aromaticity, heterocycles,
nucleic acids, DNA synthesis
sequencing, and editing

# Aromaticity Heterocycles Nucleic Acids

## By Inquisition

unique animated
workbooks
for self study
and flipped teaching

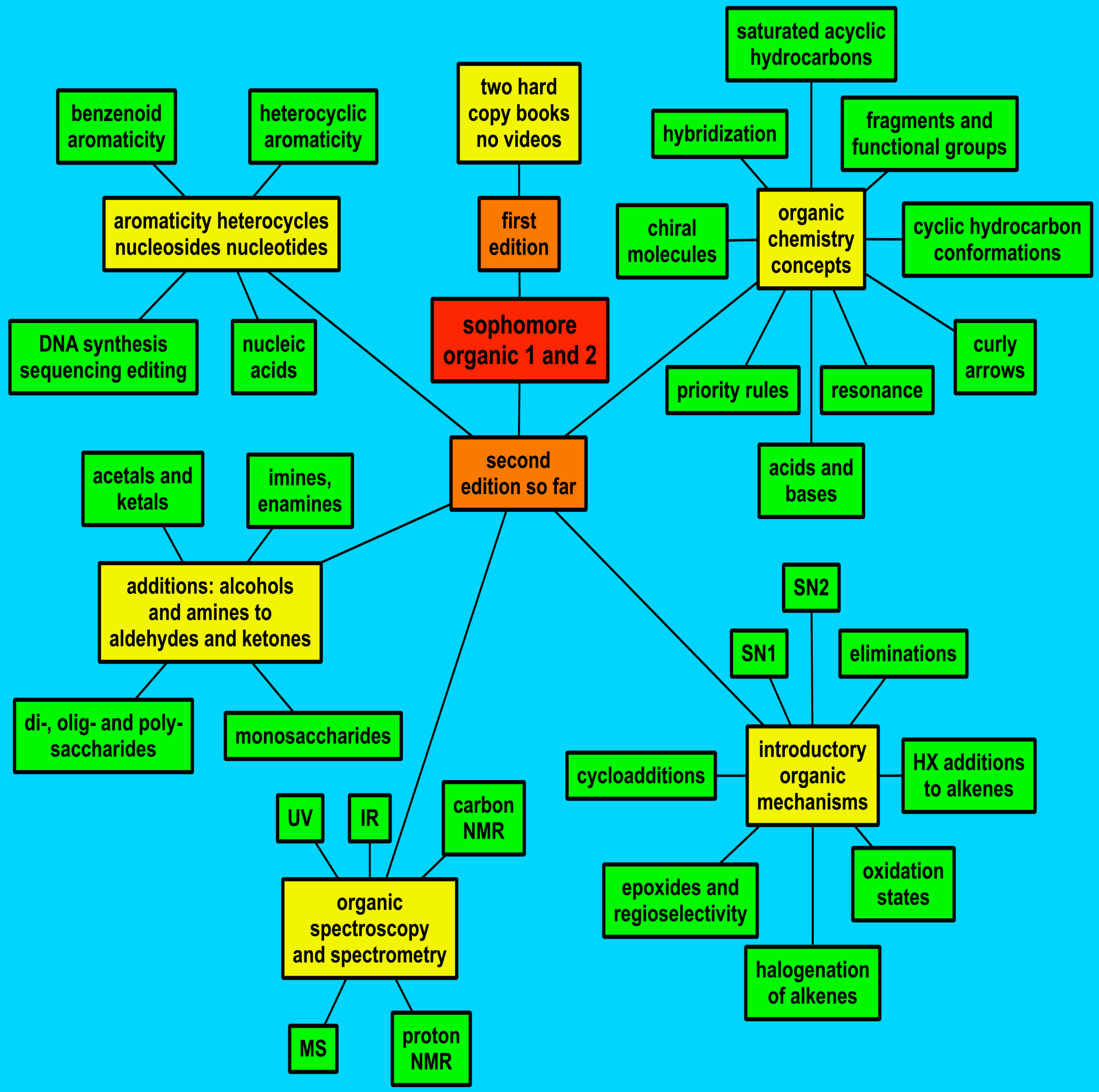
Kevin Burgess

Jan 2025

# Preface

First edition workbooks from By Inquisition Press, Sophomore Organic Chemistry 1 and 2 By Inquisition, cover typical first and second semester content, respectively. These hard copies cannot contain embedded videos. Both are for sale at byinquisition.org (link) for $35, with answers available on the same site.

I began working on the second edition of these workbooks and decided to publish in segments online (no hard copies) *with videos to explain*. Publishing in segments allows them to be used in combinations appropriate to different syllabi. This current work is the fifth segment.

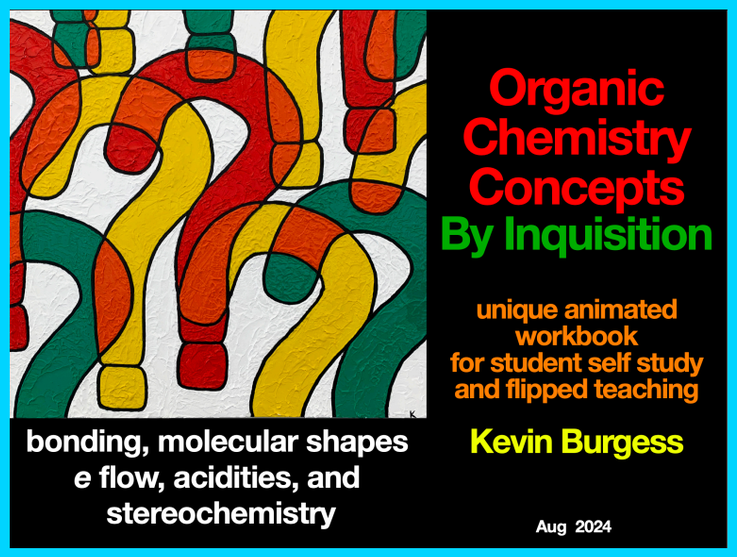


All these *modular animated eworkbooks* are on Books (Mac), the Kindle app, and VitalSource, with answers at byinquisition.org. Online versions do not work for Kindle ebook readers (because of the embedded videos) but are fine on the PC or Mac Kindle app. I think embedded animation is what Steve Jobs had in mind when he introduced iBooks, and am happy to bring it to chemistry. It is a type of instruction people can relate to, and use to learn thoroughly at their own pace. I truly hope these eworkbooks are useful.

**Organic Chemistry Concepts By Inquisition**

Amazon, Apple Books, VitalSource

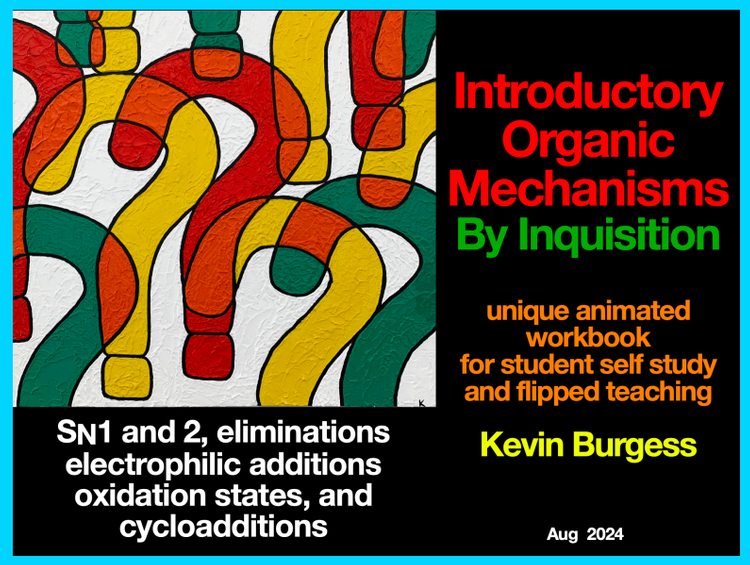
instructor copy from VitalSource




**Introductory Organic Mechanisms By Inquisition**

Amazon, Apple Books, VitalSource

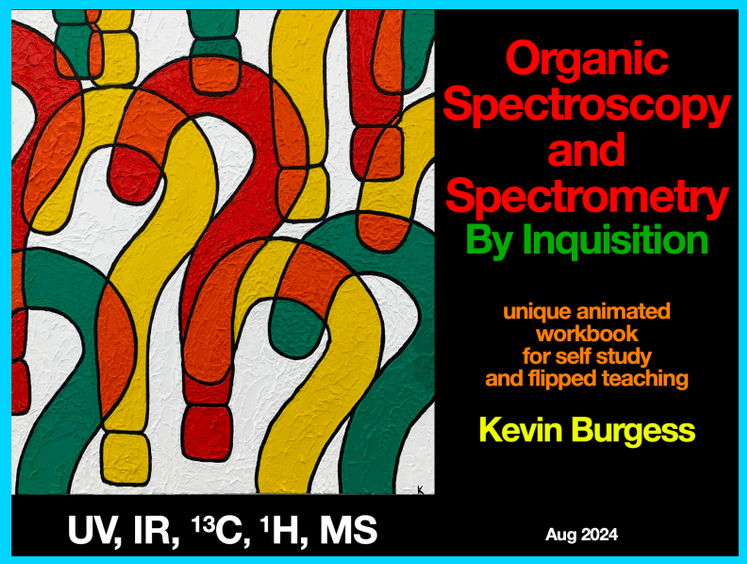
instructor copy from VitalSource




**Organic Spectroscopy and Spectrometry By Inquisition**

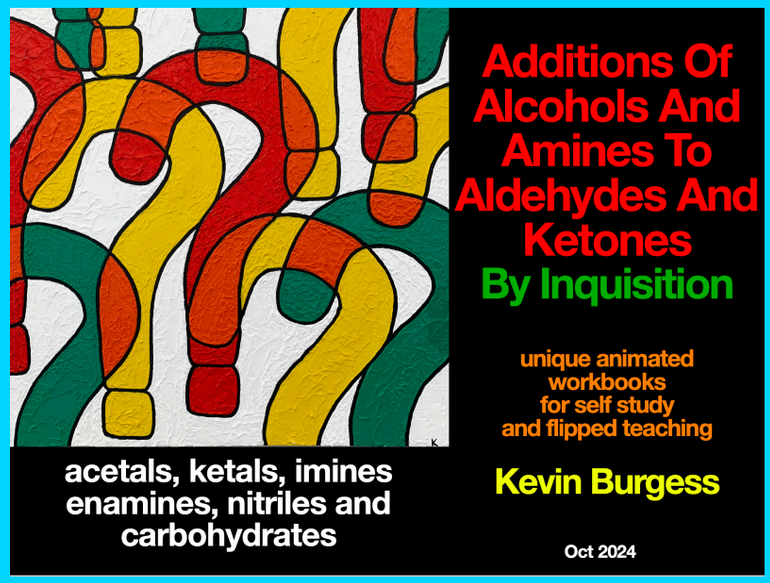
Amazon, Apple Books, VitalSource

instructor copy from VitalSource



## Additions Of Additions Of Alcohols And Amines To Aldehydes And Ketones By Inquisition

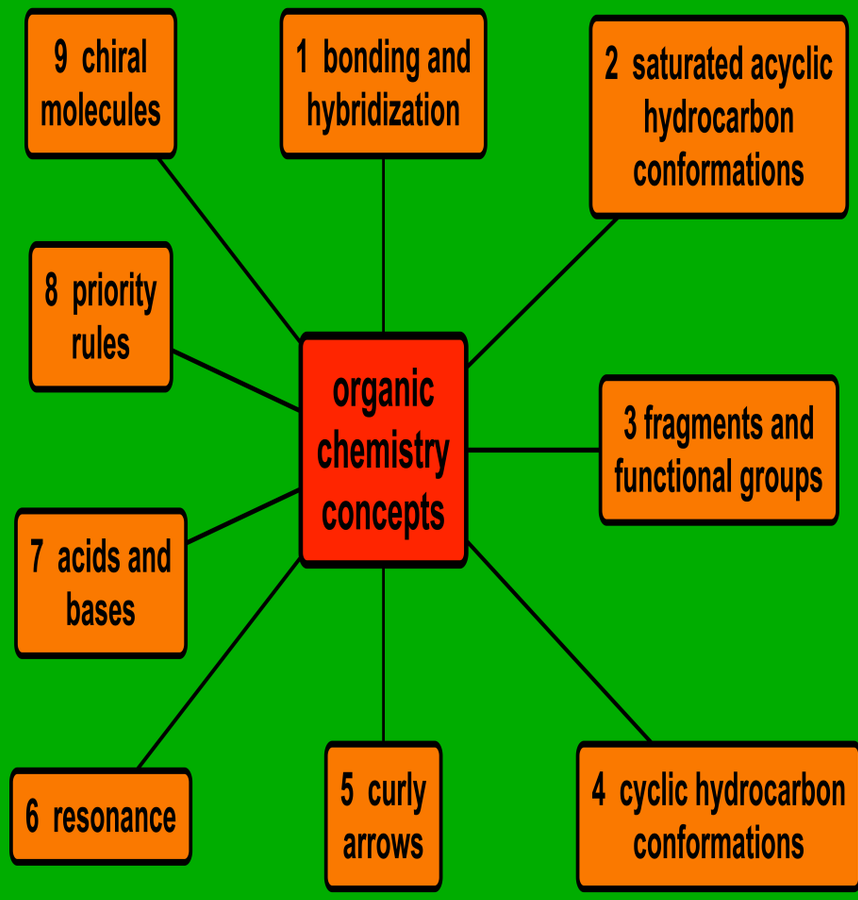
Amazon, Apple Books, VitalSource

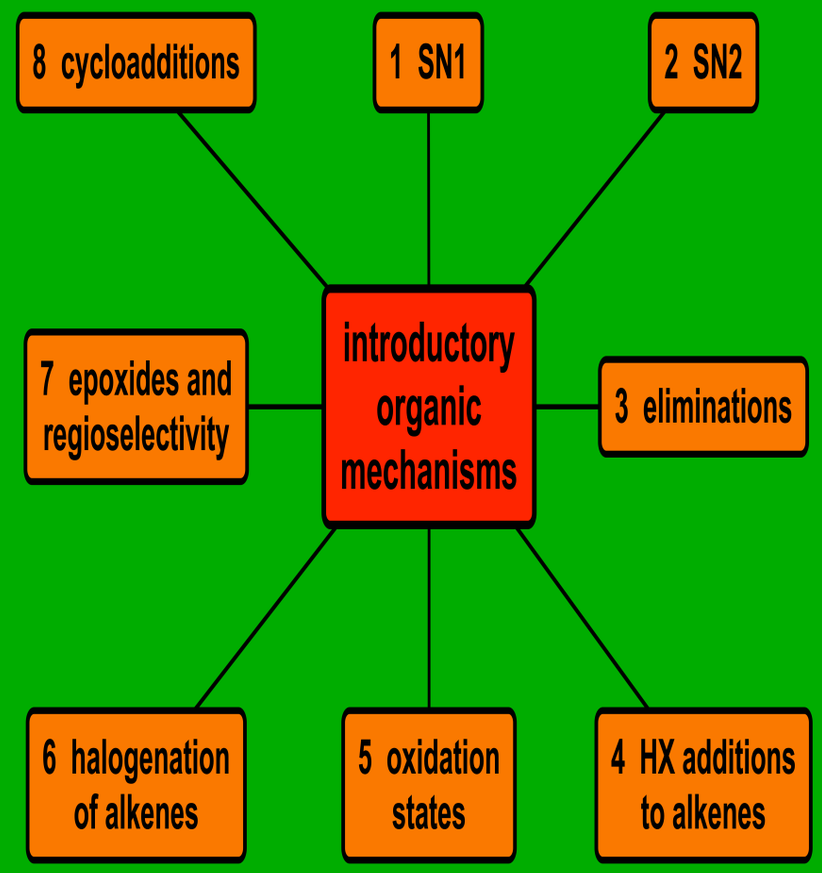


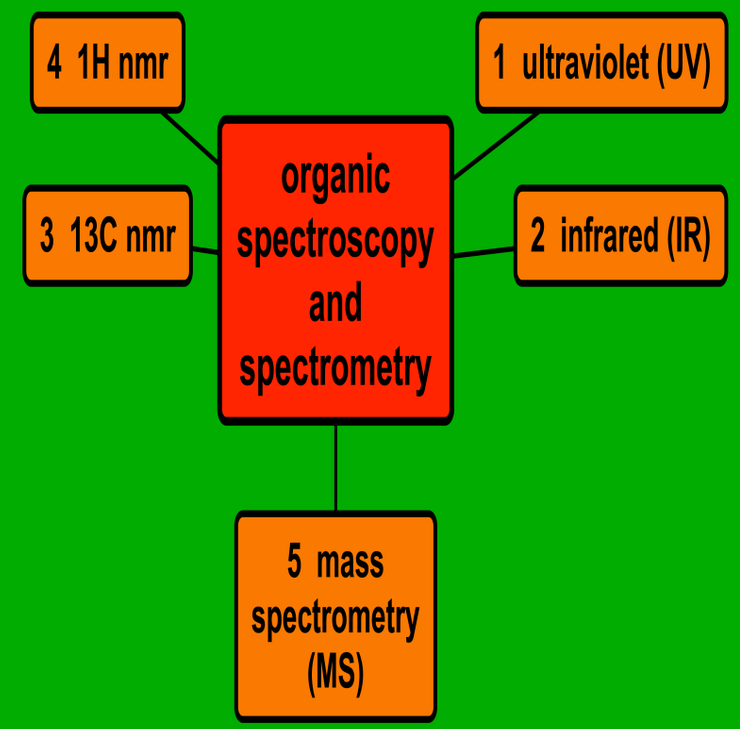

My intention is to produce modular animated eworkbooks combining subjects which mesh, to cover the whole of introductory organic chemistry. Users can select modules, and order them according to their curriculum. These eworkbooks guide students to watch video overviews, then get started on problem solving. Presenting videos with problems on the same topic avoids students having to do distracting web searches.

