphite group is protected by a base-labile 2-cyanoethyl group. Once a phosphoramidite has been coupled to the solid support-bound oligonucleotide and the phosphite moieties have been converted to the P(V) species, the presence of the phosphate protection is not mandatory for the successful conducting of further coupling reactions.
- In RNA synthesis, the 2'-hydroxy group is protected with TBDMS (*t*-butyldimethylsilyl) group or with TOM (*t*-butyldimethylsilyloxymethyl) group, both being removable by treatment with fluoride ion.

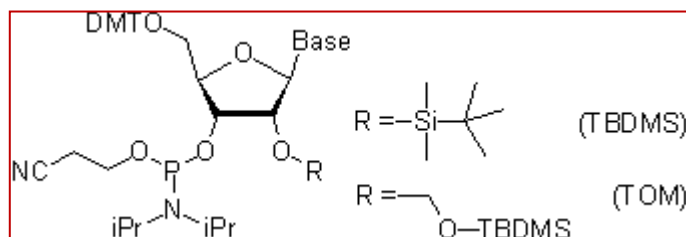


Figure 4.21: 2'-O-Protected ribonucleoside phosphoramidites.

- The phosphite moiety also bears a diisopropylamino ( $iPr_2N$ ) group reactive under acidic conditions. Upon activation, the diisopropylamino group

leaves to be substituted by the 5'-hydroxy group of the support-bound oligonucleotide.

#### **4.17.1.5.2. Non-nucleoside Phosphoramidites**

Non-nucleoside phosphoramidites are the phosphoramidite reagents designed to introduce various functionalities at the termini of synthetic oligonucleotides or between nucleotide residues in the middle of the sequence. In order to be introduced inside the sequence, a non-nucleosidic modifier has to possess at least two hydroxy groups, one of which is often protected with a DMT group while the other bears the reactive phosphoramidite moiety.

Non-nucleosidic phosphoramidites are used to introduce desired groups that are not available in natural nucleosides or that can be introduced more readily using simpler chemical designs. A very short selection of commercial phosphoramidite reagents is shown in Scheme for the demonstration of the available structural and functional diversity. These reagents serve for the attachment of 5'-terminal phosphate (**1**),  $\text{NH}_2$  (**2**), SH (**3**), aldehyde (**4**), and carboxylic groups (**5**), CC triple bonds (**6**), non-radioactive labels and quenchers (exemplified by 6-FAM amidite **7** for the attachment of fluorescein and dabcyI amidite **8**, respectively), hydrophilic and hydrophobic modifiers (exemplified by hexaethyleneglycol amidite **9** and cholesterol amidite **10**, respectively), and biotin amidite **11**.

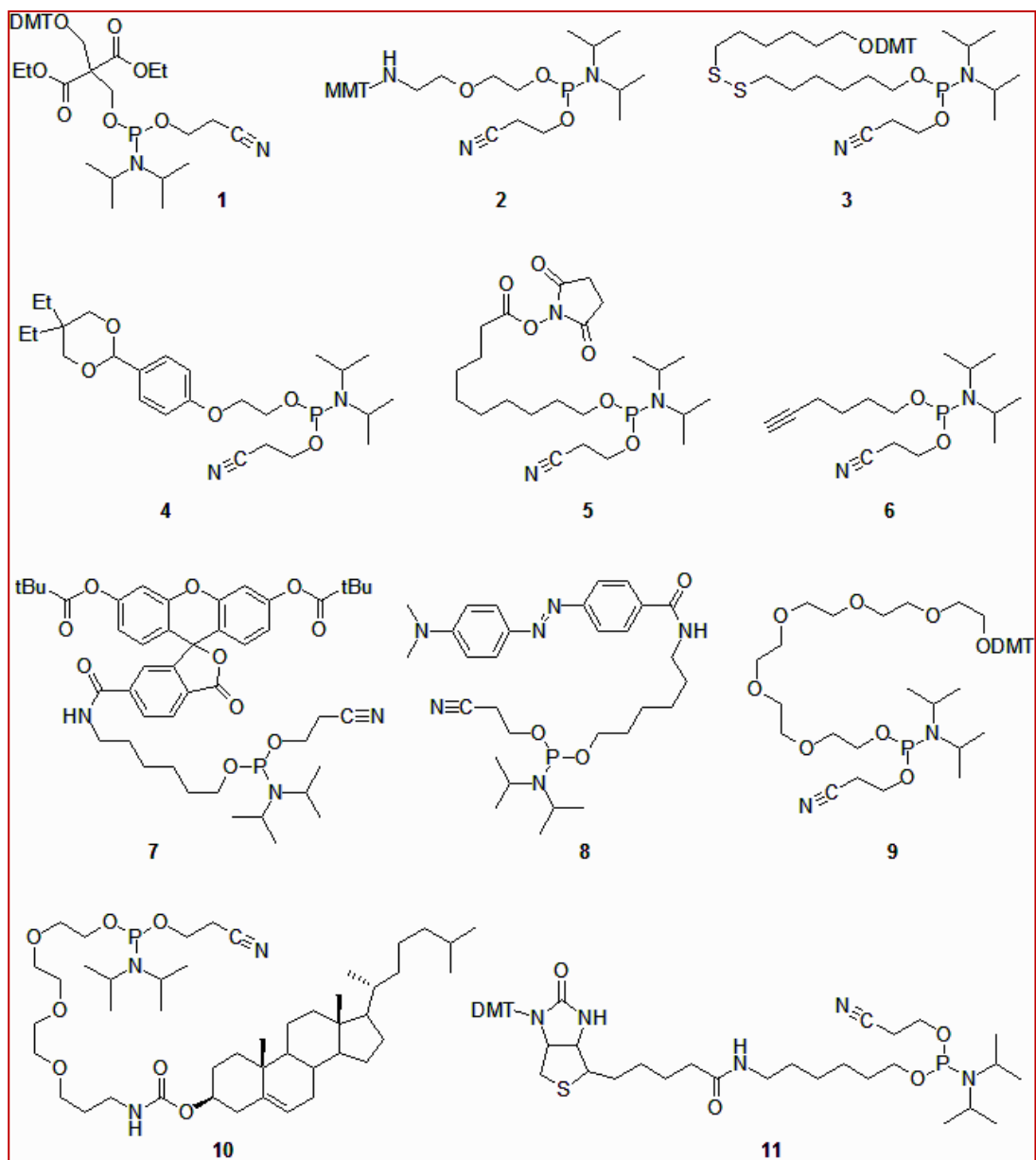
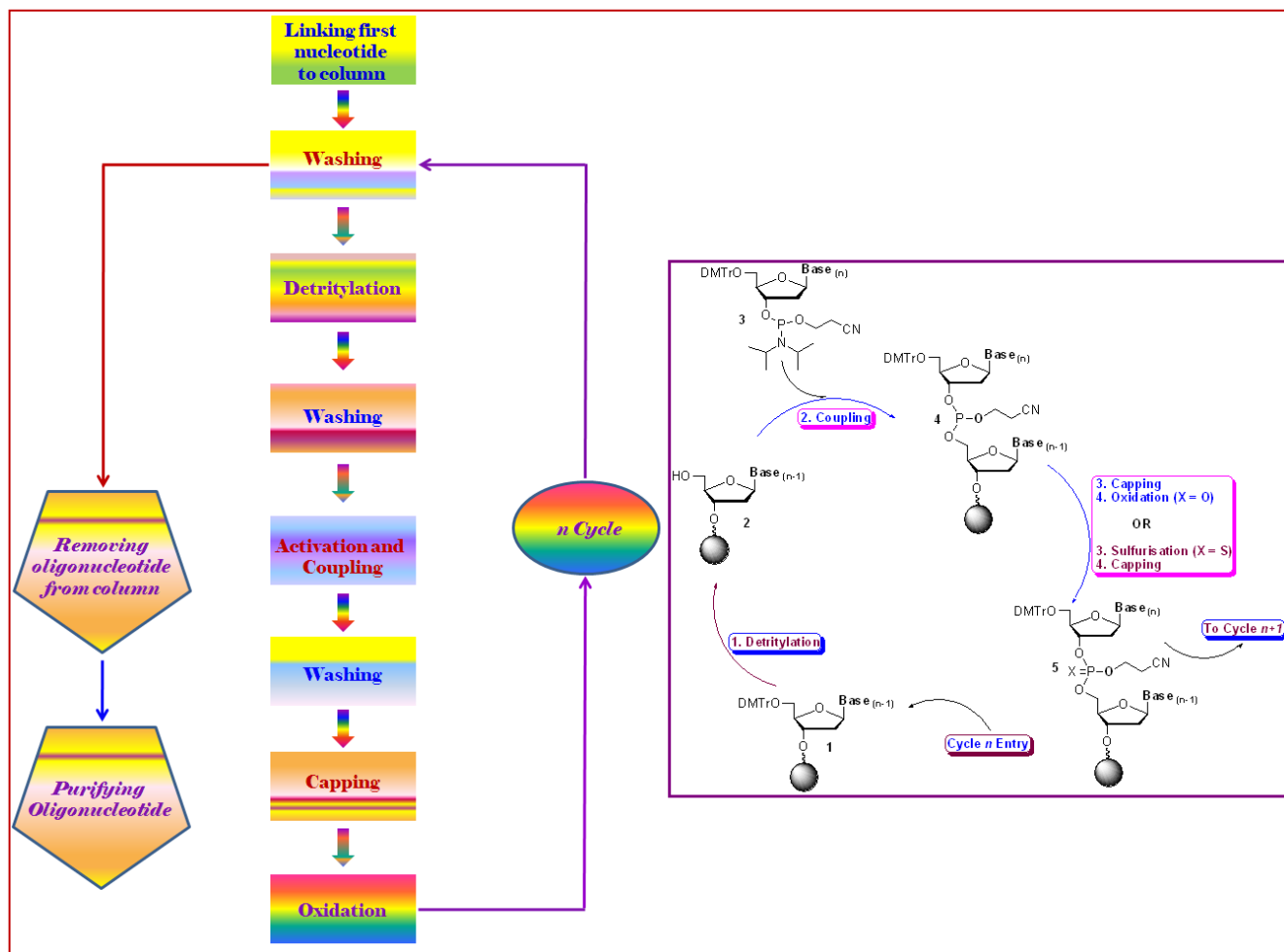


Figure 4.22: Non-nucleoside phosphoramidites for 5'-modification of synthetic oligonucleotides.

## 4.17.1.5.3. The Synthetic cycle



Scheme 4.6: Synthetic cycle for preparation of oligonucleotides by phosphoramidite method.

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. Each addition is referred to as a synthetic cycle (**Scheme 4.6**) and consists of four chemical reactions:

- **Step 1 – De-blocking (detritylation):** The DMTr group is removed with a solution of an acid, such as TCA or Dichloroacetic acid (DCA), in an inert solvent (dichloromethane or toluene) and washed out, resulting in a free 5'-terminal hydroxyl group on the solid support-bound oligonucleotide precursor.
- **Step 2 – Coupling:** A nucleoside phosphoramidite (or a mixture of several phosphoramidites) is activated by an acidic azole catalyst, 1*H*-tetrazole, 2-

ethylthiotetrazole, 2-benzylthiotetrazole, 4,5-dicyanoimidazole, or a number of similar compounds. This mixture is brought in contact with the starting solid support (first coupling) or a support-bound oligonucleotide precursor (following couplings) whose 5'-hydroxy group reacts with the activated phosphoramidite moiety of the incoming nucleoside phosphoramidite to form a phosphite triester linkage. This reaction is very rapid and requires, on small scale, about 20 s for its completion. The phosphoramidite coupling is also highly sensitive to the presence of water and is commonly carried out in anhydrous acetonitrile. Unbound reagents and by-products are removed by washing.

- **Step 3 – Capping:** After the completion of the coupling reaction, a small percentage 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 oligonucleotides with an internal base deletion commonly referred to as (n-1) short oligomers. This is done by acetylation of the unreacted 5'-hydroxy groups using a mixture of acetic anhydride and 1-methylimidazole or, less often, DMAP as catalysts. Excess reagents are removed by washing.
- **Step 4 – Oxidation:** The newly formed tricoordinated phosphite triester linkage is not natural and is of limited stability under the conditions of oligonucleotide synthesis. The treatment of the support-bound material with iodine and water in the presence of a weak base (pyridine, lutidine, or collidine) oxidizes the phosphite triester into a tetracoordinated phosphate triester, a protected precursor of the naturally occurring phosphate diester internucleosidic linkage. This step can be substituted with a sulfurization step to obtain oligonucleotide phosphorothioates. In the latter case, the sulfurization step is carried out prior to capping.

#### 4.17.1.5.4. Post-synthetic processing

After the completion of the chain assembly, the solid support-bound oligonucleotide is fully protected:

- The 5'-terminal 5'-hydroxy group is protected with DMT group;
- The internucleosidic phosphate or phosphorothioate moieties are protected with 2-cyanoethyl groups;
- The exocyclic amino groups in all nucleoside bases except for T and U are protected with acyl protecting groups.

To furnish a functional oligonucleotide, all the protecting groups have to be removed. The N-acyl base protection and the 2-cyanoethyl phosphate protection may be, and is often removed simultaneously by treatment with inorganic bases or amines. However, the applicability of this method is limited by the fact that the cleavage of 2-cyanoethyl phosphate protection gives rise to acrylonitrile as a side product. Under the strong basic conditions required for the removal of N-acyl protection, acrylonitrile is capable of alkylation of nucleic bases, primarily, at the

N3-position of thymine and uracil residues to give the respective N3-(2-cyanoethyl) adducts *via* Michael reaction. The formation of these side products may be avoided by treating the solid support-bound oligonucleotides with solutions of bases in an organic solvent, for instance, with 50% triethylamine in acetonitrile or 10% diethylamine in acetonitrile.

The solid support-bound oligonucleotides are deprotected using one of the two general approaches.

- Most often, 5'-DMT group is removed at the end of the oligonucleotide chain assembly. The oligonucleotides are then released from the solid phase and deprotected by treatment with aqueous ammonium hydroxide. This removes all remaining protection groups from 2'-deoxyoligonucleotides, resulting in a reaction mixture containing the desired product. The fully deprotected product is purified by desalting using ethanol precipitation, or size exclusion chromatography, or reverse-phase HPLC.
- The second approach is only used when the intended method of purification is reverse-phase HPLC. In this case, the 5'-terminal DMT group that serves as a hydrophobic handle for purification is kept on at the end of the synthesis. The oligonucleotide is deprotected under basic conditions as described above and, upon evaporation, is purified by reverse-phase HPLC. The collected material is then detritylated under aqueous acidic conditions. Finally, the product is desalted.
- For some applications, additional reporter groups may be attached to an oligonucleotide using a variety of post-synthetic procedures.

#### 4.17.1.5.5. Characterization

As with any other organic compound, it is prudent to characterize synthetic oligonucleotides upon their preparation. In more complex cases oligonucleotides are characterized after their deprotection and after purification. In day-by-day practice, it is sufficient to obtain the molecular mass of an oligonucleotide by recording its mass spectrum. Two methods are currently widely used for characterization of oligonucleotides: electrospray mass spectrometry (ES MS) and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF).

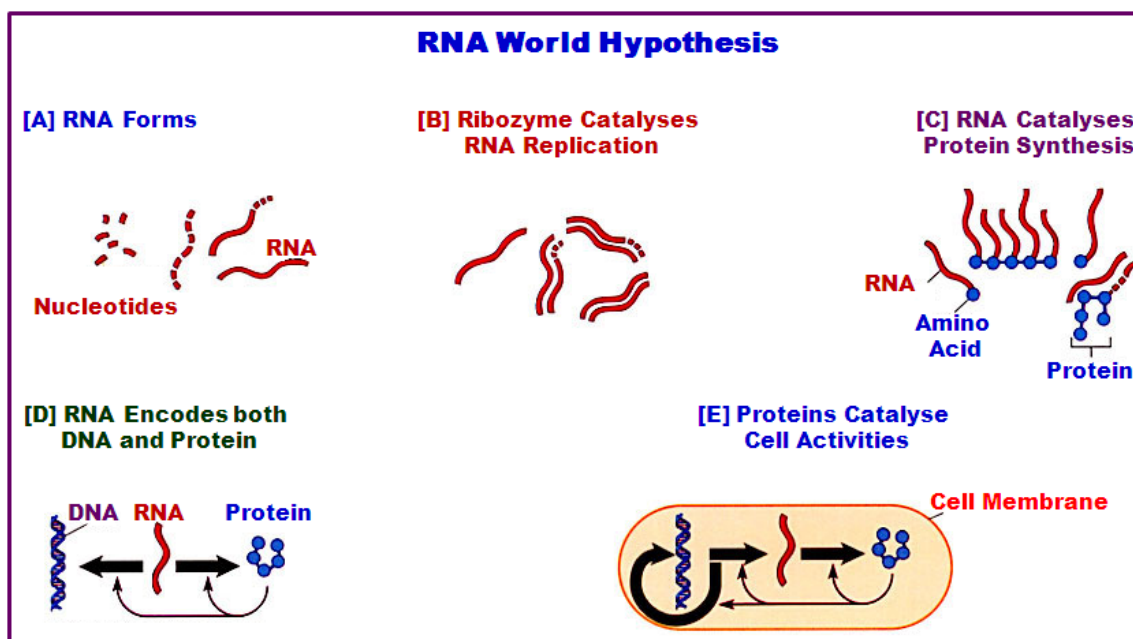


## 4.18. Catalytic RNA, siRNA, micro RNA

### 4.18.1. The RNA World

**The RNA world hypothesis** proposes that life based on ribonucleic acid (RNA) predates the current world of life based on deoxyribonucleic acid (DNA), RNA and protein. RNA is able to both store genetic information, like DNA, and catalyzes chemical reactions, like an enzyme protein. It may therefore have supported pre-cellular life and been the first step in the evolution of cellular life. The RNA world is proposed to have evolved into the DNA, RNA and protein world of today. DNA is thought to have taken over the role of data storage due to its increased stability, while proteins, through a greater variety of monomers (amino acids), replaced RNA's role in specialized biocatalysis. The RNA world hypothesis suggests that RNA in modern cells is an evolutionary remnant of the RNA world.

- First proto-organisms had both genes and enzymes made of RNA
- DNA and proteins as specialized late-comers
- DNA took over information storage
- Proteins took over enzyme activity



### 4.18.2. Central Role of RNA

#### 1. The Three Roles of RNA in Protein Synthesis

Although DNA stores the information for protein synthesis and RNA carries out the instructions encoded in DNA, most biological activities are carried out by proteins. The accurate synthesis of proteins thus is critical to the proper functioning of cells and organisms. We saw in Chapter 3 that

the linear order of amino acids in each protein determines its three-dimensional structure and activity. For this reason, assembly of amino acids in their correct order, as encoded in DNA, is the key to production of functional proteins.

Three kinds of RNA molecules perform different but cooperative functions in protein synthesis (Figure 4.24):

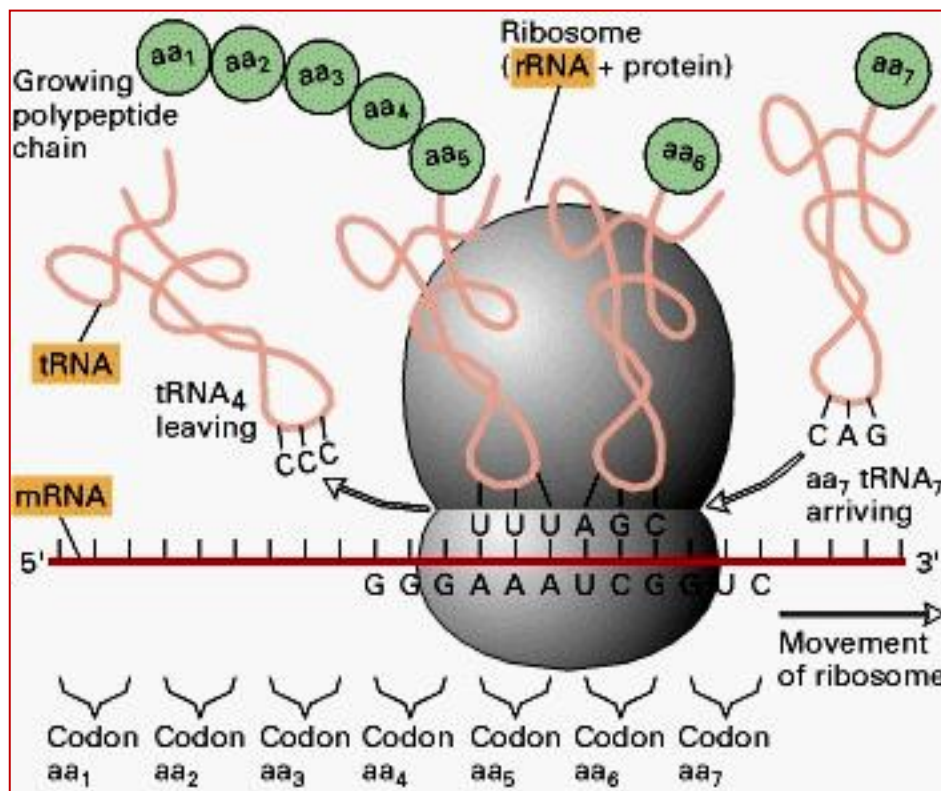


Figure 4.24: The three roles of RNA in protein synthesis

- Messenger RNA (mRNA) carries the genetic information copied from DNA in the form of a series of three-base code “words,” each of which specifies a particular amino acid.
- Transfer RNA (tRNA) is the key to deciphering the code words in mRNA. Each type of amino acid has its own type of tRNA, which binds it and carries it to the growing end of a polypeptide chain if the next code word on mRNA calls for it. The correct tRNA with its attached amino acid is selected at each step because each specific tRNA molecule contains three-base sequences that can base-pair with its complementary code word in the mRNA.
- Ribosomal RNA (rRNA) associates with a set of proteins to form ribosome. These complex structures, which physically move along an mRNA molecule, catalyze the assembly of amino acids into protein chains. They also bind tRNAs and various accessory molecules necessary for protein synthesis. Ribosomes are

composed of a large and small subunit, each of which contains its own rRNA molecule or molecules.

(d) Translation is the whole process by which the base sequence of an mRNA is used to order and to join the amino acids in a protein. The three types of RNA participate in this essential protein-synthesizing pathway in all cells; in fact, the development of the three distinct functions of RNA was probably the molecular key to the origin of life.

2. RNA viruses and viroids have genomes made of RNA
3. RNA polymerase can begin new strands
  - a. But DNA polymerase cannot start new strands
  - b. Primers made of RNA must be used to start new strands of DNA
4. Small RNA guide molecules are used for:
  - a. Removal of introns
  - b. Modification and editing of messenger RNA
  - c. Extending ends of eukaryotic chromosomes by telomerase
5. Ribozymes are RNA molecules that act as enzymes:
  - a. Most important is large ribosomal RNA
  - b. (23S rRNA in bacteria, 28S rRNA in eukaryotes)
  - c. This links the amino acids together to form the polypeptide chain
6. RNA may directly control gene expression
  - a. Small regulatory RNA may control transcription and translation
  - b. Riboswitches bind small molecules
7. RNA interference:
  - a. Novel mechanism - purpose is to destroy hostile virus RNA
  - b. Widely used in research to inactivate genes in animals and plants
  - c. Not found in bacteria

#### 4.18.2.1. RNA Enzymes - Ribozymes

**A. Definition:** Ribozyme means ribonucleic acid enzyme. A ribozyme is an RNA molecule with a well defined tertiary structure that enables it to catalyze a chemical reaction. It may also be called an RNA enzyme or catalytic RNA. Many natural ribozymes catalyze either the hydrolysis of one of their own phosphodiester bonds (self-cleaving ribozymes), or the hydrolysis of bonds in other RNAs. Some have been found to catalyze the aminotransferase activity of the ribosome. Examples of ribozymes include the hammerhead ribozyme, the VS ribozyme and the hairpin ribozyme.

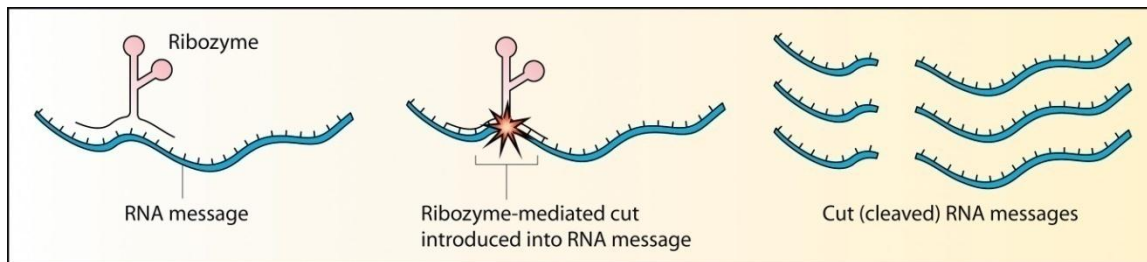
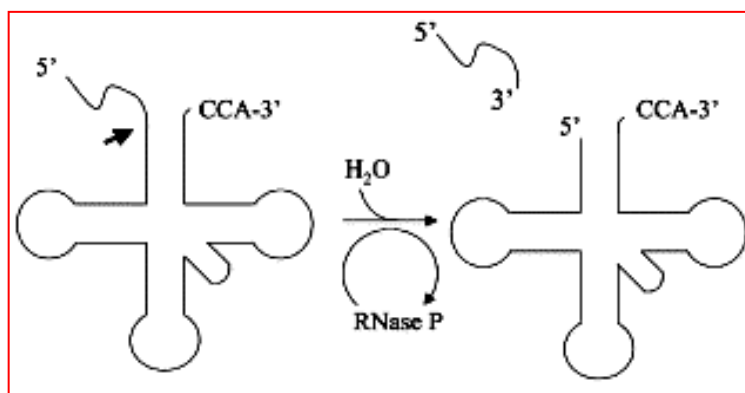


Figure 4.24: Cleavage of RNA by ribozyme

**B. Examples include:**

1. Ribosomal RNA of the large subunit of the ribosome
  - a. Links amino acids in growing polypeptide chain
  - b. Ribosomal proteins help orient the reaction components but are not catalytically active
  - c. (Previously thought that ribosomal proteins carried out the reactions and that rRNA was a scaffold to held proteins in the correct position.)
2. Ribonuclease P (a less complex ribozyme)
  - a. Processes certain transfer RNA molecules at 5'-end
  - b. Has both RNA and protein component
    - i. the RNA carries out the reaction
    - ii. the protein helps bind the tRNA
3. RNA component works slowly on its own, especially if tRNA is plentiful.



4. Self-splicing introns
  - a. They are catalytic but only act once
  - b. Technically they are not true enzymes
  - c. Genuine enzymes process large numbers of other molecules as opposed to processing themselves just once
5. Viroids (Infectious RNA molecules that infect plants)
  - a. Do not encode any proteins
  - b. RNA of some viroids carries out self-cleavage during replication
  - c. Not true enzymes - catalytic but only act once.

**C. Function:**

- a. Although most ribozymes are quite rare in the cell, their roles are sometimes essential to life. For example, the functional part of the ribosome, the molecular machine that translates RNA into proteins, is fundamentally a ribozyme, composed of RNA tertiary structural motifs that are often coordinated to metal ions such as  $Mg^{2+}$  as cofactors.
- b. RNA can also act as a hereditary molecule, which encouraged Walter Gilbert to propose that in the distant past, the cell used RNA as both the genetic material and the structural and catalytic molecule, rather than dividing these functions between DNA and protein as they are today. This hypothesis became known as the "RNA world hypothesis" of the origin of life.
- c. If ribozymes were the first molecular machines used by early life, then today's remaining ribozymes -- such as the ribosome machinery -- could be considered living fossils of a life based primarily on nucleic acids.
- d. A recent test-tube study of prion folding suggests that RNA may catalyze the pathological protein conformation in the manner of a chaperone enzyme.

**D. Naturally occurring ribozymes include:**

- a. Peptidyl transferase 23S rRNA
- b. RNase P
- c. Group I and Group II introns
- d. GIR1 branching ribozyme
- e. Leadzyme - Although initially created *in vitro*, natural examples have been found
- f. Hairpin ribozyme
- g. Hammerhead ribozyme
- h. HDV ribozyme
- i. Mammalian CPEB3 ribozyme
- j. VS ribozyme
- k. *glmS* ribozyme
- l. CoTC ribozyme

**E. Artificial Ribozymes**

- a. Since the discovery of ribozymes that exist in living organisms, there has been interest in the study of new synthetic ribozymes made in the laboratory. For example, artificially-produced self-cleaving RNAs those have good enzymatic activity have been produced. Tang and Breaker isolated self-cleaving RNAs by *in vitro* selection of RNAs originating from random-sequence RNAs. Some

of the synthetic ribozymes that were produced had novel structures, while some were similar to the naturally occurring hammerhead ribozyme.

- b. The techniques used to discover artificial ribozymes involve Darwinian evolution. This approach takes advantage of RNA's dual nature as both a catalyst and an informational polymer, making it easy for an investigator to produce vast populations of RNA catalysts using polymerase enzymes. The ribozymes are mutated by reverse transcribing them with reverse transcriptase into various cDNA and amplified with mutagenic PCR. The selection parameters in these experiments often differ. One approach for selecting a ligase ribozyme involves using biotin tags, which are covalently linked to the substrate. If a molecule possesses the desired ligase activity, a streptavidin matrix can be used to recover the active molecules.
- c. Lincoln and Joyce developed an RNA enzyme system capable of self replication in about an hour. By utilizing *in vitro* evolution of a candidate enzyme mixture, a pair of RNA enzymes emerged, in which each synthesizes the other from synthetic oligonucleotides, with no protein present.

#### F. Applications of Ribozymes

- a. Catalytic RNAs (ribozymes) are capable of specifically cleaving RNA molecules, a property that enables them to act as potential antiviral and anti-cancer agents, as well as powerful tools for functional genomic studies.
- b. Recently, ribozymes have been used successfully to inhibit gene expression in a variety of biological systems *in vitro* and *in vivo*.
- c. Phase I clinical trials using ribozyme gene therapy to treat AIDS patients have been conducted.
- d. A type of synthetic ribozyme directed against HIV RNA called gene shears has been developed and has entered clinical testing for HIV infection.

### 4.18.3. RNA Interference (RNAi)

#### 4.18.3.1. Introduction

- **Definition:** Normal cells contain double-stranded DNA and single-stranded RNA but not double-stranded RNA (dsRNA).
- **RNA interference (RNAi)**, is a technique in which exogenous, double-stranded RNAs (dsRNAs) that are complimentary to known mRNAs, are introduced into a cell to specifically destroy that particular mRNA, thereby diminishing or abolishing gene expression.



- This technology considerably bolsters functional genomics to aid in the identification of novel genes involved in disease processes and thus can be used for medicament and for delivery as therapeutics.
- RNA interference was known by other names, including co-suppression, post transcriptional gene silencing and quelling. Only after these apparently unrelated processes were fully understood did it become clear that they all described the RNAi phenomenon.
- Nomination of RNAi as the “Breakthrough of the year 2002” by the Journal *Science* prompted biologists to overhaul their vision of the cell and its evolution, and discovery of RNAi was awarded Nobel Prize to Andrew Z. Fire and Craig C. Mello in 2006 in Physiology or Medicine for their work on RNA interference in the nematode worm *C. elegans*, which they published in 1998.

#### 4.18.3.1. Effectors RNA molecules

RNAi pathways are guided by small RNAs that include:

- SiRNA:** Small interfering RNA (siRNA), sometimes known as short interfering RNA or silencing RNA, is a class of 20-25 nucleotide-long double-stranded RNA molecules.

SiRNAs can also be exogenously (artificially) introduced into cells by various transfection methods to bring about the specific knockdown of a gene of interest.

- miRNA:** microRNAs (miRNA) are single-stranded RNA molecules of about 21–23 nucleotides in length, which regulate gene expression. miRNAs are encoded by genes from whose DNA they are transcribed but miRNAs are not translated into protein (non-coding RNA); instead each primary transcript (a pri-miRNA) is processed into a short stem-loop structure called a pre-miRNA and finally into a functional miRNA. Mature miRNA molecules are partially complementary to one or more messenger RNA (mRNA) molecules, and their main function is to down-regulate gene expression.
- shRNA:** A small hairpin RNA or short hairpin RNA (shRNA) is a sequence of RNA that makes a tight hairpin turn that can be used to silence gene expression via RNA interference. shRNA uses a vector introduced into cells and utilizes the U6 promoter to ensure that the shRNA is always expressed. This vector is usually passed on to daughter cells, allowing the gene silencing to be inherited. The shRNA hairpin structure is cleaved by the cellular machinery into siRNA, which is then bound to the RNA-induced silencing complex (RISC). This complex binds to and cleaves mRNAs which match the siRNA that is bound to it.



- d. Others:** In addition to miRNAs and siRNAs, other innate RNAi effectors have been identified. One class of these is the Piwi-interacting RNAs (piRNAs). piRNAs seem to be uniquely expressed in the mammalian germline, particularly in the testes. The functional role of piRNAs is currently unclear, but a role in spermatogenesis is likely.

A number of other small RNAs associated with RNAi have been identified in different species, including trans-activating siRNAs (tasiRNAs), studied in plants and nematodes, and small scan RNAs (ScnRNAs), found in *Tetrahymena*.

#### 4.18.3.1. Mechanism of RNA interference (RNAi)

Long double-stranded RNAs (dsRNAs; >200 nt) can be used to silence the expression of target genes in a variety of organisms and cell types (e.g., worms, fruit flies, and plants). Upon introduction, the long dsRNAs enter a cellular pathway that is commonly referred to as the RNA interference (RNAi) pathway. First, the dsRNAs get processed into 20-25 nucleotide (nt) small interfering RNAs (siRNAs) by an RNase III-like enzyme called Dicer (initiation step). Then, the siRNAs assemble into endoribonuclease-containing complexes known as RNA-induced silencing complexes (RISCs), unwinding in the process. The siRNA strands subsequently guide the RISCs to complementary RNA molecules, where they cleave and destroy the cognate RNA (effector step). Cleavage of cognate RNA takes place near the middle of the region bound by the siRNA strand.

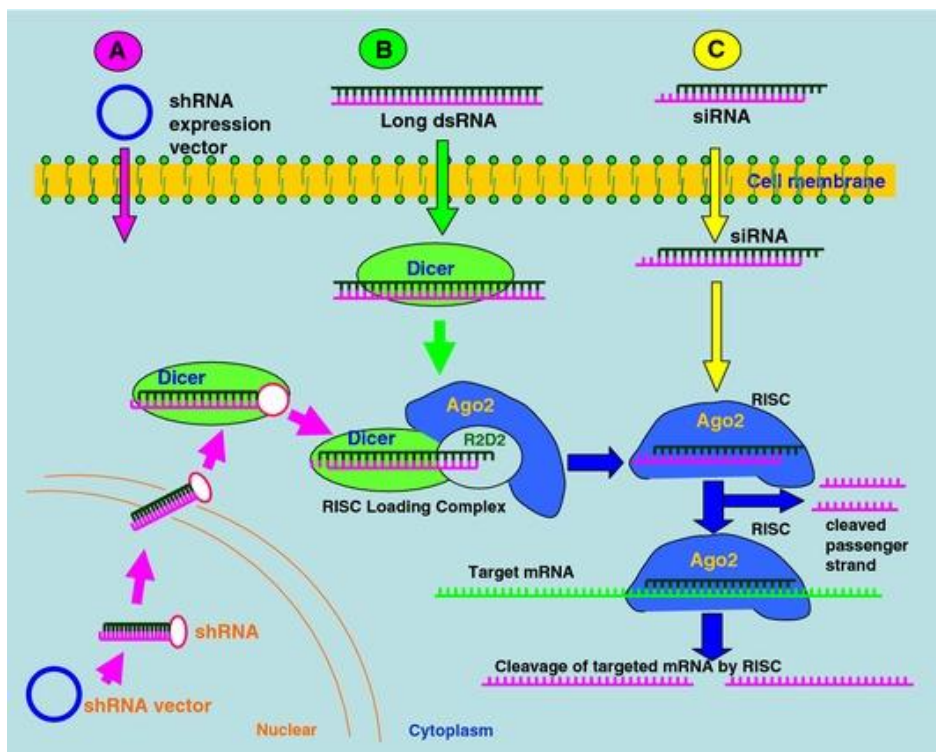


Figure 4.24: Mechanism of RNA interference (RNAi).

In mammalian cells, introduction of long dsRNA (>30 nt) initiates a potent antiviral response, exemplified by nonspecific inhibition of protein synthesis and RNA degradation. The mammalian antiviral response can be bypassed, however, by the introduction or expression of siRNAs.

**Natural role of RNAi is defending against RNA viruses**

- a. RNA viruses replicate via dsRNA intermediate (replicative form)
- b. Applies to RNA virus whether contains ssRNA or dsRNA
- c. dsRNA is seen as a sign of infection and triggers anti-viral response

**RNA interference found in:**

- a. Most eukaryotes - protozoa, invertebrates
- b. Mammals - weaker
- c. Plants – especially effective
- d. NOT found in prokaryotes
- e. [In bacteria, ribonuclease III rapidly degrades dsRNA molecules]

**4.18.3.2. Applications of in vivo siRNA delivery in disease models**

- a. RNAi technology is proving to be useful to analyze quickly the functions of a number of genes in a wide variety of organisms.
- b. RNAi technology has been successfully applied to identify genes with essential roles in biochemical signaling cascades, embryonic development, and other basic cellular process.
- c. In plants, gene knockdown-related functional studies are being carried out efficiently when transgenes are present in the form of hairpin (or RNAi) constructs. Plant endotoxins could also be removed if the toxin biosynthesis genes are targeted with the RNAi constructs, like theobromine synthase of the coffee plant was knocked down with the hairpin construct of the transgene, leading to the production of decaffeinated coffee plants.
- d. RNAi may facilitate drug screening and development by identifying genes that can confer drug resistance or genes whose mutant phenotypes are ameliorated by drug treatment, providing information about the modes of action of novel compounds.
- e. It may also be a method of choice to study the simultaneous functions of a number of analogous genes in organisms in which redundancy exists with respect to a particular function, because many of these genes can be silenced simultaneously.
- f. siRNAs have been shown to inhibit infection by human immunodeficiency virus, poliovirus, and hepatitis C virus in cultured cell lines. siRNAs can successfully be used to silence genes expressed in

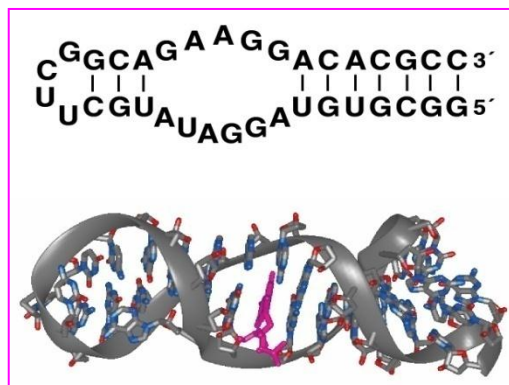
respiratory syncytial virus, an RNA virus that causes severe respiratory disease in neonates and infants. siRNA treatment has also been shown to reduce the expression of the BCR-ABL oncoprotein in leukemia and lymphoma cell lines, leading to apoptosis in these cells. With respect to future medical applications, siRNA-based therapy seems to have a great potential to combat carcinomas, myeloma, and cancer caused by over expression of an oncoprotein or generation of an oncoprotein by chromosomal translocation and point mutations.

- g. **Medicine:** RNA interference also used in therapy. It is difficult to introduce long dsRNA strands into mammalian cells due to the interferon response; the use of short interfering RNA mimics has been used for medical purpose. Among the first applications to reach clinical trials were in the treatment of macular degeneration and respiratory syncytial virus. RNAi has also been shown to be effective in the reversal of induced liver failure in mouse models.
- h. **Other probable clinical uses:** (i) antiviral therapies, like the inhibition of viral gene expression in cancerous cells, (ii) in co-receptors for HIV, (iii) in silencing of hepatitis A and hepatitis B genes (iv) inhibition of measles viral replication, (v) in treatments for neurodegenerative diseases have also been proposed, with particular attention being paid to the polyglutamine diseases such as Huntington's disease, (vi) gene therapy, (vii) in biotechnology.

#### 4.18.4. Riboswitches – mRNA Controls Itself

##### 4.18.4.1. Introduction

A riboswitch is a part of an mRNA molecule that can directly bind a small target molecule, and whose binding of the target affects the gene's activity. Thus, an mRNA that contains a riboswitch is directly involved in regulating its own activity, in response to the concentrations of its target molecule. The discovery that modern organisms use RNA to bind small molecules, and discriminate against closely related analogs, significantly expanded the known natural capabilities of RNA beyond its ability to code for proteins or to bind other RNA or protein macromolecules.



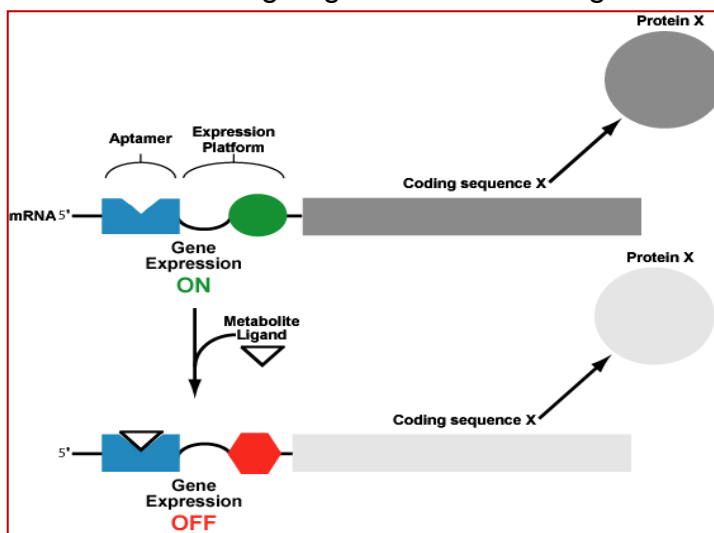
##### 4.18.4.2. RNA aptamers

- RNA aptamers are structures that bind specifically to target ligands

- Many aptamers have been generated in the laboratory for a wide range of target molecules (e.g. theophylline aptamer)
- Structural studies of aptamer-ligand complexes have provided a wealth of information regarding RNA structure and ligand interaction
- Most aptamers exhibit conformational changes upon binding to ligand (induced fit binding)

#### 4.18.4.3. Natural aptamers are the basis of riboswitches:

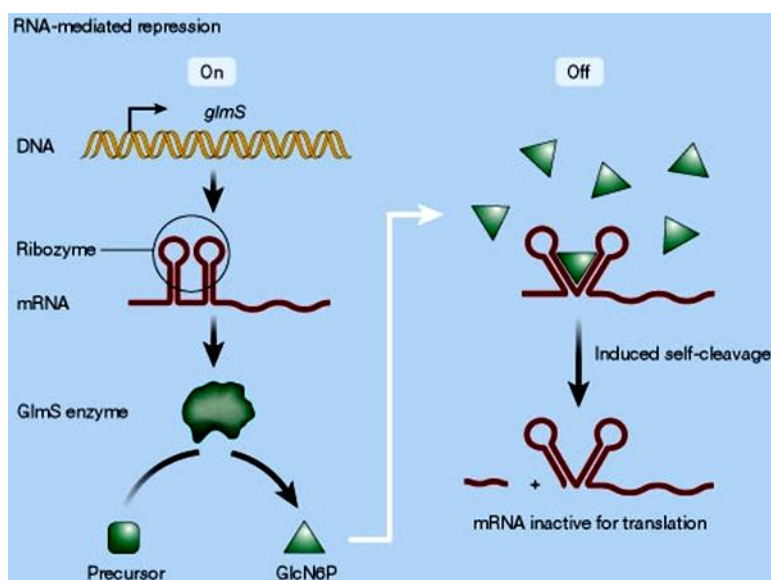
- Natural aptamers reside in the non-coding segments of messenger RNAs (mainly prokaryotic)
- Interact directly (without the need for protein) with metabolites to control gene expression
- Typically provide a feedback mechanism for controlling the expression of metabolic genes
  - metabolic product of a pathway inhibits expression of proteins required to produce the metabolite
- Most riboswitches are comprised of the aptamer and an 'expression platform' (some way of altering gene expression)



#### 4.18.4.4. Mechanisms of riboswitches

Riboswitches are often conceptually divided into two parts: an Aptamer and an expression platform. The aptamer directly binds the small molecule, and the expression platform undergoes structural changes in response to the changes in the aptamer. The expression platform is what regulates gene expression.

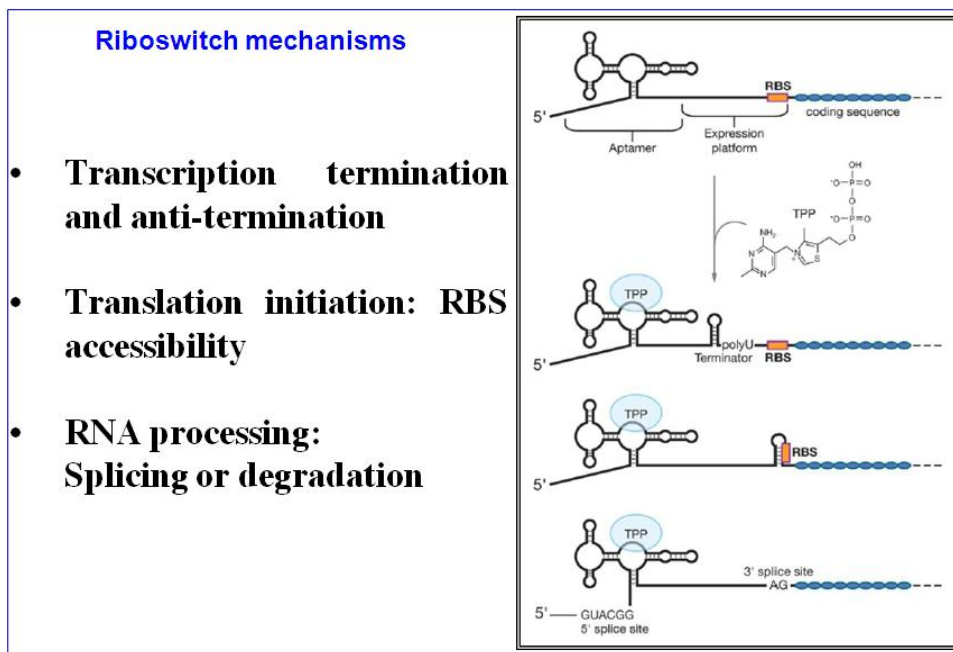
Expression platforms typically turn off gene expression in response to the small molecule, but some turn it on. The



following riboswitch mechanisms have been experimentally demonstrated.

- Riboswitch-controlled formation of rho-independent transcription termination hairpins leads to premature transcription termination.
- Riboswitch-mediated folding sequesters the ribosome-binding site, thereby inhibiting translation.
- The riboswitch is a ribozyme that cleaves itself in the presence of sufficient concentrations of its metabolite.
- Riboswitch alternate structures affect the splicing of the pre-mRNA.
- A riboswitch in *Clostridium acetobutylicum* regulates an adjacent gene that is not part of the same mRNA transcript.
- A riboswitch in *Listeria monocytogenes* regulates the expression of its downstream gene.

#### 4.18.4.4. Genetic Control Mechanisms - RIBOSWITCHES

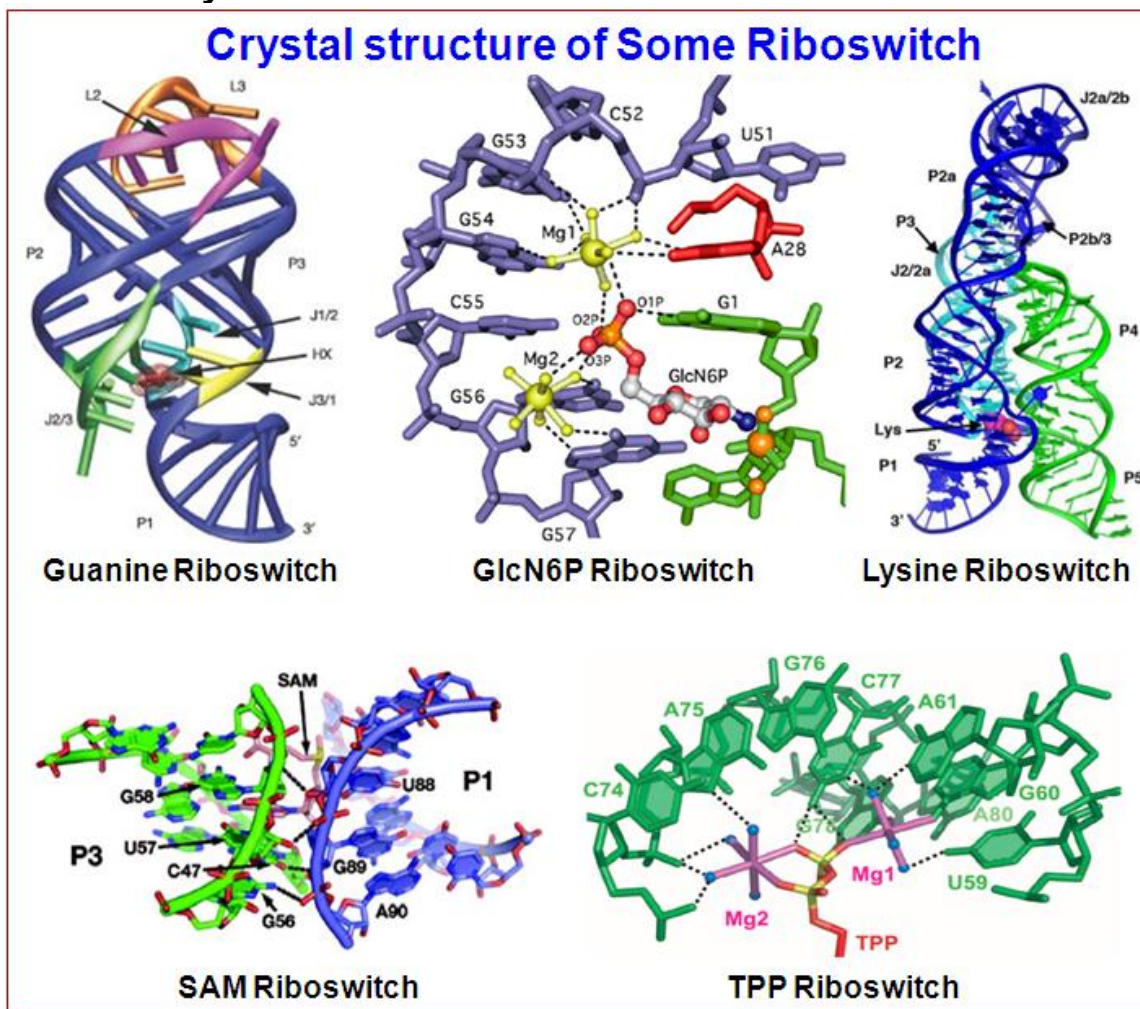


#### 4.18.4.5. Other Functional Mechanism

(a) Transcription Control termination and anti-termination, (b) Translation initiation: RBS accessibility, (c) RNA processing: Splicing or degradation



#### 4.18.4.6. Crystal structure of Some Riboswitch



#### 4.18.4.7. Types of Riboswitches Known

The following is a list of experimentally validated riboswitches, organized by ligand.

- **Cobalamin riboswitch** → regulate cobalamin biosynthesis and transport of cobalamin and similar metabolites, and other genes.
- **cyclic di-GMP riboswitches** → bind the signaling molecule cyclic di-GMP in order to regulate a variety of genes controlled by this second messenger.
- **FMN riboswitch** → regulate riboflavin biosynthesis and transport.
- **glmS riboswitch** → is a ribozyme that cleaves itself.
- **Glutamine riboswitches** → bind glutamine to regulate genes involved in glutamine and nitrogen metabolism, as well as short peptides of unknown function.

- **Glycine riboswitch** → binds glycine to regulate glycine metabolism genes, including the use of glycine as an energy source.
- **Lysine riboswitch** → binds lysine to regulate lysine biosynthesis, catabolism and transport.
- **PreQ1 riboswitches** → bind pre-queuosine<sub>1</sub>, to regulate genes involved in the synthesis or transport of this precursor to queuosine.
- **Purine riboswitches** → binds purines to regulate purine metabolism and transport.
- **SAH riboswitches** → bind S-adenosylhomocysteine to regulate genes involved in recycling this metabolite that is produced when S-adenosylmethionine is used in methylation reactions.
- **SAM riboswitches** → bind S-adenosyl methionine (SAM) to regulate methionine and SAM biosynthesis and transport. Three distinct SAM riboswitches are known: *SAM-I* (originally called *S-box*), *SAM-II* and the *S<sub>MK</sub> box riboswitch*.
- **SAM-SAH riboswitches** → bind both SAM and SAH with similar affinities. Since they are always found in a position to regulate genes encoding methionine adenosyltransferase, it was proposed that only their binding to SAM is physiologically relevant.
- **Tetrahydrofolate riboswitches** → bind tetrahydrofolate to regulate synthesis and transport genes.
- **TPP riboswitches** → binds thiamin pyrophosphate (TPP) to regulate thiamin biosynthesis and transport, as well as transport of similar metabolites. It is the only riboswitch found so far in eukaryotes.



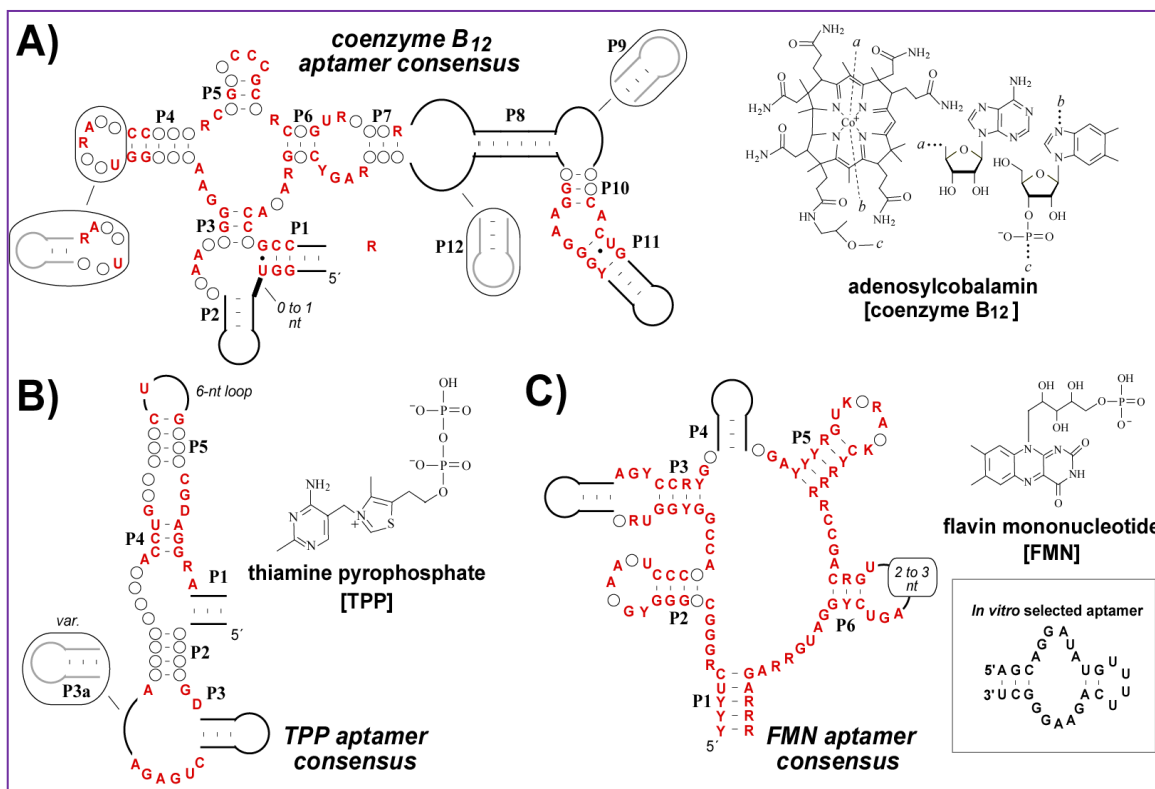


Figure 4.25a: Some riboswitches.

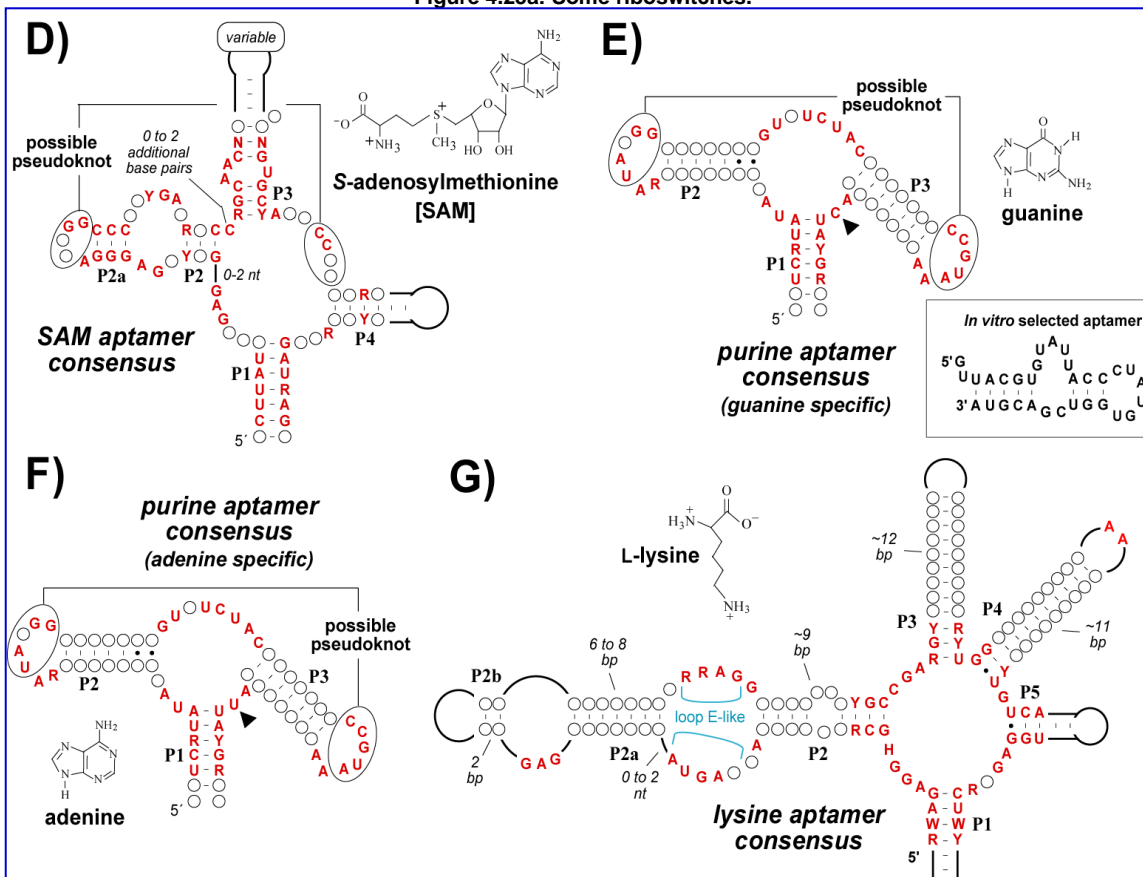


Figure 4.25a: Some riboswitches.

#### 4.18.4.8. Riboswitches and the RNA World hypothesis

Riboswitches demonstrate that naturally occurring RNA can bind small molecules specifically, a capability that many previously believed was the domain of proteins or artificially constructed RNAs called aptamers. The existence of riboswitches in all domains of life therefore adds some support to the RNA world hypothesis, which holds that life originally existed using only RNA, and proteins came later; this hypothesis requires that all critical functions performed by proteins (including small molecule binding) could be performed by RNA. It has been suggested that some riboswitches might represent ancient regulatory systems, or even remnants of RNA-world ribozymes whose binding domains are conserved.

#### 4.18.4.9. Riboswitches as antibiotic targets

Riboswitches could be a target for novel antibiotics. Indeed, some antibiotics whose mechanism of action was unknown for decades have been shown to operate by targeting riboswitches. For example, when the antibiotic pyrithiamine enters the cell, it is metabolized into pyrithiamine pyrophosphate. Pyrithiamine pyrophosphate has been shown to bind and activate the TPP riboswitch, causing the cell to cease the synthesis and import of TPP. Because pyrithiamine pyrophosphate does not substitute for TPP as a coenzyme, the cell dies.

### 4.19. Expanding the Genetic Alphabets: Background

DNA is an essential biomolecule which is responsible for encoding the complex information necessary for life. Specific pairing of dA with dT and dC with dG in duplex DNA and during polymerase-mediated replication is the basis of the genetic alphabet, itself the basis of the genetic code. However, there is no reason to assume that the requirements for duplex stability and replication must limit the genetic alphabet to only two base pairs. In addition to enabling a wide variety of biotechnology applications, an expanded genetic alphabet would enable the encoding of additional information for both *in vitro* and *in vivo* applications. Expansion of the genetic alphabet to include a third base pair, formed between two identical or different unnatural nucleotides, referred to as self-pairs and hetero pairs, respectively, would expand the informational and functional potential of DNA such as site directed oligonucleotide labeling and *in vitro* selections with oligonucleotides bearing increased chemical diversity. The use of the efficient new base pair/pairs to drive the synthesis of proteins with unnatural base pairs is also a current exciting research area to consider. Thus, translation of an expanded genetic alphabet into an expanded genetic code, provocatively, even lead to the assembly of such a system within a living cell, potentially creating a semi-synthetic organism and life with increased diversity.

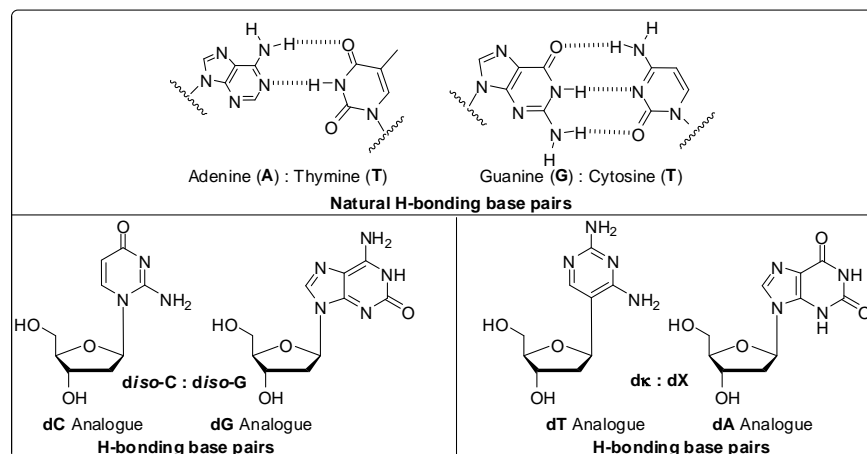


Figure 4.26: Presentation of H-bonding base pairs among natural nucleosides and analogues

After Benner first popularized the idea, an increasing amount of interest has resulted in the acceleration of progress towards this goal. Expanding the genetic alphabet requires an unnatural base pair with inter base interactions, of whatever sort, that confer stability on a DNA duplex and that is replicated by a DNA polymerase. Specifically, each unnatural triphosphate must be efficiently and selectively incorporated opposite its partner in the template to form a stable base pair. Conversely, no natural substrate should be inserted opposite the unnatural nucleotide in the template with high efficiency. Also, continued synthesis past the unnatural base pair must be efficient. Thus, the efforts toward developing a third base pair have focused on nucleobase analogues designed to pair *via* orthogonal hydrogen bonding (H-bonding, **Figure 1**), based on work of the Benner group, and more recently on predominantly non-H-bonding (**Figure 2**) analogues that pair *via* hydrophobic interactions, based on work of the Kool group.

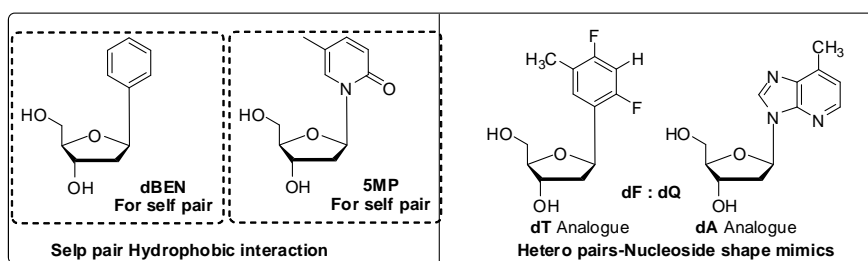


Figure 4.27: Presentation of non H-bonding base pairs among nucleoside base analogues

However, the unnatural base pairs so far have been reported has several shortcomings, including tautomerization of *iso-G* and poor recognition of *iso-C* by RNA polymerases; these shortcomings pose difficulties for mRNA preparation. Novel, hydrophobic base pairs have been developed recently, but their use in transcription is still under investigation. Thus there is a need to develop conceptually new and novel base analogues which will be recognized by both replication and transcription process with high efficiency.

### 4.19.1. Synthesis and Application of Unnatural Nucleosides

Virtually all modern molecular biology techniques require the amplification of DNA by PCR. Thus, an unnatural base pair that is compatible with both PCR amplification and *in vitro* transcription, expanding the genetic alphabet, at least in a test tube, from just two base pairs to three. It will be worthwhile to design such unnatural pair for which both efficient PCR amplification and transcription by RNA polymerase would be efficient. In addition to the incorporation of new base pairs into DNA or RNA, the unnatural base that would allow site-specific labeling of RNA will be of great importance for a variety of biotechnology applications.<sup>1</sup> For example, the ability to site-selectively modify a DNA or an RNA molecule with a fluorophore attached to the unnatural base pair should facilitate applications in both cell biology and biophysics. The third base pair should also be useful for *in vitro* selection methodologies that have already produced DNA and RNA molecules with desired properties, such as selective recognition of other molecules (aptamers) or catalysis (ribozymes). Increasing diversity of these modified DNA and RNA molecules, promises their enhanced widespread potential applications in biomedical sciences such as drug candidates.

### 4.19.2. Why Unnatural Nucleotide Bases?

The natural bases only differ in their organic structure and functional group only belongs to two different heterocyclic families. Basically they do not differ much in their properties in wider sense like in stacking ability, size, sensitivity in protonation in neutral pH etc. They only absorb light in ultraviolet region and not responsive to the visible light or can not be detected by highly sensitive fluorescence detection techniques as the intrinsic fluorescence of the naturally occurring nucleotide bases in DNA and RNA is extremely weak. These bases exhibit very short fluorescent decay times, generally in the range of a few picoseconds, and do not provide much structural information since signals are normally averaged over all bases in the oligonucleotide sequence. Although the major nucleotide bases are essentially non-fluorescent, there do exist a few naturally occurring fluorescent bases, such as wyosine (Yt) which is found in the anticodon region of tRNAPhe. The rarity of such naturally occurring fluorescent bases, however, has limited their use in fluorescence studies. Thus, in contrast to proteins, which may contain one or more naturally occurring tryptophan or tyrosine residues that can be exploited for fluorescence measurements, RNA and DNA molecules in general lack naturally occurring intrinsic fluorescent reporters.

Thus, the lack of naturally occurring fluorescent bases has spurred the development of new nucleosides which will be the probe for DNA analysis. These nucleosides can be so design as to get optimized fluorescence properties (i.e. higher quantum yields and longer lifetimes) and that can be incorporated into oligonucleotides using standard automated synthetic methods. The fluorescence properties, from DNA and RNA molecules, thus can be observed without any competing background signals mainly by two ways: (a) incorporation of unnatural nucleotide base analogs into the oligonucleotide sequence and/or (b)

incorporation of fluorophore in to the natural bases i.e. by synthesizing labeled bases.

#### **4.19.3. Designing Criteria for Modification of Nucleosides**

For probing DNA structure, the ideal fluorescent nucleoside:

- (a) Should have bright fluorescence, which is sensitive to its environment, and a large Stokes shift;
- (b) Should be amenable to phosphoramidite preparation for incorporation into oligo-nucleotides by solid-phase synthesis;
- (c) Should not disrupt duplex formation and should mimic one of the regular nucleosides; In general, maintaining Watson-Crick and Hoogsteen base pairing hydrogen bonding is an important aspect in the design and synthesis of fluorescent nucleotide base analogs.
- (d) Should behave as a regular nucleoside in its interaction with proteins and enzymes; and
- (e) Should be capable of being converted to the triphosphate and be incorporated into DNA with high efficiency by current commercial polymerases.

Thus, considering the steric hypothesis, shape complementarity and hydrophobic stabilization several base analogues for DNA replication have been reported. Researchers are hopeful to translate an expanded genetic alphabet into an expanded genetic code. Provocatively, this effort may lead to the assembly of such a system within a living cell, potentially creating a semi-synthetic organism and life with increased diversity.

#### **4.19.4. Design of Unnatural Base Pairs**

Therefore, the major efforts are focused on the development of a stable third base-pair that would be efficiently replicated with high fidelity. Recent efforts have resulted in designing and construction of a number of such base-pairs that are stable within the DNA duplex. However, to date, because of the challenging problem of enzymatic replication of such base pairs, only very few of these artificial base-pairs have been efficiently and selectively replicated. The base-pairs formed are of following categories: (a) unnatural hydrogen-bonding pattern as well as upon the shape complementarity, (b) hydrophobic forces, (c) metal-bridged base-pairing, (d) and even covalent cross-linking.