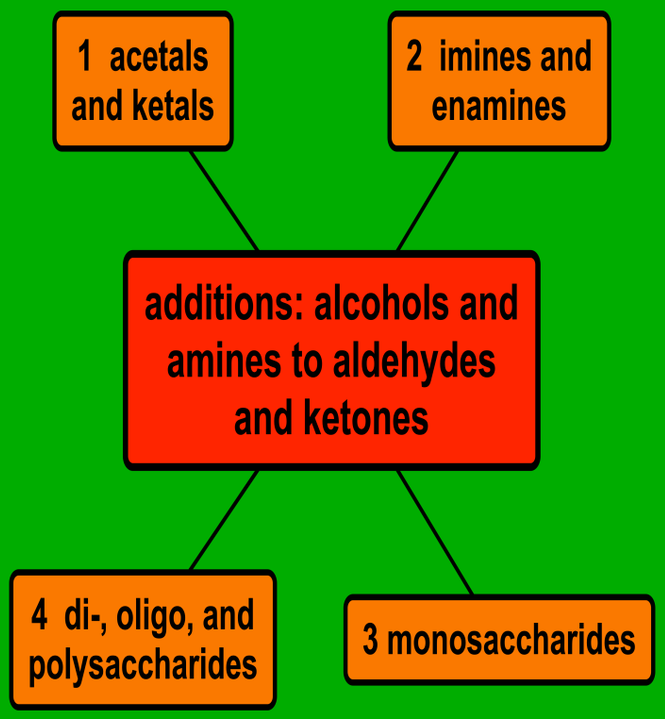
More than 90% of introductory organic chemistry students major in in life sciences, *eg* pre-med, dentistry, biochemistry, virology, immunology, chemistry *etc.. They need more about chemistry associated with the major biomolecules: lipids, carbohydrates, nucleic acids, peptides and proteins than they usually get, and less about laboratory synthetic methods.* A premed once told me she went through the whole of sophomore organic chemistry and did not learn *anything* relevant to medical school: that should not happen. Consequently, I favor syllabi introducing major biomolecules early and revisiting them often. For instance, the last eworkbook has acetals and ketals with carbohydrates. Imines, iminiums and enamines are in it to cover carbohydrate characterizations by degradation and homologation, and to help with amino sugars. Then it gets to the point: sugar chemistry. This eworkbook build on carbohydrate chemistry by introducing aromaticity, and heterocycles, then it too gets to a life sciences milestone: nucleic acids, from a chemists perspective, ie structure, DNA synthesis, sequencing and editing.

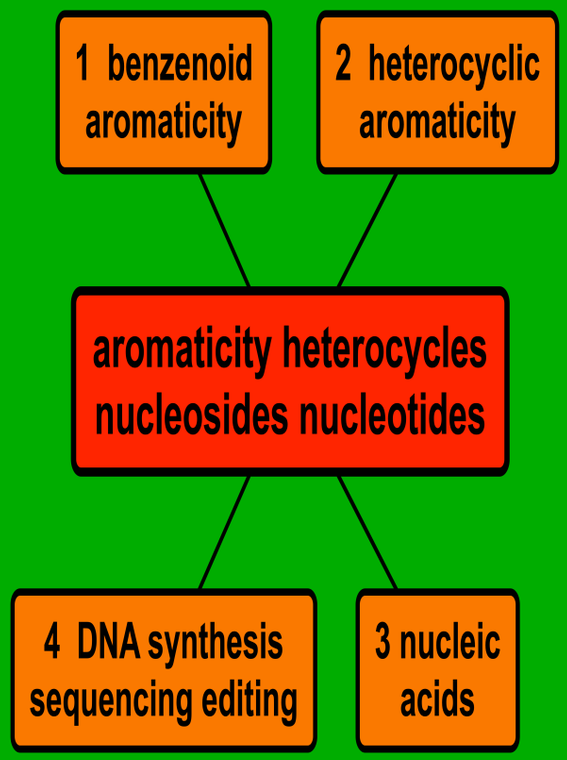
My hope is lecturers use these ebooks to ask students questions and start discussions. My ideal is students would take few notes in class, but solve problems with the professor instead. I call this semi-flipped teaching (clickable link to a video). These book should help with that, and provide a means for student self-studies.



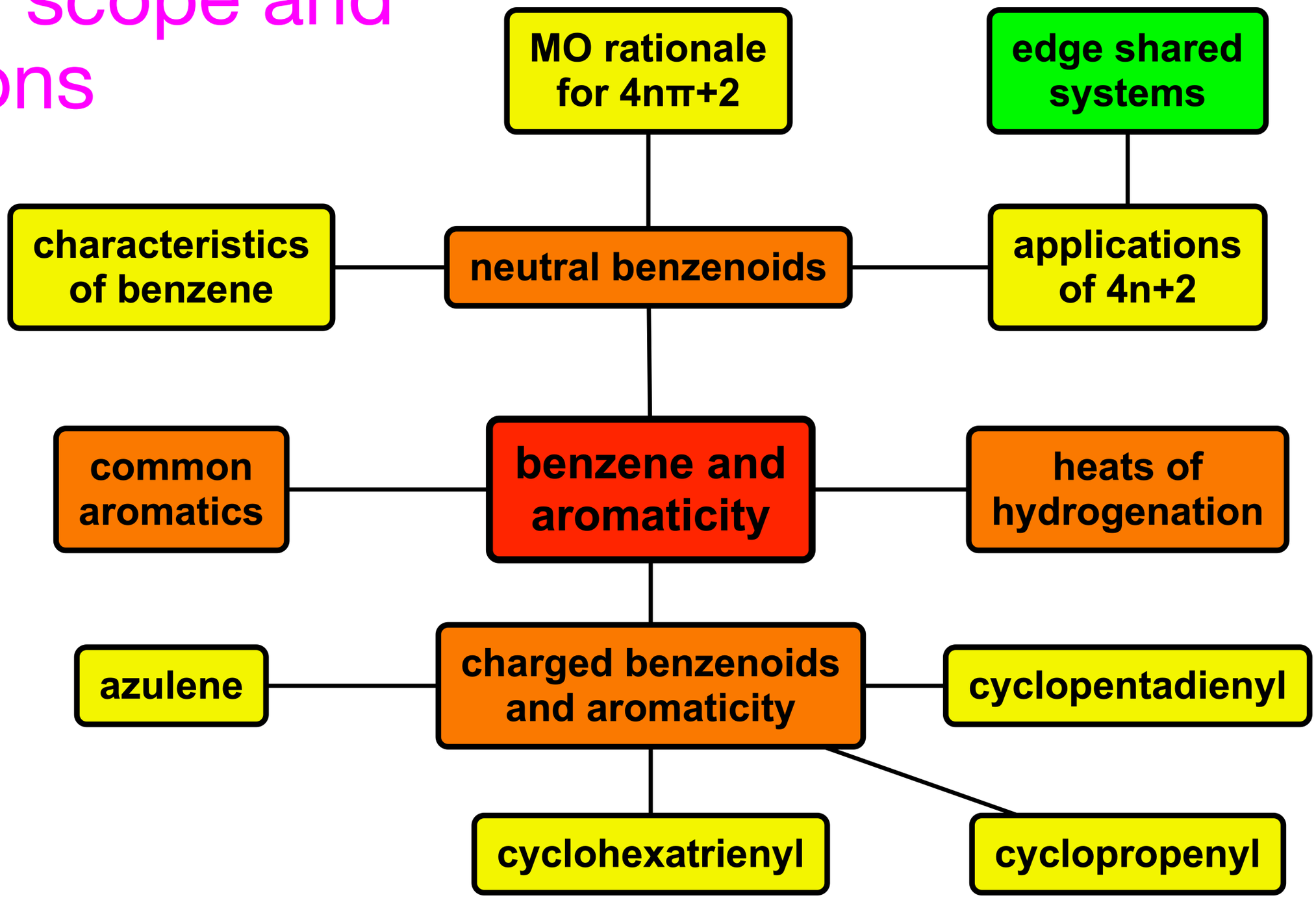




**modular animated eworkbooks from By Inquisition Press so far**

# 1 Benzenoids And Aromaticity

Hückel: scope and limitations



# B Neutral Benzenoid Compounds

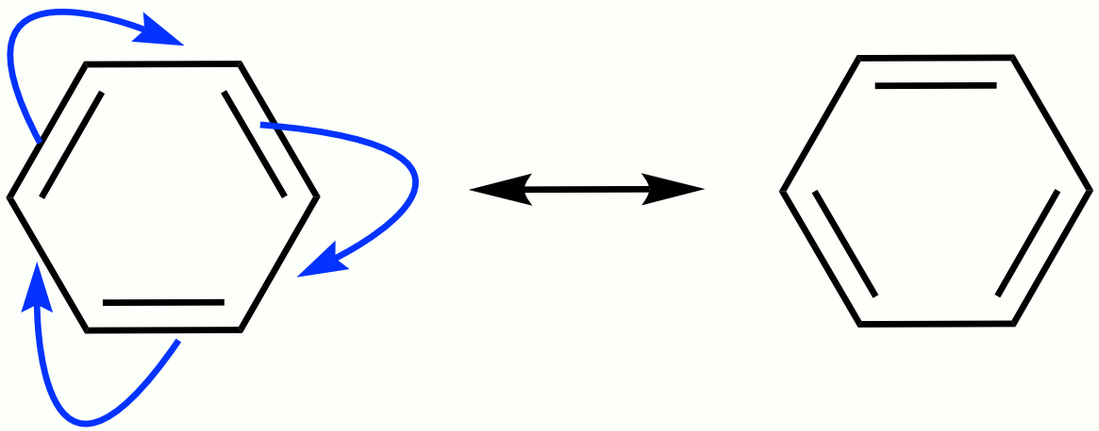
## Characteristics Of Benzene

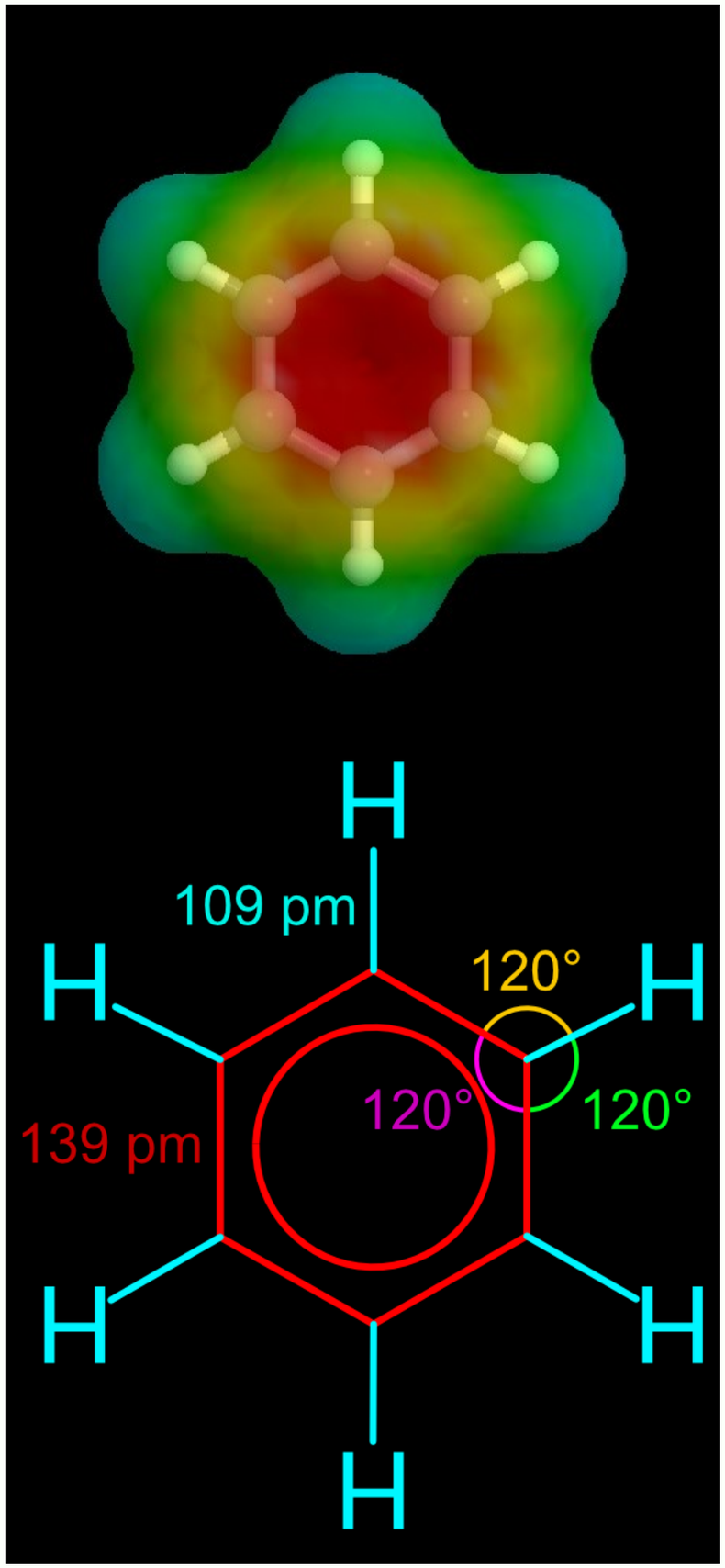
All *C*-*C* bonds in benzene are 139 pm long: between typical *C*-*C* (154 pm) and *C*=*C* (134 pm) bonds. All the C-C-C bond angles are *120°* ideal for *$sp^2$-hybridization*. Each carbon has a *p* -orbital perpendicular to the plane of the ring each containing one electron. Electron densities on carbons in benzene are *equal*. Thus benzene is a perfectly symmetrical hexagon, with six overlapping p-orbitals available to form molecular orbitals (MOs) accommodating electron density above and below it.

Benzene rings tend to make compounds containing *lipophilic* often leading to aggregation and insolubility.

Benzenoid compounds *do not* contain heteroatoms in the ring.

Draw resonance structures of benzene and illustrate these characteristics (*C*-*C* bond lengths, *C*-*C*-*C* bond angles, hybridization state) on your diagram:

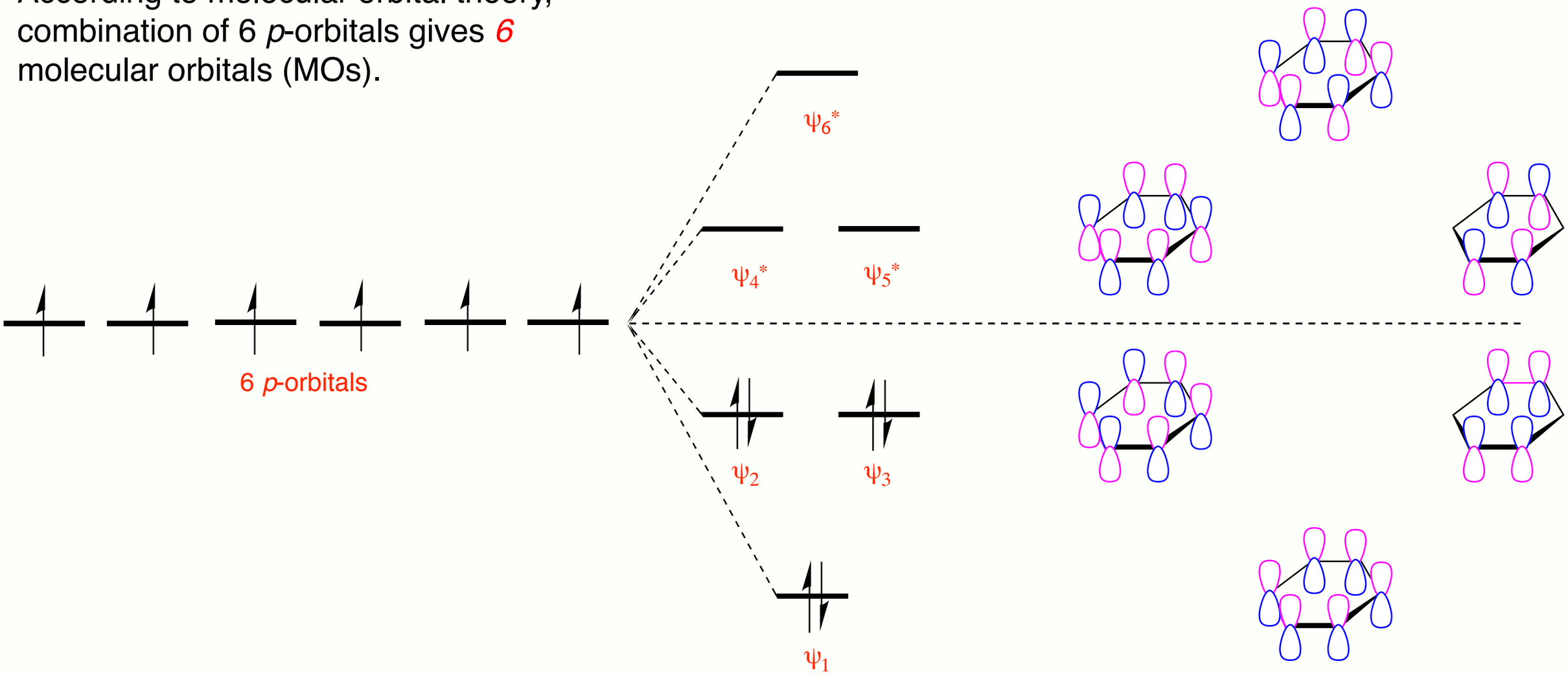




# B Neutral Benzenoid Compounds

## MO Rationale For The 4*n*+2 Rule

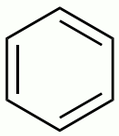
According to molecular orbital theory, combination of 6 *p*-orbitals gives *6* molecular orbitals (MOs).



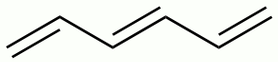

Aromatic molecules must be *cyclic*, *conjugated*, *planar*, and must have 4*n*+2 π-electrons (*n* = integer) according to *The Hückel Rule*. Based on the diagram above, explain why 4*n* + 2 might be a significant number for aromatic compounds. *The number of electrons needed to fill HOMOs of aromatic compounds tends to follow the order 2, 6, 10, 14 in the molecular diagram above, so 4n + 2 electrons are required. This occurs because the HOMOs and LUMOs and all the other orbitals above the lowest and below the highest, are degenerate*.

# C Neutral Benzenoid Compounds

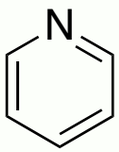
## Applying 4*n* + 2



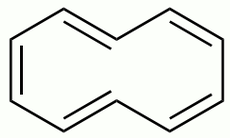
*aromatic* because it has {4(1)+2} = 6 π-electrons, follows Huckel Rule.



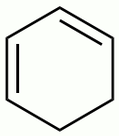
not aromatic because this is not cyclic compound.




*aromatic* because it has {4(1)+2} = 6 π-electrons, follows Huckel Rule.



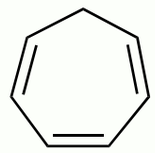
*aromatic* because because it has {4(2)+2} = 10 π-electrons, follows Huckel Rule.



not aromatic because it has 4 π-electrons, does not follow Huckel Rule.

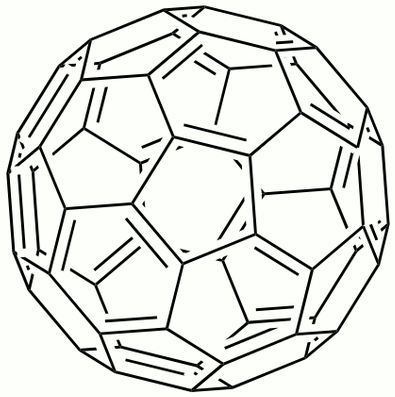
not aromatic because it has 4 π-electrons, does not follow Huckel Rule.

not aromatic because it has 4 π-electrons, does not follow Huckel Rule.