#### 4.19.4.1 Artificial Base-Pairs Based on Hydrogen Bonding

##### 4.19.4.1.1. Modification of Heterocyclic (Purine or Pyrimidine) Bases

Specific base-pairing between nucleic acids is essential to the accurate duplication and expression of genetic information. It is made possible, in part, by unique hydrogen bonding complementarity, which depends critically on the tautomeric states of the bases. Tautomerization reverses the polarity by interconverting H-bond donors and acceptors, and the different electronegativities of oxygen and nitrogen exocyclic substituents are consistent with observed equilibrium constants of ca  $10^4$  in favour of the amino tautomers of A and C, and the keto forms of G and T. Modification of these groups though, perturbs the equilibrium, it is made possible by suitable design and with the help of synthetic chemistry. Thus, the compounds, which are mutagenic *in vivo*, can function as analogues of both A, G, C and T, depending on the opportunities for H-bonding available in a given environment. The concept of hydrogen-bonding patterns and shape complementarity was pioneered by Benner (e.g., isoguanosine, *iso-G*; isocytidine, *iso-C*; **Figure 4.28**). Further development has led to the introduction of other donor-acceptor (**D-A**) purine-pyrimidine pairs and finally to a generalization of the Watson-Crick nucleobases pairs.

A number of nucleotide base analog probes are now available, some even commercially, for incorporation into oligonucleotides for biophysical and biochemical studies. These nucleotide base probes share common chemical features, like faithful mimicry of native base structure and optimization of fluorescence quantum yields and lifetimes. Each base analog also exhibits some unique fluorescence, structural or chemical properties, which should be considered when deciding on the optimal probe for use in studies of a particular nucleic acid system. We are summarizing below some of the nucleotide base analogs available to date.

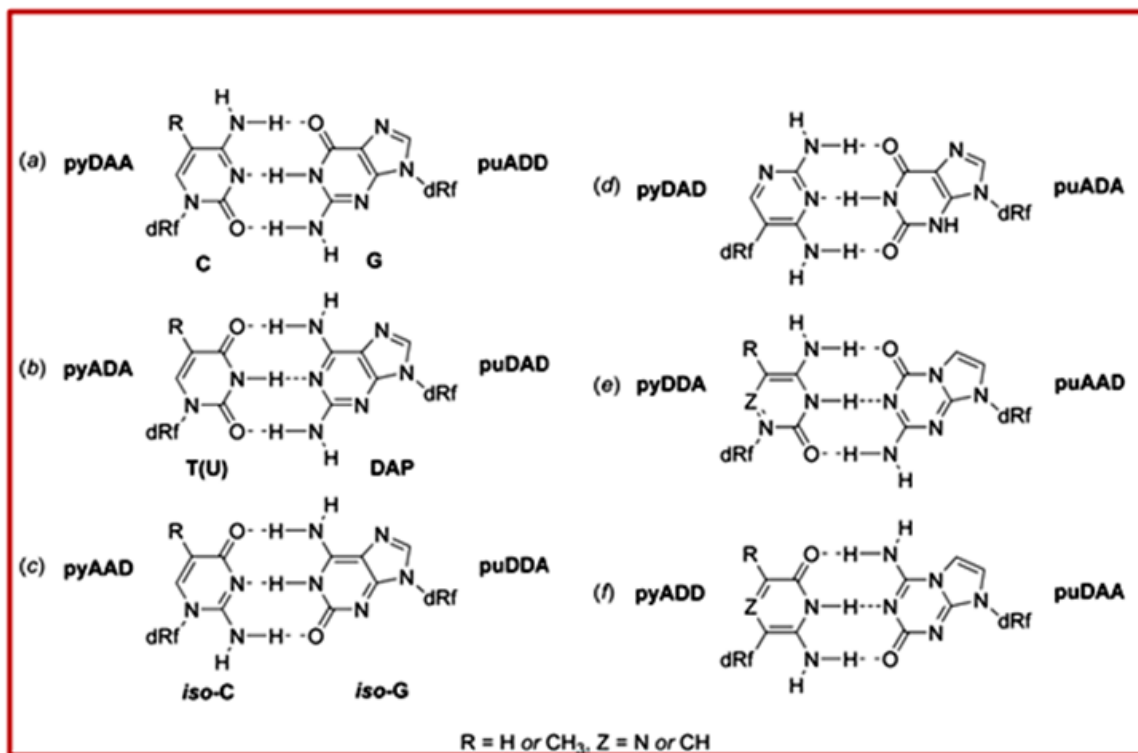


Figure 4.28: Several base pairs with H-bond

#### 4.19.4.1.2. Purine base Analogues

##### 4.19.4.1.2.1. Adenine Base Analogs

**2-aminopurine (2-AP) as Adenine Base Analogue:** More than 30 years ago, Stryer and colleagues first demonstrated that 2-AP (**Figure 4.29**), a structural isomer of adenine, was strongly fluorescent. Being structurally similar to adenine (6-aminopurine) 2-AP is a non-perturbing substitution and forms thermodynamically equivalent base pairs with thymine in DNA helices and uracil in RNA helices. In contrast to adenine, 2-AP is also capable of forming a base pair with cytosine at a substantially higher frequency, which is the basis for 2-AP's mutagenicity. The introduction of 2-AP into both DNA and RNA oligonucleotide sequences in a site-specific manner is straightforward and the quantum yield of 2-AP is highly sensitive to its microenvironment and insensitive to base pairing and other H-bonding interactions, thus, allows to be used to detect subtle conformational changes in nucleic acids.

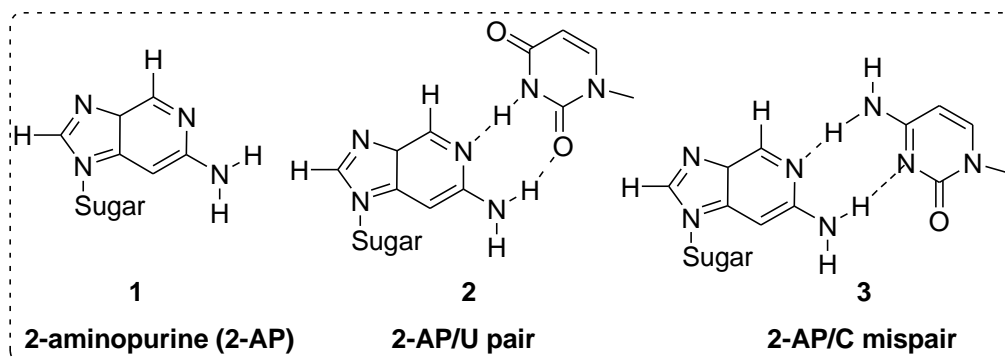


Figure 4.29: (a) Structure of the highly fluorescent nucleotide base analog 2-aminopurine (2-AP), 1; 2-AP-uracil pair, 2 and 2-AP-cytosine mispair, 3.

In 1988, Andre Chollet *et al.*, have introduced the base analogue 2-aminoadenine (2,6-diaminopurine, D) at selected positions into synthetic oligodeoxyribonucleotides and DNA by the combined use of chemical and enzymatic methods. 2-aminoadenine substitution for adenine introduces changes in the minor groove of DNA and creates an additional hydrogen bond in the Watson-Crick base pair with thymine. Oligonucleotide hybridization probes containing 2-aminoadenine showed increased selectivity and hybridization strength during DNA- DNA hybridization to phage or genomic target DNA. Properties of the base analogue with respect to DNA modifying enzymes were examined. 2-aminoadenine was used to probe minor groove determinants during the treatment of DNA by 12 restriction endonucleases. Inhibition of cleavage was found for several restriction enzymes.

Recently, Tom Brown, has synthesized the heterocyclic base 7-aminopropargyl-7-deaza-2,6-diaminopurine (D) and incorporated into oligodeoxynucleotides. They have shown that D:T has similar thermodynamic stability to G:C and is a stable analogue of A:T.

**ATP/AMP Analogs:** Other fluorescent analogs of adenine have been developed and used primarily as probes of AMP/ATP binding by enzymes (Figure 4.30). These other analogs include etheno-ATP and lin-benzo-AMP, which are analogs of ATP and AMP, respectively. Another ATP analog, formycin 5'-triphosphate, has been used as a substrate analog for adenylate cyclase. These analogs have been used primarily in studies of nucleotide cofactor binding to enzymes.



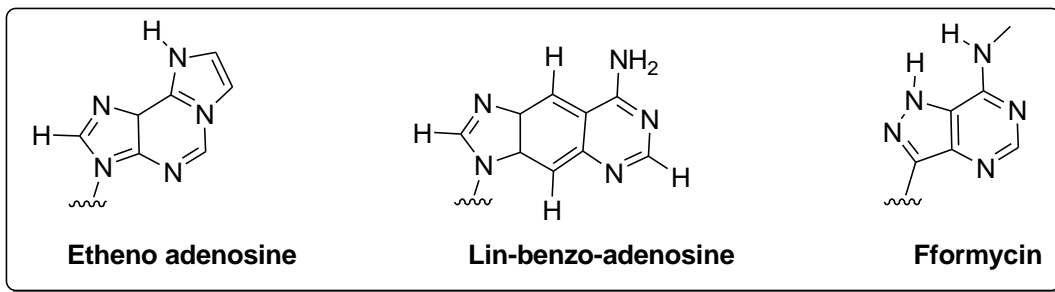


Figure 4.30: Structures of adenine base analogs used to construct fluorescent nucleoside and nucleotide analogs- Etheno-adenosine, lin-benzo-adenosine, and formycin.

**Pteridineadenine analogs:** Pteridines are a class of bicyclic planar compounds that are structurally similar to purines. Two pteridine adenine analogs, 4-amino-6-methyl-8-(2-deoxy- $\beta$ -D-ribofuranosyl)-7(8*H*)-pteridone (6-MAP) and 4-amino-2,6-dimethyl-8-(2'-deoxy- $\beta$ -D-ribofuranosyl)-7(8*H*)-pteridone (DMAP) (Figure 4.31) have also been synthesized and phosphoramidite also incorporated into DNA oligonucleotides for fluorescence measurements. Like the pteridine guanine analogs, these analogs display significant quenching of fluorescence intensity, increased complexity of decay curves and decreased mean fluorescence lifetimes when incorporated into oligonucleotides. The pteridine analogs have been shown to be minimally disruptive of DNA structure as evidenced by the similarity of melting point temperatures of pteridine-containing oligonucleotides with unmodified control oligonucleotides.

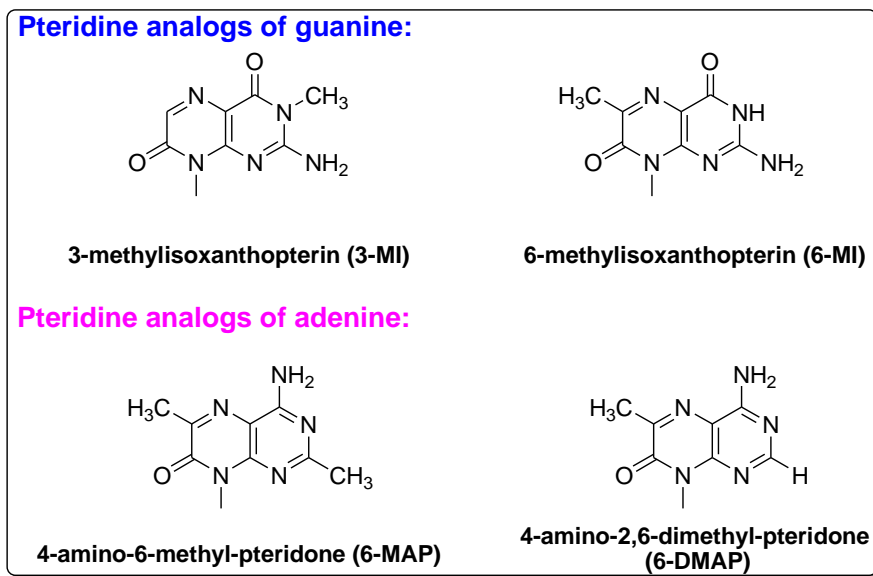


Figure 4.31: (5). Two pteridine analogs of guanine, (A) 3-MI and (B) 6-MI and two adenine analogs synthesized from the pteridine class of compounds, (C) DMAP and (D) 6-MAP, which have been incorporated into short DNA oligonucleotides for structural and enzymatic studies.

#### 4.19.4.1.2.2. Guanine Base Analogs

**Pteridine Guanine Analogue:** Hawkins and colleagues have described the synthesis of several pteridine base analogs, some of which are highly fluorescent with  $\Phi_F$ 's ranging from 0.03 to 0.88. Among the available pteridines, two guanine analogs, 3-methylisoxanthopterin (3-MI) and 6-methylisoxanthopterin (6-MI) have been synthesized as phosphoramidites and incorporated into DNA oligonucleotides. Like 2-AP, incorporation of these probes into oligonucleotides significantly quenched their fluorescence signal and the degree of quenching correlated with the number and proximity of purines in the oligonucleotide. Incorporation also resulted in a shift in absorbance, emission and decay-associated spectra for 6-MI. An increase in the complexity of the fluorescence decay curve and a decrease in the mean lifetime were also observed for both the 3-MI and 6-MI probes. Formation of double stranded oligonucleotides did not substantially increase the degree of quenching but generally increased the complexity of decay curves and decreased mean fluorescence lifetimes.

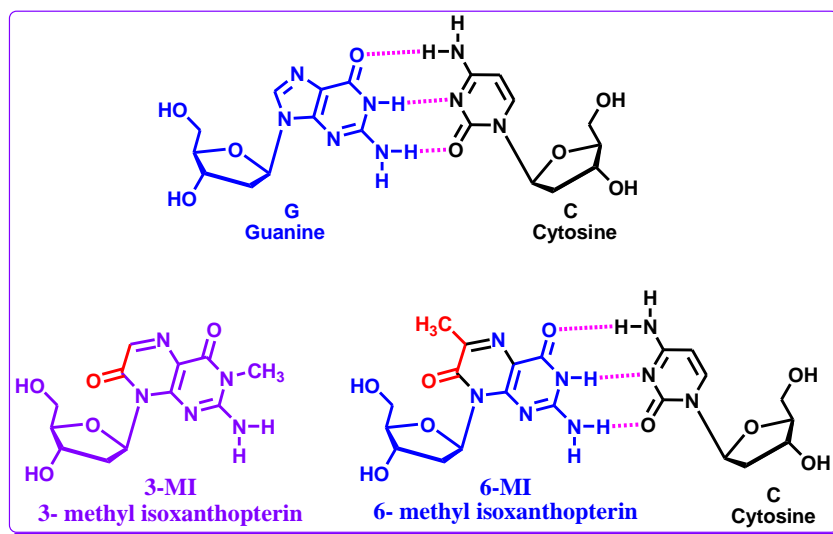


Figure 4.32: 3-MI and 6-MI as pteridine guanine analogues

#### 4.19.4.1.3. Pyrimidine base Analogues

##### 4.19.4.1.3.1. Thymine Analogue

**Benzo[g]quinazoline based thymine base analogs:** Godde and colleagues have synthesized bases with extended aromatic domains that increase third strand binding through stacking interactions. One of the polycyclic aromatic base analogs of thymine, benzo[g]quinazoline-2,4-(1*H*,3*H*)-dione (Figure 4.33), was found to display strong fluorescence emission centered at 434 nm ( $\Phi_F \sim 0.82$ ) and two major excitation maxima (260 and 360 nm). Formation of the triple helical structure using a third oligopyrimidine Hoogsteen strand that contained this fluorescent thymine analog, resulted in a shift of the fluorescence emission

maximum to shorter wavelengths and a decrease in fluorescence intensity. In a duplex, it did not produce any significant changes in fluorescence properties. Thus, the sensitivity of this base analog to the helical conformation allowed selective detection of triplex over duplex formation.

**Thymine analogue:** Another thymine analog 5-methyl-2-pyrimidinone (Figure 4.33) has been synthesized and used in early studies of DNA duplexes. This base analog does not pair well with adenine, however, using time-resolved fluorescence decay measurements it was shown that the predominant state of the base in the context of a DNA oligonucleotide is stacked so that its fluorescence is efficiently quenched.

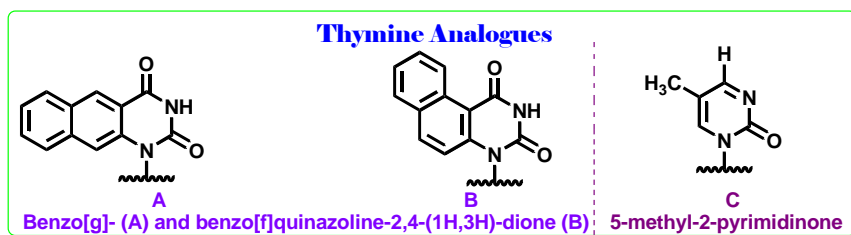


Figure 4.33: benzo[g]/[f]quinazoline-2,4-(1H,3H)-dione and 5-methyl-2-pyrimidinone as Thymine analogue

#### 4.19.4.1.3.2. Cytosine Analogue

**Benzo[g]quinazoline based cytosine base analog:** A 2'-O-Me ribonucleoside derivative of 4-amino-1Hbenzo[g]quinazoline-2-one has also been synthesized based on the same heterocyclic benzo[g]quinazoline design and used as a novel fluorescent cytosine base analog probe. This cytosine base analog exhibits a fluorescence emission centered at 456 nm, characterized by four major excitation maxima (250, 300, 320 and 370 nm) and a fluorescence quantum yield  $\Phi_F = 0.62$  at pH= 7.1. The fluorescence emission of this probe shifted from 456 to 492 nm when pH was decreased from 7.1 to 2.1. The pKa (4.0) of the probe was close to that of cytosine (4.17). This probe was used to detect the protonation state of base triplets in triple stranded structures.

**Tricyclic cytosine (tC):** Norden and colleagues have described a new cytosine base analog, 3,5-diaza-4-oxophenothiazine or tricyclic cytosine (tC), which can form a specific base pair with guanine (**Figure 4.34**). Like the benzo[g]quinazoline base analogs, this base maintains its relatively high quantum yield ( $\Phi_F = 0.20$ ) even after incorporation into single and double stranded oligonucleotides, like artificial peptide nucleic acid (PNA) biopolymers and RNA-DNA duplexes.

Elaboration of the tC(O) scaffold can yield a nitroxide spin labeled compound that may be used for EPR measurements and the “G-clamp” which has increased binding affinity to guanine (A, C, D) (**Figure 4.34**).

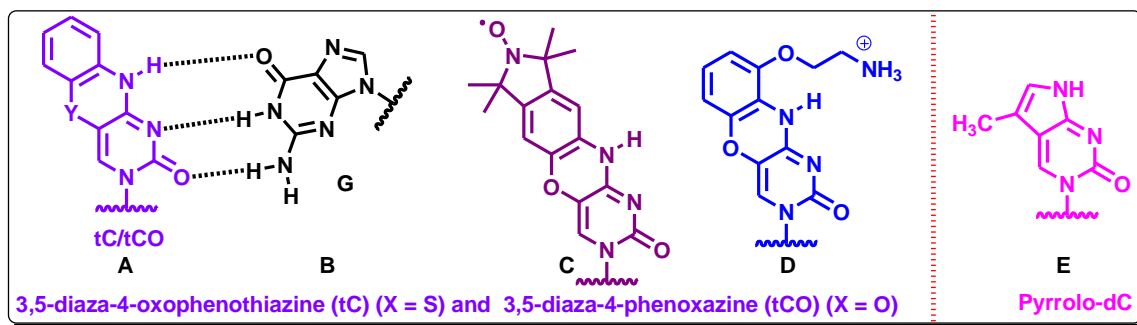


Figure 4.34: 3,5-diaza-4-oxophenothiazine (tC) (X = S) and 3,5-diaza-4-phenoxazine (tCO) (X = O) in hybridization with guanine (A, C, D). The structure of pyrrolo-dC (E).

**Pyrrolo-dC: cytosine Analog:** In another study, a new highly fluorescent analog of cytosine, pyrrolo-dC, (**Figure 4.34**) has been introduced to characterize the transcription bubble in elongation complexes of T7 RNA polymerase. Pyrrolo-dC has excitation and emission maxima at 350 nm and 460 nm, respectively, which, like the previously described analogs, allows selective excitation in the presence of native nucleic acid bases and proteins. This base analog can pair with guanine and, like 2-AP, pteridine and hydrocarbon base analogs, shows significant quenching of fluorescence when incorporated into single and double stranded DNA. The quenching can be used to monitor local melting of the **G:C** base pairs in a DNA helix and can serve as a complementary probe to 2-AP, which reports on melting of AT base pairs.

#### 4.19.4.1.4. Xanthosine Analogue

In addition to these cytosine and thymine analogs, the synthesis of the fluorescent analog of the rare base xanthosine (5-aza-7-deazaxanthine) has also been reported (**Figure 4.35**). This base was designed based on the supposition that the rare tRNA constituent wyosine carries the 5-aza-7-deazapurine substructure, and it is this structure that makes the base fluorescent. This base analog can be excited at 250 nm and displays two emission maxima at 410 and 580 nm.

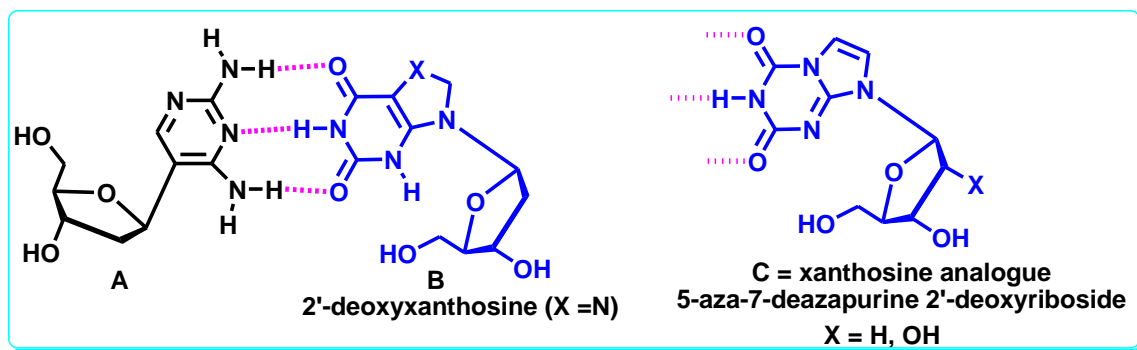


Figure 4.35: The Xanthosine analogue

#### 4.19.4.2 Extended Base-Pairs/Base-Pairs Based on Four H-Bonds/Halogen-Bonding Base-Pair:

Most, recently Kool and coworkers are interested in designing much more size expanded nucleobases for new genetic bases having interesting fluorescent properties. Thus, they reported the synthesis, and properties of DNA-like molecules in which the base pairs are expanded by benzo homologation. The resulting size-expanded genetic helices are called xDNA ("expanded DNA") and yDNA ("wide DNA"). The large component bases are fluorescent, and they display high stacking affinity. When singly substituted into natural DNA, they are destabilizing because the benzo-expanded base pair size is too large for the natural helix. However, when all base pairs are expanded, xDNA and yDNA form highly stable, sequence-selective double helices. The size-expanded DNAs are candidates for components of new, functioning genetic systems. In addition, the fluorescence of expanded DNA bases makes them potentially useful in probing nucleic acids (**Figure 4.36a-b**).

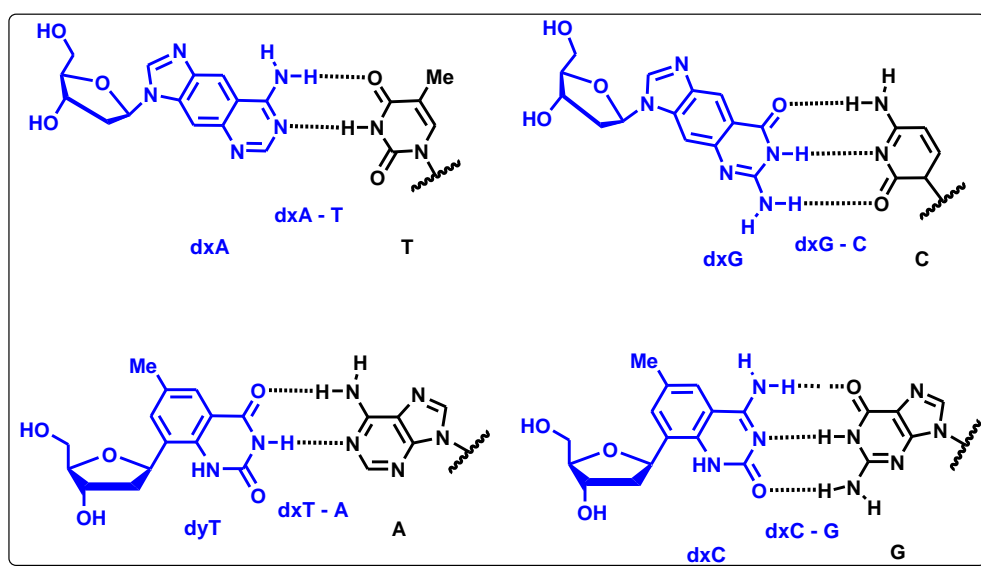


Figure 4.36a: Structures and H-bonding pattern of xDNA bases.

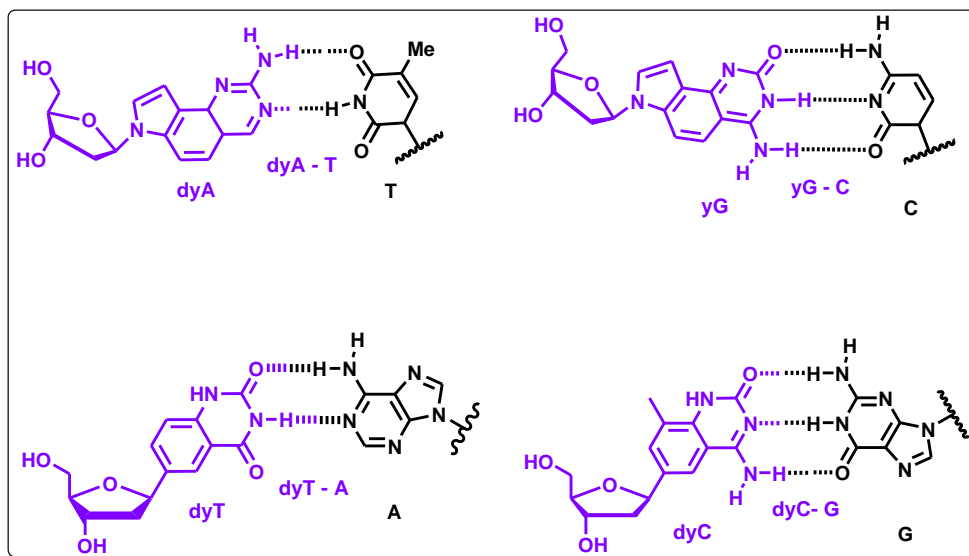


Figure 4.36b: Structures and H-bonding pattern of yDNA bases.

In a homologous way they again expanded the size of the bases by inclusion of another benzene ring to the pairs. This would yield 4.8 Å of expansion, giving hypothetical xxDNA and yyDNA ("double-wide DNA") designs (Figure 4.37). The possible self-assembly of the double-wide DNA helix establishes a new limit for the size of information-encoding, DNA-like molecules, and the fluorescence of yyDNA bases suggests uses as reporters in monomeric and oligomeric forms.

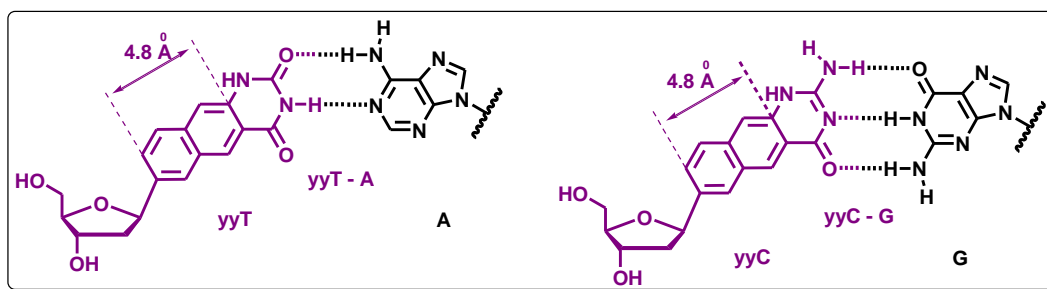


Figure 4.37: The base pairs of doublewide DNA

Another type of extended DNA pair was recently reported by Inouye, based on the acetylene-linked C-nucleosides bearing pyrimidine heterocycles capable of forming Watson-Crick-like H-bonds ( $iG^*-iC^*$  and  $A^*-T^*$ ) (Figure 4.38).

Two interesting pairs of extended imidazolopyridopyrimidines and naphthyridine C-nucleosides ( $ImO^N/NaN^O$  and  $ImN^O/NaO^N$ ) were prepared by Minakawa and Matsuda (Figure 4.38). Their four-H-bond interactions ensure extra stable duplexes and moderately selective recognition and incorporation by Klenow fragment polymerase.

Hirao *et al.* have reported the nucleobases **y** and **v** (**Figure 4.38**) and showed that **dyTP** is incorporated into DNA template opposite to **dx** more efficiently than any natural substrate. The **dyTP** was incorporated into the template opposite to **ds** with higher efficiency and a 3-fold higher selectivity than opposite to **dx**. The efficiency and fidelity of **y–v** pairing were as high as in natural base-pairs. The **y–s** pair was used for in vitro incorporation of chlorotyrosine into a protein proving a new genetic codon system.

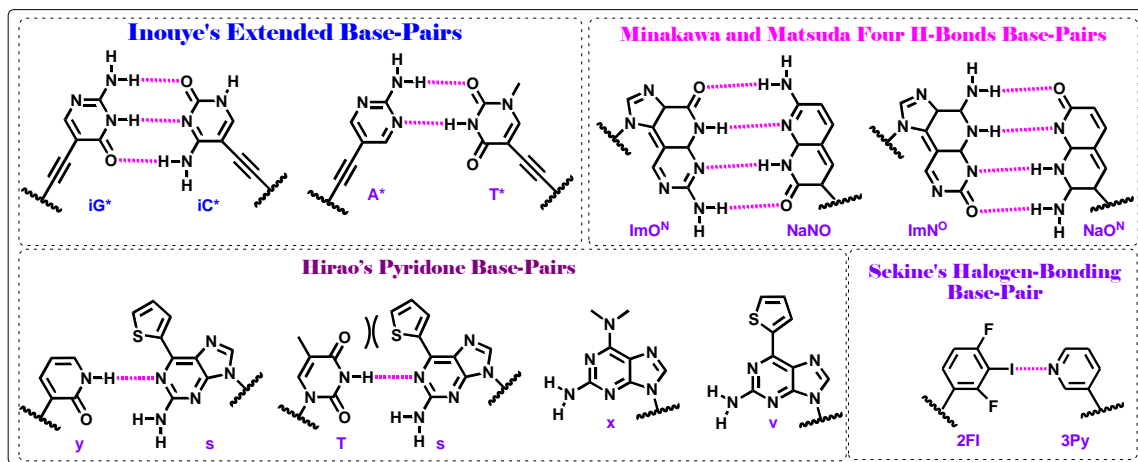


Figure 4.38: Few more unnatural base-pairs that are paired via H-bonding

An interesting alternative to H-bonds was studied by Sekine, who endeavored to employ halogen bonding for specific base-pairing of artificial nucleosides. Out of several halogenated benzene and pyridine pairs, the most stable and selective binding was found for difluoriodobenzene (**2FI**) and pyridine (**3Py**) pair, but no evidence for the halogen bond was provided (**Figure 4.38**).

It is quite clear that unnatural base-pairs that are paired via H-bonding patterns offer promising approaches to expanding the genetic alphabet. However, there are some inherent limitations to their selective replication in DNA, due to the population of minor tautomeric forms of the unnatural bases that form mispairs.

#### 4.19.4.3 Artificial Base-Pairs Based on Hydrophobic Interactions- Replacement of Heterocyclic (Purine or Pyrimidine) Bases by Polycyclic Aromatic Hydrocarbons

Kool and colleagues have taken a different approach to develop novel fluorescent nucleotide base analogs for incorporation into nucleic acids. In their synthetic design, instead of modifying an existing heterocyclic nitrogenous-containing nucleoside base, they simply replace it with aromatic polycyclic hydrocarbon structures. Like the modified natural bases, these hydrocarbon aromatic structures can be incorporated with only small perturbation to the natural DNA structure and allow for close interactions including stacking within

the DNA helix. In general, the hydrocarbon base moieties are designed to be non-polar, weakly hydrogen bonding and planar aromatic groups. Synthetic routes to 4-methyl-1H-indole, phenanthrene and pyrene fluorophores conjugated at the C(1) position of 2-deoxy-D-ribose (**Figure 4.39**) have been reported. In addition, the derivative with a pyrene moiety at C(1) in the  $\beta$ -D-form has been shown to be of use in DNA diagnostics strategies, where it efficiently forms excimers with neighboring pyrene labels. Pyrene derivatives in the  $\beta$ -D-form stabilize DNA helices markedly due to low polarity and have been incorporated into the DNA helix enzymatically by DNA polymerase. The synthesis of additional fluorescent aromatic hydrocarbon nucleoside analogs with novel fluorescence characteristics, like shifted ranges of excitation and emission wavelength maxima, has more recently been reported. These fluorophores included terphenyl, stilbene, terthio- phene, benzoterthiophene (**Figure 4.39**) and were conjugated to deoxyribose at the C(1) position in the  $\beta$ -D- form. The emission spectral maxima for the free nucleoside forms of these base analogs ranged from 345 nm to 536 nm, with fluorescence quantum yields that ranged from 0.025 to 0.67. Synthesis of the phosphoramidite derivative of each of these conjugated nucleoside base analogs was also reported, allowing straightforward incorporation using standard automated oligonucleotide synthesis. In addition to the conjugated fluorophore nucleosides, a C-nucleoside with cyclohexene at the 1- position was also synthesized as a non-fluorescent spacer. The rationale for synthesizing this spacer was to attempt to disrupt ideal stacking conditions that results in a high degree of quenching of the aromatic hydrocarbon base analogs conjugated at C(1) when incorporated into oligonucleotides.