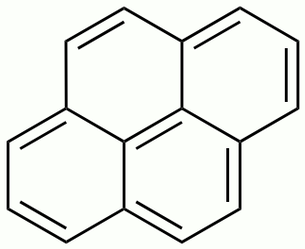


not aromatic because of nonplanarity of the methylene bridge

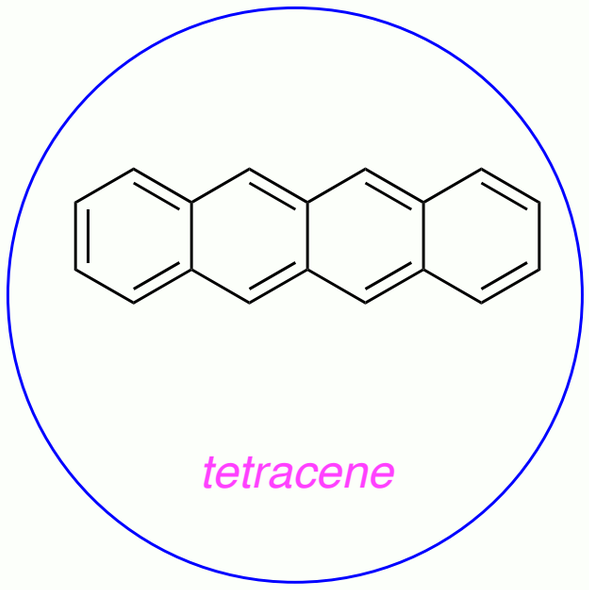
# B Neutral Benzenoid Compounds



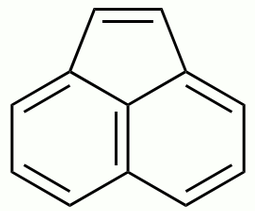
$C_{60}$



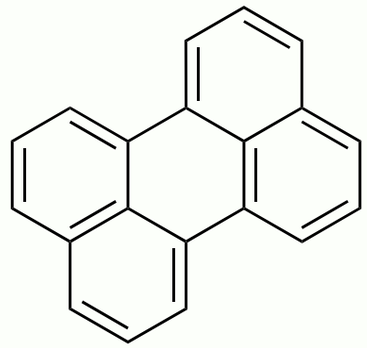
*pyrene*



*tetracene*



*acenaphthalene*



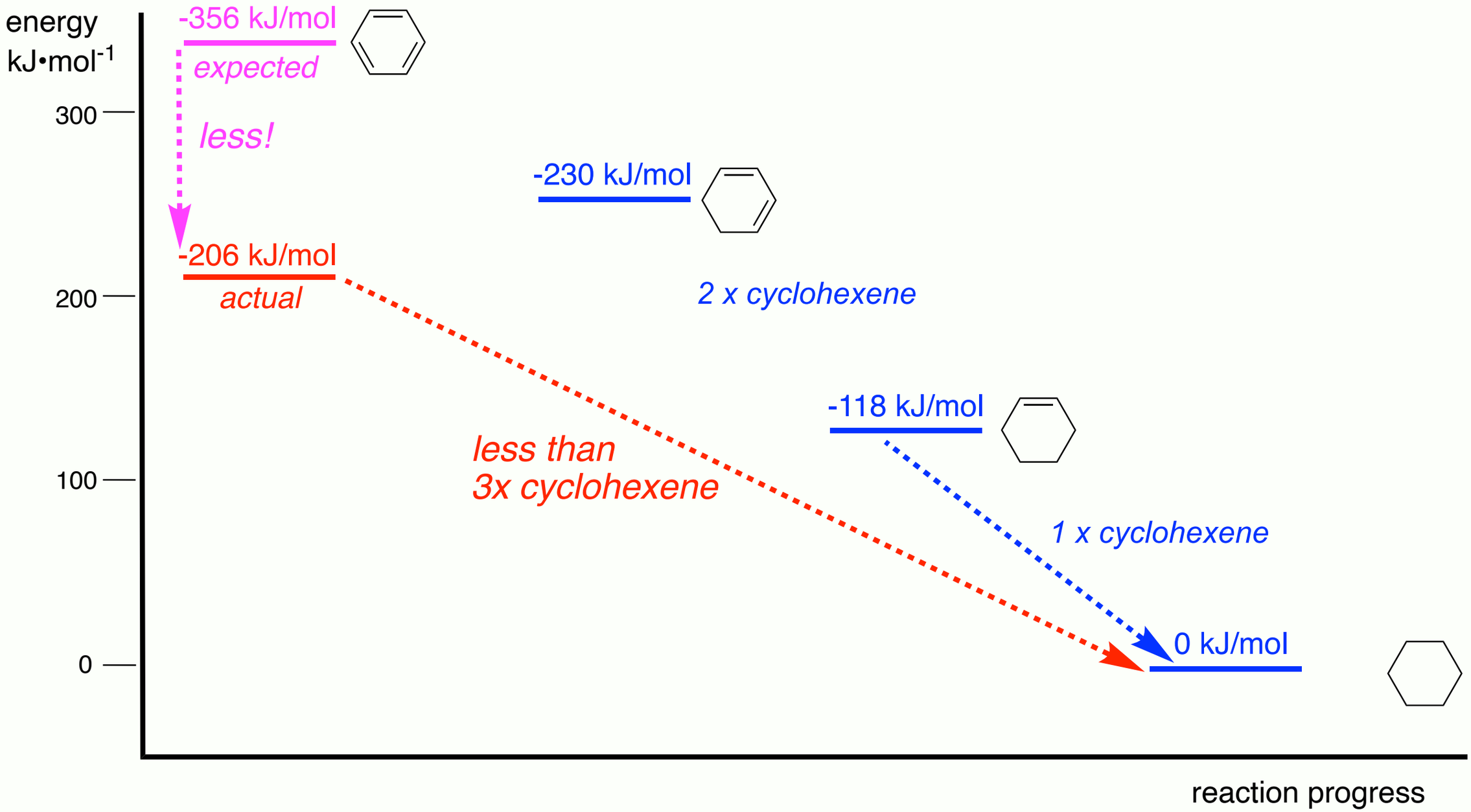
*perylene*

The 4n + 2 rule *is not* inviolable, especially for extended systems.

# C Heats Of Hydrogenation

Energies *liberated* when hydrogen is added across a *C=C* bond are *heats of hydrogenation*.

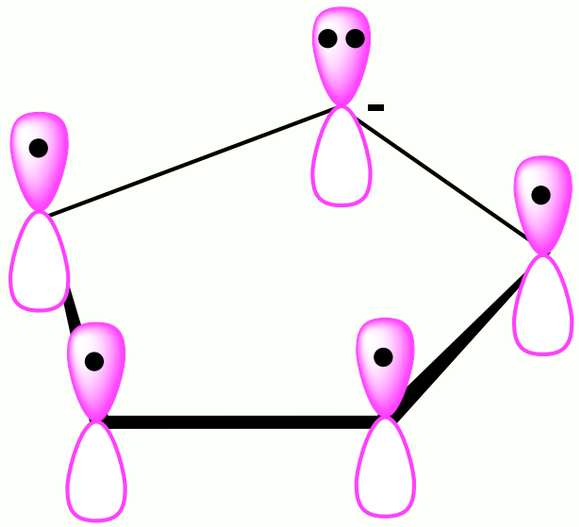
Hydrogenations of different compounds to give the *same* product *can* be used to gauge starting material relative stabilities. Heats of hydrogenation in the following series are cyclohexene (-118 kJ/mol); 1,3-cyclohexadiene (-230 kJ/mol). Hydrogenation of these molecules gives *cyclohexane*. Benzene is *more* stable than expected because its heat of hydrogenation is -206 kJ/mol, *ie less* than expected.



# D Charged, Conjugated, Carbocyclic Systems

## Cyclopentadienyl Systems

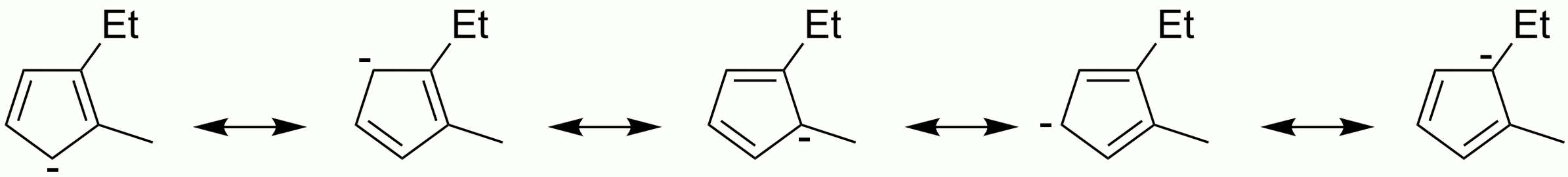
The cyclopentadienyl *anion* is *aromatic*.




*cyclopentadienyl anion*

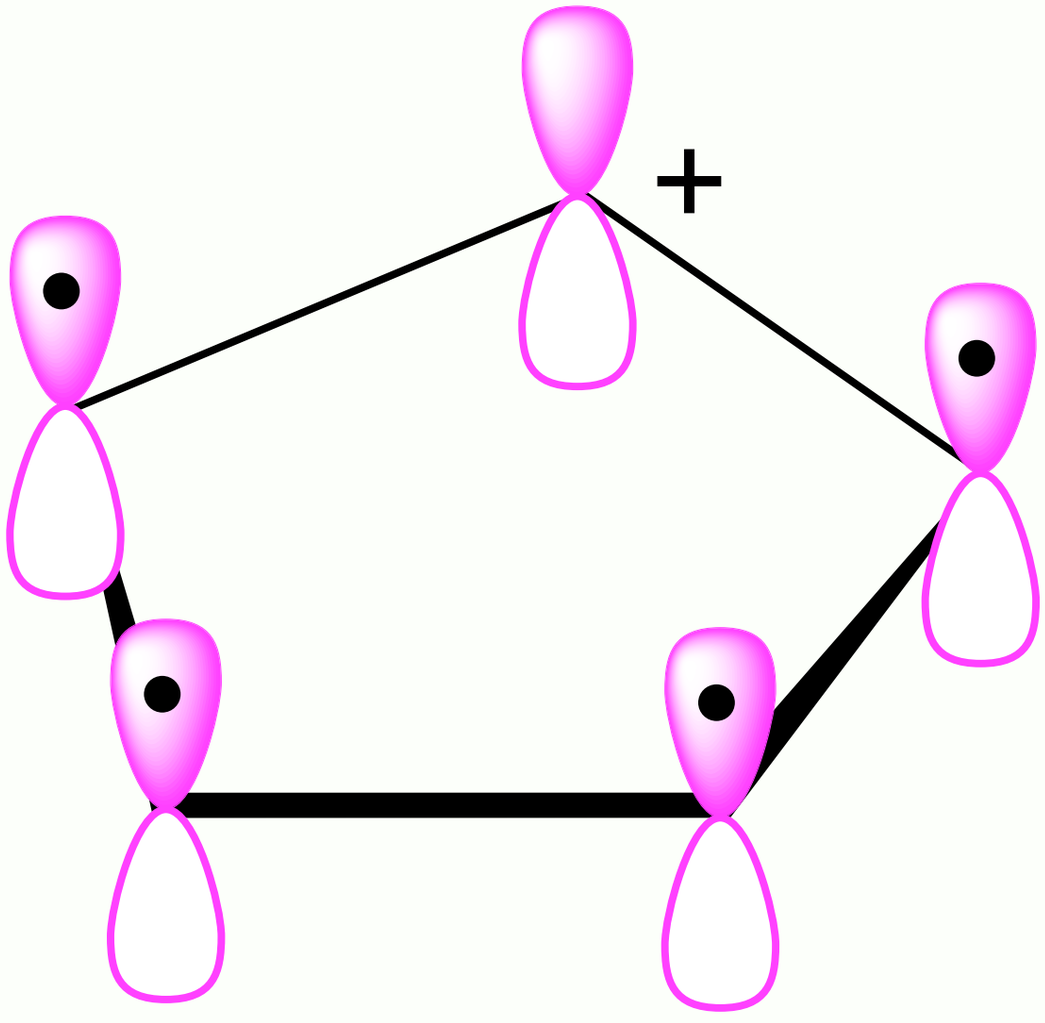
*6πe aromatic*

Resonance structures of the 1-ethyl-2-methylcyclopentadienyl anion.



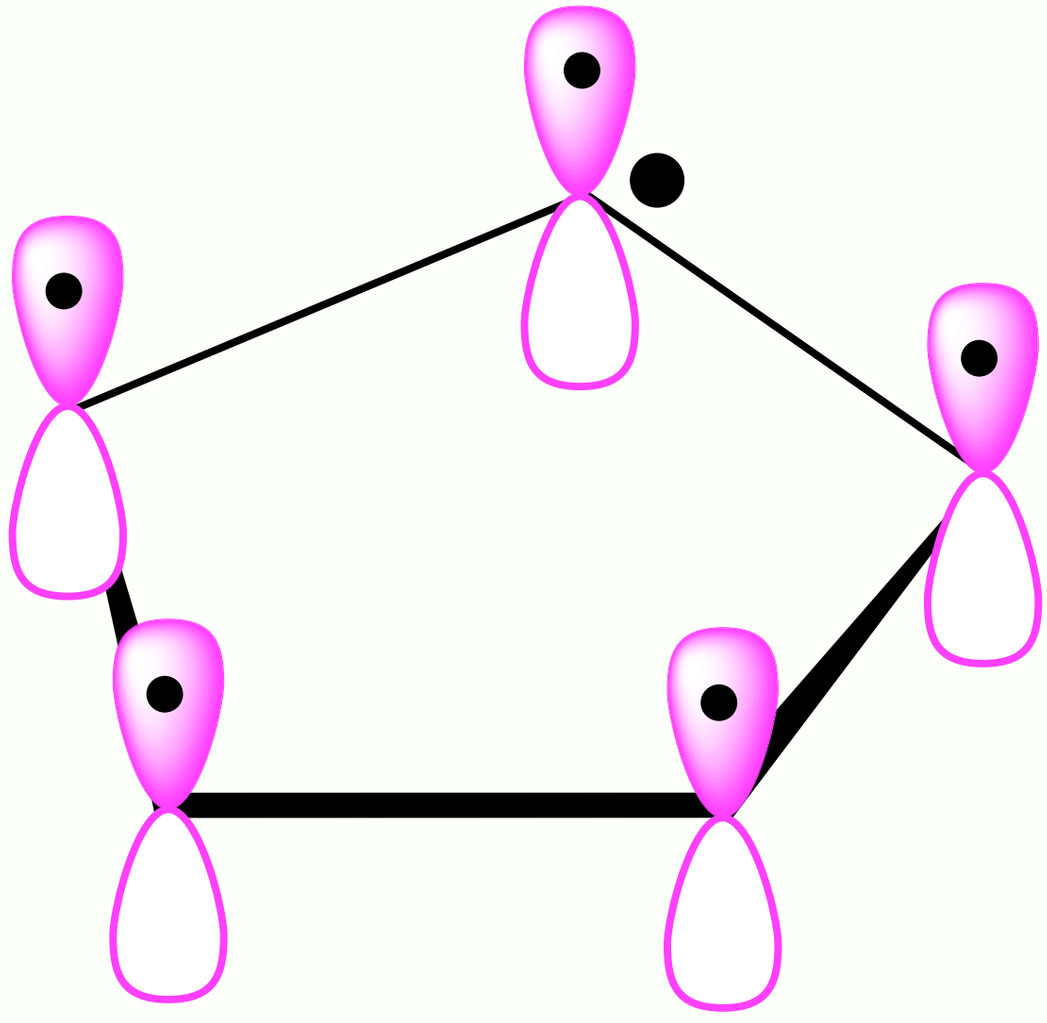
# D Charged, Conjugated, Carbocyclic Systems

Cyclopentadienyl *cation p*-orbitals. Number of electrons in each orbital that contributes to the delocalized system above the ring of: cyclopentylium cation (left) and cyclopentyl radical (right).




*cation*

*4πe, not aromatic*



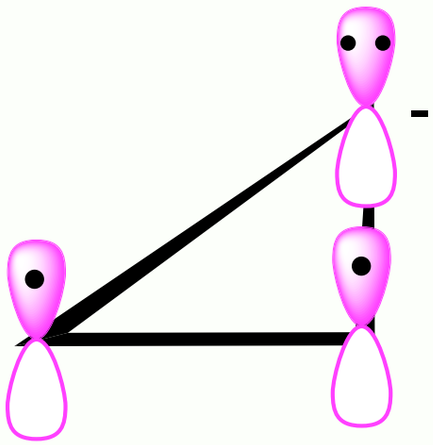
*neutral radical*

*5πe,not aromatic*

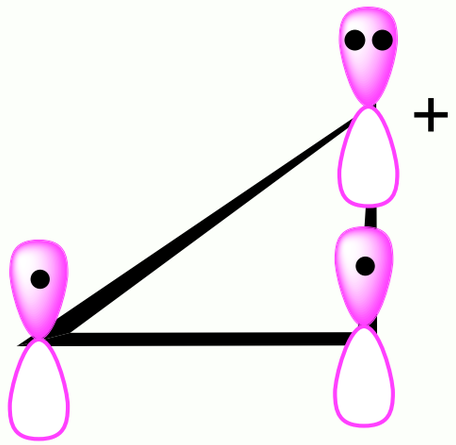
# D Charged, Conjugated, Carbocyclic Systems

## Cyclopropenyl Systems

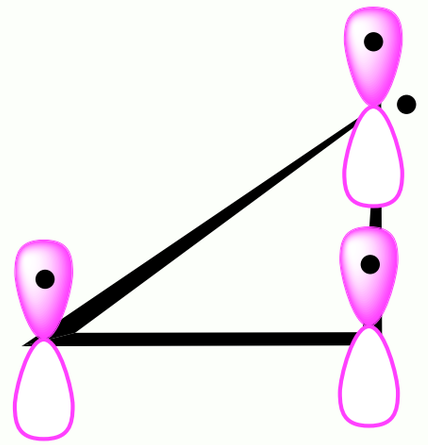
*p*-Orbitals for three cyclopropene derivatives, and electrons each orbital contributes to the delocalized system above the ring.