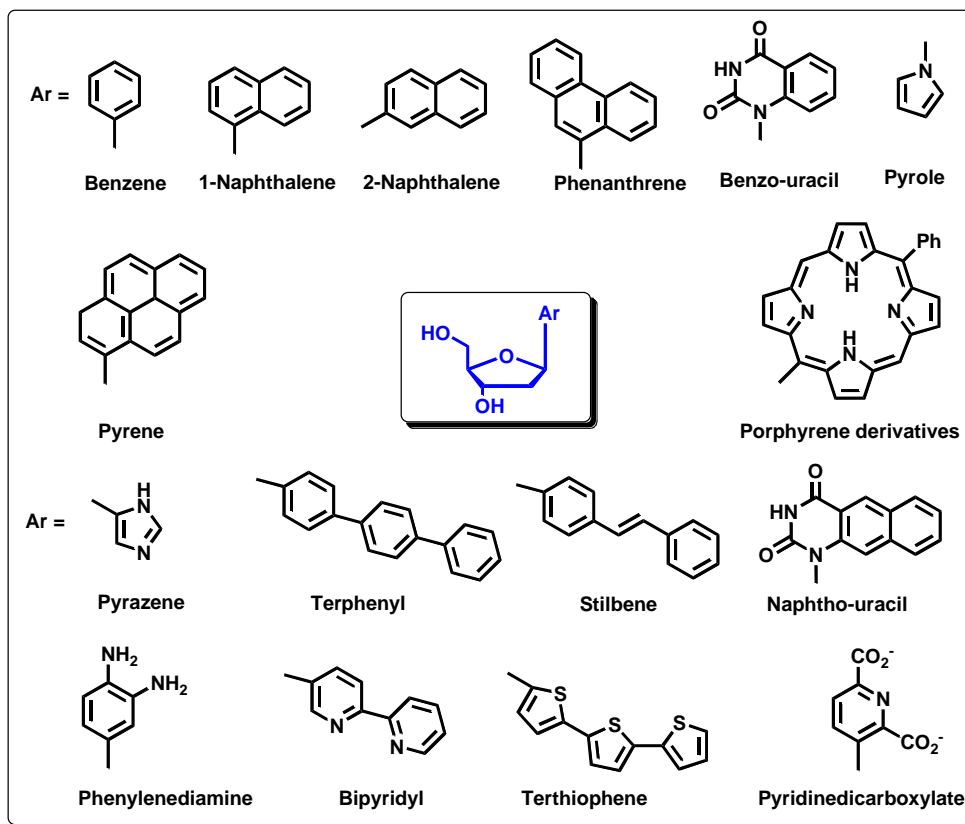


Figure 4.39: Structures of a series of polycyclic aromatic planar hydrocarbon and oligothiophene base analogs that display a range of fluorescent properties:

4.19.4.4 Fluorescently Labeled Nucleosides and oligonucleotide probes: Covalent Attachment of Probes on to the Natural Bases

As discussed earlier, the third strategy is to install the fluorescence property into the natural nucleosides for biochemical application is the attachment of a fluorophore covalently through a rigid carbon linker producing what is called is fluorescently labeled nucleosides. Oligonucleotides containing such labeled nucleosides can widely be used in a variety of genomic assays including gene quantitation, allelic discrimination, expression analysis, and SNP typing. Many assays use such fluorescently labeled oligonucleotide probes in which a biochemical event (e.g., hybridization) causes an increase in fluorescence intensity. The fluorescently labeled nucleobases, in contrast, are attached through a carbon linker to oligonucleotides and can move more freely in ways that may be independent of the overall complex. Such intrinsic motions of a dye can complicate the interpretation of some fluorescence measurements, yet a rigid carbon linker will have the capability to restrict this motion to some extent and these fluorophore are so design as to achieve intercalation or groove binding properties.

The lack of naturally occurring fluorescent bases has spurred the development of new nonnatural fluorescent nucleosides/analogues which will be the probe for DNA analysis. These nucleosides can be so design as to get optimized fluorescence properties (like, higher quantum yields and longer

lifetimes) and that can be incorporated into an oligonucleotide using standard automated synthetic methods. The fluorescence properties, from DNA and RNA molecules, thus can be observed without any competing background signals mainly by two ways: (a) incorporation of unnatural nucleotide base analogs into the oligonucleotide sequence and/or (b) incorporation of fluorophore into the natural bases *i.e.* by synthesizing labeled bases.

Since several years Saito and coworkers are engaged in designing covalently labeled nucleobases for SNPs typing in homogeneous solution. The labeled nucleosides upon incorporation into an oligonucleotide are capable of detecting the change of microenvironment within the DNA and can report the presence of opposite base in a target DNA *via* a generation of an enhanced fluorescence signal.

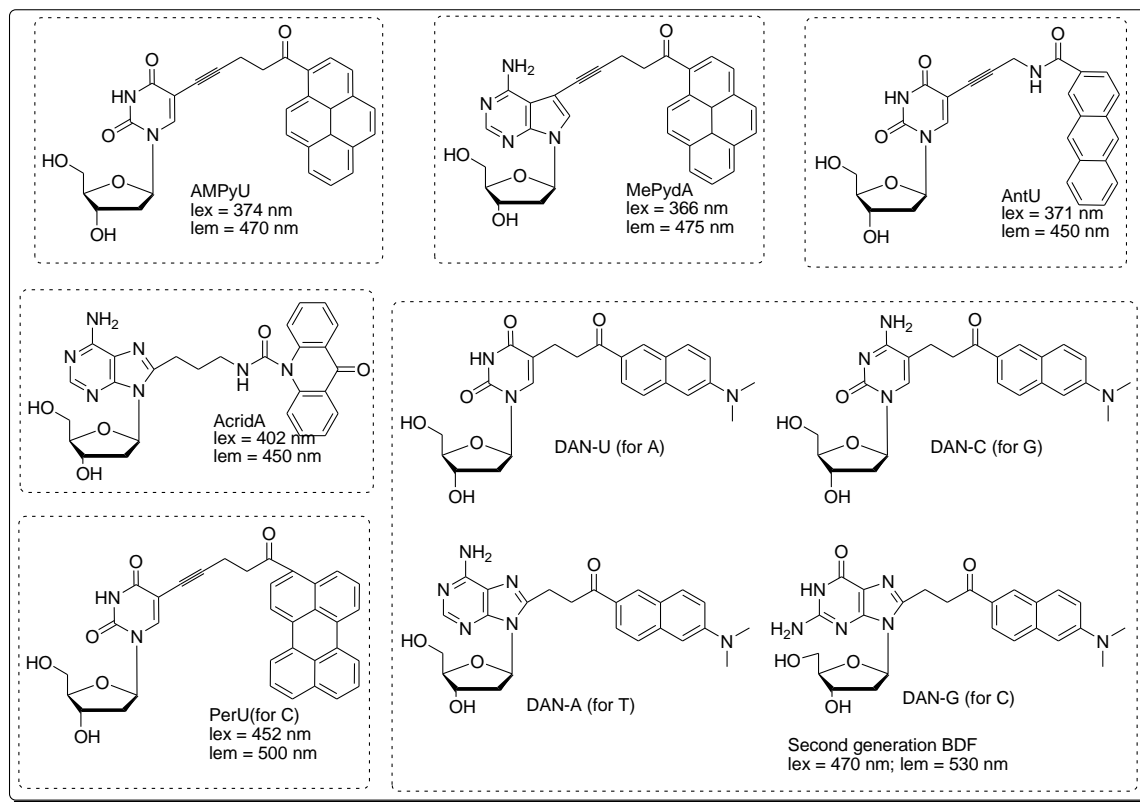


Figure 4.40: Few examples of covalently conjugated fluorescent nucleosides

As for example, anthracene-based fluorescent probe would be valuable in sensing the microenvironment inside and outside the DNA duplexes and thereby they might be potent candidate to develop new base discriminating fluorescent (BDF) nucleosides for detection of matched base opposite to the BDF base in a complementary oligonucleotide sequence by a sharp change in fluorescence intensity.

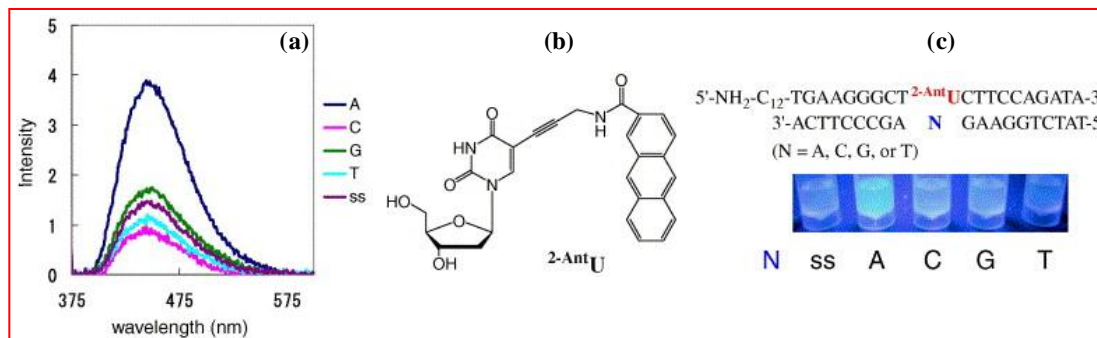


Figure 4.41: (a) Fluorescence spectra of  $\text{ODN}_{abl}^{2\text{-AnthU}}$  hybridized with  $2.5 \mu\text{M}$   $\text{ODN}_{abl}$  ( $N = A$ , full match),  $\text{ODN}_{abl}$  ( $N = T, G, \text{ or } C$ ), and single-stranded  $\text{ODN}_{abl}^{2\text{-AnthU}}$ . Excitation wavelength was at 371 nm. (b) Structure of BDF base  $2\text{-AnthU}$  (c) Comparison of the fluorescence for the bases opposite  $2\text{-AnthU}$  (50 mM sodium phosphate, 0.1 M sodium chloride, pH 7.0, room temperature). “ss” denotes a single-stranded BDF probe. The sample solutions were illuminated with a 365 nm transilluminator.

## 4.20. Single Nucleotide Polymorphism (SNPs) and Introducing Hap Map Project

### 4.20.1. Introduction-The SNPs

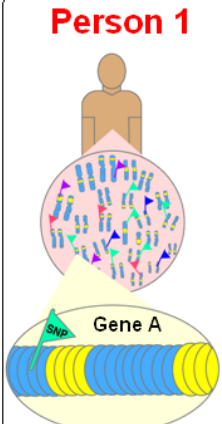
DNA is most widely thought of as the central storehouse of genetic information in each cell and obviously that is a major function. However, it serves several other biological functions, like polyfunctional switch for control of gene expression as well. Thus, chemists and biomedical researchers have used the DNA structure as a starting point of design of many new molecules with new functions such as to act as a ligand for nucleic acids or as sensors and probes for genetic analysis.

The completion of human genome project has set the stage to the biomedical researcher all over the world to find out the variations in the human genome through the genetic analysis. The genetic variations affect the human health largely. In particular, genetic variation accounts for some of the differences between individuals, such as eye color, length, and blood group. However, they may also predispose some people to disease and explain why some respond better to certain drugs than others. Therefore, detecting genetic differences between individuals and determining their impact on human health are fundamental in genomic research. Thus, it will be worthwhile to identify the DNA sequence variants conferring susceptibility or resistance to common human diseases with improved methods on a genome-wide scale. The mapping of genetic variations will lead to the understanding of the role of mutations in simple as well as complex genetic diseases. Among all other genetic variations, single nucleotide polymorphism (SNP), a single base-pair mutation that occurs at a specific site in the DNA sequence, and present in at least 1% of the population, is the most common one. Scientists think that the SNP is the key enabler in the realization of the concept of personalized medicine, the target of recent research.

### 4.20.1. Why SNPs are so Important?

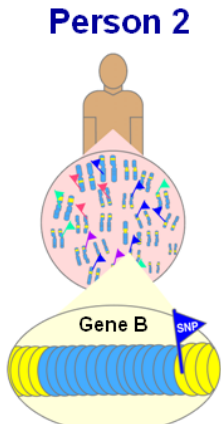
## WHY SNPs ARE SO IMPORTANT ?

**Person 1**



SNP marks Gene A

**Person 2**



SNP may cause Gene B to make altered protein

▶ = SNP variations in DNA

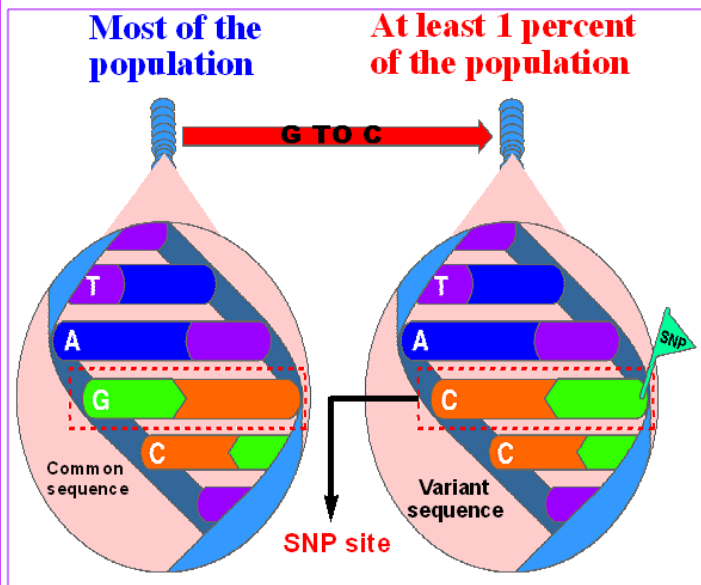
1. SNPs can cause silent, harmless, harmful, or latent effects.
2. Most SNPs occur in noncoding regions and do not alter genes.
3. The remaining SNPs occur in coding regions. They could alter the protein made by that coding region, which in turn could influence a person's health.

“SNP is the key enabler in the realization of the concept of personalized medicine”.

**So, let us have a look how proteins are changed in presence of SNPs**

Among all other genetic variations, single nucleotide polymorphism (SNP), a single base-pair mutation that occurs at a specific site in the DNA sequence, and present in at least 1% of the population, is the most common one. Scientists think that the SNP is the key enabler in the realization of the concept of personalized medicine, the target of recent research.

## SNPs - THE MOST COMMON TYPE OF VARIATION



In the DNA sequence **TAGC**, a SNP occurs when the **G** base changes to a **C**, and the sequence becomes **TACC**

Tiny variations in human genome --- single nucleotide polymorphisms, or **SNPs** ("snips"), ---- help to answer large questions and to catalogue the unique sets of changes involved in different **DISEASES OR IN CANCERS**.

1. Scientists see SNPs as a potential tool to improve diseases **DIAGNOSIS** and **TREATMENT** planning.
2. They suspect that SNPs may play a role in the different responses to treatments seen among cancer patients.
3. And they think that SNPs may also be involved in the different levels of individual cancer risk observed.

## 4.20.2. Important of SNPs

### SNPs IN CODING REGIONS AND THE EFFECT ON HEALTH

(1) SNPs -- No Changes in Protein; (2) SNPs -- Subtle, Harmless Changes in Protein; (3) Harmful Changes in Protein: Mutations; (4) Subtle Changes in Protein --- Switch on Under Certain Conditions

#### SNPs -- No Changes in Protein

DNA SNP C to G

mRNA: G A C → G A G

RNA Codon: CUG → CUC

Leucine

Protein: Leucine to Leucine

No change in shape

#### Harmful Changes in Protein: Mutations--- Sickle cell anemia. Hemoglobin beta gene Hemoglobin molecule not carry oxygen

DNA SNP T to A

mRNA: C T A → C A A

RNA Codon: G A U → G U U

Aspartic acid → Valine

Protein: Aspartic acid to Valine

Change in shape

### SNPs AND DRUG INTERACTION

Drug

Absorption in the breast

Drug in breast tissue

Metabolism in the liver

Excretion in the kidney

Transportation in the blood

Drug in bloodstream

Transporter

**SNPs explain :**

1. Person's uniqueness in physical appearance
2. Disease susceptibility
3. Different response to a specific drug treatment
4. Different side effects in response to the same drug
5. Proteins- drug interaction

### OTHER IMPORTANT FEATURES OF SNPs

- ❖SNPs in Noncoding Regions: Markers for Genes
- ❖SNP Profiles and Response to Drug Therapy

**Thus,**

- ❖SNP Profiles may help to identify Cancer Genes
- ❖Scientists are also using SNPs to calculate risk factors associated with cancer in large populations

### 4.20.3. Various SNPs Typing Protocols

Therefore, because of their significant biological and clinical importance, recently, there has been a growing interest in single nucleotide polymorphisms (SNPs) typing. A great deal of effort has been paid by various research groups all over the world for the development of SNP genotyping technologies since last few years. As a result, a lot of different SNP typing protocols have become available for us however there is no single chemistry, platforms and protocols that meet all research needs. The technology, which will offer high sensitivity, reproducibility, accuracy, less laborious, less time consuming and low cost of genotyping SNPs will be the better alternatives than the presently available methods. The presently available methods generally consider the difference in hybridization efficiency or the difference in enzymatic recognition between the nucleotide probes.

Fluorescence techniques are now a days, find widespread application in biochemical and biophysical studies of macromolecules, specially, in studies of nucleic acids, fluorescence spectroscopy provides an important tool for the detection and probing of structure, dynamics and interactions. Fluorescence based detection schemes are now used routinely in numerous applications, such as DNA sequencing, in situ genetic analysis through techniques like fluorescence in situ hybridization (FISH), and high throughput screening (HTS). However, there is a concomitant need to develop methods for the rapid and routine analysis of specific nucleic acid sequences so that this information can be understood and interpreted. In recent development for the detection of single base alterations, the methods that utilize fluorescence-labeled short oligonucleotide probes to reveal the presence of complementary sequences offer an attractive alternative.

The single base discrimination in nearly all reported methods is achieved directly or indirectly on the basis of difference in hybridization efficiency between matched and mismatched target DNA/probe DNA duplex. However, as far as the detection relies on the hybridization events, such DNA probes suffer from shortcomings in their selectivity. The differences in the hybridization efficiency are varied with sequence context and often very small for the detection of a single base mismatch in long target DNA. In order to realize high sensitivity, the hybridization and washing conditions are carefully selected to minimize undesirable responses from the mismatched hybridization probes. From these points of view, the alternative probes that do not rely on the hybridization events are highly desirable.

A tremendous progress has been achieved due of intense research efforts, especially in the post genome era, by several research groups in designing chemistry, platforms and protocols for DNA detection and SNPs typing, to reach the goal of personalized medicine. Thus, a new SNPs typing method, MagSNiPer, based on single base extension, magnetic separation, and chemiluminescence was developed by Yohda *et al.* Recently, a chip based detection of unlabeled DNA targets was developed by our group using base-discriminating fluorescent nucleobase (BDF) labeled probe. Using this system

Saito and co-workers were able to discriminate a matched and mismatched base by selective fluorescence emission.

Peptide nucleic acids (PNA) besides their application in antisense and antigene therapy are now a day widely used in detection of DNA. Thus, a label-free DNA detection with electrochemical impedance spectroscopy using PNA probes was developed. Gold nanoparticles modified with locked nucleic acid (LNA) system was developed and found that they form stronger duplexes with a single stranded DNA target and offer better discrimination against single base pair mismatches than analogous DNA probes. The LNA nanoparticle probes have also been used to detect double stranded DNA through triplex formation and it enabled in simultaneous detection and identification of multiple DNA targets.

Thus, it is clear that several methods have been developed focusing on optical detection using oligonucleotides probes, electrochemical sensing, quantum dots or enhanced absorption of light by oligonucleotide-modified nanoparticles. Because of its high sensitivity, selectivity, and low cost, label-free electronic methods promise more attraction for DNA detection now a day. Nanomaterials possess unique properties that are amenable to biosensor applications; they are one-dimensional structures that are extremely sensitive to electronic perturbations, readily functionalized with biorecognition layers, and compatible with many semiconducting manufacturing processes. Thus, one-dimensional silicon nanowires and indium oxide nanowires have shown promising performance in DNA detection. There are also several reports on electrochemical detection of DNA hybridization using multi-walled carbon nanotube electrodes.



#### 4.20.4. Need for an Alternative Protocols

### SNPs GENOTYPING: SEARCHING FOR A SIMPLE ALTERNATIVE DETECTION METHOD

- SNPs (“snips”) --- **Most Common Variation in human genome**
- Attractive target for better understanding the **genetic basis of complex diseases**, and to realize the **potential of pharmacogenetics**
- After completion of Human Genome project the world wide aim is to:
  - ❖ Identify SNPs,
  - ❖ Mapping of SNPs,
  - ❖ Profile making of all SNPs in the human genome.
- A lot of different SNP typing protocols, chemistry and **platforms** available
- But, there is no single protocol available that meets all research needs because:
  - ❖ (a) some need strict hybridization conditions;
  - ❖ (b) some are of High cost; or
  - ❖ (c) some are operationally not simple

**\*So, there is a high need for searching an easy and simple method for homogeneous SNPs typing\***

#### 4.20.5. Background of International Hap Map Project (This part is taken From Wikipedia, the free encyclopedia)

Combinations of different genes and the environment play an important role in the development and progression of common diseases such as diabetes, cancer, heart disease, stroke, depression and asthma, or in the individual response to a particular drug medication. This is quite different from that of rarer Mendelian diseases. To find the genetic factors involved in these diseases, one could in principle obtain the complete genetic sequence of several individuals, some with the disease and some without, and then search for differences between the two sets of genomes. This approach is currently infeasible because of the cost of full genome sequencing. The Hap Map project proposes a shortcut.

Although any two unrelated people share about 99.5% of their DNA sequence, some people may have an **A** at a particular site on a chromosome while others have a **G** instead. Such a site is known as a single nucleotide polymorphism (SNP), and each of the two possibilities is called an allele. The

Hap Map project focuses only on common SNPs, those where each allele occurs in at least 1% of the population.

Each person has two copies of all chromosomes, except the sex chromosomes in males. For each SNP, the combination of alleles a person has is called a genotype. Genotyping refers to uncovering what genotype a person has at a particular site. The Hap Map project chose a sample of 269 individuals and selected several million well-defined SNPs, genotyped the individuals for these SNPs, and published the results.

The alleles of nearby SNPs on a single chromosome are correlated. This means that if the allele of one SNP for a given individual is known, the alleles of nearby SNPs can often be predicted. This is because each SNP arose in evolutionary history as a single point mutation, and was then passed down to descendants surrounded by other, earlier, point mutations. SNPs that are separated by a large distance are typically not very well correlated, because recombination occurs in each generation, mixing the allele sequences of the two chromosomes. A sequence of consecutive alleles on a particular chromosome is known as a haplotype.

To find the genetic factors involved in a particular disease, one can proceed as follows. First a certain region of interest in the genome is identified, possibly from earlier inheritance studies. In this region one then locates a set of tag SNPs from the Hap Map data; these are SNPs that are very well correlated with all the other SNPs in the region, so that knowledge of the alleles of the tag SNPs in an individual will determine the individual's haplotype with high probability. Next, one determines the genotype for these tag SNPs in several individuals, some with the disease and some without. By comparing the two groups, one can then determine the likely locations and haplotypes that are involved in the disease.

### **The International Hap Map Project**

The International Hap Map Project is an organization that aims to develop a haplotype map (Hap Map) of the human genome, which will describe the common patterns of human genetic variation.

The International Hap Map Project is a multi-country effort to develop a haplotype map (Hap Map) of the human genome and to identify and catalog genetic similarities and differences in human beings. Using the information in the Hap Map, researchers will be able to find genes that affect health, disease, and individual responses to medications and environmental factors. The Project is collaboration among scientists and funding agencies from Japan, the United Kingdom, Canada, China, Nigeria, and the United States. All of the information generated by the Project will be released into the public domain.

The goal of the International Hap Map Project is to compare the genetic sequences of different individuals to identify chromosomal regions where genetic

variants are shared. By making this information freely available, the Project will help biomedical researchers find genes involved in disease and responses to therapeutic drugs. In the initial phase of the Project, genetic data are being gathered from four populations with African, Asian, and European ancestry. Ongoing interactions with members of these populations are addressing potential ethical issues and providing valuable experience in conducting research with identified populations.

The International Hap Map Project is collaboration among researchers at academic centers, non-profit biomedical research groups and private companies in Canada, China, Japan, Nigeria, the United Kingdom, and the United States. It officially started with a meeting on October 27 to 29, 2002, and was expected to take about three years. It comprises two phases; the complete data obtained in Phase I was published on 27 October 2005. The analysis of the Phase II dataset was published in October 2007. The Phase III dataset was released in spring 2009.

#### **4.20.6. The Homogeneous DNA Detection Methods**

As we have seen earlier section, much work has been therefore devoted to the development of methods that allow for detection of hybridization event. The hybridization assays that are most commonly used are heterogeneous assay systems. However, it has got several shortcomings like non specific adsorption and involved tedious washing protocols, which could be circumvent by use of homogeneous assay system involving solution phase study. In addition, homogeneous assay system will provide a means to perform real time measurements both in vitro and as well as in vivo. Towards this end, various research groups have exerted a great deal of effort on the development of SNP genotyping technologies over the last few years. As a result, a lot of different SNP typing protocols have become available for us however there is no single protocol that meets all research needs. Therefore an alternative probe, strategy or concept that does not rely on the hybridization events is highly demanded.

As a result of various research efforts, molecular beacon (MB) and its several modified forms were developed as effective genetic probes. A reporter incorporation PCR technique was invented utilizing non-natural bases for detection of nucleic acid targets and genotyping SNPs. Nakatani *et.al.* have reported a new chemical approach to practical SNP typing based on allele-specific PCR integrated with a new concept of DNA-labeling by ligand-inducible secondary structure. For efficient multiplex SNP typing, DigiTag2 assay was developed. The method has the potential to analyze nearly any SNP with high accuracy and reproducibility. Visual detection of SNPs is highly effective. Thus, a colorimetric and visual detection of nucleic acids and SNPs using a DNA probe acting like split DNAzyme was developed recently. Peptide nucleic acids (PNA) besides their application in antisense and antigene therapy are now a day widely used in detection of DNA. Ferrocene-labeled peptide nucleic acid (Fc-PNA) was also developed for electrochemical detection of DNA. Thus, it is clear that several

methods have been developed focusing on optical detection using oligonucleotide probes in homogeneous genotyping SNPs.

Saito group is actively engaged in developing base-discriminating fluorescent (BDF) nucleosides and the oligonucleotide probes for SNPs typing via a generation of fluorescence signal. A number of conceptually new techniques for DNA detection have also been developed recently by this group. The concept of the BDF probes is based on the fluorescence change of the BDF base itself in response to the bases on a complementary strand, not on whether the probe is hybridized as illustrated in **Figure 4.41**. When hybridized with fully matched complementary target, the fluorophore has to reside outside the groove exposing itself to the more polar aqueous environment and as a result strong fluorescence signal is observed. On the contrary, because of lack of Watson-Crick base pairing in the mispaired position in the mismatched duplexes, the fluorophore remains inside the duplex facing highly hydrophobic environment thereby exhibit very weak fluorescence. The homogeneous SNP typing method using BDF probes is a powerful alternative to conventional SNP typing methods presently used.

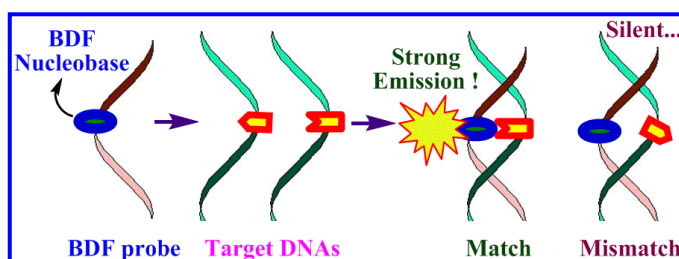


Figure 4.41: Schematic illustration for homogeneous SNP typing using BDF probe

## 4.20.6.1. DNA Detection Using BDF Probes

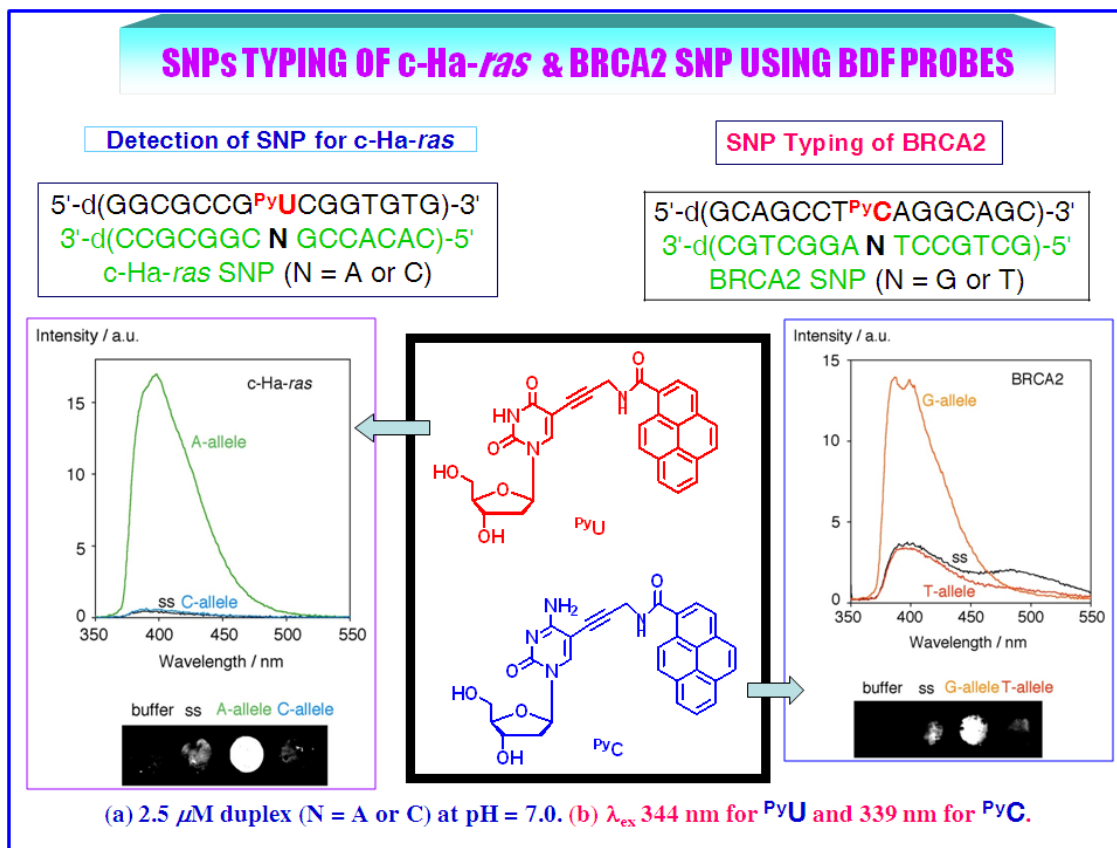


Figure 4.42: Pyrene labeled Probe for detection of SNP of c-Ha-ras and BRCA2 gene sequence

Bag *et al.* have developed, perylene labeled oligonucleotides ODN [d(5'-CGC AAC<sup>Per</sup>UCAACGC-3')] and d(5'-CGCAATPerUTAACGC-3'), which are capable of detecting mismatched cytidine base (C) from the corresponding complementary target, selectively, at wavelength 500 nm having a long trail to 590 nm via an enhancement of fluorescence (**Figure 4.43**). The fluorescence enhancement was observed only when the base opposite to <sup>Per</sup>U is cytidine in irrespective of the sequences. Thus this intelligent fluorescence nucleoside can be used for the detection of T-C SNP.

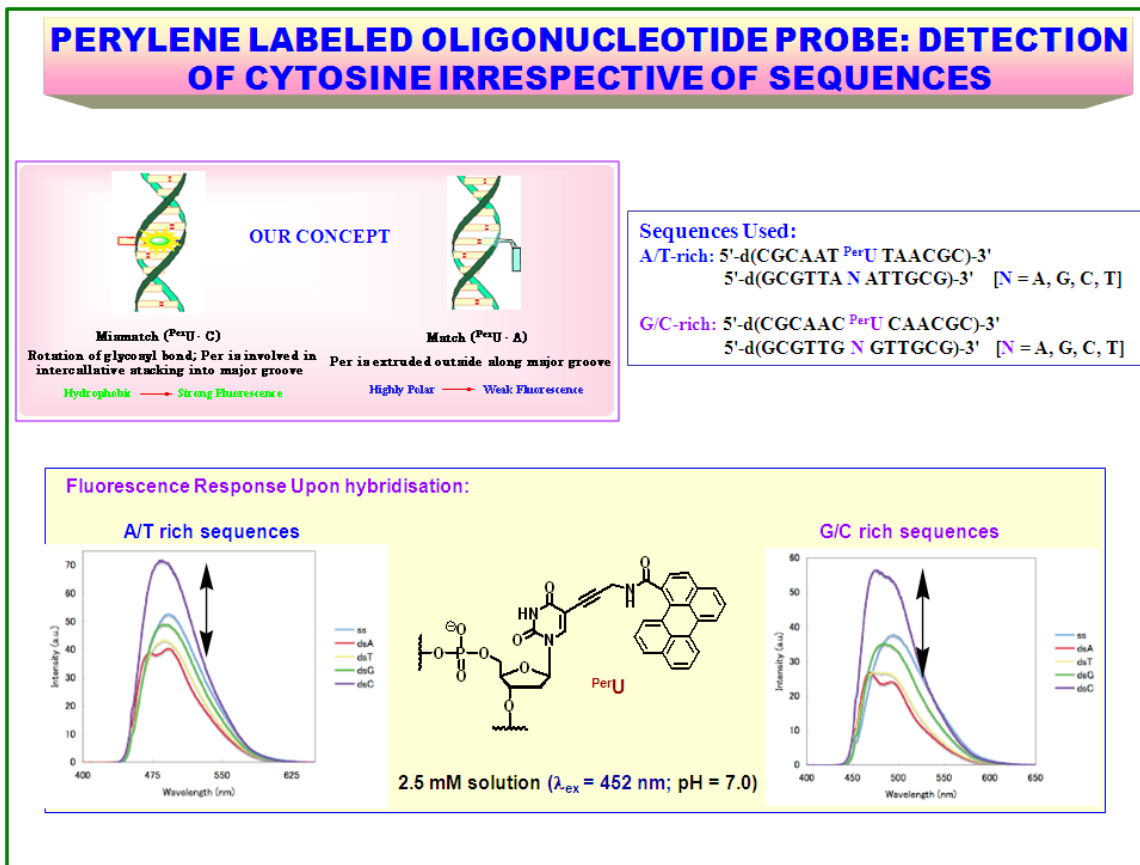
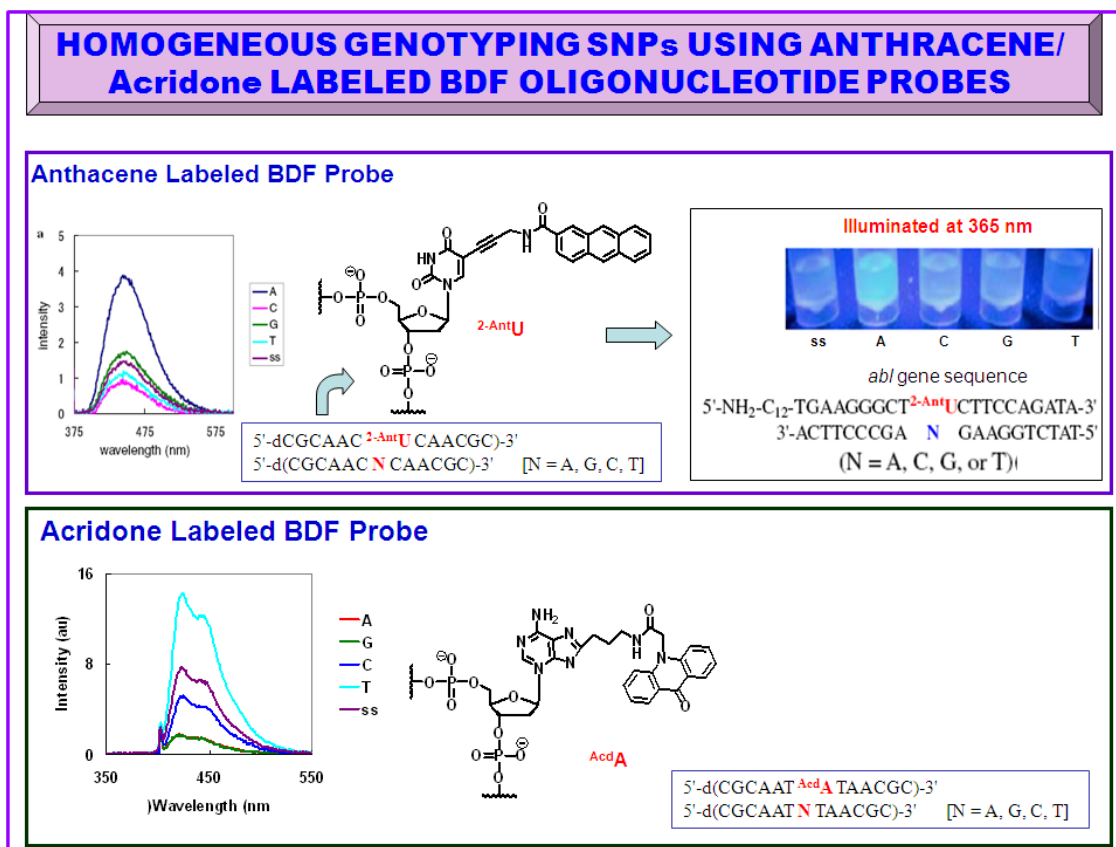


Figure 4.43: Perylene labeled Probe for detection of T/C mismatched DNA sequence.



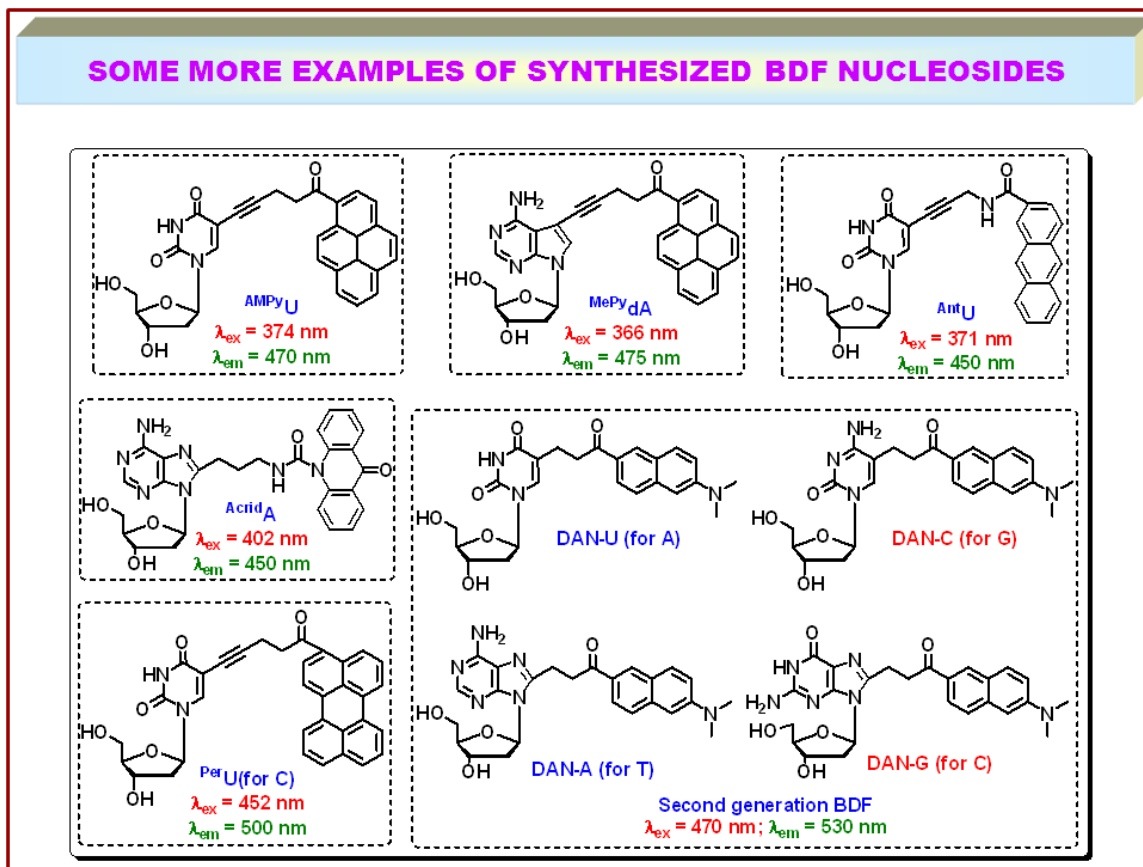


Figure 4.45: More examples of BDF probes.

#### 4.20.6.2. DNA Detection Using Other Fluorescently Labeled Probes

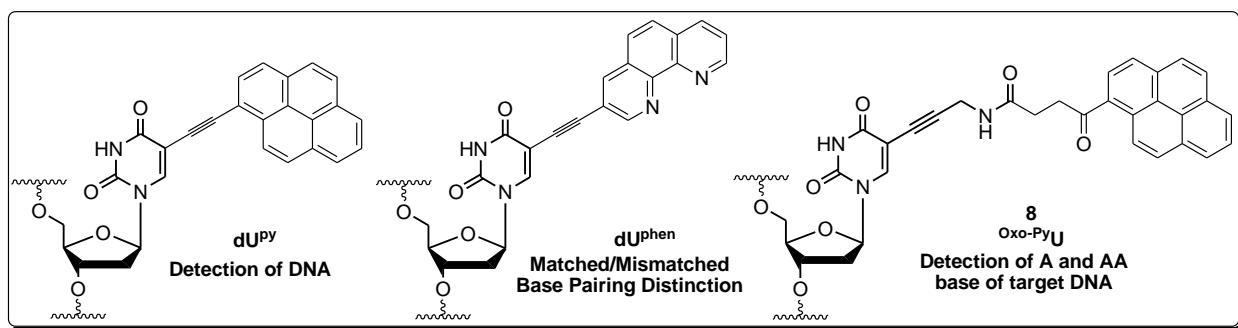


Figure 4.46: Examples of some fluorescently labeled Nucleoside/Oligonucleotide probes.

Bag, *et al.* engaged in designing labeled probe for SNPs detection via “Just Mix-and Read strategy based on the matched/mismatched and fluorescence read out strategy. The merits of homogeneous SNPs typing using labeled probe are: (a) Fluorescence labeling of the targets is not required, (b) hybridization condition is very simple-just mixing and no annealation is needed, (c) no



washing, drying or centrifugation step is necessary. Thus, this concept would be very important and useful for devising new SNP typing methods. Due to the fluorescence change of the labeled nucleobases, the bases on the complementary strands can be fluorometrically read out without any need of annealing, separation and washing steps. Therefore, this concept and principle of SNP typing method using labeled probes will constitute a very powerful homogeneous assay that does not require enzymes or time-consuming steps, and avoids hybridization errors. Furthermore, a combination of these probes also facilitates the SNP typing of a heterozygous sample. Thus homogeneous SNP typing method using labeled probes would be a powerful alternative to conventional SNP typing methods presently used.

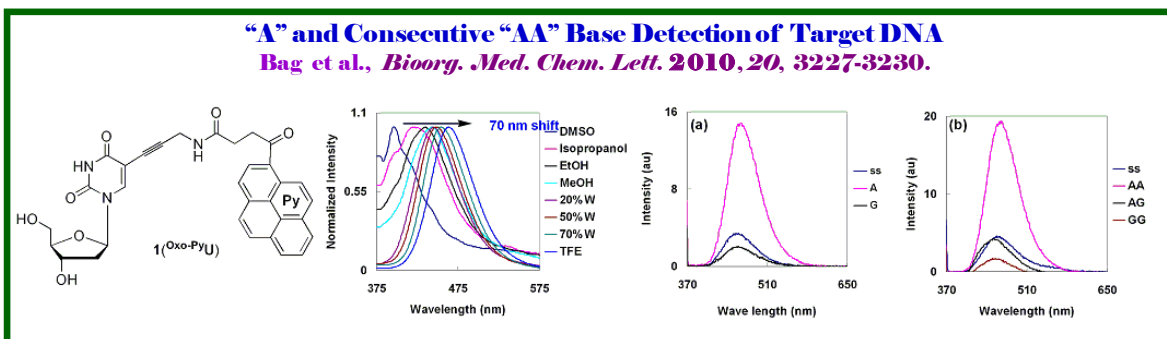


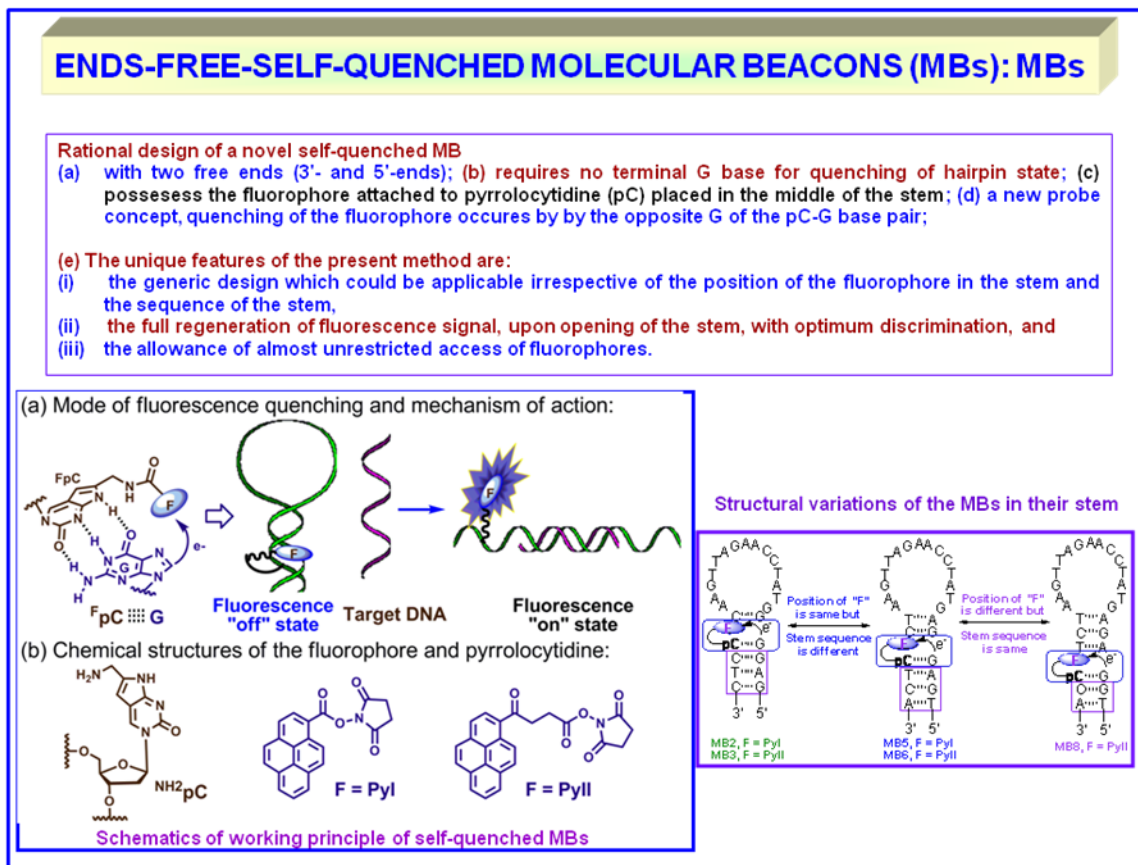
Figure 4.47: Oxo-pyrene labeled probe for DNA detection.

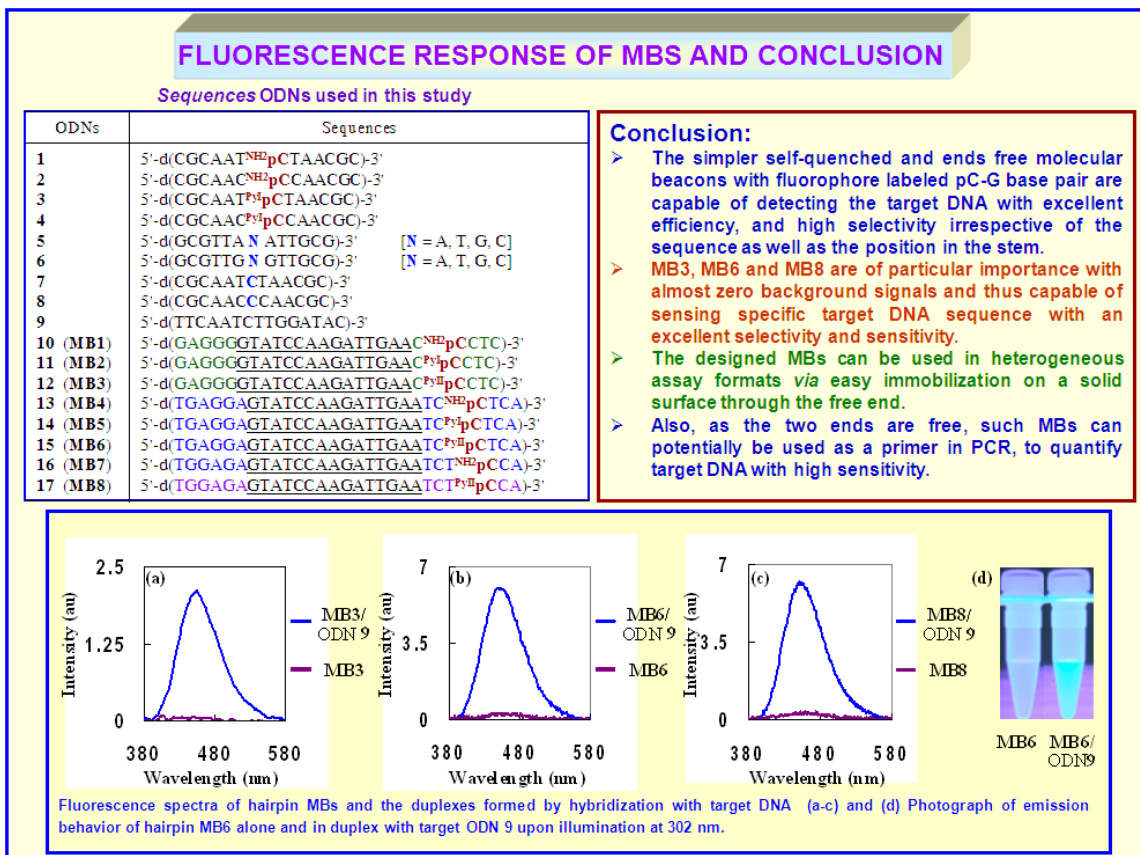
#### 4.20.6.3. DNA Detection Using Self-Quenched MB Probes

All the recent advances in the design of quencher free molecular beacon mainly based on the introduction of a fluorophore at the stem terminus. For efficient quenching of the hairpin state, these MBs also require at least one guanosine (G) base as a quencher at the opposite stem terminus, thus limiting their generality of the design and the use for multiplex detection. There has been no report so far for the design of quencher-free MBs having no terminal G but with the fluorophore in the middle of the stem and with two free ends (3'- and 5'-ends) for introducing other functionalities. Therefore, Saito *et al.* thought that it will be worthwhile and of greatest versatility if we can develop such a new type MBs having two free termini for further applications and modifications.

They have designed a novel self-quenched MB with two free ends (3'- and 5'-ends) that require no terminal G for quenching at hairpin state and with the fluorophore attached to pyrrolocytidine (pC) placed in the middle of the stem. The generation of signal “off” hairpin state originates from the quenching of the fluorophore by the opposite base G of the pC-G base pair due to their close proximity (**Figure 4.48**). The unique features of this method are; (a) the generic design which could be applicable irrespective of the position of the fluorophore in the stem and the sequence of the stem, (b) the full regeneration of fluorescence signal, upon opening of the stem, with optimum discrimination, and (c) the allowance of almost unrestricted access of fluorophores. Therefore, with its

simplicity, sensitivity, and specificity, this strategy holds a great promise in applications such as in heterogeneous assay and high-throughput DNA analysis.





**Figure 4.49: Fluorescence Response of quencher free molecular beacon probe in presence of different target DNA sequences.**

It was concluded from their findings that the fluorescence of the pyrene-hooked pyrrolocytidine is efficiently quenched by the opposite G. The sensitivity of designed MBs in response to the duplex formation with the target DNA was tested by the different signal generated from the fluorophores. Thus, their result shows that an efficient quenching of hairpin state lead to the fluorescence “off” state and upon duplex formation due to no contact with G, resulted in an efficient detection of the target DNA with full regeneration of fluorescence signal (**Figure 4.48**). It was also shown that PypC-G base pair could be used irrespective of the stem sequence and the position within the stem.

### 4.20.7. Microarray Based DNA Detection

#### 4.20.7.1. Introduction

A DNA microarray also known as gene chips, DNA chip, or biochip is a collection of microscopic DNA spots attached to a solid surface. DNA microarrays are used to measure the expression levels of large numbers of genes simultaneously or to genotype multiple regions of a genome. Each DNA spot contains picomoles (10–12 moles) of a specific DNA sequence, known as probes or reporters. A probe can be a short section of a gene or other short DNA sequence that are used to hybridize a complementary DNA (cDNA) or cRNA

sample that are called target under high-stringency conditions. The Probe-Target Hybridization event is usually detected by a quantifiable signal generation from the fluorophore-, silver-, or chemiluminescence-labeled targets to determine relative nucleic acid sequences in the target (**Figure 4.50**). Since an array can contain tens of thousands of probes, a microarray experiment can accomplish many genetic tests in parallel. Therefore arrays have dramatically accelerated many types of investigation in genomics.

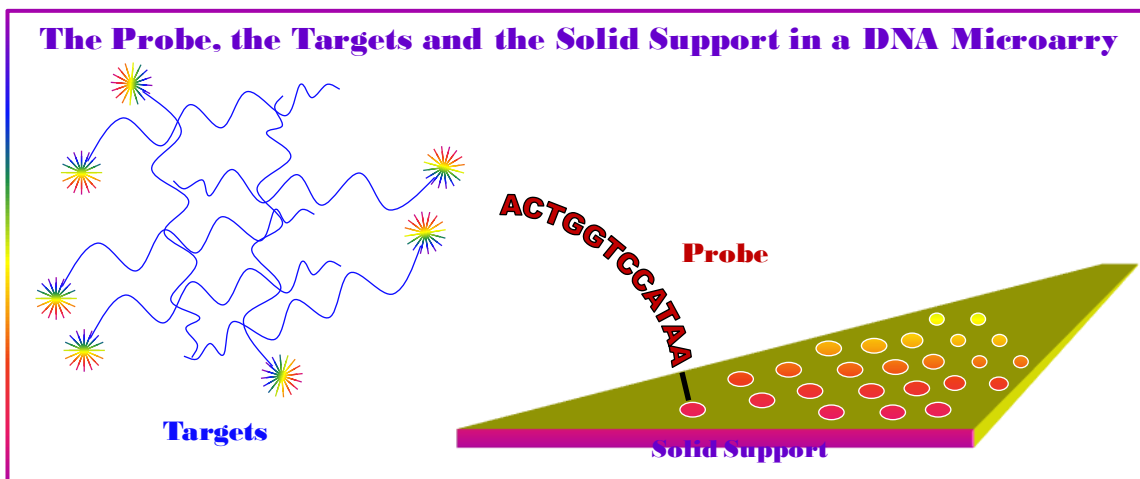


Figure 4.50: The definition of Target and Probe in a Microarray.

In standard microarrays, the probes are synthesized and then attached to a solid surface by a covalent bond to a chemical matrix such as epoxy-silane, amino-silane, lysine, polyacrylamide or others. The solid surface in general is a glass or a silicon chip, in which case they are collectively known as an Affy chip when an Affymetrix chip is used. Other microarray platforms, such as Illumina, use microscopic beads, instead of the large solid support. Alternatively, microarrays can be constructed by the direct synthesis of oligonucleotide probes on solid surfaces. DNA arrays are different from other types of microarray only in that they either measure DNA or use DNA as part of its detection system.

#### 4.20.7.2. Uses of Microarray in DNA Detection

- DNA microarrays can be used to measure changes in expression levels, to detect single nucleotide polymorphisms (SNPs), or to genotype or resequence mutant genomes.
- Revolution in the analysis of genetic information
- Hybridization is a highly parallel search by each molecule for matching partner on an affinity matrix.
- Specificity and affinity of complementary base pairing.
- Use of glass as a substrate, fluorescence for detection and the development of new technologies for synthesizing or depositing DNA have allowed the miniaturization of DNA arrays with increases in information content.

- "DNA chip"/ "DNA microarray" allows to collect more information about DNA sequences in an afternoon than an army of scientists could collect in years using earlier several techniques.
- DNA chips promise to carry the science of understanding genomes to a whole new level, and to bring tools for getting DNA-sequence information out of research labs into doctors' offices, the better to tailor-fit medical treatments to an individual's particular genetic makeup.
- **Principle Major technologies**
  - cDNA probes (> 200 nt), usually produced by PCR, attached to either nylon or glass supports
  - Oligonucleotides (25-80 nt) attached to glass support
  - Oligonucleotides (25-30 nt) synthesized in situ on silica wafers (Affymetrix)
  - Probes attached to tagged beads.

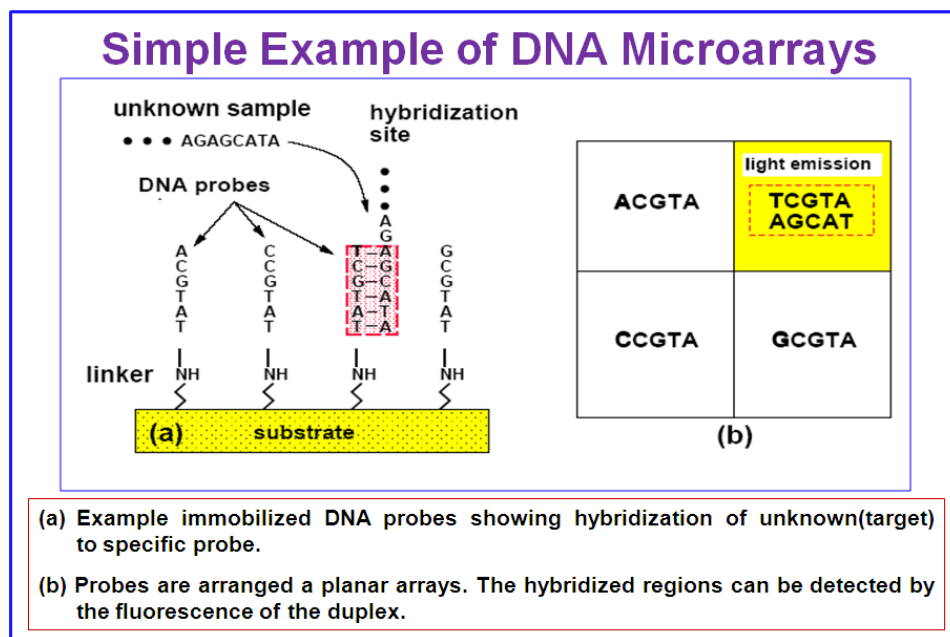
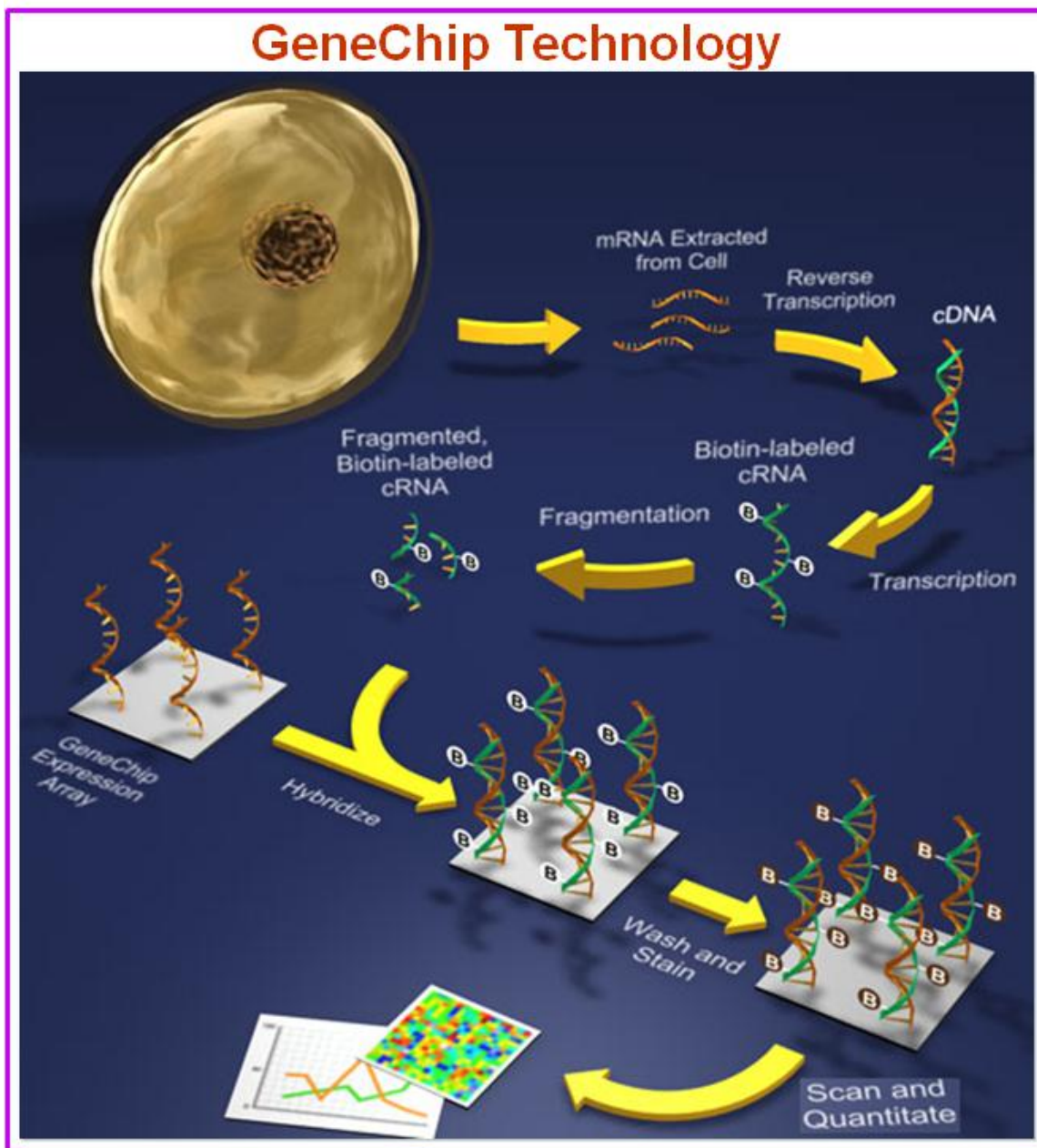


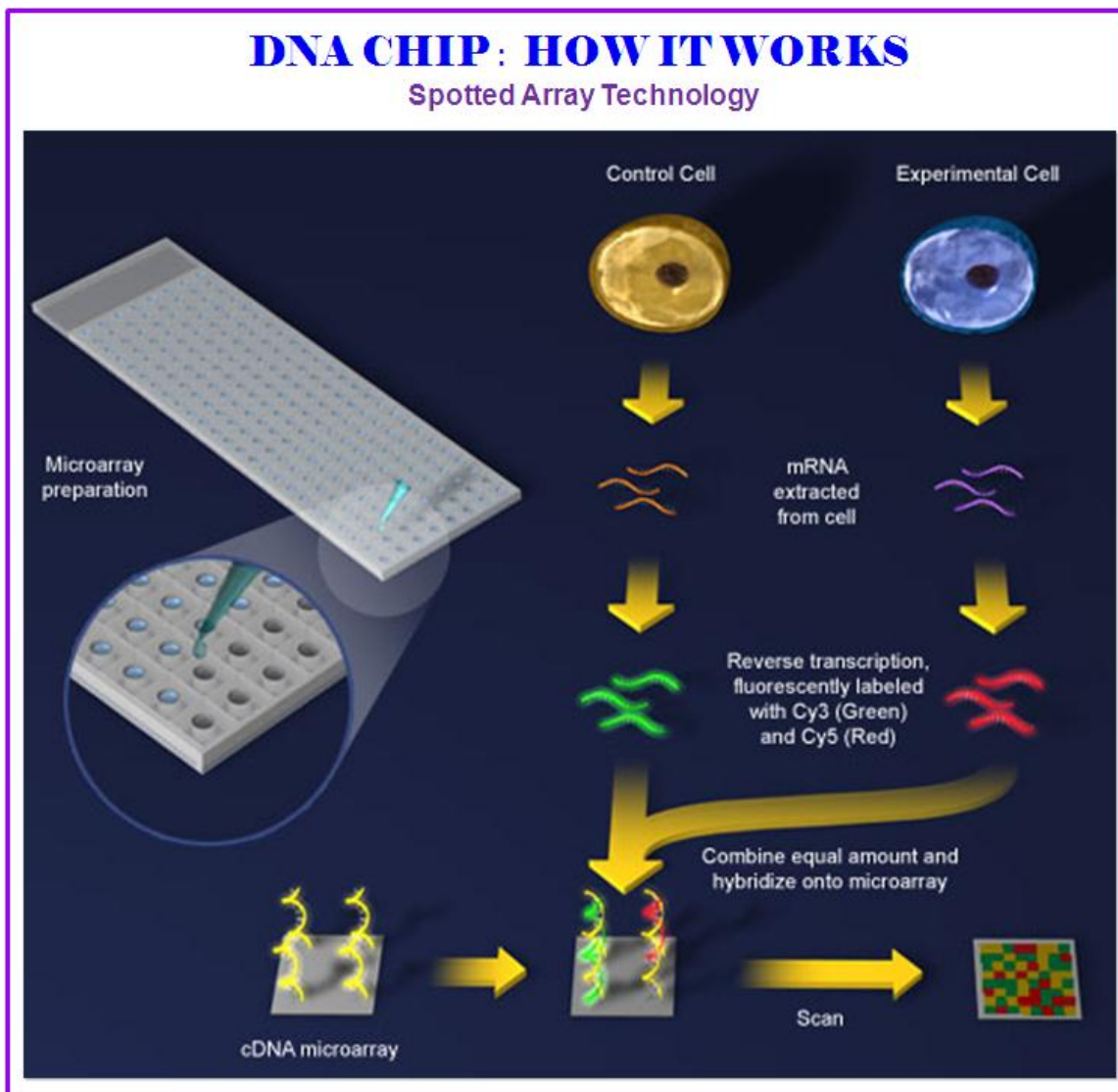
Figure 4.51: A simple example of DNA Microarray.

### 4.20.7.3. Examples Microarray Explaining Their Working Principles

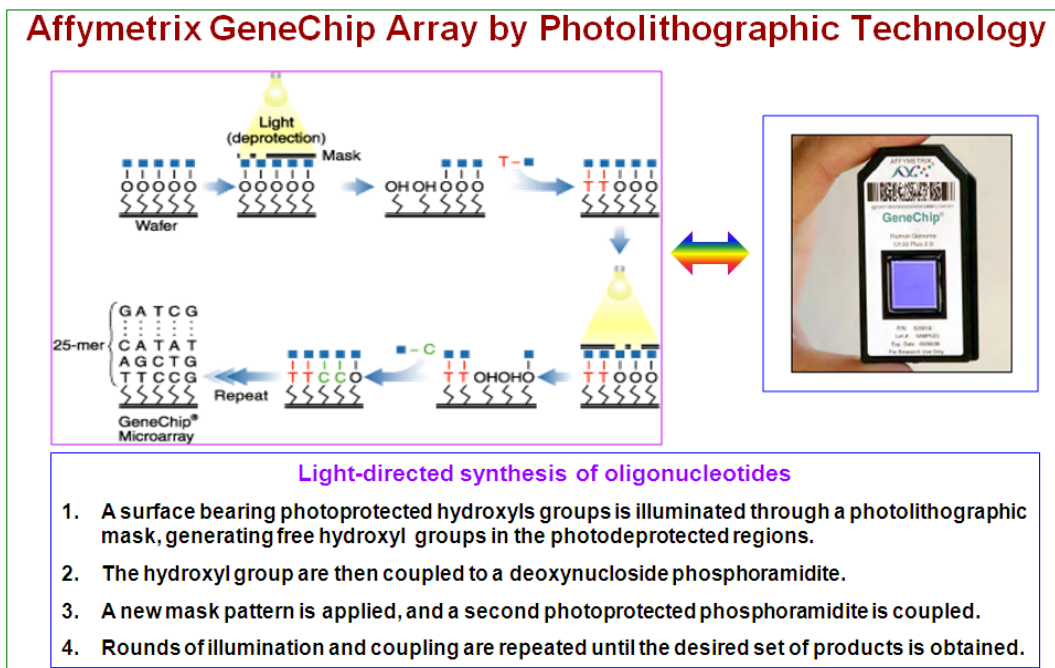
#### 4.20.7.3.1. The Gene Chip Technology







### 4.20.7.3.2. The Affymetrix Gene Chip Array Technology



### 4.20.7.4. The Principle of Microarray Technology

The main principle behind microarray technology is the hybridization between two DNA strands. The property of complementary nucleic acid sequences to specifically pair with each other by forming hydrogen bonds between complementary nucleotide base pairs is the main force of recognizing complementary partner. A high number of complementary base pairs in a nucleotide sequence allow tighter non-covalent bonding between the two strands. After washing off of non-specific bonding sequences, only strongly paired strands will remain hybridized. So fluorescently labeled target sequences that bind to a probe sequence generate a signal that depends on the strength of the hybridization determined by the number of paired bases, the hybridization conditions such as temperature, and washing after hybridization. Total strength of the signal, from a spot, depends upon the amount of target sample binding to the probes present on that spot. Microarrays use relative quantification in which the intensity of a spot is compared to the intensity of the same spot under a different condition, and the identity of the spot is known by its position. The figure below explains the steps involved in a microarray.