*4πe, not aromatic*

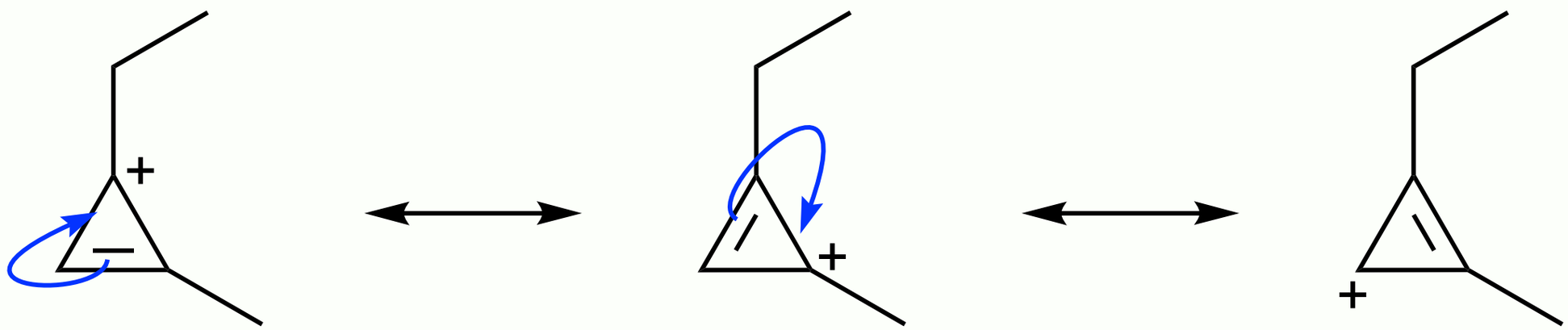



*2πe, aromatic*



*5πe, not aromatic*

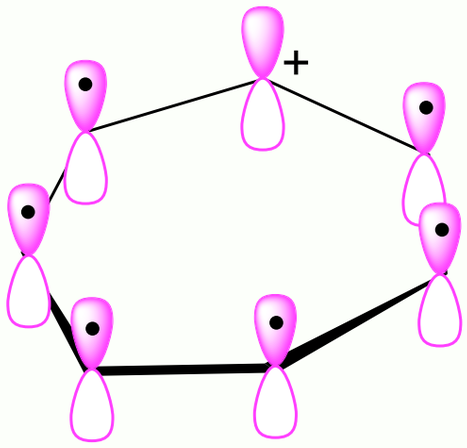
Draw all resonance structures.



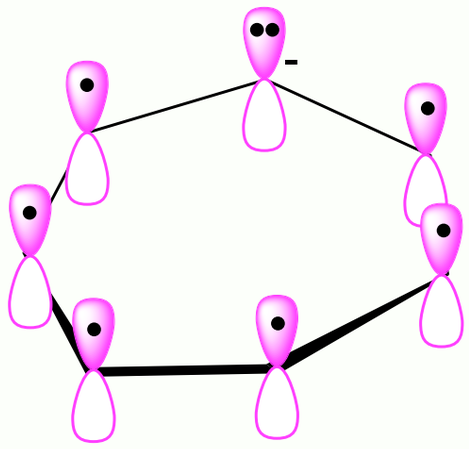
# D Charged, Conjugated, Carbocyclic Systems

## Cycloheptatrienyl Systems

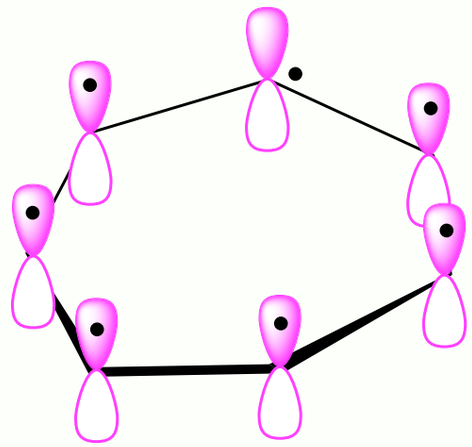
*p*-Orbitals for three cycloheptatrienyl derivatives.



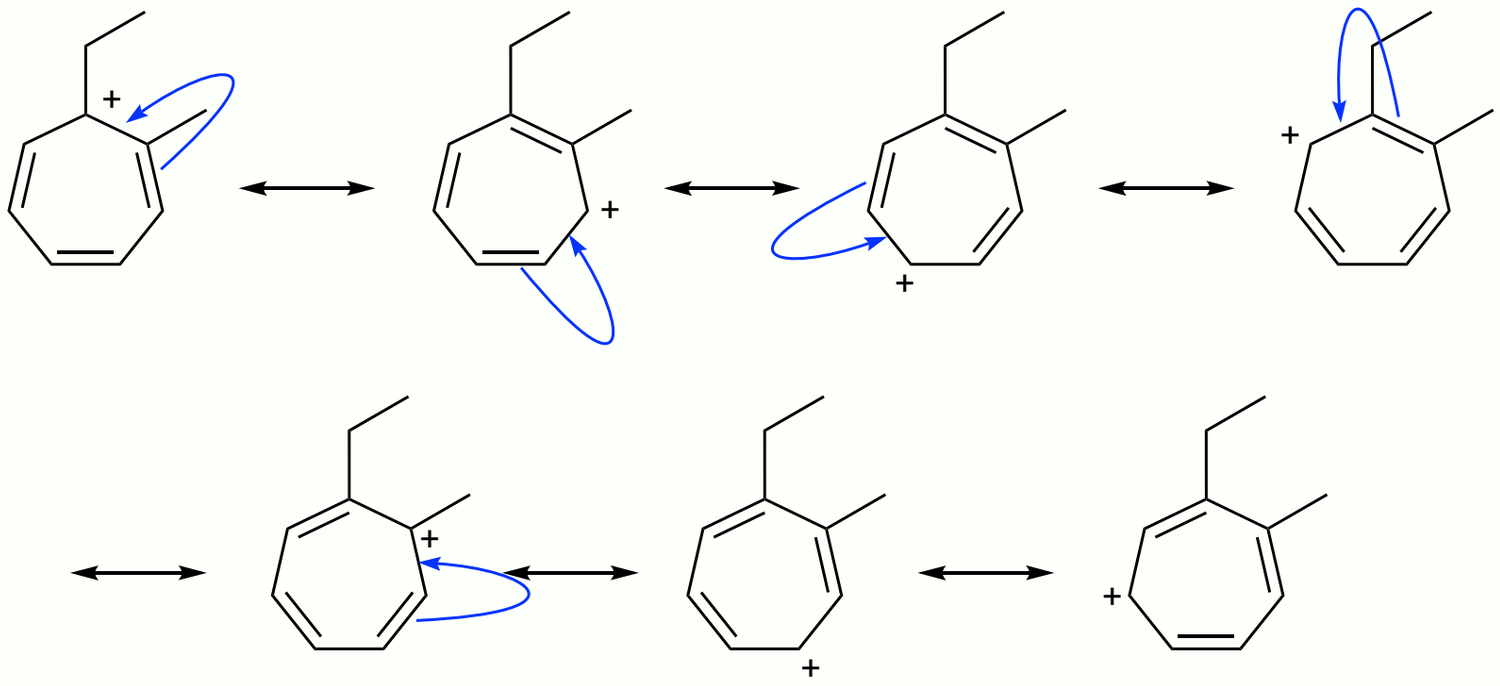

*cation*
*6πe, not aromatic*




*anion*
*6πe, not aromatic*



*radical*
*6πe, not aromatic*

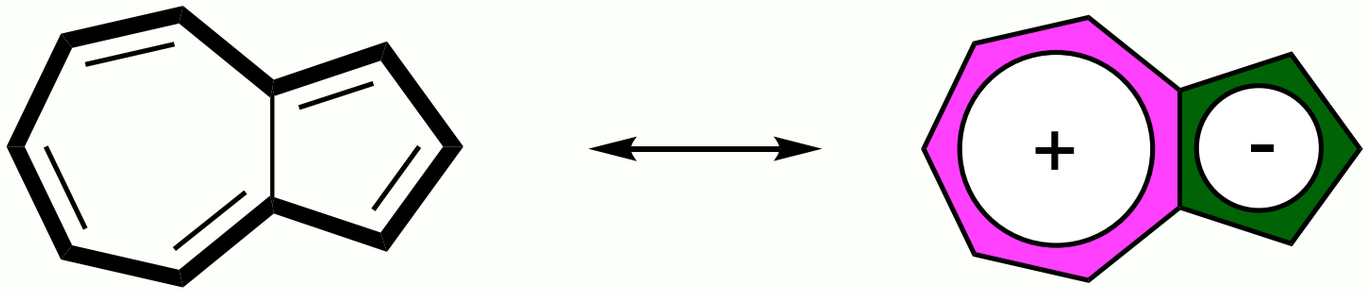



Cycloheptatrienyl cations are called tropylium. There are *7* resonance structures of 1-ethyl-2-methyl tropylium.

# D Charged, Conjugated, Carbocyclic Systems

## Azulene

Draw azulene resonance structures to explain why it is polar, with a negative charge in the five-membered ring and a positive one in the six.



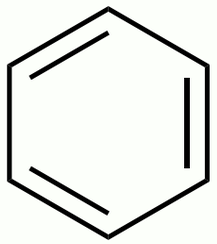

*aromatic but not 7 or 5 membered rings*

*7 and 5 membered rings aromatic*

# E Common Aromatic Compounds

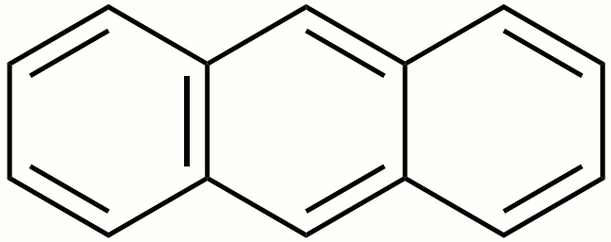
As the name implies, aromatic compounds tend *to* have distinctive odors. They also react *differently to* aliphatic compounds. Industrially they can be formed by distillation from *oil* or by heating petroleum to a high temperature over *a catalyst*.

Examples of aromatic compounds here (do not use examples from next question).



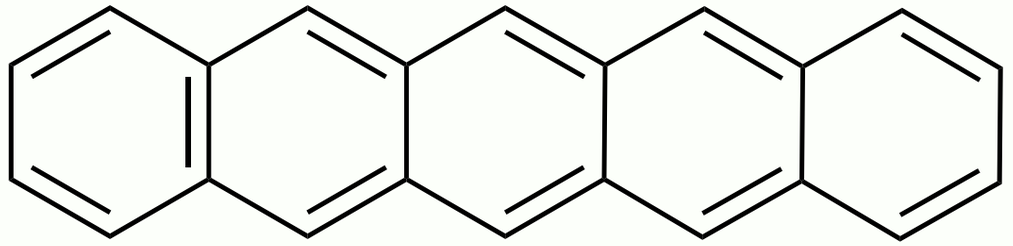
6πe

*benzene*



18πe

*your example 1*

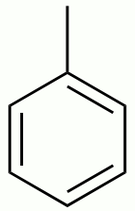


22πe

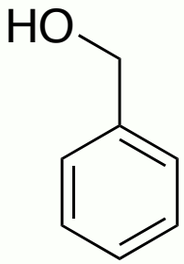
*your example 2*

# E Common Aromatic Compounds

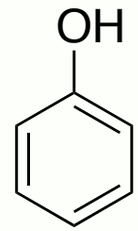
Show structures for the following common aromatic compounds (can be found on Wiki).



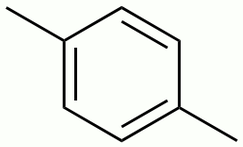
*toluene*



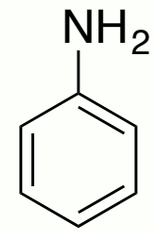

*benzyl alcohol*



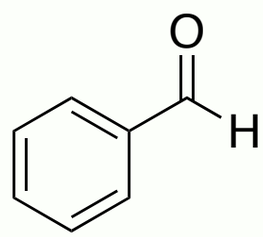

*phenol*



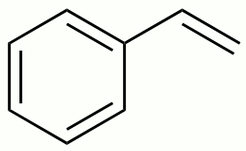
*para-xylene*



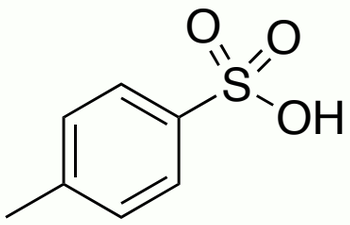

*aniline*



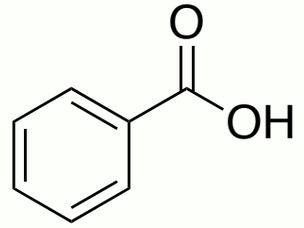

*benzaldehyde*



*styrene*

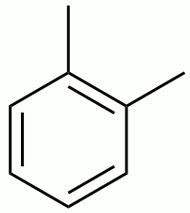



*tosic acid*

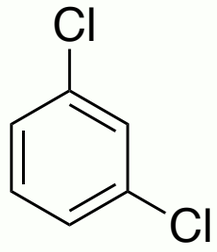



*benzoic acid*

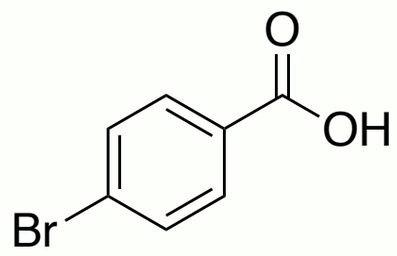
# E Common Aromatic Compounds



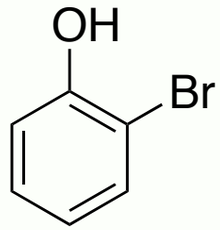
*1,2-dimethylbenzene*



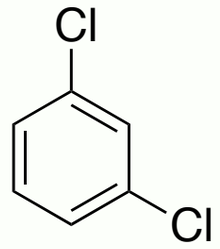

*meta-dichlorobenzene*



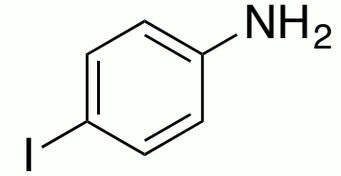

*para-bromobenzoic acid*




*ortho-bromophenol*

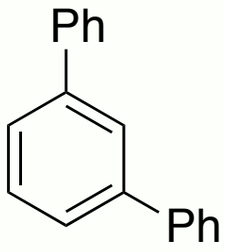



*1,3-dichlorobenzene*

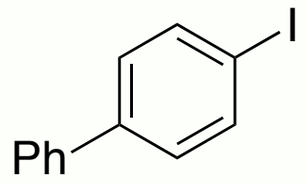



*4-iodoaniline*


*2-bromostyrene*



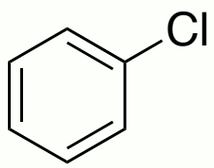

*1,3-diphenylbenzene*



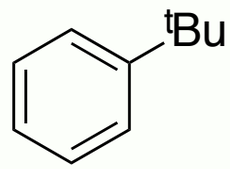

*para-phenyliodobenzene*

# E Common Aromatic Compounds

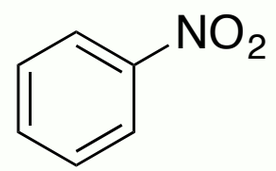
Rigorous names for *monosubstituted* benzenes are based upon “benzene” provided the longest chain substituent does not have more *C*-atoms than the benzene ring, *ie* 6.




*chlorobenzene*




*tert-butylbenzene*



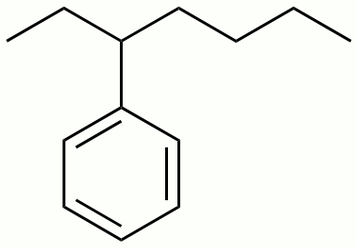

*nitrobenzene*

Benzenoid compounds *do not* contain heteroatoms in the ring.

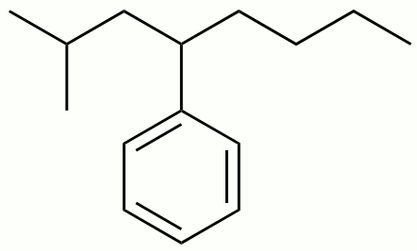
Benzene rings tend to make compounds *lipophilic* often leading to aggregation and insolubility in water.

# E Common Aromatic Compounds

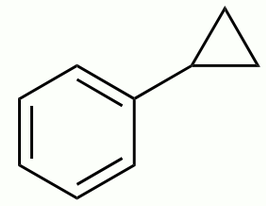
If the substituent has more carbon atoms than a benzene ring, then the $C_6H_5$- unit is a *phenyl* substituent (Ph).



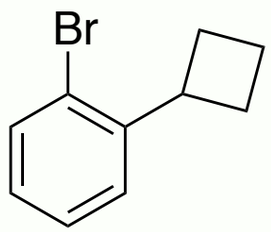
*3-phenylheptane*



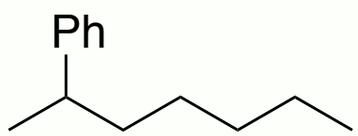
*2-methyl-4-phenyloctane*



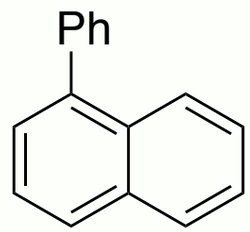
*phenylcyclopropane*




*1-bromo-2-cyclobutylbenzene*




*2-phenylheptane*

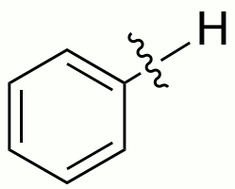



*1-phenylnaphthalene*

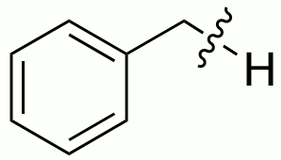
One of those names is wrong!  Give it the correct name: *cyclopropylbenzene*.

# E Common Aromatic Compounds

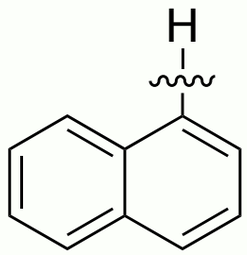
Draw jagged lines that separate a hydrogen atom from:




*phenyl group in benzene*



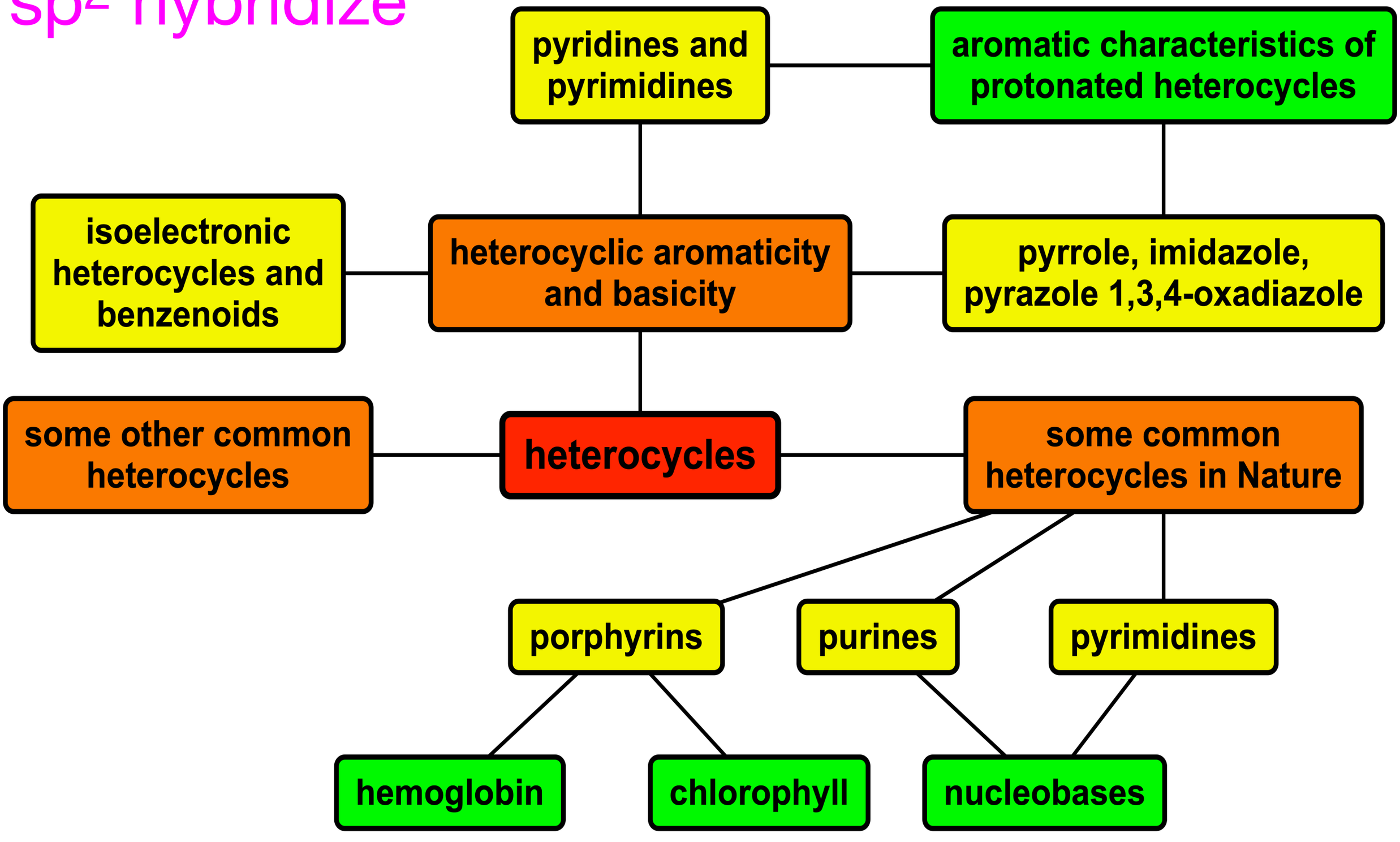

*benzyl group in toluene*




*1-naphthyl group in naphthalene*

# 2 Heterocycles And Aromaticity

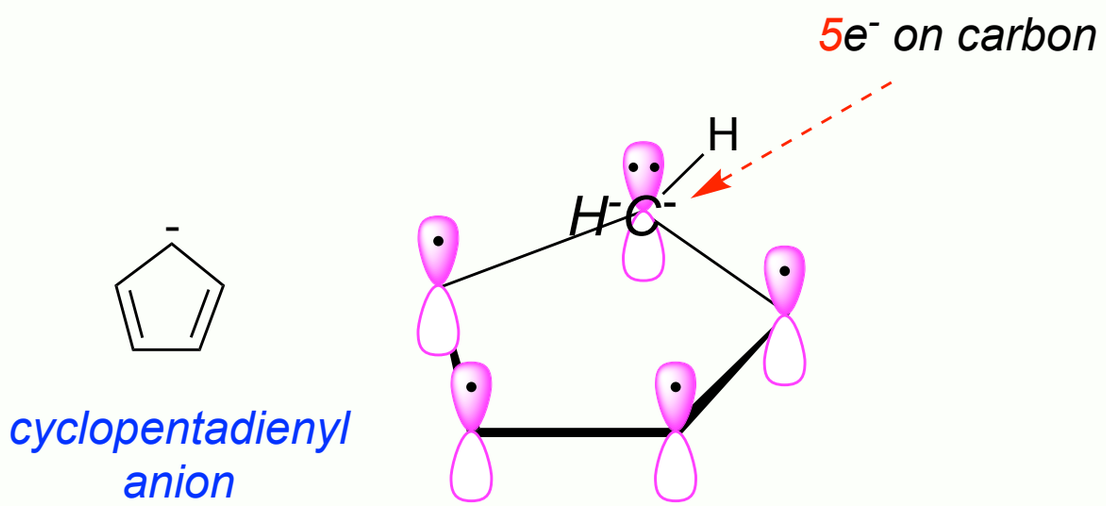
aromaticity can drive heteroatoms to $sp^2$ hybridize

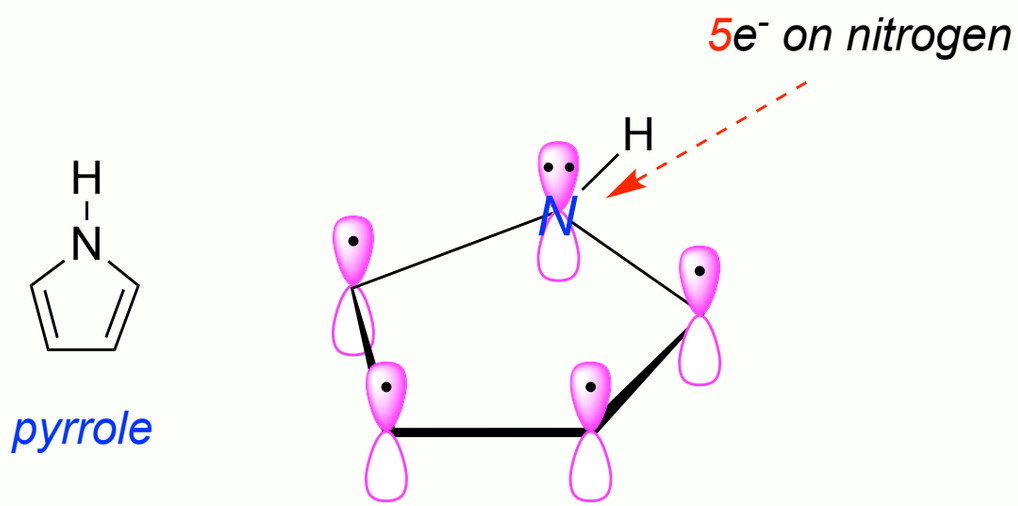


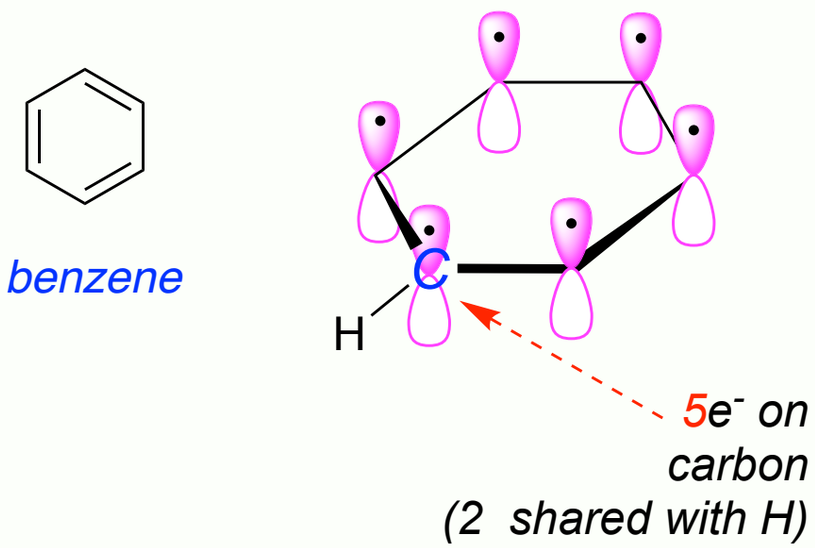
# B Heterocyclic Aromaticity And Basicity

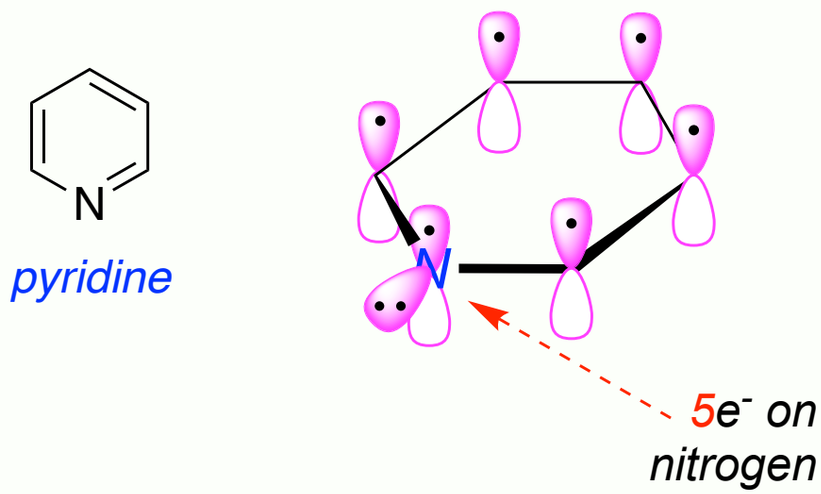
## Some Heterocycles And Benzenoids Are *Isoelectronic*

Numbers of electrons on the atoms below indicate they *are* isoelectronic. Cyclopentadienyl anion *is* aromatic, and pyrrole *is* aromatic.



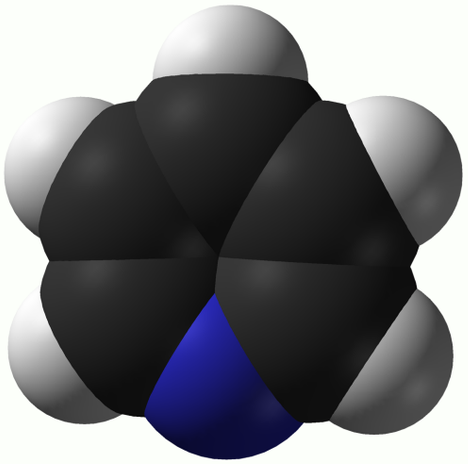



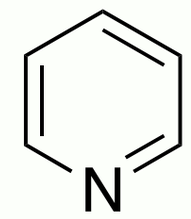



# B Aromaticity And Basicity Of Heterocycles

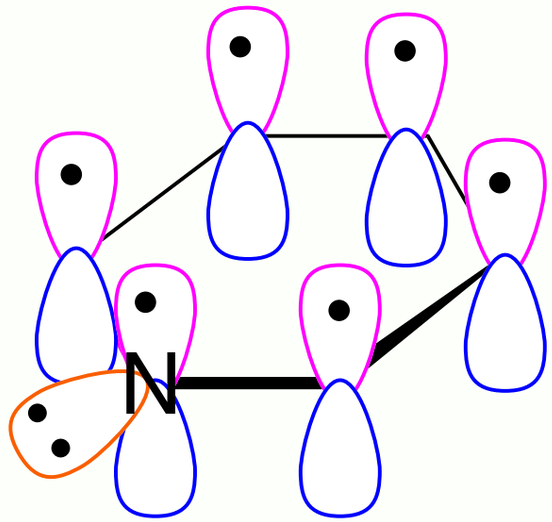
## Pyridines And Pyrimidines

In pyridine, the *N*-atom is *sp*$^2$-hybridized with *a lone pair* occupying one of the lobes. There is *1* electron in the unhybridized *p*-orbital so it is *aromatic*.





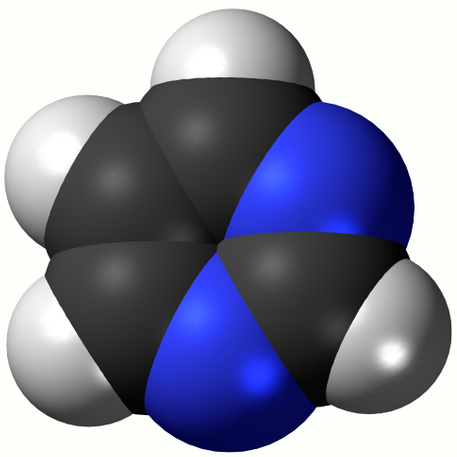

*pyridine*

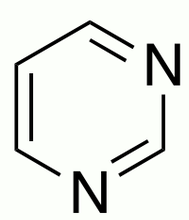


# B Aromaticity And Basicity Of Heterocycles

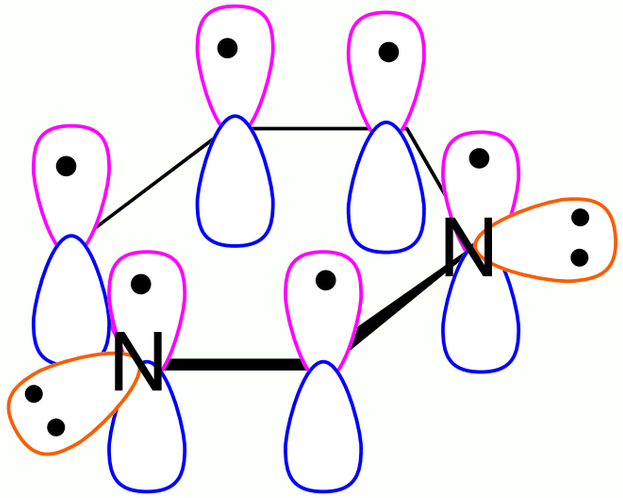
In 1,3-pyrimidine, the *N*-atoms are *equivalent because of resonance*, $sp^2$-hybridized with a *lone pair* occupying one of the lobes, and there is/are *1* electron(s) in the unhybridized *p*-orbital.

1,3-Pyrimidine is *aromatic*.






*1,3-pyrimidine*

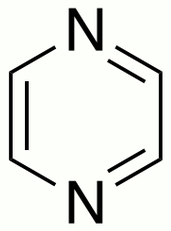



Pyrimidines are *less* basic than pyridines because of the electron withdrawing influences of two nitrogen atoms in one ring.

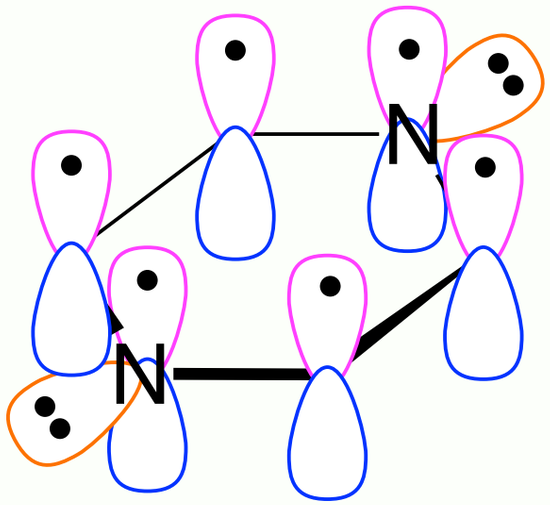
# B Aromaticity And Basicity Of Heterocycles

1,4-Pyrimidine's *N*-atoms are *equivalent because of resonance*, $sp^2$-hybridized with *a lone pair* occupying one of the lobes, and there is *1* electron in the unhybridized *p*-orbital.

1,4-Pyrimidine is *aromatic*.




*1,4-pyrimidine*



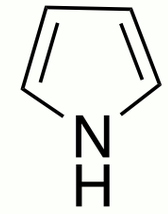
# B Aromaticity And Basicity Of Heterocycles

## Pyrrole, Imidazole, Pyrazole, And 1,3,4-Oxadiazole

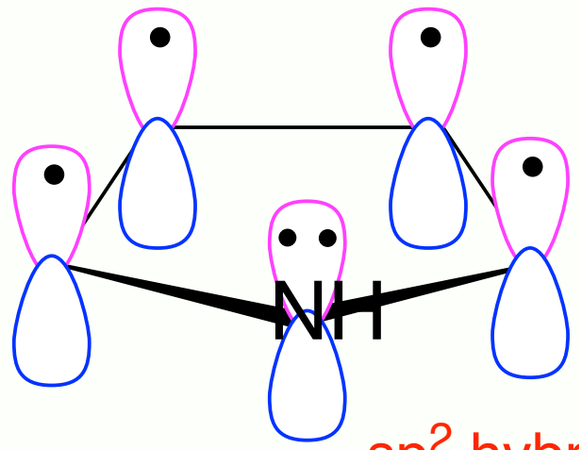
In pyrrole, the *N*-atom is *sp*$^2$-hybridized, and there are *2* electron(s) in the unhybridized *p*-orbital.

Pyrrole is *aromatic*.

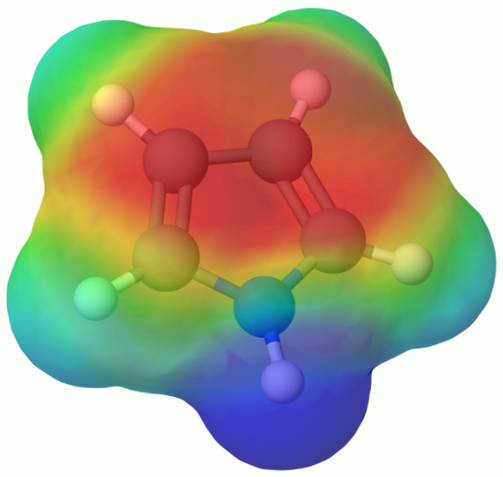
Pyrrole's nitrogen *cannot* be sp$^3$-hybridized and simultaneously give a flat conjugated aromatic system.