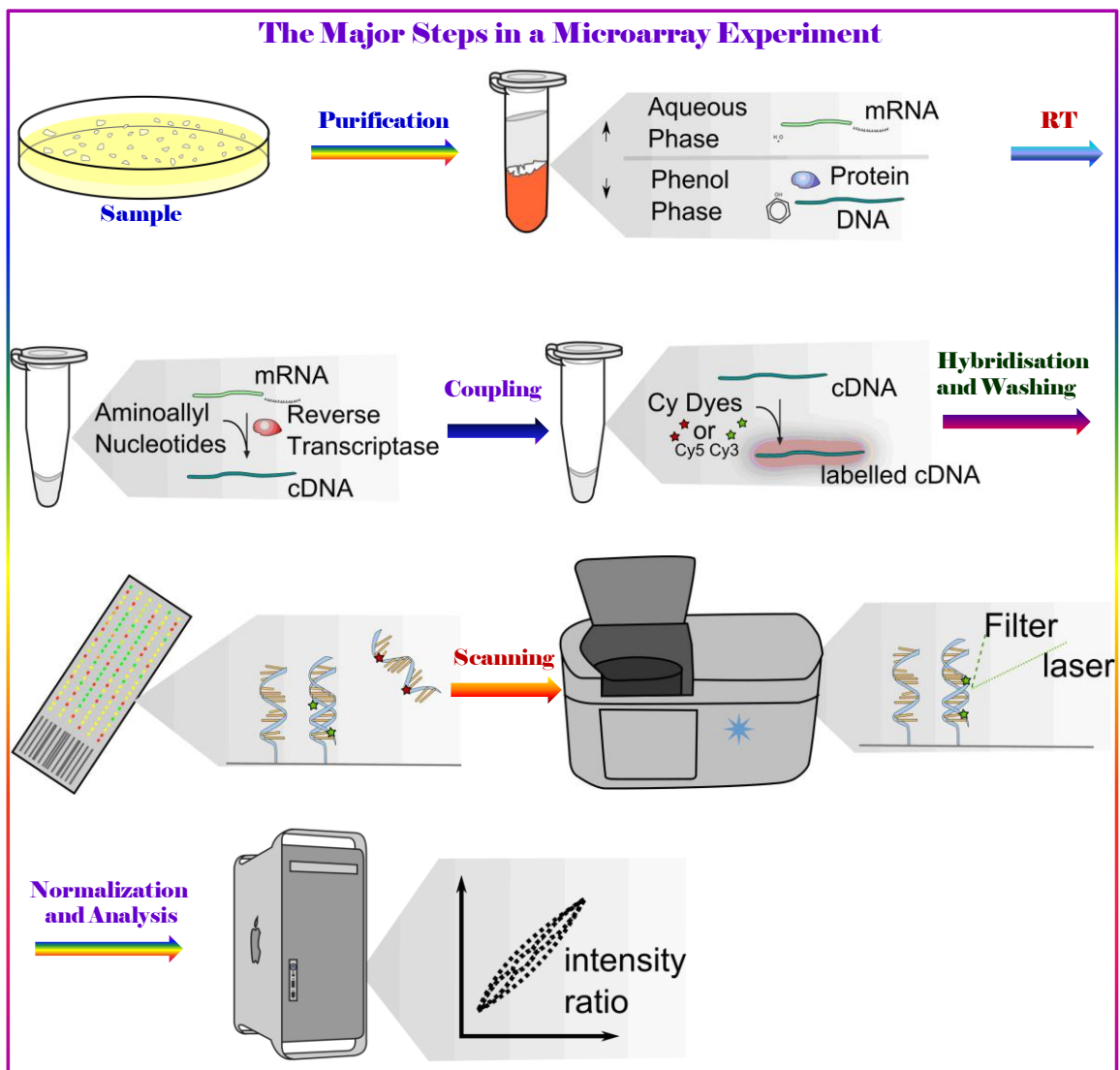


Figure 4.52: The major step involved in a microarray experiment

## 4.20.7.5. A Labeled Free Strategy as an Alternative Microarray Technology

**SEARCH FOR AN ALTERNATIVE DNA CHIP**

**Traditional chip based detection techniques**

● = Fluo. Tag- Cy3, Cy5 etc.-----

**DNA CHIP MAKERS**

Several companies—Affymetrix, Agilent Technologies, Amesham BioSciences, BD Clontech, Illumina

**WHY TO SEARCH FOR AN ALTERNATIVE?**

**Traditional chip based detection techniques Facing Problems—**

1. Require PCR and Fluorescence labeling of Targets
2. Thus, mishybridization
3. Stringent hybridization conditions
4. Background errors
5. SNPs detection is not easy
6. High cost, time consuming

Thus, it will be worthwhile if one can produce a technique in which a labeled probe is capable of detecting unlabeled target ODNs in chip based techniques.

**Microarray with Unlabeled Target-An Alternative Efficient Method**

**Traditional Method-Target is Labeled**

• Labeling is indispensable

• Mishybridization produces large errors

**BDF Method- Probe is Labeled**

• Labeling is not required

• Mishybridization produce only small S/N errors

## 4.21. Basics of Peptide Nucleic Acids

### 4.21.1. Introduction

Peptide Nucleic Acid (PNA) an artificially created DNA analogue was invented by Dr. Nielsen, Egholm, Berg, and Buchardt in 1991. The phosphate ribose ring of DNA was replaced with the polyamide backbone in PNA. Despite a radical structural change, PNA is capable of sequence-specific binding in a helix form to its complementary DNA or RNA sequence. Due to its superior binding affinity and chemical/biological stability, PNA has been widely applied in the field of biology.

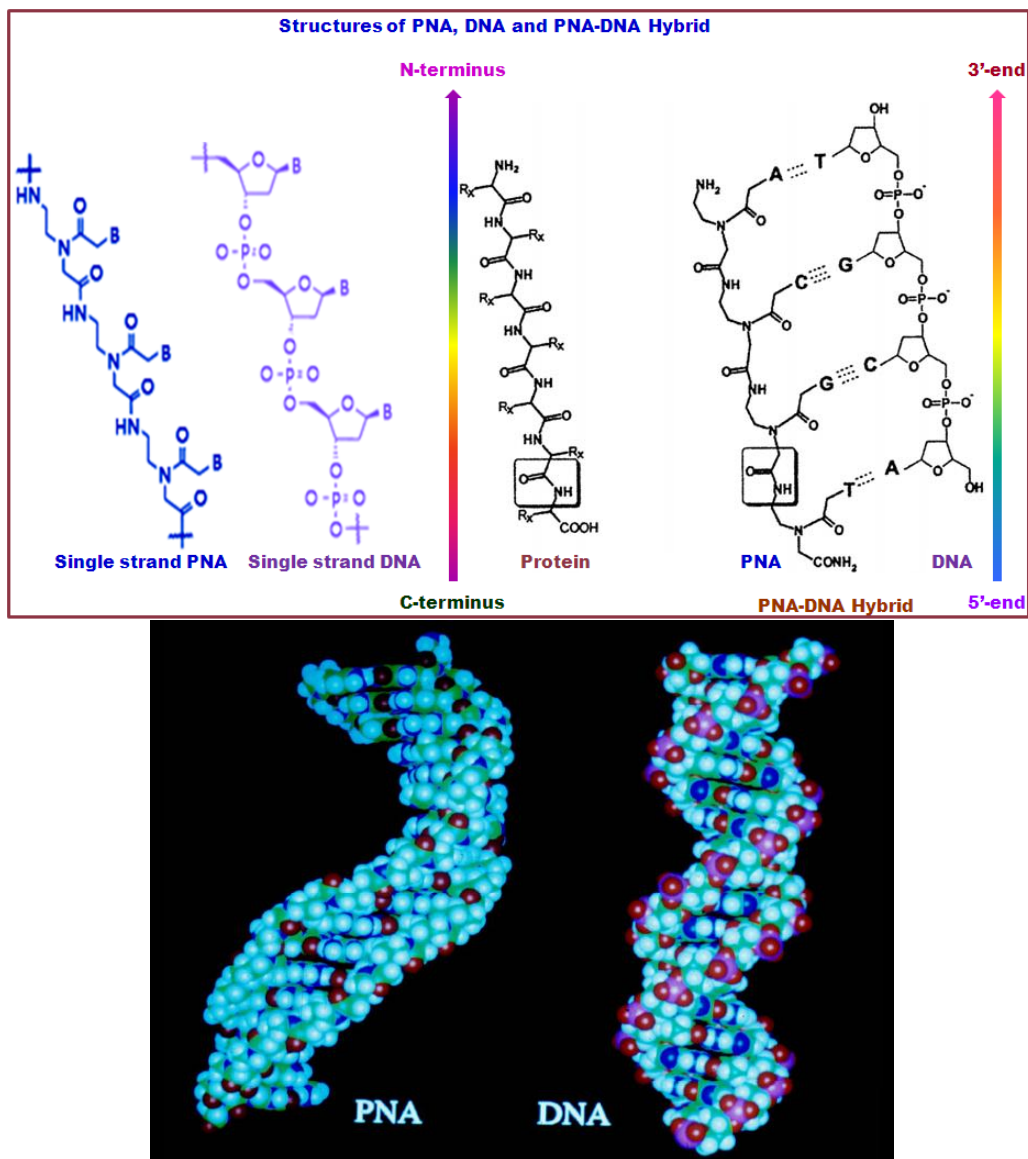


Figure 4.53: The Structures of PNA, DNA, and Protein.

### 4.21.2. Main Features of PNA

The main features of PNA are:

- High binding affinity to its complementary DNA or RNA
- Differentiation of single-base mismatch by high destabilizing effect
- High chemical stability to temperature and pH
- High biological stability to nuclease and protease
- Salt independence during hybridization with DNA sequence
- Triplex formation with continuous homopurine DNA

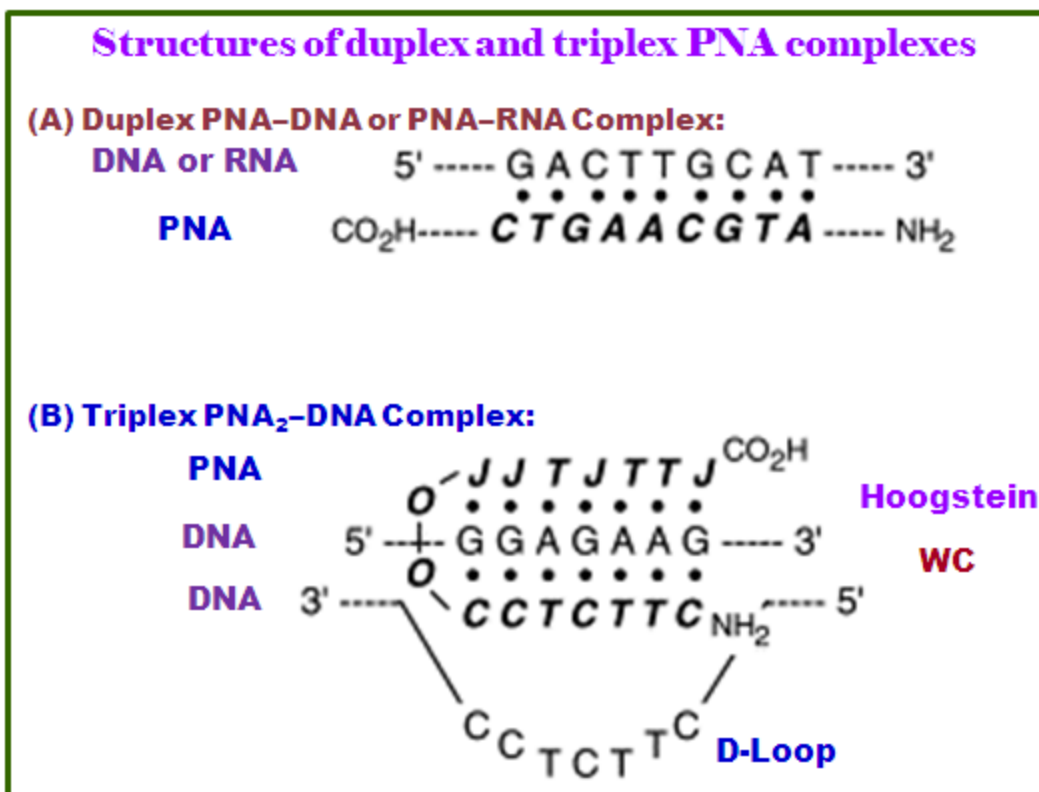


Figure 4.54: PNA binding modes for targeting double stranded DNA.

### 4.21.3. Comparison of PNA and DNA

Criteria	PNA	DNA
Hybridization affinity with DNA	At least 1 °C higher $T_m$ per base	---
Hybridization rate with DNA	100 - 5000 times faster	---
Salt concentration for hybridization	Independent	Dependent
$T_m$ for each single mismatch	Lowering 15°C	Lowering 10°C
Chemical stability	Stable	Unstable or moderate
Biological stability	Stable to nuclease and protease	Degradation by nuclease
Thermal stability	Good	Moderate
Water solubility	Restricted solubility enhanced by use of appropriate linkers	Soluble
Probe length for diagnostic use	13 - 18 bases	20-30 bases

### 4.21.4. Application of PNA

- miRNA Inhibitors:** PNA has stronger binding affinity to its complementary RNA than to DNA as DNA does. This principle can be exactly applied to miRNA to inhibit its activity. This has been experimentally proved.
- Bio-drugs for Antigene and Antisense therapy:** The stronger binding properties and biological stability of PNA imply that a small quantity of PNA can be effective for therapeutic applications. Triplex invasion of a PNA shows good potential as antigene material. As a third generation molecule in antisense therapy, there has been experimental data that shows good effect *in vitro* and *in vivo*.
- Molecular tools for molecular biology and functional genomics:** PNA is used as tools of molecular biology and functional genomics research. PNAs can be used in many of the applications where traditional synthetic DNA or RNA have been used, but with the added benefits of tighter binding, greater specificity and stability.

- **Molecular probes for diagnostics and detection:** Binding properties of PNA give more specific, more sensitive, and more accurate result for the detection of target sequences. Some of the commercialized PNA products available on the market are probes that detect genetic disease and or detect viral or bacterial infections.
- **SNP detection:** PNA probe can be used for SNP detection.
- **Biosensor:** PNA probe are also used for nucleic acid biosensor

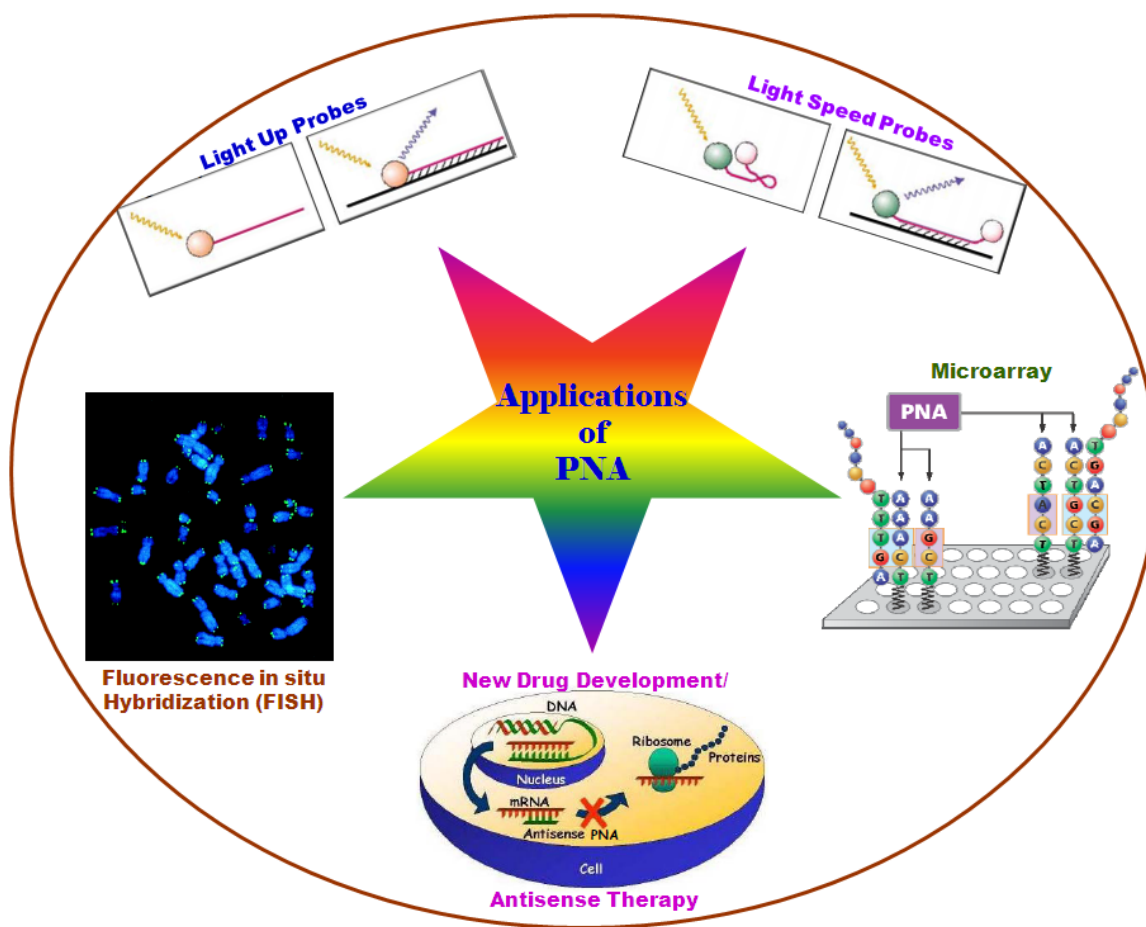
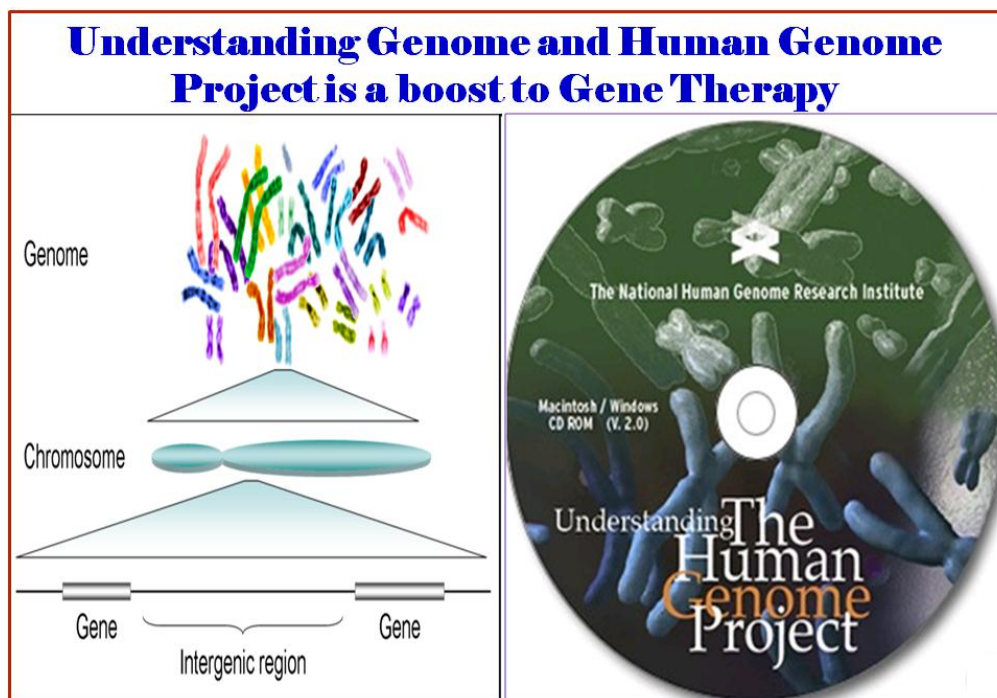


Figure 4.54: Schematic presentation of different applications of PNA.

## 4.22. Concepts of Antigen/Antisense Therapy



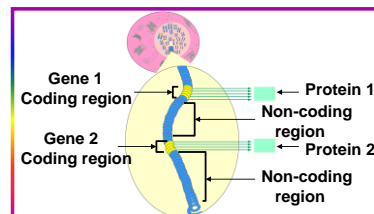


## 4.22.1. Introducing Gene and Gene Defects

### Introducing Gene and Gene Defects

#### What are Genes:

- No two individuals are same in this world, not even identical twins. Large variation in phenotypic characters.
- A gene is a part (A coding region) of DNA molecule, and humans have about 30,000 genes.
- Gene is biological unit of heredity. Gene determine obvious traits, such as hair and eye color, as well as more subtle characteristics such as the ability of the blood to carry oxygen.
- Genes carry 'instructions' that allow the cells to produce specific functional proteins.

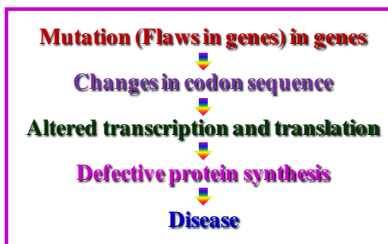
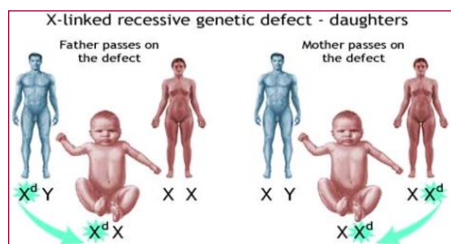


#### What Genes can do?

- Genes, which are carried on chromosomes, are the basic physical and functional units of heredity. Genes are specific sequences of bases that encode instructions on how to make proteins. It's the proteins that perform most life functions and even make up the majority of cellular structures.

#### Why Genetic Disorders?

- When genes are altered so that the encoded proteins are unable to carry out their normal functions, genetic disorders can result.



#### All of us carry some defective Genes, some are apparent and many in apparent :

- Each of us carries about half a dozen defective genes. We remain blissfully unaware of this fact unless we, or one of our close relatives, are amongst the many millions who suffer from a genetic disease. About one in ten people has, or will develop at some later stage, an inherited genetic disorder, and approximately 2,800 specific conditions are known to be caused by defects (mutations) in just one of the patient's genes.

#### We Inherit from Parents:

- Most of us do not suffer any harmful effects from our defective genes because we carry two copies of nearly all genes, one derived from our mother and the other from our father. The only exceptions to this rule are the genes found on the male sex chromosomes. Males have one X and one Y chromosome, the former from the mother and the latter from the father, so each cell has only one copy of the genes on these chromosomes.

#### Law of Inheritance:

- In the majority of cases, one normal gene is sufficient to avoid all the symptoms of disease. If the potentially harmful gene is recessive, then its normal counterpart will carry out all the tasks assigned to both. Only if we inherit from our parents two copies of the same recessive gene will a disease develop.



## 4.22.2. What is Gene Therapy? Why is It Used?

### What is gene therapy? Why is it used?

- **Gene therapy** is the insertion of genes into an individual's cells and tissues to treat a disease, such as a hereditary disease in which a deleterious mutant allele is replaced with a functional one. Although the technology is still in its infancy, it has been used with some success.
- It is the introduction of normal genes into cells that contain defective genes to reconstitute a missing protein product
- It is used to correct a deficient phenotype so that sufficient amounts of a normal gene product are synthesized → to improve a genetic disorder
- It is a technique for correcting defective genes that are responsible for disease development.



- The most common form of gene therapy involves the insertion of functional genes into an unspecified genomic location in order to replace a mutated gene, but other forms involve directly correcting the mutation or modifying normal gene that enables a viral infection.
- Although the technology is still in its infancy, it has been used with some success.
- **Gene Therapy is Experimental:** Advances in understanding and manipulating genes have set the stage for scientists to alter a person's genetic material to fight or prevent disease. Gene therapy is an experimental treatment that involves introducing genetic material (DNA or RNA) into a person's cells to fight disease.
- **Majority are Trails:** Gene therapy is being studied in clinical trials (research studies with people) for many different types of cancer and for other diseases. It is not currently available outside a clinical trials.

### **What is Gene Therapy? Why is it used?**

#### **Introduction:**

- Gene therapy is the insertion of genes into an individual's cells and tissues to treat a disease, such as a hereditary disease in which a deleterious mutant allele is replaced with a functional one. Although the technology is still in its infancy, it has been used with some success.
- It is the introduction of normal genes into cells that contain defective genes to reconstitute a missing protein product
- It is used to correct a deficient phenotype so that sufficient amounts of a normal gene product are synthesized → to improve a genetic disorder
- It is a technique for correcting defective genes that are responsible for disease development.
- The most common form of gene therapy involves the insertion of functional genes into an unspecified genomic location in order to replace a mutated gene, but other forms involve directly correcting the mutation or modifying normal gene that enables a viral infection.

#### **Gene Therapy is Experimental:**

- Advances in understanding and manipulating genes have set the stage for scientists to alter a person's genetic material to fight or prevent disease. Gene therapy is an experimental treatment that involves introducing genetic material (DNA or RNA) into a person's cells to fight disease.

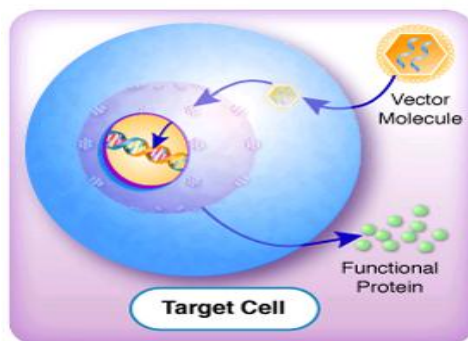
#### **Majority are Trails:**

- Gene therapy is being studied in clinical trials (research studies with people) for many different types of cancer and for other diseases. It is not currently available outside a clinical trials.

### What is Gene Therapy? Why is it used?.....Contd.

#### Achievement of Gene therapy:

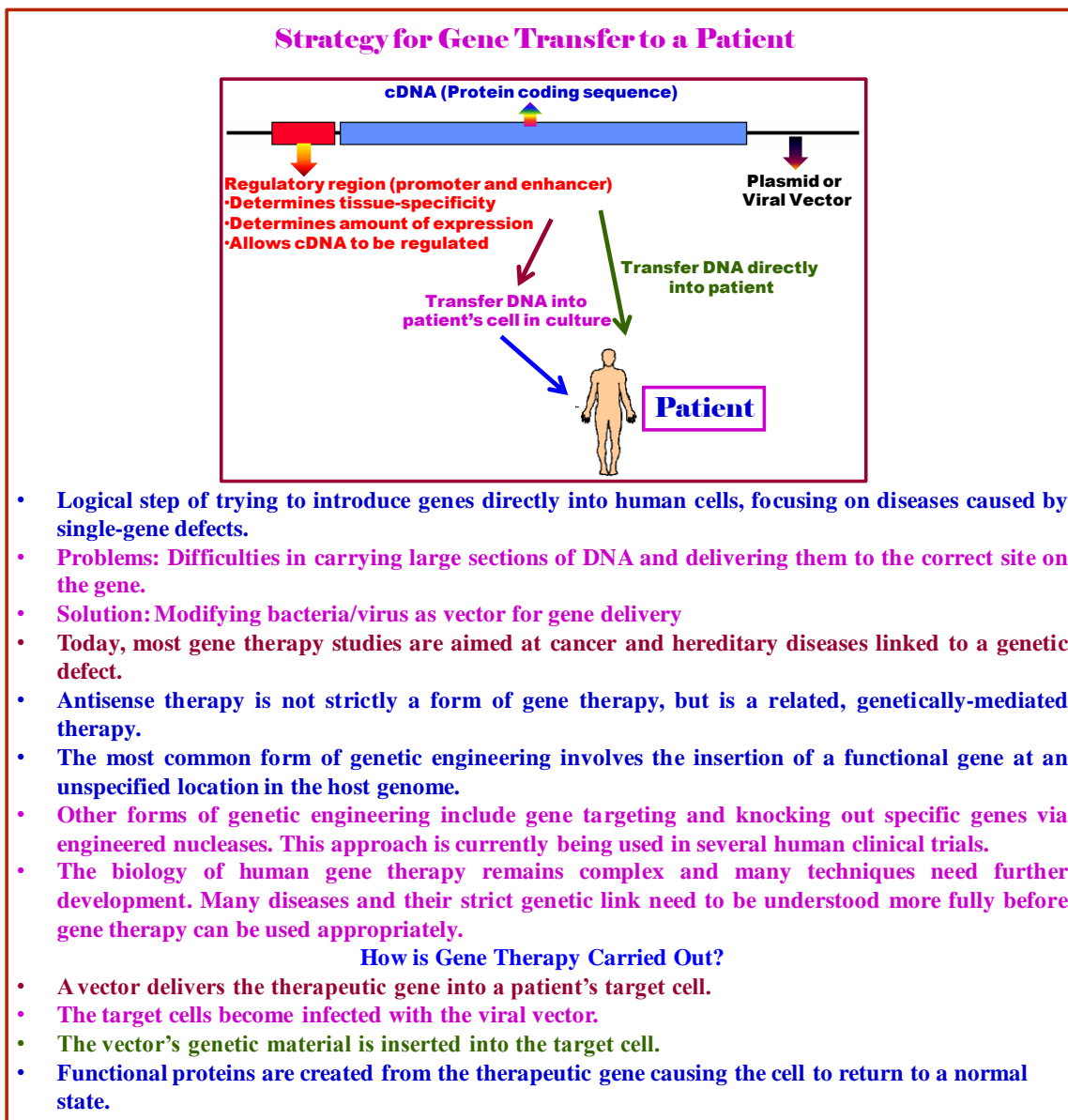
- Replacing a mutated gene that causes disease with a healthy copy of the gene.
- Inactivating, or “knocking out,” a mutated gene that is functioning improperly.
- Introducing a new gene into the body to help fight a disease.



#### Making the new Genetic Material Functional:

- Gene that is inserted directly into a cell usually does not function. Instead, a carrier called a vector is used to introduce the therapeutic gene into the patient's target cells. The most common vector that is used is a virus that has been genetically altered to carry normal human DNA. Viruses cause diseases in humans by encapsulating and delivering the genes into cells.

### 4.22.3. Strategy of Gene Transfer



#### 4.22.4. Different Facets of Gene Therapy

##### **Gene Therapy: Its Different Facets**

##### **Uses of gene Therapy:**

- **Replace missing or defective genes.**
- **Deliver genes that speed the destruction of cancer cells.**
- **Supply genes that cause cancer cells to revert back to normal cells.**
- **Deliver bacterial or viral genes as a form of vaccination.**
- **Provide genes that promote or impede the growth of new tissue.**
- **Deliver genes that stimulate the healing of damaged tissue.**

##### **Genes are Medicine:**

- **Gene therapy is ‘the use of genes as medicine’. It involves the transfer of a therapeutic or working gene copy into specific cells of an individual in order to repair a faulty gene copy. Thus it may be used to replace a faulty gene, or to introduce a new gene whose function is to cure or to favourably modify the clinical course of a condition.**

##### **Goal of Gene Therapy:**

- **A normal gene may be inserted into a non-specific location within the genome to replace a non-functional gene. This approach is most common.**
- **An abnormal gene could be swapped for a normal gene through homologous recombination.**
- **The abnormal gene could be repaired through selective reverse mutation, which returns the gene to its normal function.**
- **The regulation (the degree to which a gene is turned on or off) of a particular gene could be altered.**

##### **Principles of Gene therapy:**

- **A normal gene may be inserted into a non-specific location within the genome to replace a non-functional gene. This approach is most common.**
- **An abnormal gene could be swapped for a normal gene through homologous recombination.**
- **The abnormal gene could be repaired through selective reverse mutation, which returns the gene to its normal function.**
- **The regulation (the degree to which a gene is turned on or off) of a particular gene could be altered.**

##### **Gene Therapy Depends on Delivery of Corrective Genes:**

- **Viral vectors are a tool commonly used by molecular biologists to deliver genetic material into cells. This process can be performed inside a living organism (*in vivo*) or in cell culture (*in vitro*). Viruses have evolved specialized molecular mechanisms to efficiently transport their genomes inside the cells they infect.**

#### 4.22.5. Categories of Gene Therapy: Somatic and Germ Line Gene Therapy

##### Categories of Gene Therapy: Somatic and Germ Line Gene Therapy

- Gene therapy can target somatic (body) or germ (egg and sperm) cells. In somatic gene therapy the recipient's genome is changed, but the change is not passed on to the next generation; whereas with germ line gene therapy the newly introduced gene is passed on to the offspring.

##### Three categories of somatic cell gene therapy:

1. *Ex vivo* – cells removed from body, incubated with vector and gene-engineered cells returned to body.

- Example of *ex vivo* somatic cell gene therapy
- Usually done with blood cells because they are easiest to remove and return.
- Sickle cell anemia.

2. *In situ* – vector is placed directly into the affected tissues.

- Example of *in situ* somatic cell gene therapy
- Infusion of adenoviral vectors into the trachea and bronchi of cystic fibrosis patients.
- Injection of a tumor mass with a vector carrying the gene for a cytokine or toxin.
- Injection of a dystrophin gene directly into the muscle of muscular dystrophy patients.

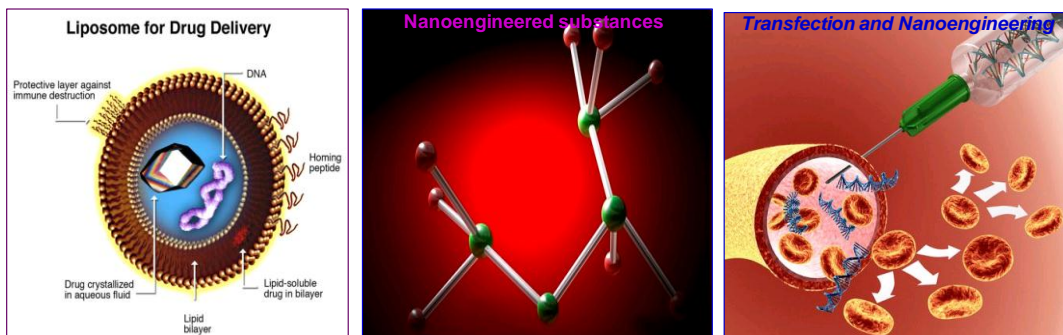
3. *In vivo* – vector injected directly into the blood stream.

- Example of *in-vivo* somatic cell gene therapy
- No clinical examples.
- *In vivo* injectable vectors must be developed.