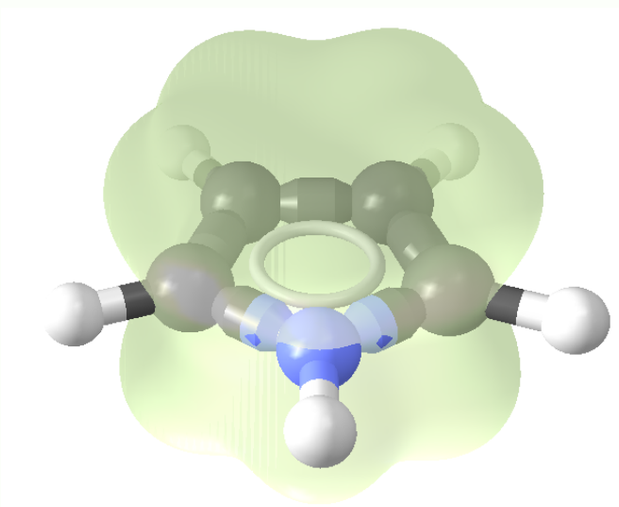



*pyrrole*




sp$^2$ hybridized

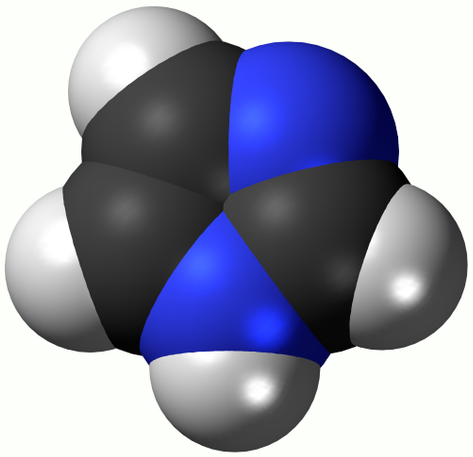




# B Aromaticity And Basicity Of Heterocycles

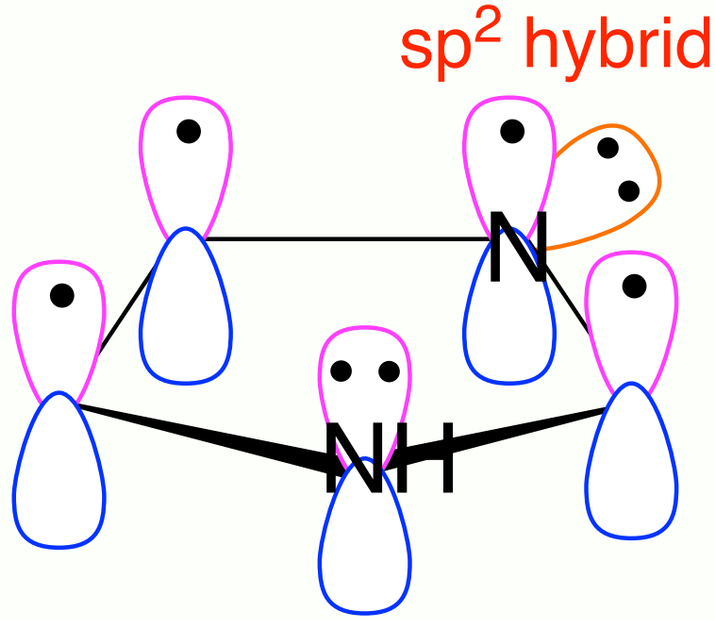
Imidazole contains two *different* *N*-atoms. They *can* interconvert via tautomerism.

The nitrogen atoms in imidazole *are* both $sp^2$-hybridized, and *only one* contributes 2*e* to the aromatic system at any instant, consequently this heterocycle *is* basic.






*imidazole*

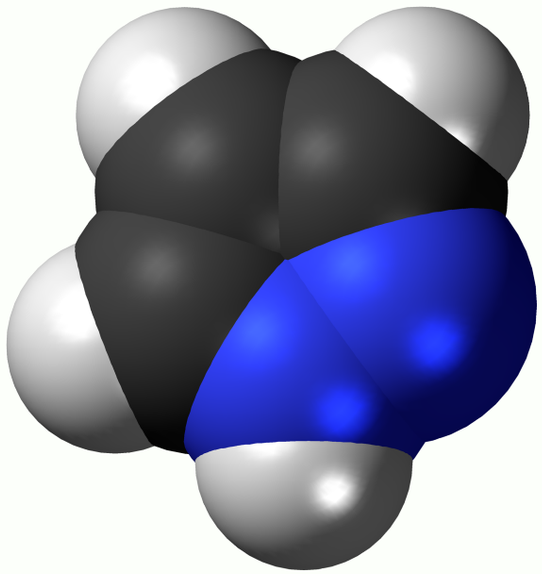


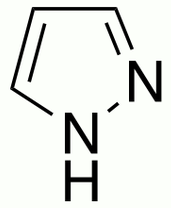
# B Aromaticity And Basicity Of Heterocycles

Pyrazole has two *different* *N*-atoms. Their types *can* interconvert via tautomerism.

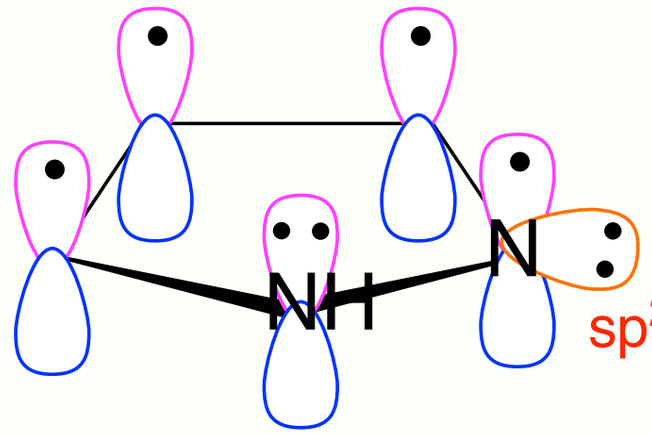
Pyrazole nitrogen atoms *are* both $sp^2$-hybridized, and *one* contributes 2*e* to the aromatic system at any instant.

Pyrazole *is* easily protonated relative to pyrrole.






*pyrazole*

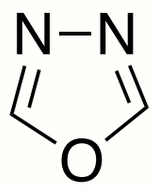



In nitrogen heterocycles like pyrrole and imidazole it is possible for a *N*-atom to be $sp^3$ hybridized, but it preferentially $sp^2$ hybridizes to gain *aromatic stabilization*.

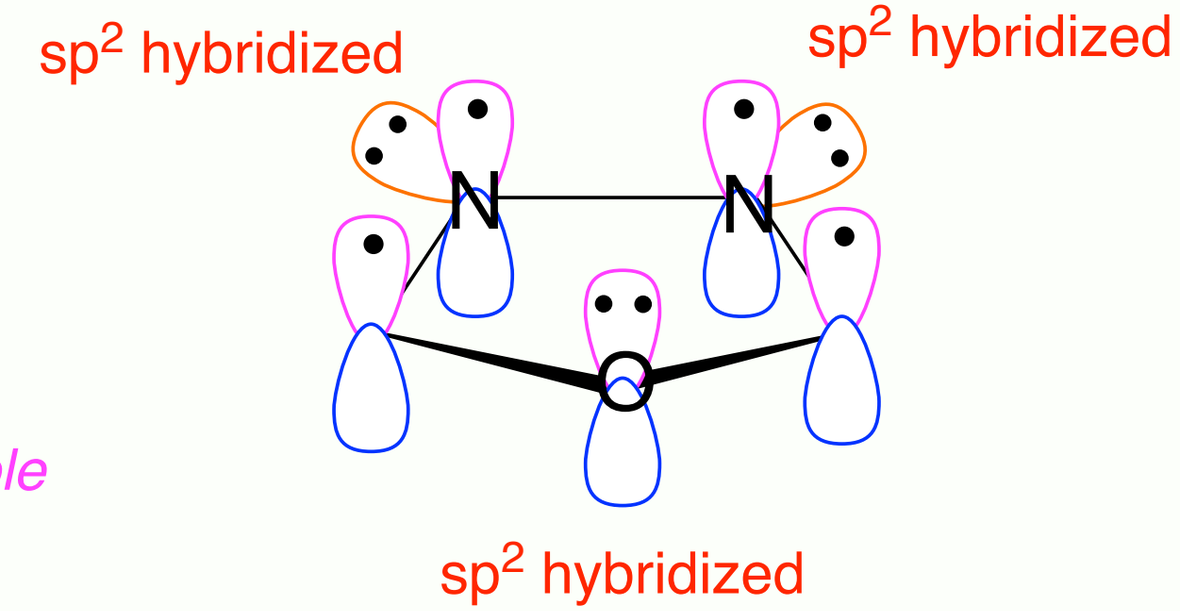
# B Aromaticity And Basicity Of Heterocycles

*N*-Atoms in 1,3,4-oxadiazole must be *sp*$^2$-hybridized.

The *O*-atom is *sp*$^2$-hybridized and contributes *2 e*$^-$ to the aromatic system.




*1,3,4-oxadiazole*



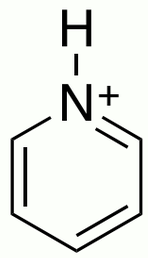

Protonation of a lone pair *in a sp*$^2$*-hybridized orbital* within an aromatic heterocycle means the atom *stays sp*$^2$, hence protonation *does not* remove electrons from the 4*n* + 2 delocalized system.

The *N*-atom contributing that lone pair is a relatively *basic* center because aromatic stabilization *is not* lost on protonation.

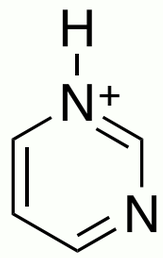
# B Aromaticity And Basicity Of Heterocycles

## Aromatic Characteristics Of Protonated Heterocycles

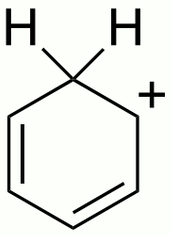
Indicate numbers of π-electrons and the following systems are aromatic or not.



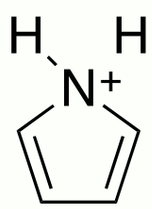
*aromatic* because it has 6 $\pi e^-$.



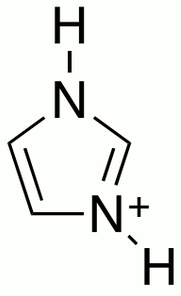
*aromatic* because it has 6 $\pi e^-$.



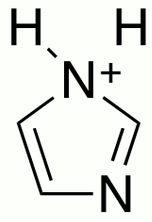
*not aromatic* because it has 4 $\pi e^-$.



*not aromatic* because it has 4 $\pi e^-$.

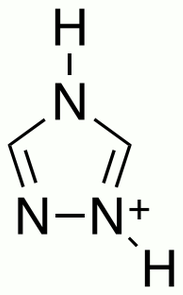


*aromatic* because it has 6 $\pi e^-$.

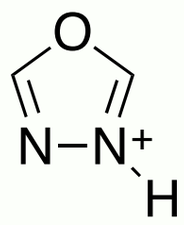


*not aromatic* because it has 4 $\pi e^-$.

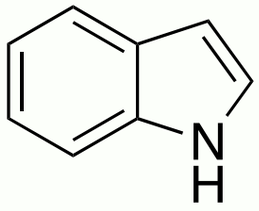
## B Aromatic Characteristics Of Protonated Heterocycles



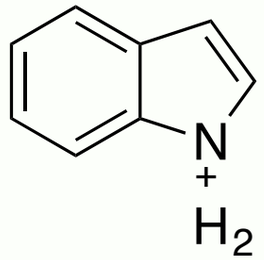
*aromatic* because it has 6 $\pi e^-$.



*aromatic* because it has 6 $\pi e^-$.



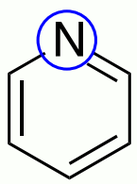
*aromatic* because it has 10 $\pi e^-$.



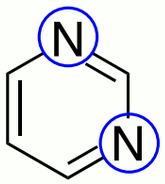
*not aromatic* because it has 8 $\pi e^-$.

# B Aromatic Characteristics Of Protonated Heterocycles

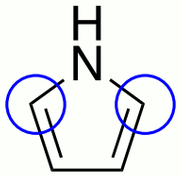
In the last examples, the *N*-atom is, in fact, *not* where indole protonates, *it protonates preferentially on C3*.



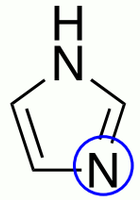

*pyridine*



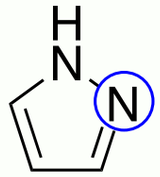

*pyrimidine*



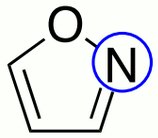

*pyrrole*




*imidazole*

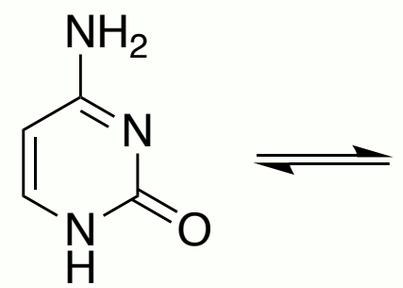



*pyrrazole*

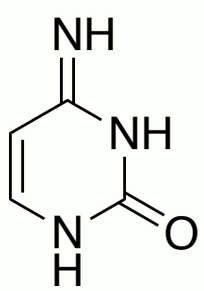



*oxazole*

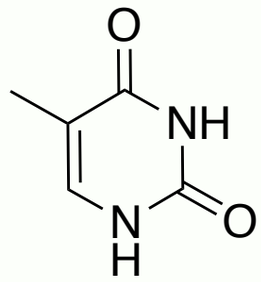
# C A Few Common Heterocycles In Nature



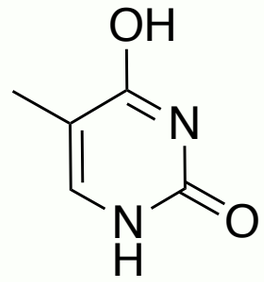
*cytosine*



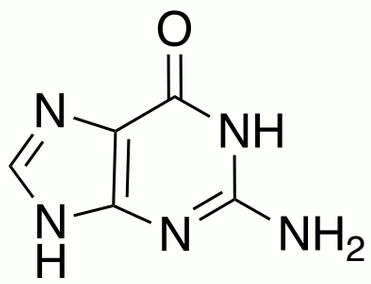
*keto*

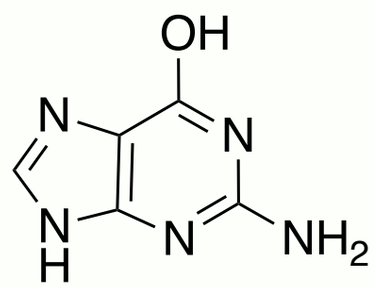


*thymine*



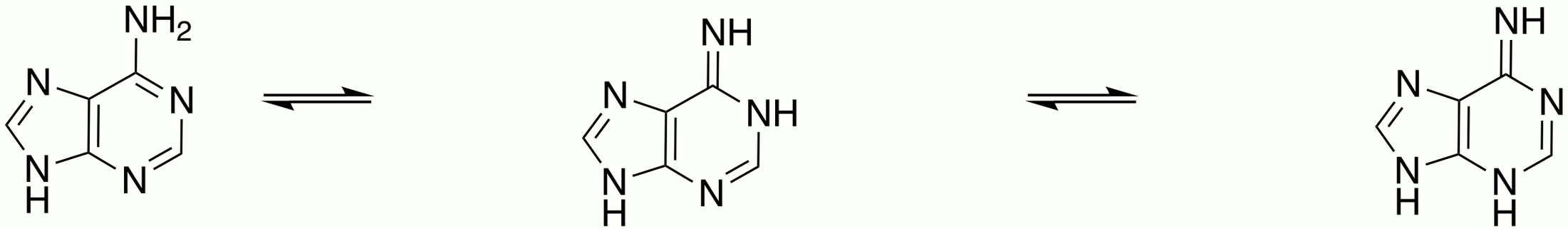
*enol*





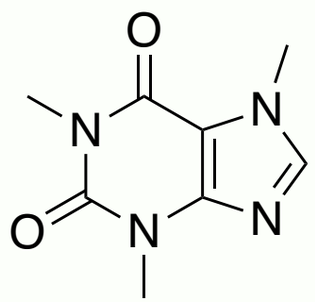
*enol*

# C A Few Common Heterocycles In Nature




*keto* *keto*

Caffeine *cannot* tautomerize.

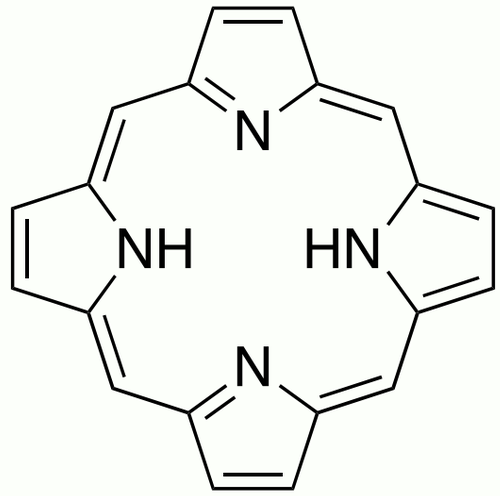



*caffeine*

## C A Few Common Heterocycles In Nature

Porphyrins involve *2* basic pyridine-like nitrogen atoms, *2* non-basic pyrrole-like nitrogen atoms, and *26* π-electrons; they *are* aromatic.

In the presence of base it may be easily deprotonated *2* times to create a ligand ideally suited to complex with transition metals in their M(2+) oxidation states.



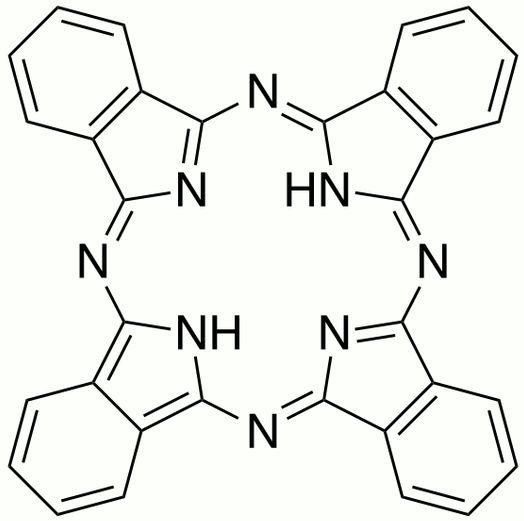

*porphyrin*

## C A Few Common Heterocycles In Nature

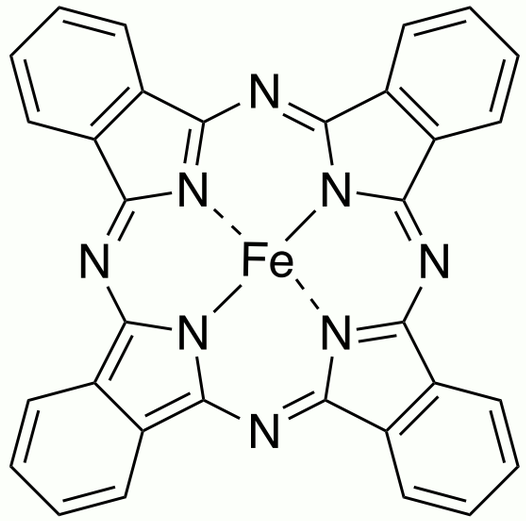
Iron complexes like this are found in *hemoglobin*, while the magnesium complexes feature in *chlorophyll* (circle all that apply).

Complexes like those are *strongly UV absorbing, fluorescent, and capable of redox chemistry*.

Complete the diagram on the right below to show the complex of phthalocyanine with $Fe^{2+}$.

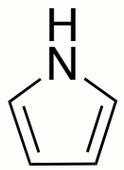



*phthalocyanine*

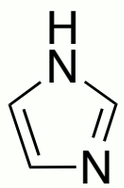



*Fe-phthalocyanine*

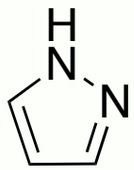
# D Some Other Common Heterocycles



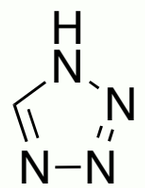
*pyrrole*



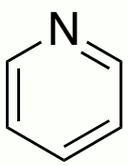
*imidazole*



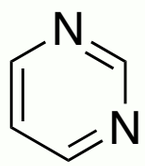
*pyrrazole*



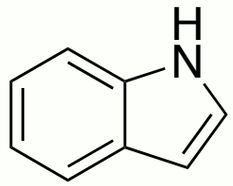
*tetrazole*



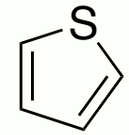
*pyridine*



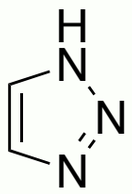
*1,3-pyrimidine*



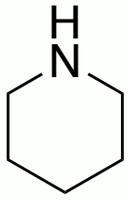
*indole*



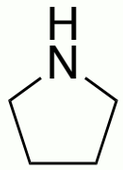
*thiophene*



*1,2,3-triazole*



*piperidine*

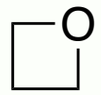


*pyrrolidine*

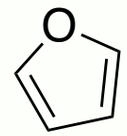
*aziridine*

# D Some Other Common Heterocycles

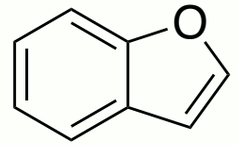

*oxirane*



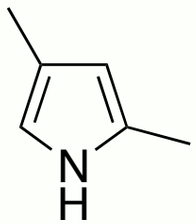

*oxetane*



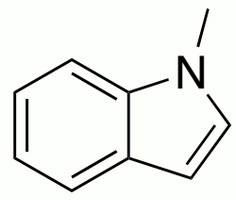

*furan*



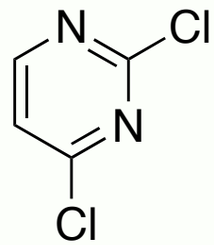

*benzofuran*



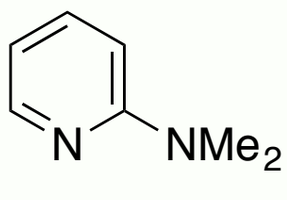

*2,4-dimethylpyrrole*



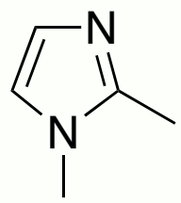

*1-methylindole*



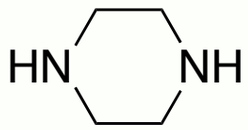

*2,4-dichloropyrimidine*



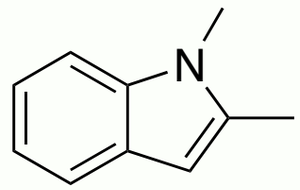

*2-dimethylaminopyridine*



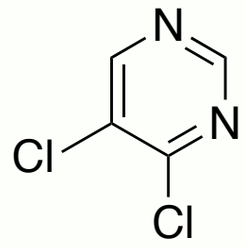

*1,2-dimethylimidazole*




*piperazine*

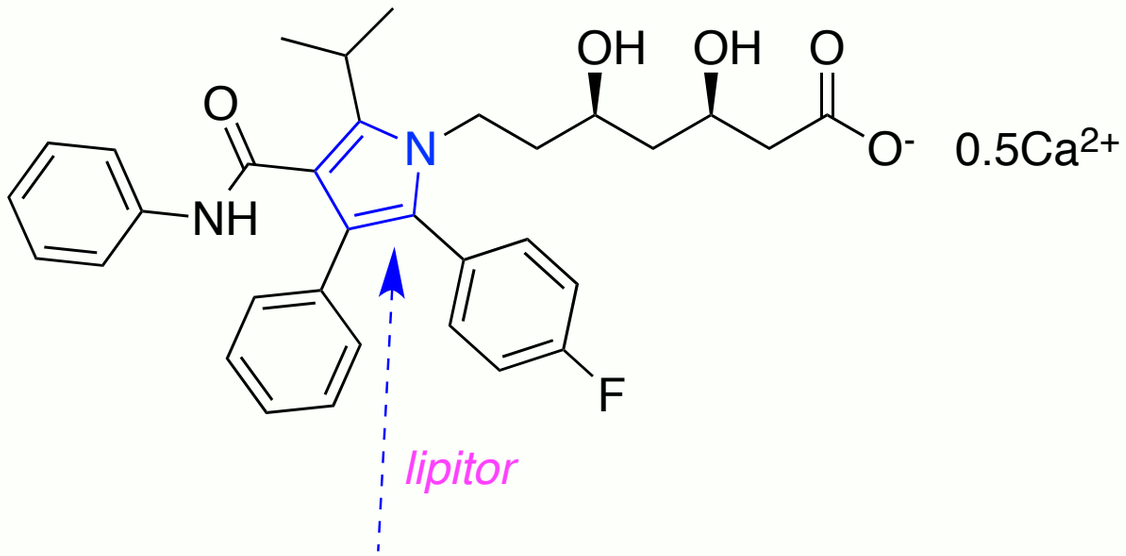



*1,2-dimethylindole*



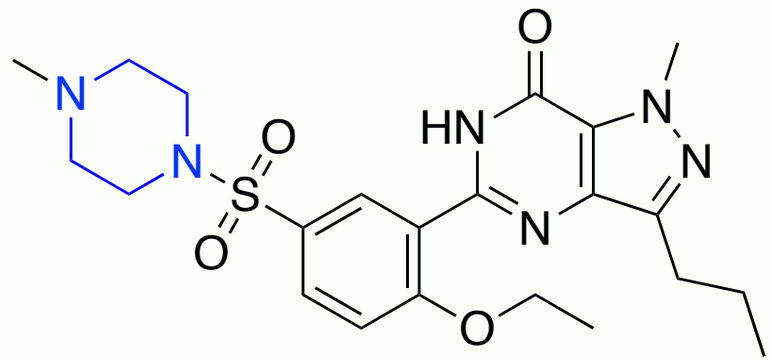

*4,5-dichloropyrimidine*

# D Some Other Common Heterocycles



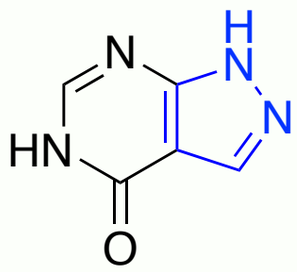

*lipitor*

*pyrrole*



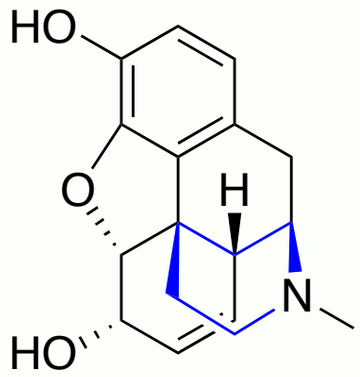

viagra

*piperazine*



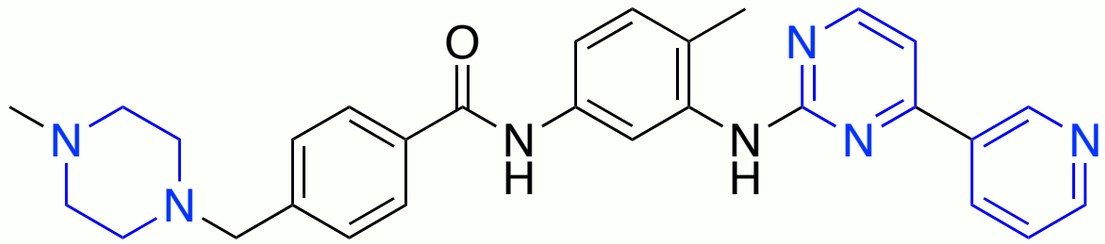

*allopurinol*

*pyrrazole*



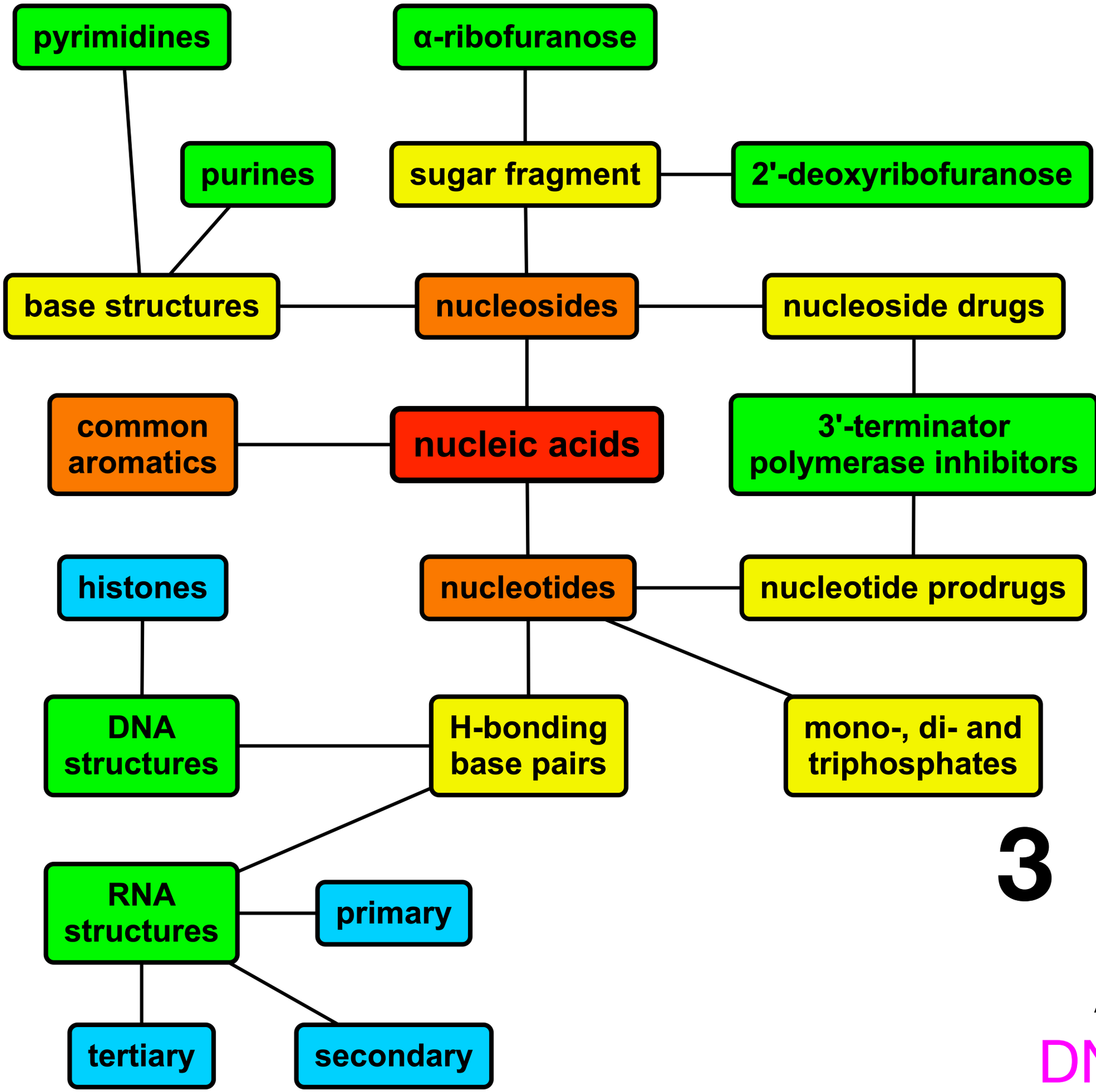

*morphine*

*piperidine*




*imatinib*

*piperazine* and *pyrimidine* and *pyridine*

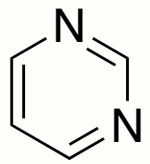



# 3 Nucleic Acids

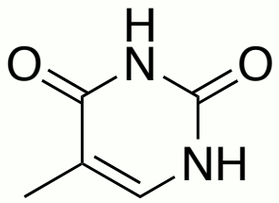
DNA and RNA

# B Nucleoside Structures

Two heterocycles in DNA, and in RNA, are 1,3-pyrimidine structures, and another two are purine derivatives. Memorizing DNA and RNA base structures is easier starting with the parent heterocycles.

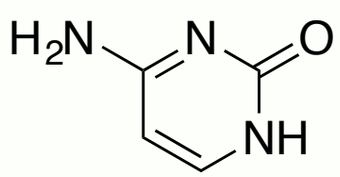



*pyrimidine*



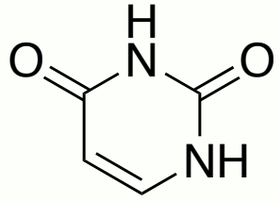

*thymine*

*DNA*



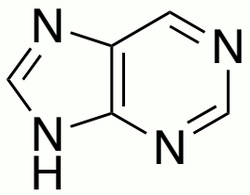

*cytosine*

*both*




*uracil*

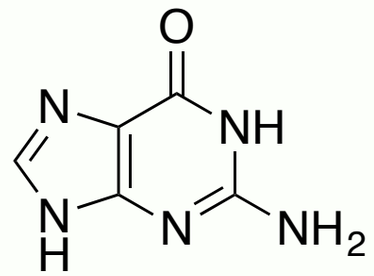
*RNA*




*purine*


*adenine*

*both*




*guanine*

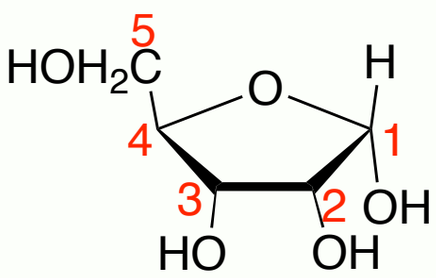
*both*

# B Nucleoside Structures

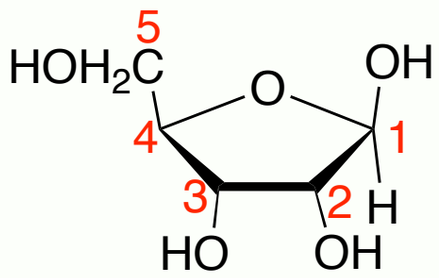
Sugar fragments in RNA are derived from *ribose* in a *furanose* form.

When the heteroatom at the anomeric position is *cis*-oriented to the $CH_2OH$ group in furanopentoses that is a α-anomer.

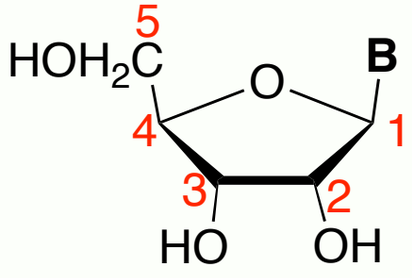
Bases in DNA and RNA are α-anomers.



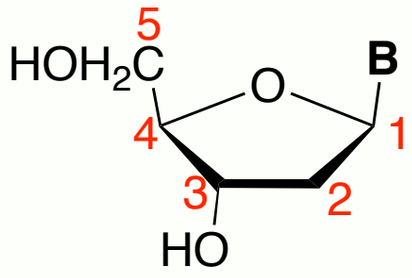

*β-D-ribofuranose*




*α-D-ribofuranose*



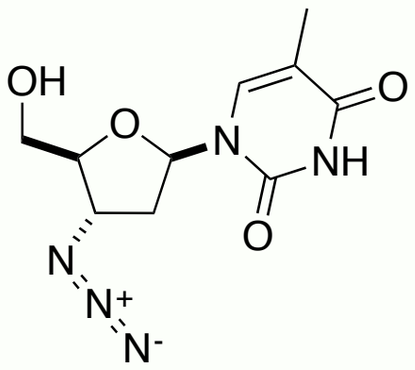

*generic RNA*



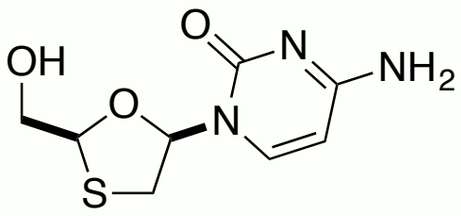

*generic DNA*

# C Nucleoside 3'-Terminator Drugs

Viruses are vulnerable to nucleoside drugs which become incorporated into replicating viral *DNA arresting* its replication. This is like incorporation of dideoxynucleotides (ddNTPs) in sequencing, covered in the next section. All the following drugs are polymerase chain terminators because they *do not* have 3'-hydroxyls which polymerase use to elongate.

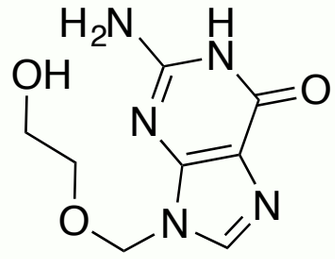


*Zidovudine*

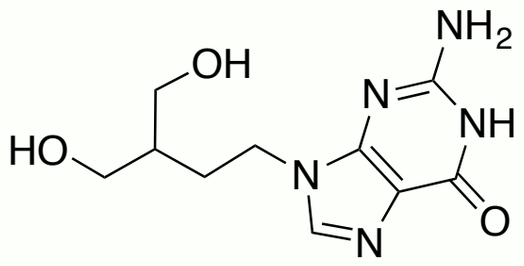


*Lamivudine*

*treatment of HIV*



*Acyclovir*



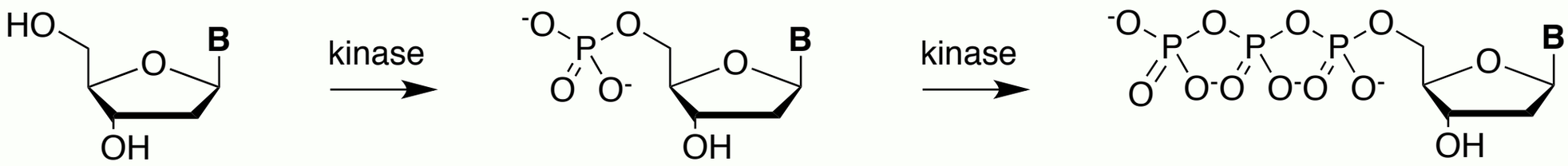
*Penciclovir*

*treatment of herpes*

*Nucleosides* comprise carbohydrate and heterocyclic base components *without* a phosphate.