## 4.22.6. Types of Vectors Used in Gene Therapy

### Types of Vectors for Gene Therapy

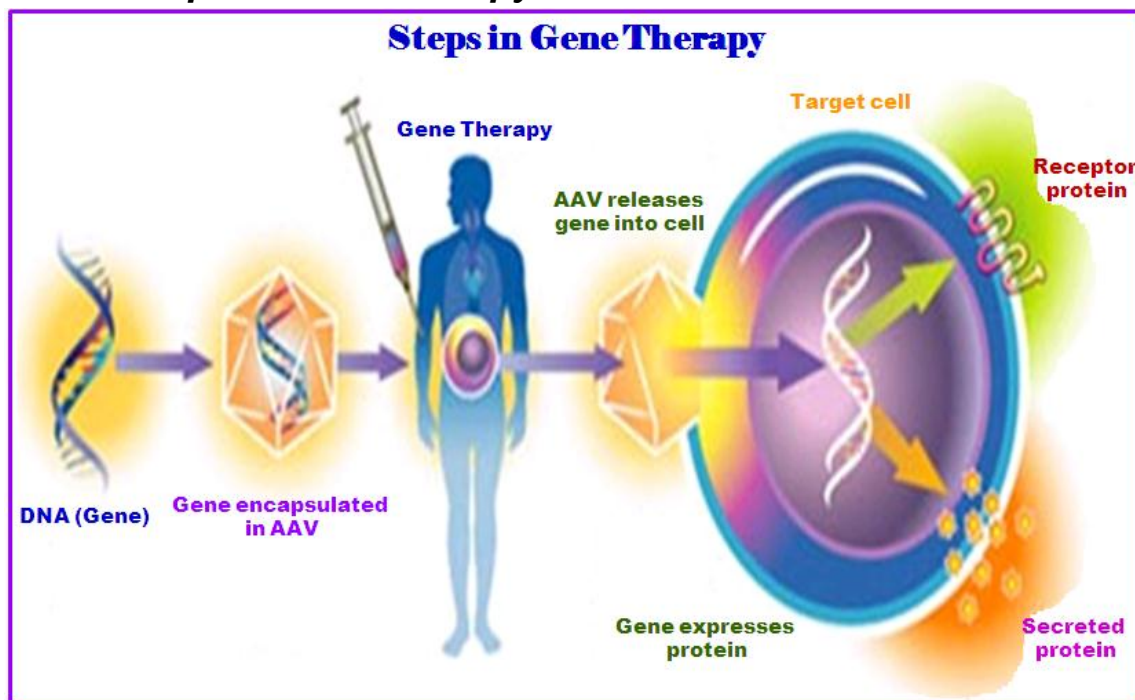
- **RNA viruses (Retroviruses):-** (a) Murine leukemia virus (MuLV); (b) Human immunodeficiency viruses (HIV); (c) Human T-cell lymphotropic viruses (HTLV).
- **DNA viruses:-** (a) Adenoviruses; (b) Adeno-associated viruses (AAV); (c) Herpes simplex virus (HSV); (d) Pox viruses.
- **Non-viral vectors:-** (a) Liposomes; (b) Naked DNA; (c) Liposome-polycation complexes; (d) Peptide delivery systems.
  - **Nonviral approach involves the creation of an artificial lipid sphere with an aqueous core. This liposome, which carries the therapeutic DNA, is capable of passing the DNA through the target cell's membrane.**
  - **Liposomes: DNA/lipid complexes are easy to prepare and there is no limit to the size of genes that can be delivered. Because carrier systems lack proteins, they may evoke much less immunogenic responses. More importantly, the cationic lipid systems have much less risk of generating the infectious form or inducing tumorigenic mutations because genes delivered have low integration frequency and cannot replicate or recombine.**



- **Nanoengineered substances:** Nonviral substances such as Ormosil have been used as DNA vectors and can deliver DNA loads to specifically targeted cells in living animals. (Ormosil stands for *organically modified silica or silicate*).
- **Transfection and Nanoengineering:** Transfection is the process of introducing nucleic acids into cells by non-viral methods. The term "transformation" is preferred to describe non-viral DNA transfer in bacteria and non-animal eukaryotic cells; "transduction" is often used to describe virus-mediated DNA transfer.



### 4.22.7. Steps in Gene Therapy



### 4.22.8. Uses of Gene Therapy

#### Uses of gene therapy:

- Replace missing or defective genes;
- Deliver genes that speed the destruction of cancer cells;
- Supply genes that cause cancer cells to revert back to normal cells;
- Deliver bacterial or viral genes as a form of vaccination;
- Provide genes that promote or impede the growth of new tissue; and;
- Deliver genes that stimulate the healing of damaged tissue.

#### Genes are Medicine

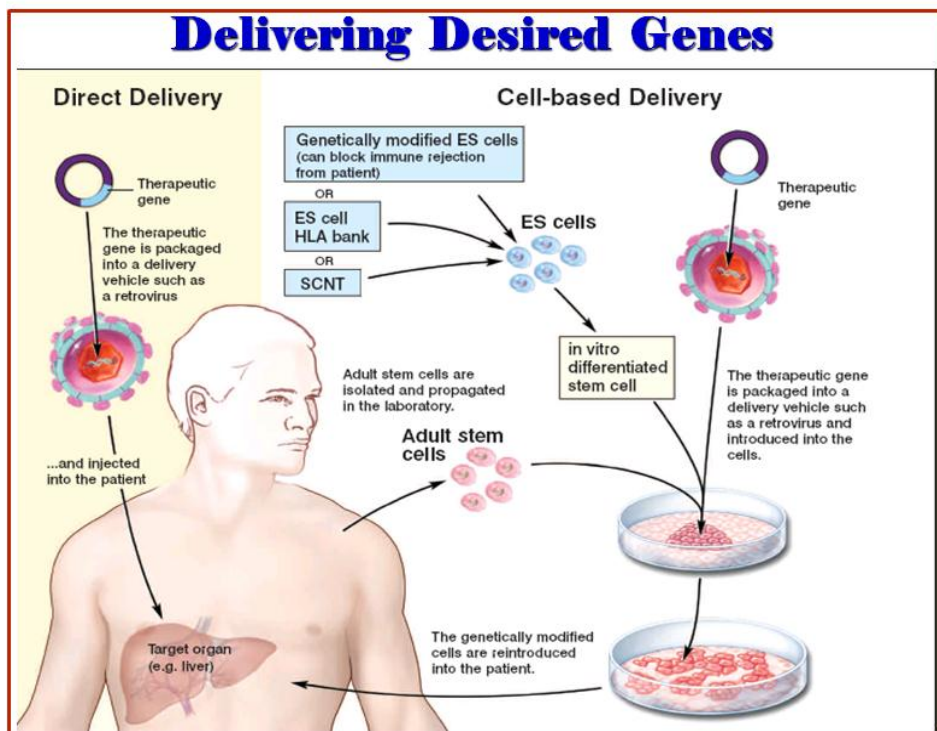
- **Gene therapy is 'the use of genes as medicine'. It involves the transfer of a therapeutic or working gene copy into specific cells of an individual in order to repair a faulty gene copy. Thus it may be used to replace a faulty gene, or to introduce a new gene whose function is to cure or to favourably modify the clinical course of a condition.**

#### Goal of Gene therapy

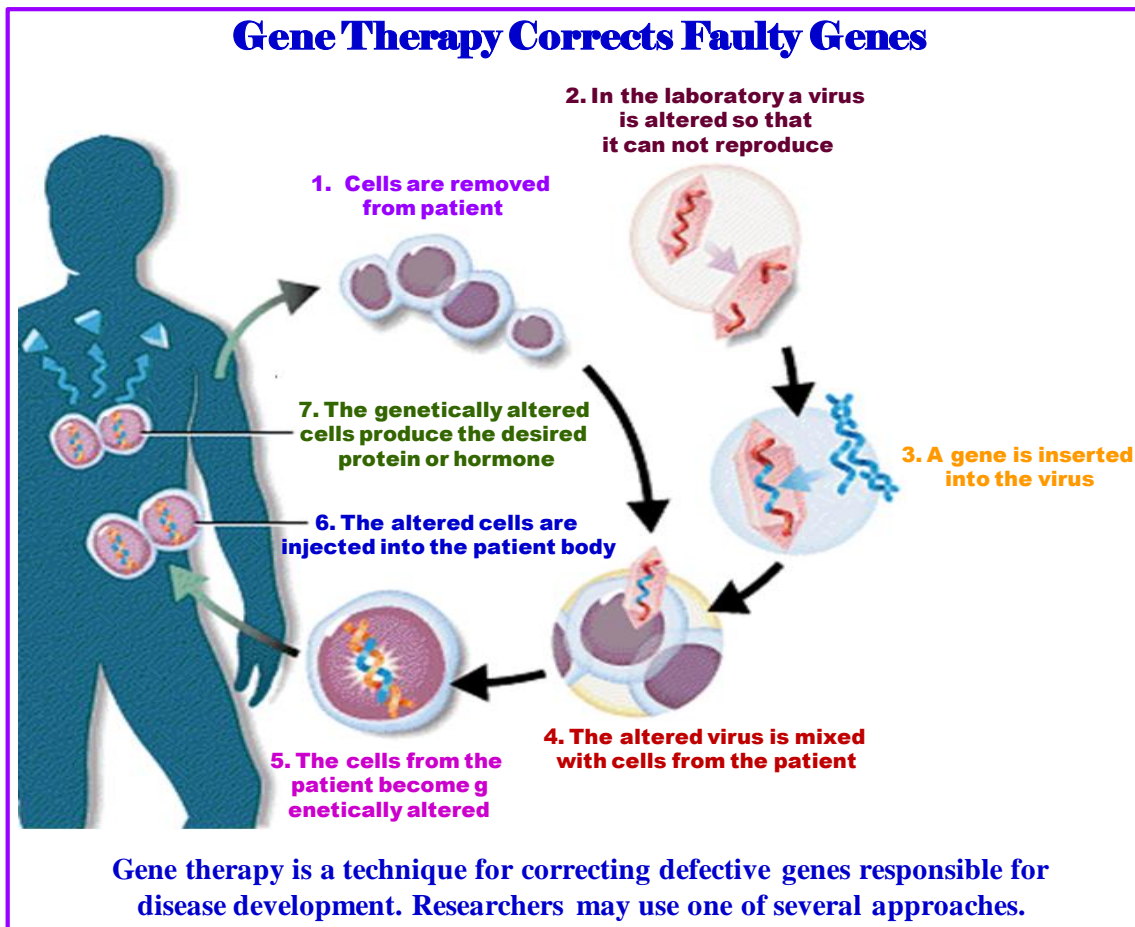
- A normal gene may be inserted into a non-specific location within the genome to replace a non-functional gene. This approach is most common.
- An abnormal gene could be swapped for a normal gene through homologous recombination.
- The abnormal gene could be repaired through selective reverse mutation, which returns the gene to its normal function.
- The regulation (the degree to which a gene is turned on or off) of a particular gene could be altered.



### 4.22.9. Gene Therapy Delivers Desired Genes



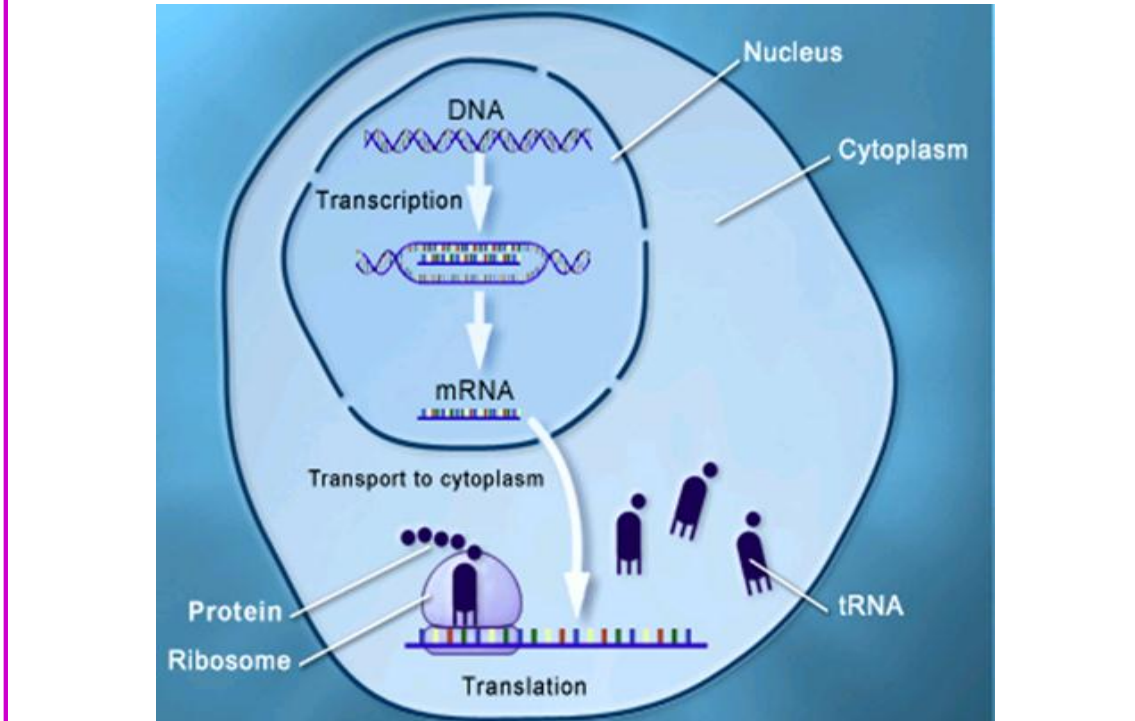
### 4.22.10. Gene Therapy Corrects Faulty Genes



#### 4.22.11. Gene Therapy Delivers Protein

### Gene Therapy delivers Proteins

Today, gene therapy is the ultimate method of protein delivery, in which the delivered gene enters the body's cells and turns them into small "factories" that produce a therapeutic protein for a specific disease over a prolonged period.



## 4.22.12. Problems of Gene Therapy

### Limitations/Problems of Gene Therapy

#### Gene control/regulation:

- Most viral vectors are unable to accommodate full length human genes containing all of their original regulatory sequences
- Human cDNA often used → much regulatory information is lost (e.g. enhancers inside introns).
- Often promoters are substituted → therefore gene expression pattern may be very different
- Random integration can adversely affect expression (insertion near highly methylated heterogeneous DNA may silence gene expression).

#### Expense:

- Costly because of cell culturing needs involved in *ex vivo* techniques
- Virus cultures for *in vivo* delivery
- Usually the number of patients enrolled in any given trial is <20
- More than 5000 patients have been treated in last ~12 years worldwide

#### Risks associated with current gene therapy:

- Viruses can infect more than one type of cells. Viral vectors may alter more than the intended cells. Or the new gene might be inserted into the wrong location in the DNA, causing cancer or other damage.
- When DNA is injected directly into a tumor there is a chance that some DNA could be introduced into germ cells, producing inheritable changes.
- The gene might be over-expressed (toxicity); the viral vector could cause inflammation or immune reaction; the virus could be transmitted to other individuals or the environment.

#### Ethical issues and Laws:

- How can “good” and “bad” uses of gene therapy be distinguished?
- Who decides which traits are normal and which constitute a disability or disorder?
- Will the high costs of gene therapy make it available only to the wealthy?
- Could the widespread use of gene therapy make society less accepting of people who are different?
- Should people be allowed to use gene therapy to enhance basic human traits such as height, intelligence, or athletic ability?

### **Limitations/Problems of Gene Therapy**

#### **Gene delivery:**

- **Limited tropism of viral vectors**
- **Dependence on cell cycle by some viral vectors (i.e. mitosis required)**

#### **Limitation of Direct Gene Induction:**

- **The simplest method is the direct introduction of therapeutic DNA into target cells. This approach is limited in its application because it can be used only with certain tissues and requires large amounts of DNA.**

#### **Duration of gene activity:**

- **Non-integrating delivery will be transient (transient expression)**
- **Integrated delivery will be stable**

#### **Short Lived:**

- **Hard to rapidly integrate therapeutic DNA into genome and rapidly dividing nature of cells prevent gene therapy from long time**
- **Would have to have multiple rounds of therapy**

#### **Patient safety (Viral Vectors):**

- **Immune hyperresponsiveness (hypersensitivity reactions directed against viral vector components or against transgenes expressed in treated cells)**
- **Integration is not controlled → oncogenes may be involved at insertion point → cancer?**
- **Patient could have toxic, immune, inflammatory response. Also may cause disease once inside**

#### **Immune Response:**

- **New things introduced leads to immune response**
- **Increased response when a repeat offender enters**

#### **Multigene Disorders:**

- **Heart disease, high blood pressure, Alzheimer's, arthritis and diabetes are hard to treat because you need to introduce more than one gene**
- **May induce a tumor if integrated in a tumor suppressor gene because insertional mutagenesis.**

### **Limitations of Gene Therapy**

- **Gene delivery**
  - **Limited tropism of viral vectors**
  - **Dependence on cell cycle by some viral vectors (i.e. mitosis required)**
- **Duration of gene activity**
  - **Non-integrating delivery will be transient (transient expression)**
  - **Integrated delivery will be stable**
- **Patient safety**
  - **Immune hyperresponsiveness (hypersensitivity reactions directed against viral vector components or against transgenes expressed in treated cells)**
  - **Integration is not controlled → oncogenes may be involved at insertion point → cancer?**
- **Gene control/regulation**
  - **Most viral vectors are unable to accommodate full length human genes containing all of their original regulatory sequences**
  - **Human cDNA often used → much regulatory information is lost (e.g. enhancers inside introns).**
  - **Often promoters are substituted → therefore gene expression pattern may be very different**
  - **Random integration can adversely affect expression (insertion near highly methylated heterogeneous DNA may silence gene expression).**
- **Expense**
  - **Costly because of cell culturing needs involved in *ex vivo* techniques**
  - **Virus cultures for *in vivo* delivery**
  - **Usually the number of patients enrolled in any given trial is <20**
  - **More than 5000 patients have been treated in last ~12 years worldwide**

### 4.22.13. Progress toward Gene Therapy

#### **Progress Toward Gene Therapy**

##### **First approved Gene Therapy:**

- On September 14, 1990 at the U.S. National Institutes of Health, W. French Anderson M.D. and his colleagues R. Michael Blaese, M.D., C. Bouzaid, M.D., and Kenneth Culver, M.D., performed the first approved gene therapy procedure on four-year old Ashanthi DeSilva. Born with a rare genetic disease called severe combined immunodeficiency (SCID).

##### **A success story:**

- Over the last 20 years, the initial thoughts of gene therapy have been transformed into reality with more than 175 clinical trials and 2,000 patients already treated . Yet with all the trials, there is still no conclusive evidence for efficacy.
- As of early 2007, she was still in good health, and she was attending college. Some would state that the study is of great importance despite its indefinite results, if only because it demonstrated that gene therapy could be practically attempted without adverse consequences.

##### **Treating Parkinson's Disease:**

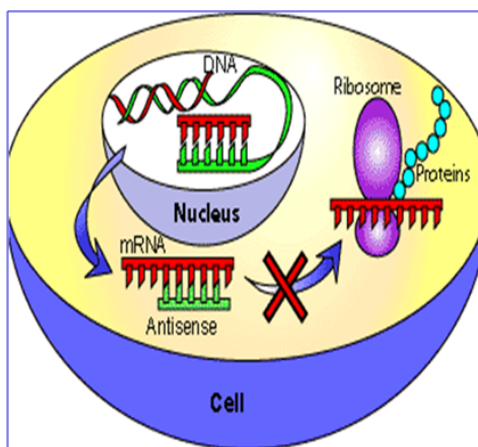
- Neurologix a biotech company announced that they have successfully completed its landmark Phase I trial of gene therapy for Parkinson's Disease.
- This was a 12 patient study with four patients in each of three dose escalating cohorts. All procedures were performed under local anesthesia and all 12 patients were discharged from the hospital within 48 hours of the procedure, and followed for 12 months. Primary outcomes of the study design, safety and tolerability, were successfully met. There were no adverse events reported relating to the treatment.

## 4.22.14. Antisense Gene Therapy

### 4.22.14.1. Introduction

#### Antisense Therapy

- Antisense therapy is a form of treatment for genetic disorders or infections.
- Antisense therapy refers to the inhibition of translation of mRNA into the corresponding protein (cause of disease) by using a single-stranded oligonucleotide which binds strongly to that specific mRNA via Watson-Crick base pairing (ON•RNA duplex).
- A strand of nucleic acid (DNA, RNA, PNA, or a chemical analogue) can bind to the messenger RNA (mRNA) produced by the genetic sequence of a particular gene which is known to cause a particular disease → Thus, that gene is inactivated, effectively turning that gene "off".
- This is because mRNA has to be single stranded for it to be translated.
- Alternatively, the strand might be targeted to bind a splicing site on pre-mRNA and modify the exon content of an mRNA.
- The synthesized nucleic acid is termed an "anti-sense" oligonucleotide because its base sequence is complementary to the gene's messenger RNA (mRNA), which is called the "sense" sequence.
- Thus, a sense segment of mRNA "5'-AAGGUC-3'" would be blocked by the anti-sense mRNA segment "3'-UCCAG-5'".



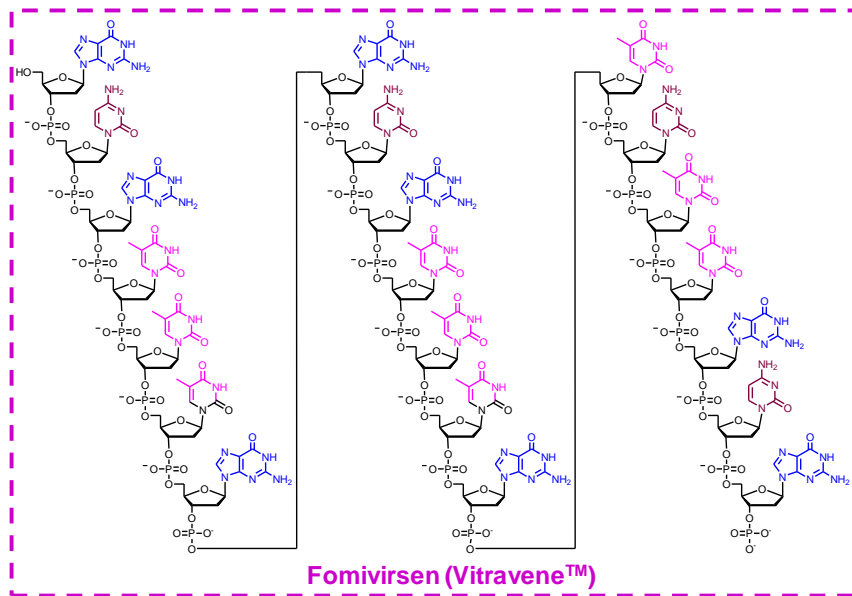
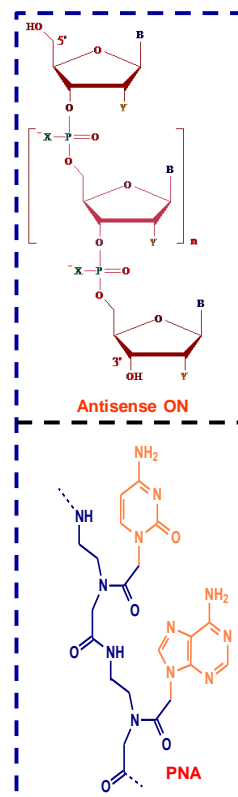
- Antisense therapy is not strictly a form of gene therapy, but is a genetically-mediated therapy and is often considered together with other methods
- Antisense drugs are being researched to treat :
  - Cancers (including lung cancer, colorectal carcinoma, pancreatic carcinoma, malignant glioma and malignant melanoma).
  - Diabetes.
  - Amyotrophic lateral sclerosis (ALS).
  - Duchenne muscular dystrophy and diseases such as asthma and arthritis with an inflammatory component.
  - Most potential therapies have not yet produced significant clinical results, though one antisense drug, fomivirsen (marketed as Vitravene), has been approved by the U.S. Food and Drug Administration (FDA) as a treatment for cytomegalovirus retinitis.



## 4.22.14.2. The Antisense Oligonucleotides

### The Antisense Oligonucleotides

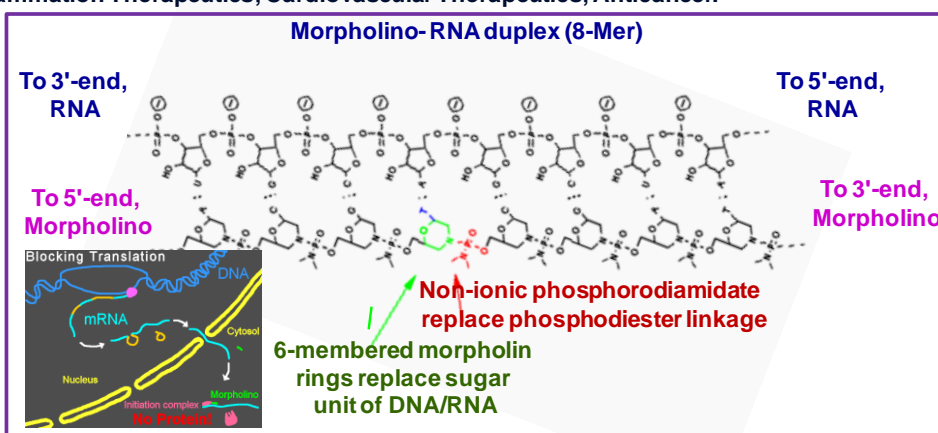
- The completion of human genome project and the Hap-Map project have set the stage for the development of the gene therapy/antisense technology.
- **Antisense Oligonucleotides** are unmodified or chemically modified ssDNA, RNA or their analogs. They are specifically designed to interact (hybridize) with natural genetic material (mRNA) by Watson-Crick binding, thereby stopping the production of disease related protein (translation)
- Zamecnik and Stephenson in 1978 were the first to introduce the concept of exploiting antisense compounds as therapeutic agents → Sequence: A-A-T-G-G-T-A-A-A-T-G-G
- First Approved Antisense Drug: Substantial development in antisense technology led to the approval of the first antisense drug fomivirsen (Vitravene™) for the treatment of AIDS-related CMV retinitis.
- Now, 50 new antisense compounds have entered phase I/II, and in some cases phase III trials.
- Like DNA, antisense molecules carry much more information than a typical drug molecule.
- The property of high-affinity nucleic acid binding because of lack of electrostatic repulsion by the PNA oligomers and their non-degradable property by RNase H or other RNases, PNAs can bind mRNA and inhibit splicing or translation initiation and elongation. Thus, they can also be used as potent antisense drug.



### 4.22.14.3. The Applications of Antisense Oligonucleotides

#### The Antisense Oligonucleotides and Their Applications

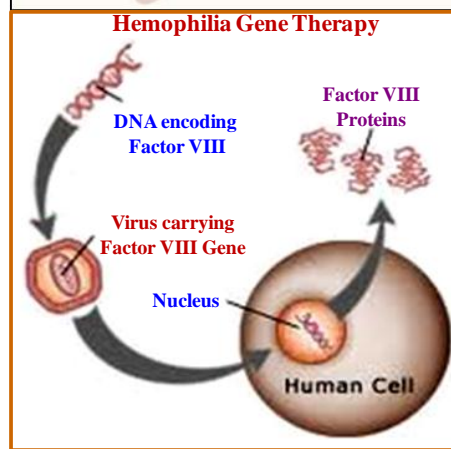
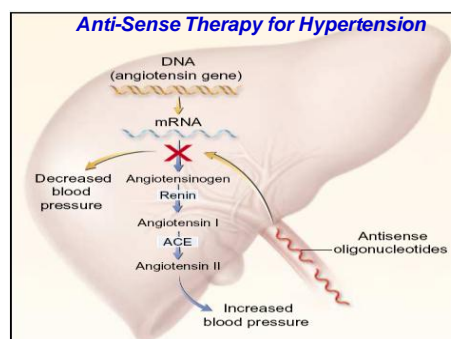
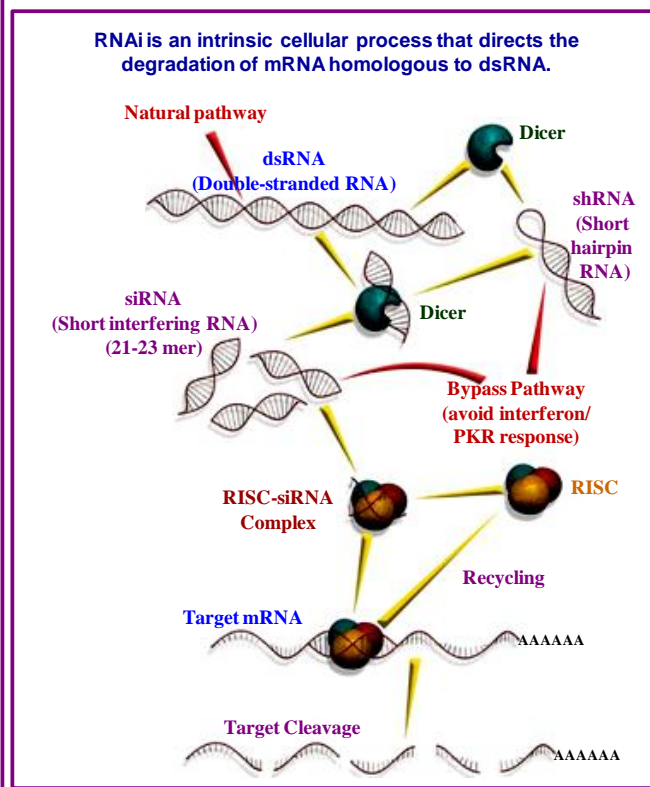
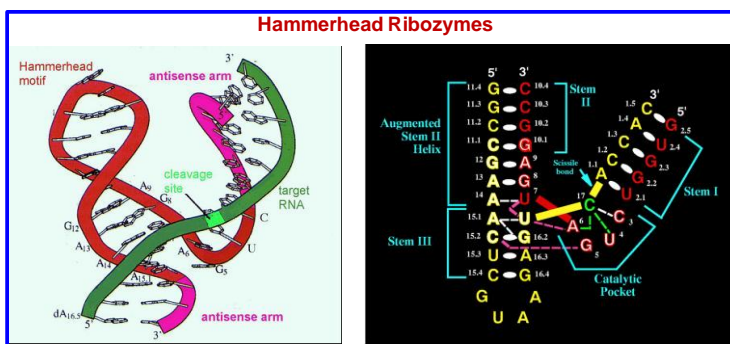
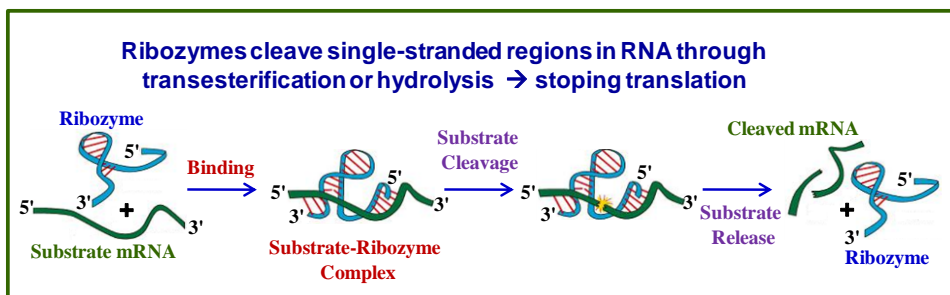
- Functional Genomics and Target Validation:**  
Antisense oligonucleotides can be used to selectively manipulate the expression of chosen gene or genes. The process results in :
  - A pharmacophore with a well-understood mechanism of action.
  - Well characterized distribution and a safe side effect profile which could be used as a human therapeutic.
- Potential Therapeutic Applications of Antisense Oligonucleotides**  
Major areas of therapeutic applications include: Antiviral; Antibacterial; CNS Therapeutics; Inflammation Therapeutics; Cardiovascular Therapeutics; Anticancer.



- Some Antisense Drugs: Approved/in Clinical Trial:**
  - Fomivirsen (Vitravene™):** Approved by U.S. FDA in Aug 1998 for treating treatment cytomegalovirus retinitis (against DNA-virus).
  - Morpholino:** In early 2006, used for treating Hemorrhagic fever viruses (against RNA): AVI-6002 and AVI-6003: In late 2008, filed Investigational New Drug (IND) These drugs, are novel analogs based on AVI's PMO antisense chemistry in which anti-viral potency is enhanced by the addition of positively-charged components to the morpholino oligomer chain.
  - AP 12009:** In 2006. It is a phosphorothioate antisense oligodeoxynucleotide specific for the mRNA of human transforming growth factor TGF-beta2. Tested in patients with high grade gliomas.
  - HIV/AIDS:** Since in 2004, researchers in the US have been conducting research on using antisense technology to combat HIV. In February 2010 researchers reported success in reducing HIV viral load. The patient T-cells modified with an RNA antisense strand and complexed with the HIV viral envelope protein.
  - Mipomersen:** It is a 'second-generation' antisense oligonucleotide; the nucleotides are linked with phosphorothioate linkages, and the sugar parts are deoxyribose in the middle part of the molecule and 2'-O-methoxyethyl-modified ribose at the two ends. It is used for high cholesterol treatment. In 2010 mipomersen successfully completed phase 3 trials for some types of high cholesterol.

### 4.22.14.4. Various Examples of Antisense Gene Therapy

#### Some Examples of Mechanism Gene/Antisense-Gene Therapy



## 4.22.15. Antigen Gene Therapy

### 4.22.15.1. Introduction

Since the complete mapping of the human genome, there has been a considerable interest in exploring various strategies that would bring about specific gene silencing. Although siRNA technology has made an enormous step toward drug development which effectively shut down gene expression at the mRNA level (antisense therapy) but suffer from the problem of prolonged shut down of gene expression. Thus, treatment targeting at DNA level (anti-gene therapy) have come up to cure gene. Thus, a DNA-modifying agent has to so design as to bring about a site-specific DNA modification, as in the form of a double-strand binding or breaking, may result in a mutation that would knockout the gene's function.

Antigene-based therapy require the targeting molecule to (a) enter the nucleus, (b) bind DNA, (c) stop transcription and/or (d) elicit specific DNA damage without hampering with normal cellular function. In 1987, Le Doan et al., and et al. were the first to show the potential of attaching a DNA-modifying agent to a triplex-forming oligonucleotide (TFO) to achieve site-specific DNA damage to double-stranded DNA (dsDNA).

**Triplex-Forming Oligonucleotides (TFO):** TFOs are defined as DNA/RNA or DNA/RNA analogs that have the propensity of forming a triplex with a given dsDNA target molecule. They are typically composed of a stretch of homopurines or homopyrimidines, a prerequisite that is usually required for efficient triplex formation. Ben Gaied et al., in 2009 showed that modified nucleobases or appended DNA intercalators that, when included into a TFO, allow the generation of stable triplex helices even with a dsDNA target. In addition, several DNA/RNA analogs have been shown to improve triplex formation, thus show their potency as antigene agents. Other TFO-enhancing elements include nucleotides such as locked nucleic acids.

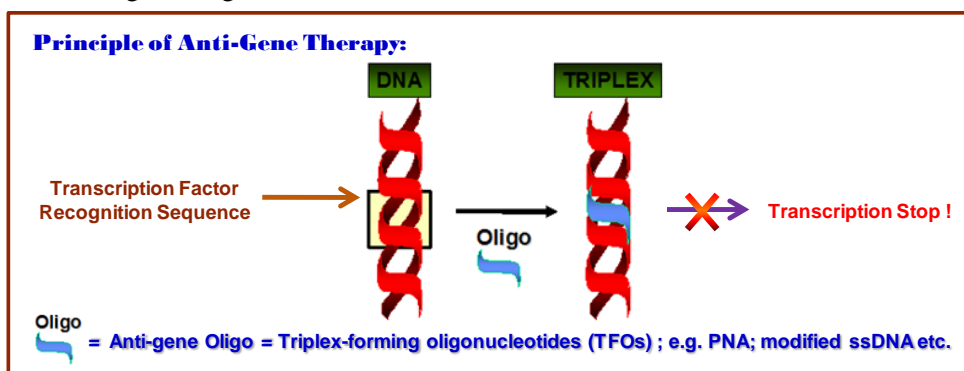
**Peptide nucleic acid (PNA):** Peptide nucleic acid (PNA), introduced Nielsen in 1991, is a DNA mimic composed of a peptide like backbone. These DNA analogs have found many applications in the antigen field because of their high propensity to bind to duplex DNA via strand invasion. In a PNA-DNA-PNA triplex, one PNA strand forms Hoogsteen hydrogen bonds with the DNA target, whereas the other PNA invades the DNA duplex, forming Watson–Crick hydrogen bonds with the same DNA target while displacing the complementary DNA strand.

**Photoactivated TFO conjugates:** Photoactivated TFO conjugates are also found to be potential candidates for improving the antigen-based performance of oligonucleotides. By binding the target gene and damaging it is the basic working principle of such photoactivated TFO. As for example, psoralens (tricyclic compounds containing furan and pyron rings) are bifunctional photoreagents that can intercalate into double-stranded nucleic acids and upon UV irradiation undergo a (2+2) cycloaddition reaction with pyrimidine bases.

PNA-Benzophenone upon photo irradiation at 330–400 nm form PNA-DNA crosslinks DNA-Pyrene conjugate at 355nm breaks dsDNA.

### Anti-Gene Therapy

- **Definition:** Treating the diseases by blocking transcription of DNA using a single-stranded oligonucleotide sequence that hybridizes with the specific gene is called antigene therapy.
- **A single stranded oligo bind to a specific gene** which is causative to disease → Thus formed triplex can no longer able to produce mRNA (no transcription) → No production of disease causing protein → disease is stopped at DNA level.
- Unlike existing antisense therapies that target RNA, an antigene drug is a triplex-forming oligonucleotide that recognizes and attaches directly to a specific DNA sequence.
- By attaching a photoreactive agent to the antigene and delivering light energy to the attachment site, the light-sensitive drug complex becomes activated, triggering a cleavage or cross-linking reaction. This photo-induced, site-specific DNA damage effectively silences the gene target.



- **Triplex-forming oligonucleotides (TFOs) Bind DNA in a sequence-specific manner at polypurine / polypyrimidine sites via specific Hoogsteen H-bonding interaction → inhibition of transcription.**
- **Antigene therapy is based on the recognition and binding of a single oligonucleotide strand to a dsDNA *i.e.* a part of the disease related gene.**



## 4.22.16. The Future Prospects of Gene Therapy

### The Future of Gene Therapy

- Current uses of gene therapy focus on treating or curing existing conditions. In the future, the focus could shift to prevention. As more of the human genome is understood, medicine will know more about which genes contribute to or cause disease. With that knowledge in hand, gene therapy could be used to head off problems before they occur.
- The most likely candidates for future gene therapy trials will be rare diseases.

#### Lesch-Nyhan syndrome:

- Lesch-Nyhan syndrome, a distressing disease in which the patients are unable to manufacture a particular enzyme. This leads to a bizarre impulse for self-mutilation, including very severe biting of the lips and fingers. The normal version of the defective gene in this disease has now been cloned.

#### Gene therapy of pain: emerging strategies and future directions:

- Gene therapy to alleviate pain could appear surprising and perhaps not appropriate when opioids and other active molecules are available. However, the possibility of introducing a therapeutic protein into some targeted structures, where it would be continuously synthesised and exert its biological effect in the near vicinity of, or inside the cells, might avoid some drawbacks of "classical" drugs.

#### Pain – Cancer a major research area:

- Numerous other molecules involved in pain processing or associated with chronic pain have been identified and the gene-based techniques might be particularly adapted for the evaluation of the possible therapeutic interest of these new potential targets.

#### Creating 47<sup>th</sup> Chromosome:

- Researchers are also experimenting with introducing a 47<sup>th</sup> artificial chromosome to the body. It would exist autonomously along side of the other 46, not affecting their workings or causing any mutations.
- It would be a large vector capable of carrying substantial amounts of genetic information and the body's immune system would not attack it.



## 4.23. Goal for Personalized Medicine

### 4.23.1. Introduction of the Concept of Personalized Medicine

The mapping of the human genome and cataloguing of all the disease related polymorphisms are the major scientific milestones that have opened the door to new approaches to understand at molecular level and treat disease. Cancer and cardiovascular disease are two areas in which genomics are showing promise for treatment advances, although challenges remain. Among all other polymorphism, SNPs are the key enabler to realize the concept of personalized medicine. Personalized medicine is a term used in science and medicine that holds significant promise of improved disease treatment. Personalized medicine is holding out the promise of administering medicines explicitly tailored to a person's specific genome or metabolism. Personalized medicine is the drugs that cure disease without adverse effects. "Personalized medicine" is now echoing across the health care sector, with the hope that genomic sciences will revolutionize the medical health care planning by individualizing and thereby optimizing it. Scientists, physicians and the pharmaceutical industry are actively developing ways to customize medical treatments to suit our unique genetic signatures. This emerging science is called personalized medicine.

Therefore, personalized medicine is the use of new methods of molecular analysis to better manage a patient's disease or predisposition toward a particular disease. It aims to realize optimal medical outcomes by helping physicians and patients to choose the disease management approaches which can likely to work best in the context of the patient's unique genetic and environmental profile. Such approaches may include genetic screening programs that more precisely diagnose diseases and their sub-types, or help physicians select the type and dose of medication best suited to a certain group of patients thereby minimize or nullifying the side effect.

Thus, in short personalized medicine can be defined as the use of information from a patient's genotype to:

- initiate a preventative measure against the development of a disease or condition, and
- select the most appropriate therapy for a disease or condition that is particularly suited to that patient.





Figure 4.55: Drugs for future.

### ***4.23.2. Genomic Variations/Biomarkers Are the Foundation of Personalized medicine***

People vary from one another in many ways — what they eat, the types and amount of stress they experience, exposure to environmental factors. Therefore their DNA also varies. Many of these variations in DNA sequences play a role in health and disease. For example, the natural variations found in our genes could influence our risk of developing a certain disease, and the degree to which it progresses. Variations in several genes can influence a patient's response to a particular drug medication. The completion of the Human Genome Project has set the stage to answer a number of questions like: What makes us similar and what makes us different both in physical appearance and in predisposition to a disease and in response to a particular drug medication? Using the information of Hap Map project, researchers are now able to find genes that affect health and disease as well as individual responses to medications and environmental factors.



Figure 4.56: Right gene-Right Drug.

Thus, strong assurances and long term goals help in the long and painstaking job to unlock the mysteries of the human body. We now are grateful to the genomic researchers for the fact that there are between 30,000 and 40,000

protein-coding genes in the human genome, not the more than 1,00, 000 previously thought. From an evolutionary perspective, this represents only about twice as many as in the worm or fly genome. Genomics has yielded new molecular targets and therapeutic interventions in the world of cancer which have already been exploited, especially in patient selection, diagnosis and treatment. Genomics is also showing promise in cardiovascular, and many other diseases, with substantial challenges remaining. We are also better able to quantify the role of genes and genetic variation in disease and have learned that environmental influences are far more important than specific alterations in the genetic codes of common diseases. Scientists see SNPs as a potential tool to improve diseases DIAGNOSIS and TREATMENT planning. They suspect that SNPs may play a role in the different responses to treatments seen among cancer patients and they think that SNPs may also be involved in the different levels of individual cancer risk observed. Thus, genetic variations/SNPs are attractive target for better understanding the genetic basis of complex diseases, and to realize the PERSONALISED MEDICINE.

Personalized medicine hopes to use these variations to develop new safe and effective treatment planning for genetically defined sub-groups of patients. Treatments may include administration of drug therapy as well as recommendations for lifestyle changes that can delay onset of a disease or reduce its impact.

#### 4.23.2. Examples of SNPs, Mutations, and Chromosomal Alterations

Gene Symbol	Description	SNPs	Clinical Phenotype
<i>CYP2C9</i>	Cytochrome P450 2C9	multiple	variable metabolism of CYP substrates in the liver
<i>TPMT</i>	thiopurine methyl transferase	multiple	hematopoietic thiopurine toxicity
<i>UGT1A1</i>	UDP-glycosyl transferase 1A1	multiple, in promoter & encoding regions	UGT1A1*28 variants associated with increased irinotecan toxicity
<i>VKORC1</i>	vitamin K epoxide reductase complex I	multiple, i.e. 1639G>A in promoter	variable anticoagulant effect of warfarin
t(9,22) translocation	t(9,22) BCR-ABL translocation	translocation	Gleevee (imatinib) effective against chronic myeloid leukemia with translocation
<i>ERBB2</i>	ERBB2, HER/Neu	overexpression of protein	Herceptin for breast cancer with ERBB2 overexpression
<i>EGFR</i>	epidermal growth factor receptor	exon 18-21 mutations	human lung cancers with mutations response better for Iressa (gefitinib)

#### 4.23.3. Aspects of Personalized Medicine --- Differ from Traditional Medicine

Continuous work in genomics, proteomics and metabolomics will undoubtedly bring major discoveries about the inner workings of the human body, the genetic variations, the disease related genes and disease susceptibility. It also definitely will provide more molecular targets, and promising diagnostic and therapeutic interventions for a genetic disease. Research is going on to understand how people differ in their susceptibility to certain diseases. Researchers are developing genetic tests that can predict a person's risk of developing some common diseases such as heart disease or asthma for example. Now a day, doctors hope that studying and correlating genetic variations/SNP profiles in populations, will give relation between genetic variations/SNPs and specific responses to disease/or treatment. Thus, in near future, after a doctor has diagnosed breast cancer in a new patient, he or she may also request a SNP profile and use the information to help advice his patient about her treatment options. A doctor, in the future, might prescribe a drug developed to prevent the onset of a disease and recommend lifestyle changes the at-risk individual should make. Thus in a nut shell, the following table can explain the future's treatment planning.

Past and Present	Example	Present and Future	Example
Diagnosis – Disease by Symptoms	High Blood Pressure – Many Causes	Diagnosis and Prognosis -Disease by Mechanisms	Breast Cancer – HER2 Gene and Oncotype Dx
Treatment Guidelines – Disease Uniformity	Non-Hodgkin's Lymphoma – Many Cancers of Immune System	Customized Guidelines – Disease Heterogeneity	Subclass of B-Cell and T-Cell – Use of Rituximab if CD-20 Positive
Patient Uniformity – One Size Fits All Dosing	Oral Warfarin Anticoagulation --5 mg per day	Patient Variability – Genetic-Guided Dosing	Genotypes Defined by 2C9 and VKORC– 0.5 to 6 mg/day
Industry Blockbuster Model	Few with Sales Between \$5 – \$10 Billion	Mixed Blockbuster and Mini-Buster Model	Many with Sales Between \$1 --\$5 Billion
Lack of Physician and Patient Awareness	Absence of Formal Education -Access to Information	Patient Empowerment and Societal Expectations	deCode SNP Analysis, Paternity Testing Kits, Safe

#### 4.23.4. How Can Diagnostics “Personalize” Medicine?

##### How Can Diagnostics “Personalize” Medicine?

###### 1. Definition of Pharmacogenetics (PGx):

- Pharmacogenetics is the science of how an individual's genotype affects their body's response to drugs.

Joint initiative of IITs and IISc – Funded by MHRD

Page 147 of 161

###### 2. Definition of Biomarkers:

- Biomarkers are typically in the causal pathway of disease pathology or drug pharmacology.

###### 3. Current Role of Biomarkers in Drug Selection and Use:

- A recent study of FDA-approved drug labels found that:
  - 121 drug labels contained pharmacogenomic information

#### 4.23.5. The Paradigm of Personalized Medicine

The paradigm of personalized medicine can be illustrated as follows:

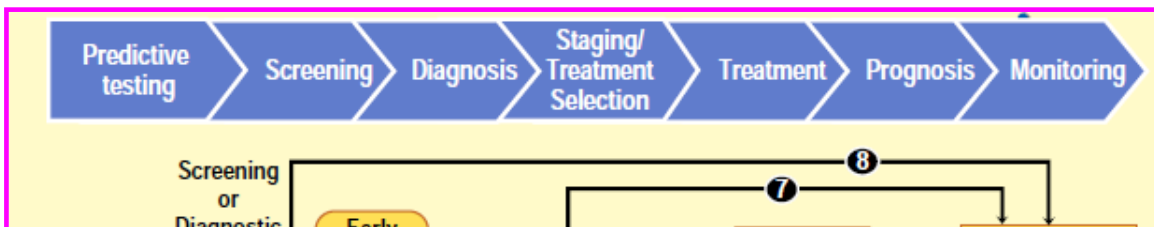


Figure 4.57: Schematics of “Clinical Analytical Framework” for assessing future medical technologies

This arrow reflects the current and projected flow of healthcare services, and changing points of intervention, as medicine becomes more personalized. Early detection testing will continue based on large population risk (e.g., mammograms). The new forms of risk assessment will be incorporated such as the determination of the patients who carry the genetic variation that increases their risk for developing cancer will be done. Though true prevention must occur before disease symptoms are present, better risk assessment enables more targeted monitoring. As for example the patient with the genetic variation should have more frequent mammograms. Then the symptom-driven diagnosis will be done which could possibly identify disease subtypes that cannot be clinically determined. Such diagnosis may or may not lead to targeted therapy. However, in any case we may also benefit from improvements in monitoring a patient's response to a particular therapy.

#### **4.23.6. The Three Promises of Personalized Medicine**

Personalized medicine has the potential to change the way we think about, identify and manage health problems. It is already having an exciting impact on both clinical research and patient care, and this impact will grow as our understanding and technologies improve.

It is already clear that personalized medicine promises three key benefits:

- 1. Better Diagnoses and Earlier Interventions:** Molecular analysis could determine precisely which variant of a disease a person has, or whether an individual is susceptible to drug toxicities, to help guide treatment choices. For preventive medicine, such analysis could improve the ability to identify which individuals are predisposed to develop a particular condition-and guide decisions about interventions that might prevent it, delay its onset or reduce its impact.
- 2. More Efficient Drug Development:** A better understanding of genetic variations could help scientists to identify new disease subgroups or their associated molecular pathways. Thus, it would help in designing drugs

that target those subgroups. Molecular analysis could also help to select patients for inclusion in/ exclusion from the late stage clinical trials which would help to approve drugs that might otherwise be abandoned because they appear to be ineffective in the larger patient population.

3. **More Effective Therapies:** Currently, physicians often have to use trial and error to find the most effective medication for each patient. As we learn more about which molecular variations best predict how a patient will react to a treatment, and develop accurate and cost-effective tests, doctors will have more information to guide their decision about which medications are likely to work best. Testing is already being used to find the one in four women likely to respond to a particular breast cancer drug. In the future, tests will help in identifying the one in ten patients who for tumor-specific molecular reasons will benefit from a new lung cancer drug. In addition, testing will help to predict the best dosing schedule or combination of drugs for a particular patient.

#### **4.23.7. Case Study of Personalized Medicine: Genomic Testing in Cardiovascular Disease: Antithrombotic Drugs**

##### **Genomic Testing in Cardiovascular Disease: Changing Landscape for Antithrombotic Drugs**

- **Plavix (clopidogrel)** => widely used antithrombotic drug.
  - At higher doses, antithrombotic drugs have higher risk of bleeding events.
- New drug Effient, **Prasugrel**, launched in 2009/10.
  - “Black box” warning re-bleeding events.
  - Faces competition from generic clodipogrel.
- **May 2009:** FDA updates Plavix label to include CYP2C19 testing.
  - 30%-40% Americans carry marker for poor response to Plavix.
  - As this unfolds it will be one of the more important case studies in personalized

#### 4.23.8. Few Personalized Medicine Drugs, & Today's Treatment Recommendation

Disease/Condition Area	Diagnostic Tests	Treatment Options
Cancer		
Breast cancer	HER2, Oncotype Dx, Mammaprint,	Herceptin, Tamoxifen, hormonal therapies

Colorectal cancer	EGFR, KRAS, ERCC, RRM1	Camptosar, Erbitux, Vectibix
Non-small cell lung cancer (NSCLC)	EGFR, KRAS, UGT1A1	Avastin, Gemzar, Iressa, Tarceva
Acute myeloid leukemia	CD33, FLT3, inv16	Mylotarg, multiple chemotherapies
Non-Hodgkin's lymphoma	CD20, MALT	Bexxar, Rituxan, Zevalin
<b>Cardiovascular disease</b>		
Adverse events from warfarin therapy	CYP2C9, VKORC1, PGx Predict	Coumadin
Homozygous familial hypercholesteremia	LDLR	Atorvastatin
Heart transplant rejection	AlloMap	Immunosuppressive therapy
<b>HIV</b>		
HIV treatment resistance	HLA-B*5701, CCR5	HIV therapies

#### 4.23.9. Future of Personalized Medicine

Despite a few successes, patients should not expect anything more than a small number of additional tailored interventions. They should not expect the cures for all common diseases. First, because the interplay between genes, proteins, cell metabolism and the numerous environmental exposures are simply too complex to expect simple solutions, like a targeted molecule. Secondly, getting any new therapy to market takes at least a decade.

However, from a patient's perspective, "personalized medicine" is not only unrealistic, but also has the potential to do more than just raise false hope and expectations. The term "personalized medicine" does not impress much in public.

To make the concept of personalized medicine a success we need to give more impact on the followings:

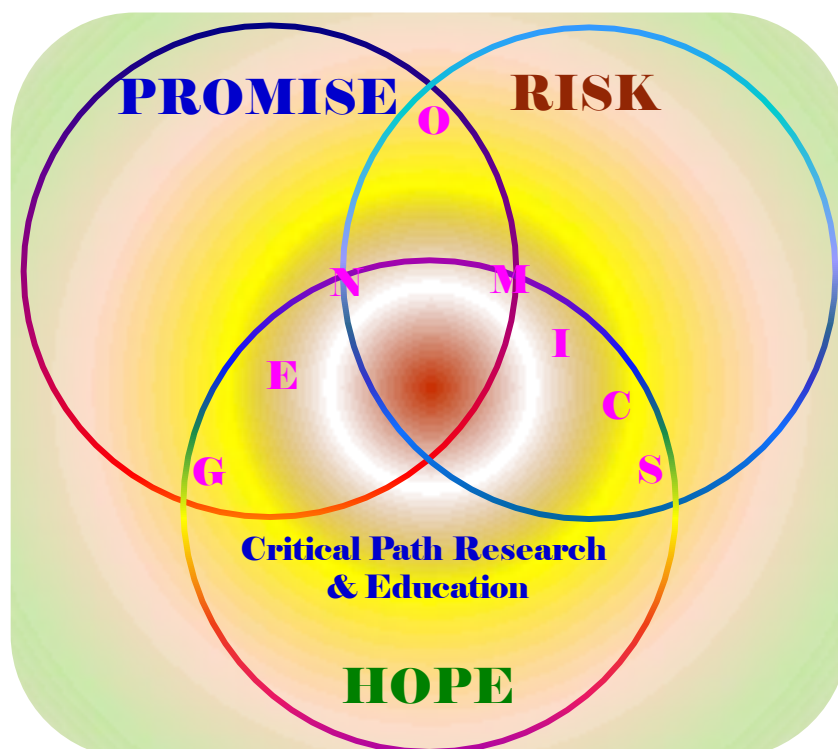
- Better communication between patients and health providers is needed.
- The optimal use of electronic genetic information systems will help to follow the exact way of diagnosis and treatment planning.
- Encouraging patients to make lists of things to discuss with their physicians is important
- Coaching the patients is necessary on how to participate effectively in their medical consultation which will help to improve overall well-being as well as survival.



- Labeling of “personalized medicine” to broaden the focus of scientifically validated interventions from genes is necessary.
- Avoiding any misinterpretations and overstated claims in our scientific discussion as well as in our choices of words and labels related to personalized medicine.
- Increase funding is necessary for cutting-edge biological, behavioural, social and organizational scientific investigation designed to better understand the human life and to acquire knowledge about it as a mixture of biology, circumstances and needs.

Scientists hope that the day will come when “personalized medicine” will be the image in the worlds of both science and medicine. However, the full adoption of the concept need the care from several sectors and its success lies on many things as is mentioned below.

#### 4.23.9.1. Role of FDA in Supporting the Future Direction of Personalized Medicine



#### 4.23.9.2. Intellectual Property Rights and Personalized Medicine

##### Intellectual Property Rights and Personalized Medicine

##### A. Pharmacogenetics Claim:

- Methods of treatment based on genetic information (SNPs) of an individual using suitable dosages of medications can be claimed.

##### B. SNP Claim:

- Claims can be drawn to isolated SNPs in DNA.

##### C. Methods Correlating SNPs and Diseases and Treatment:

- Methods of treatment of diseases can be claimed based on genetic information (SNPs) of an individual using correlations of particular SNPs.

##### A. Pharmacogenetics Claim:

- Method of treating a human being a thrombosis with a dosage of **Warfarin**

## Intellectual Property Rights and Personalized Medicine

### C.2. Methods of Treating Diseases that Correlate with SNPs:

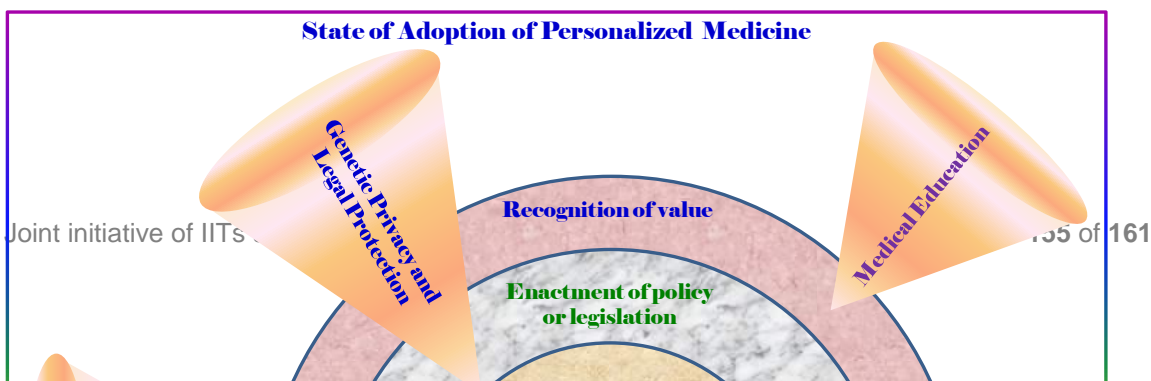
- A method for treating a human subject having breast cancer comprising:
  - a. obtaining a nucleic acid sample from said human subject;
  - b. subjecting the sample to PCR and identifying the nucleotide present at position 101 of SEQ ID NO:1; and
  - c. treating the human subject with “breast cancer drug X” when a cytosine is detected at position 101 of SEQ ID NO:1.

### C.3. Enabling Methods of Treating Diseases that Correlate with SNPs:

- The specification tells that SEQ ID NO:1 is a variant of the *ERBB2* gene having an A (adenine) to C (cytosine) mutation at position 101 (A101>C) but can not distinguish among patient populations tested.

### 4.23.9.3. State of Adoption of Personalized Medicine

The implementation of personalized medicine requires a support/influence of several sectors as is represented in the diagram below. Concentric circles and range represent stages of implementation for each sector from public or stakeholder recognition of the value of personalized medicine, the establishment of supporting policies and laws, the launch and execution of smaller scale pilot programs and projects, to the final stage of full implementation and widespread use. Full implementation of personalized medicine can only be achieved when all sectors converge toward the center. **Senator Edward M. Kennedy (D – Mass.)** *Remarked by stating* "...We are in a new era of the life sciences, and the truth of that statement can be seen in fields from medical imaging, to new biologic drugs, and even to the use of DNA technology to improve our environment and reduce greenhouse gasses. But in no area of research is the promise greater than in the field of personalized medicine."----**Senator Edward M. Kennedy (D – Mass.)** *Remarks on the Senate's Consideration of the Genetic Information Nondiscrimination Act. April 24, 2008.*



**Figure 4.57: State of adoption of Personalized Medicine showing the role of several sectors for its full implementation.**

The actions of healthcare organizations, industry, government institutions, Congress, and the presidential administrations see that personalized medicine to come to full shape, requires an association of laws, regulatory and insurance reimbursement policies, healthcare information technology, medical education, and research investment. The progress of each element of that infrastructure is going on a different pace. However, to make progress of certain elements, such as insurance reimbursement and medical education, requires a substantial effort to change mindsets and create new policies (**Figure 4.57**).

Thus, the “personalized medicine” will be known, simply, as medicine when-

1. all of the elements of infrastructure fall into place,
2. we begin to classify and treat diseases not just by their most obvious signs and symptoms, but also by their molecular profiles,
3. physicians combine their knowledge and judgment with a network of linked databases that help them interpret and act upon a patient’s genomic information,
4. insurance companies pay for tests and treatments that anticipate the needs of the patient as much as it reacts to them.

5. and when regulators insist on using all information available to the physician, including genetic tests, to ensure the safety and efficacy of an approved drug.

#### 4.24. Selected References

1. Stryer, L.; Berg, J. M.; Tymoczko, J. L. In *Biochemistry*, 5<sup>th</sup> Ed. (Hardcover).
2. Lehninger Principles of Biochemistry, 5<sup>th</sup> Ed. by Nelson and Cox.
3. (a) McLaughlin, L. W.; Wilson, M.; Ha, S. B. Use of Nucleoside Analogues to probe Biochemical Processes. In *Comprehensive natural product*

- chemistry, Vol. 7 (DNA and Aspects of Molecular Biology)*; Barton, D. H. R., Nakanishi, K., Meth-Cohn, O., Eds.; Elsevier: New York, 1999; pp 252-284.
4. (a) Wojciechowski, F.; Leumann, C. J. *Chem. Soc. Rev.* **2011**, *40*, 5669. (b) Jan Stambasky, Michal Hocek and Pavel Kotovsky *Chem. Rev.*, **2009**, *109*, 6729.
  5. (a) Kool, E. T.; Morales, J. C.; Guckian, K. M. *Angew. Chem. Int. Ed.* **2000**, *39*, 990. (b) Kool, E. T. *Acc. Chem. Res.* **2002**, *35*, 936. (c) Teo, Y. N.; Wilson, J. N.; Kool, E. T. *J. Am. Chem. Soc.* **2009**, *131*, 3923. (d) Schweitzer, B. A.; Kool, E. T. *J. Am. Chem. Soc.* **1995**, *117*, 1863. (e) Kool, E. T. *Annu. Rev. Biochem.* **2002**, *71*, 191. (f) Jarchow-Choy, S. K.; Sjuvarsson, E.; Sintim, H. O.; Eriksson, S.; Kool, E. T. *J. Am. Chem. Soc.* **2009**, *131*, 5488.
  6. (a) Geyer, C. R.; Battersby, T. R.; Benner, S. A. *Structure* **2003**, *11*, 1485. (b) Benner, S. A. *Nature* **2003**, *421*, 118. (c) Benner, S. A.; Sismour, A. M. *Nature Rev. Genet.* **2005**, *6*, 533. (d) Yang, Z.; Hutter, D.; Sheng, P.; Sismour, A. M.; Benner, S. A. *Nucleic Acids Res.* **2006**, *34*, 6095.
  7. Piccirilli J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature* **1990**, *343*, 33.
  8. Hirao, I.; Kimoto, M.; Mitsui, T.; Fujiwara, T.; Kawai, R.; Sato, A.; Harada, Y.; Yokoyama, S. *Nature Methods* **2006**, *3*, 729.
  9. Doi, Y.; Chiba, J.; Morikawa, T.; Inouye, M. *J. Am. Chem. Soc.* **2008**, *130*, 8762.
  10. Delaney, J. C.; Gao, J.; Liu, H.; Shrivastav, N.; Essigmann, J. M.; Kool, E. T. *Angew. Chem. Int. Ed.* **2009**, *48*, 4524.
  11. (a) Seo, Y. J.; Matsuda, S.; Romesberg, F. E. *J. Am. Chem. Soc.* **2009**, *131*, 5046. (b) Seo, Y. J.; Matsuda, S.; Romesberg, F. E. *J. Am. Chem. Soc.* **2009**, *131*, 5046. (c) Matsuda, S.; Fillo, J. D.; Henry, A. A.; Rai, P.; Wilkens, S. J.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Schultz, P. G.; Spraggon, G.; Romesberg, F. E. *J. Am. Chem. Soc.* **2007**, *129*, 10466.
  12. (a) Kempe, T.; Sundquist, W. I.; Chow, F.; Hu, S. L. *Nucleic Acids Res.* **1985**, *13*, 45. (b) Zischler, H.; Nanda, I.; Schafer, R.; Schmid, M.; Epplen, J. T. *Hum. Genet.* **1989**, *82*, 227 (c) Hustedt, E. J.; Spaltenstein, A.; Kirchner, J. J.; Hopkins, P.B.; Robinson, B. H. *Biochemistry* **1993**, *32*, 1774. (d) Haugland, R. P. *The Handbook: A Guide to Fluorescent Probes and Labeling Technologies*; Molecular Probes. **2005**. (e) Okamoto, A.; Saito, Y.; Saito, I. *J. Photochem. Photobiol. C* **2005**, *6*, 108. (f) Bag, S. S.; Kundu, R.; Matsumoto, K; Saito, Y.; Saito, I. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 3227. (g) Sinkeldam, R. W.; Greco, N. J.; Tor, Y. *Chem. Rev.* **2010** *110*, 2579 (h) Teo, Y. N.; Kool, E. T. *Chem. Rev.* **2012**, *112*, 4221.
  13. (a) Bag, S. S.; Kundu, R.; Matsumoto, K, Saito, Y.; Saito, I. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 3227. (b) Sinkeldam, R. W.; Greco, N. J.; Tor, Y. *Chem. Rev.* **2010**, *110*, 2579.
  14. (a) Dias, N.; Stein, C. A. *Mol Cancer Ther* **2002**, *1*, 347-355. (b) Scanlon, K. J. *Curr. Pharmaceutical Biotechnol.* **2004**, *5*, 000-000.

15. (a) [http://www.ornl.gov/sci/techresources/Human\\_Genome/project/journals/journals.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/project/journals/journals.shtml). (b) "About the Human Genome Project: What is the Human Genome Project". (c) The Human Genome Management Information System (HGMS). 2011-07-18. [http://www.ornl.gov/sci/techresources/Human\\_Genome/project/about.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/project/about.shtml). Retrieved 2011-09-02. [http://www.ornl.gov/sci/techresources/Human\\_Genome/home.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml). (d) [www.hapmap.org](http://www.hapmap.org)
16. Carlson, Bruce (2008-06-15). "SNPs — A Shortcut to Personalized Medicine". Genetic Engineering & Biotechnology News (Mary Ann Liebert, Inc.) 28 (12). <http://www.genengnews.com/gen-articles/snps-a-shortcut-to-personalized-medicine/2507/>. Retrieved 2008-07-06. "(subtitle) Medical applications are where the market's growth is expected."
17. Sachidanandam, R.; Weissman, D.; Schmidt, S. C.; Kakol, J. M.; Stein, L. D.; Marth, G.; Sherry, S.; Mullikin, J. C. *et al.* "A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms". *Nature* **2001**, *409*, 928–33.
18. (a) [www.personalizedmedicinecoalition.org](http://www.personalizedmedicinecoalition.org). (b) [http://www.ostp.gov/galleries/PCAST/pcast\\_report\\_v2.pdf](http://www.ostp.gov/galleries/PCAST/pcast_report_v2.pdf)

## 4.25. Assignments

1. Double stranded DNA is composed of canonical base pairs.
  - a. What are they?
  - b. What is the orientation between the two strands of dsDNA?
  - c. What is the consequence of the two glycosidic bonds holding the bases in each base pair not being on opposite sides of the helix?

2. Answer the following questions with A (A form DNA), B (B form DNA) or Z (Z-DNA).
  - a. The double helical form adopted by double stranded RNA -----.
  - b. Left-handed helix ----- .
  - c. Approximately 10 base pairs per turn -----.
  - d. Plane of base pair perpendicular to helix axis -----.
  - e. Accommodates syn glycosyl bond conformation -----.
  - f. Is stabilized by cytosine methylation and by supercoiling -----.
  - g. Formation is sequence dependent -----.
  - h. Approximately 11 base pairs per turn -----.
  - i. Helix axis runs through the base pairs -----.
  - j. Helix axis runs through the major groove -----.
3. Write down the basic characteristic features of genetic code.
4. X-ray diffraction studies showed that the double-stranded DNA is in helical conformation in which the rise per base pair ( $Z$ ) = 0.32 nm and the pitch of the helix is  $P = 3.36$  nm. Calculate the other parameters of this helix: (a) the number of base pairs per turn, (b) the mean rotation per base pair ( $\theta$ ), and (c) the true repeat?
5. Single nucleotide polymorphisms (SNPs) are a kind of missense mutation and thought as the key enabler of the goal for the development of “Personalised Medicine”. How are SNPs responsible to affect human health?
6. It is proposed that the primordial genetic material could have been peptide nucleic acids (PNAs). Draw the structure of a PNA dimeric unit with the DNA bases A and G.
7. Explain that a PNA/DNA duplex is more stable than a DNA/DNA duplex.
8. Fluorophore/Quencher containing dual labeled PNA molecular beacon can function as an efficient hybridization probe for DNA detection in a similar way as is done by DNA molecular beacon. Can you draw the structure of any one molecular beacon probe (with labeling the different parts) before and after hybridization with a target DNA?
9. Gene therapy is the current research focus mainly based on antigene and antisense nucleic acids. What do you mean by antigene and antisense therapy?
10. In which way DNA synthesis by polymerase is different from the solid phase synthesis?
11. Any DNA sequence can be written by using and arranging four genetic alphabets, A, G, C, and T. How many different 8-mer sequences of DNA are there?
12. State the main differences between DNA and RNA considering the structure and biological role.
13. How tRNA can recognize more than one codon for a specific amino acid? - Explain with example and structure of tRNA anticodon and mRNA codon.