*Nucleotides* are nucleosides except with *5' O* incorporated into a *phosphate*.

Substrates for DNA replicating enzymes are *nucleotide triphosphates*. The nucleosides shown above enter cells and are converted to triphosphates by *kinase* enzymes as illustrated below.




*cell permeability likely /unlikely*

*cell permeability likely /unlikely*

*cell permeability likely /unlikely*

# C Nucleoside Drugs

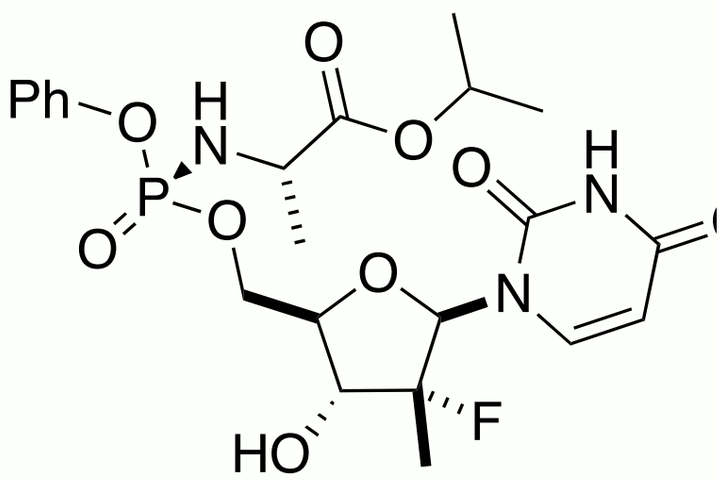
Intracellular formation of nucleotide triphosphates does not occur for all nucleosides. This limits development of new antivirals.

Cell membranes are partially composed of lipids (fats) with phosphate head groups, hence they tend to be impermeable to *negatively* ionized species. Triphosphates cannot permeate because they are *negative*. Even nucleotide monophosphates *do not* permeate easily into cells, so these *do not* make effective anti-viral drugs. The first step, formation of *mono*-phosphates is the difficult one in many cases. However, there is a clever prodrug approach to get monophosphate derivatives into cells to bypass this.

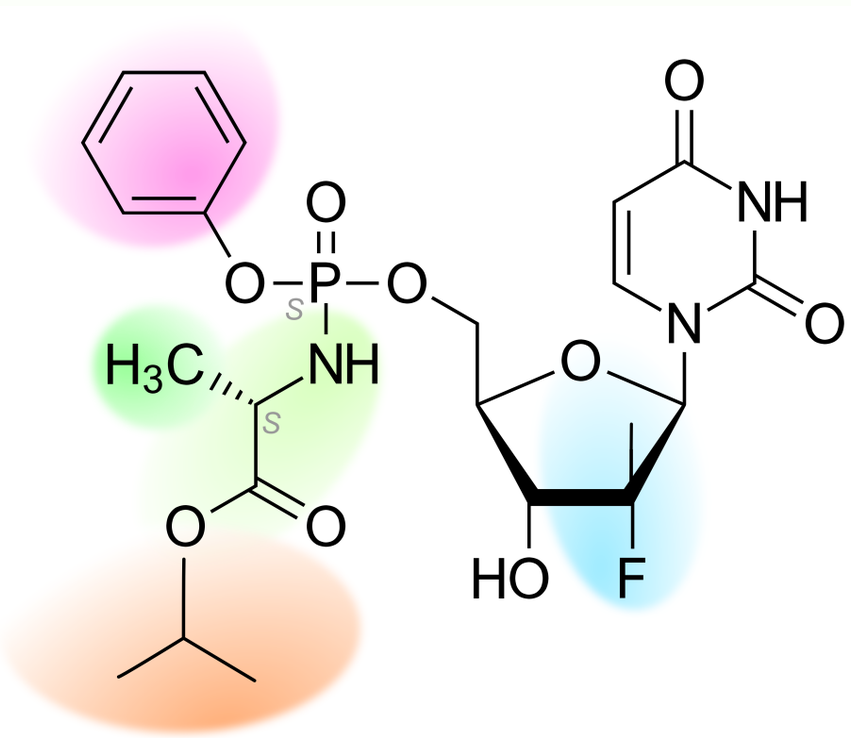
## Nucleotide Prodrugs

A solution to the problem outlined above is to design *neutral*, cell *permeable* nucleoside derivatives which form nucleoside monophosphates *inside* cells.

Nucleotide *prodrugs* are introduced protected, cell permeable, forms, then converted to the active (but cell impermeable) ones once inside the cell. One such drug is Sofosbuvir, illustrative of a class called *ProTides* (on Wiki). They use intracellular esterases (enzymes which hydrolyze esters) to remove a hydrophobic group, facilitating a chemical hydrolysis, then another one mediated by a *P-amidase* (an enzyme which hydrolyzes *N*-bonds to phosphorus).

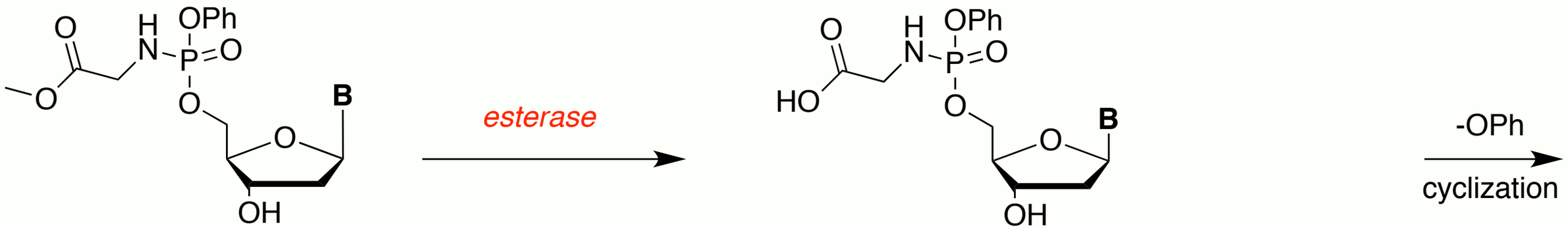



*Sofosbuvir*
*treatment of hepatitis C*

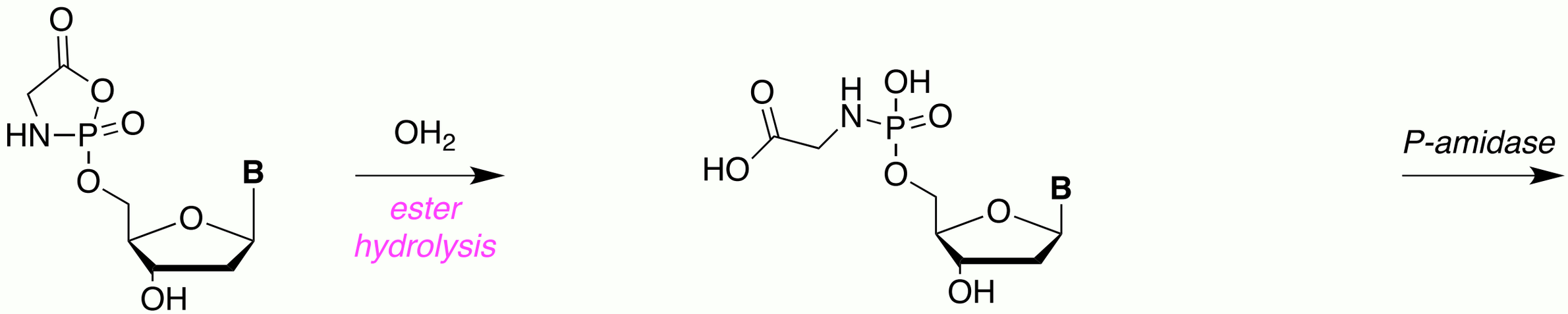


# C Nucleoside Drugs

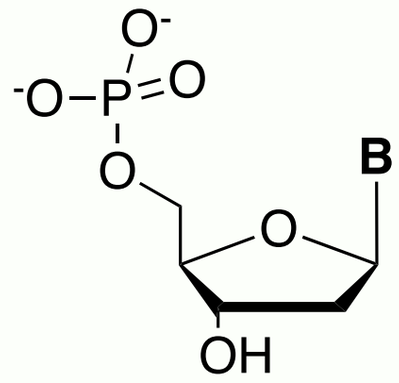
Complete the following to rationalize how drugs like Sofosbuvir are converted to their active monophosphate forms inside cells.



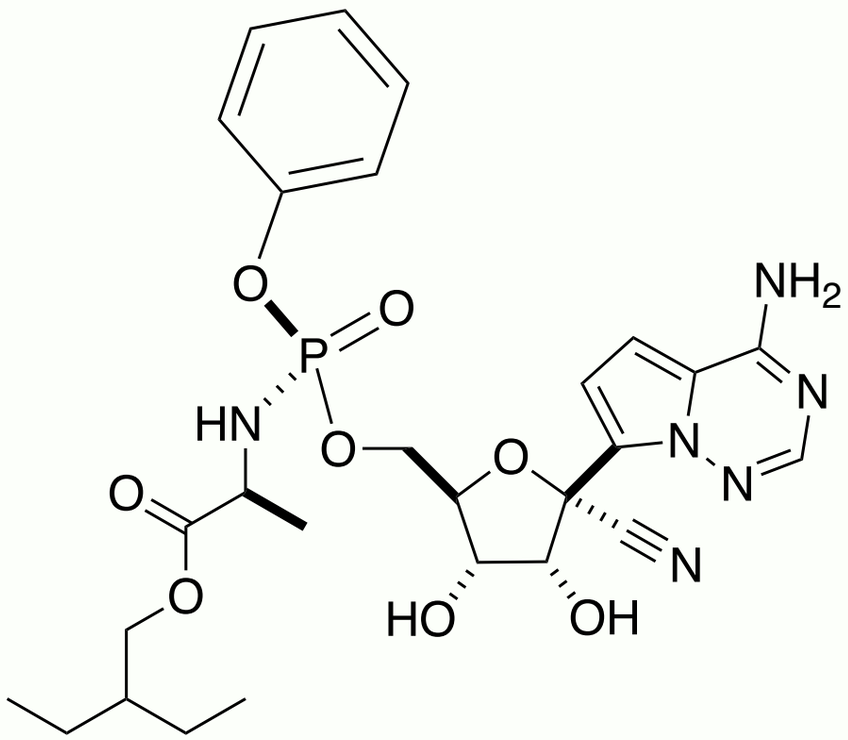

*cell permeable*



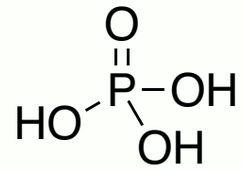
# C Nucleoside Drugs



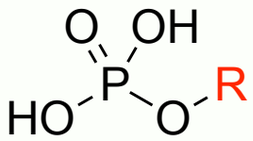

Remdesivir was one of the first pharmaceuticals used to treat COVID-19, caused by the SARS-CoV-2 virus.  It *is* a prodrug, because it slips into the cell like a Trojan Horse carrying the viruses' enemy, an antiviral nucleotide.



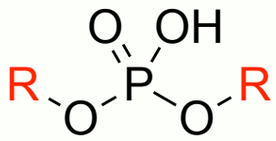
# D Phosphate Derivatives Of Nucleosides: Nucleotides



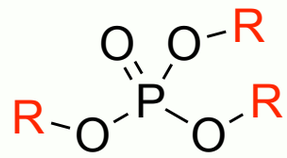
*phosphoric acid*



*phosphomonoester*



*phosphodiester*



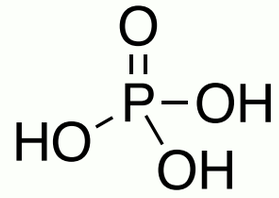
*phosphotriester*

Addition of a water molecule to an alkyl phosphate gives one molecule of an alcohol, and one phosphate.

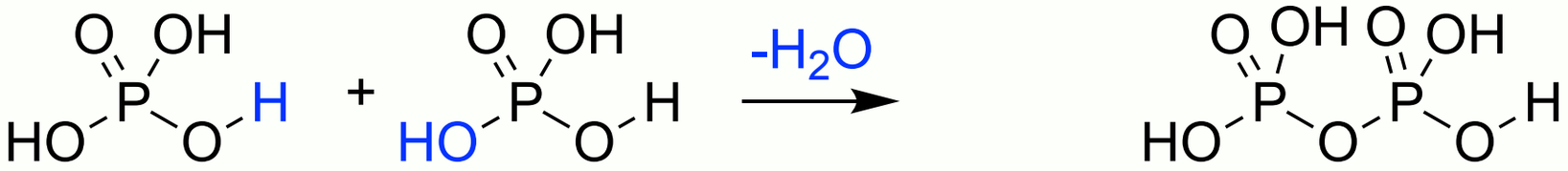
Addition of two molecules of water to an alkyl diphosphate gives *2* alcohols and *1* phosphates.

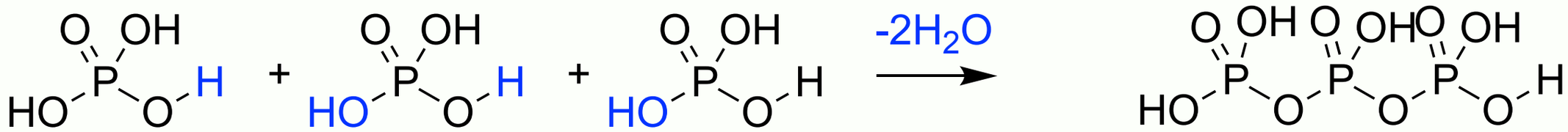
Addition of three molecules of water to an alkyl triphosphate gives *3* alcohols and *1* phosphate.

# D Phosphate Derivatives Of Nucleosides: Nucleotides



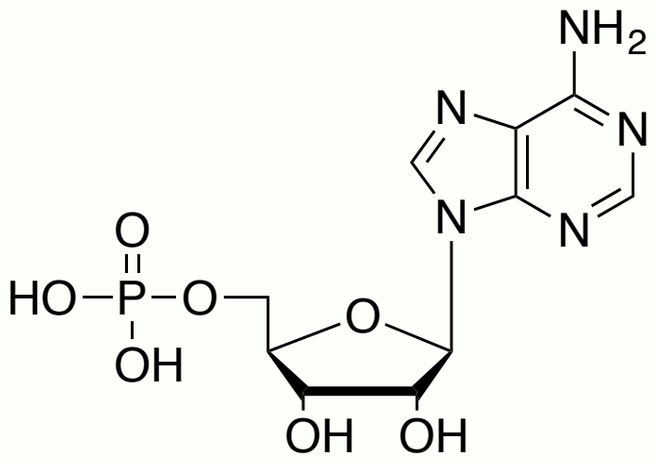
*phosphoric acid*



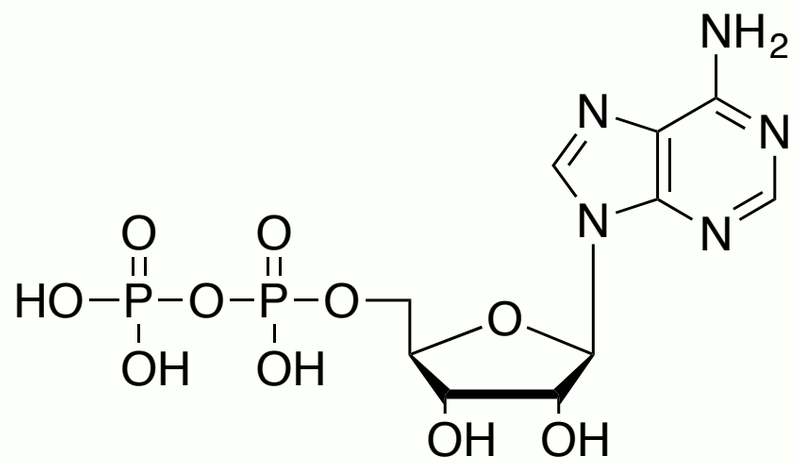



# D Phosphate Derivatives Of Nucleosides: Nucleotides

Mono-, di-, and tri-phosphates are abbreviated to “MP”, “DP”, “TP” suffixes. Phosphates referred to by these abbreviations are on the 5’-*O* unless something specifically indicates otherwise.

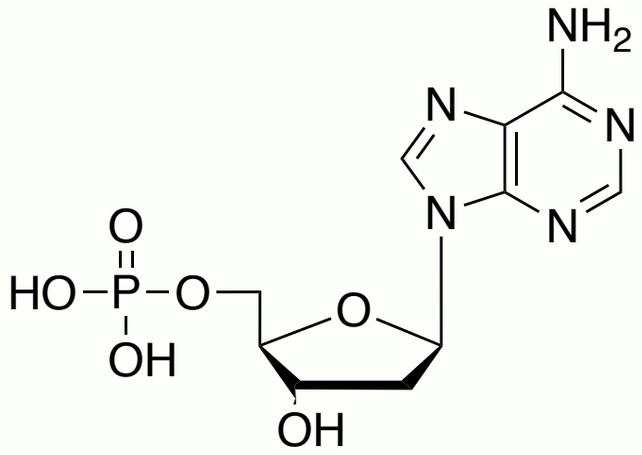



*AMP*

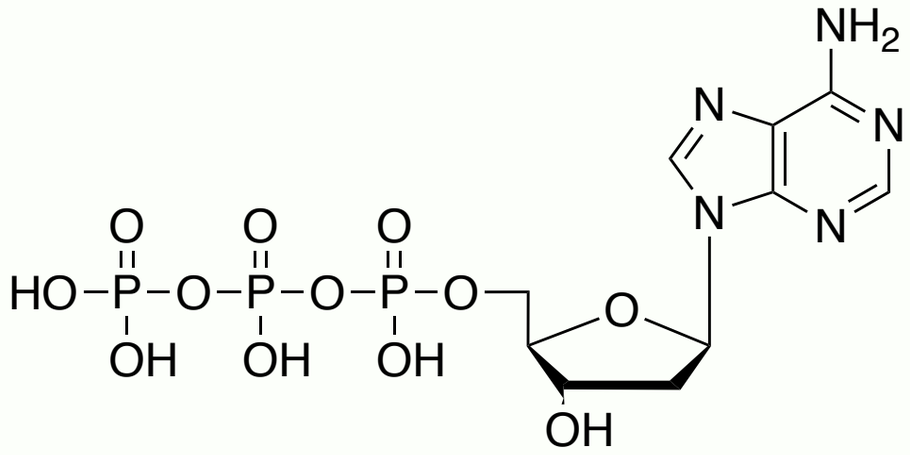



*ADP*

# D Phosphate Derivatives Of Nucleosides: Nucleotides

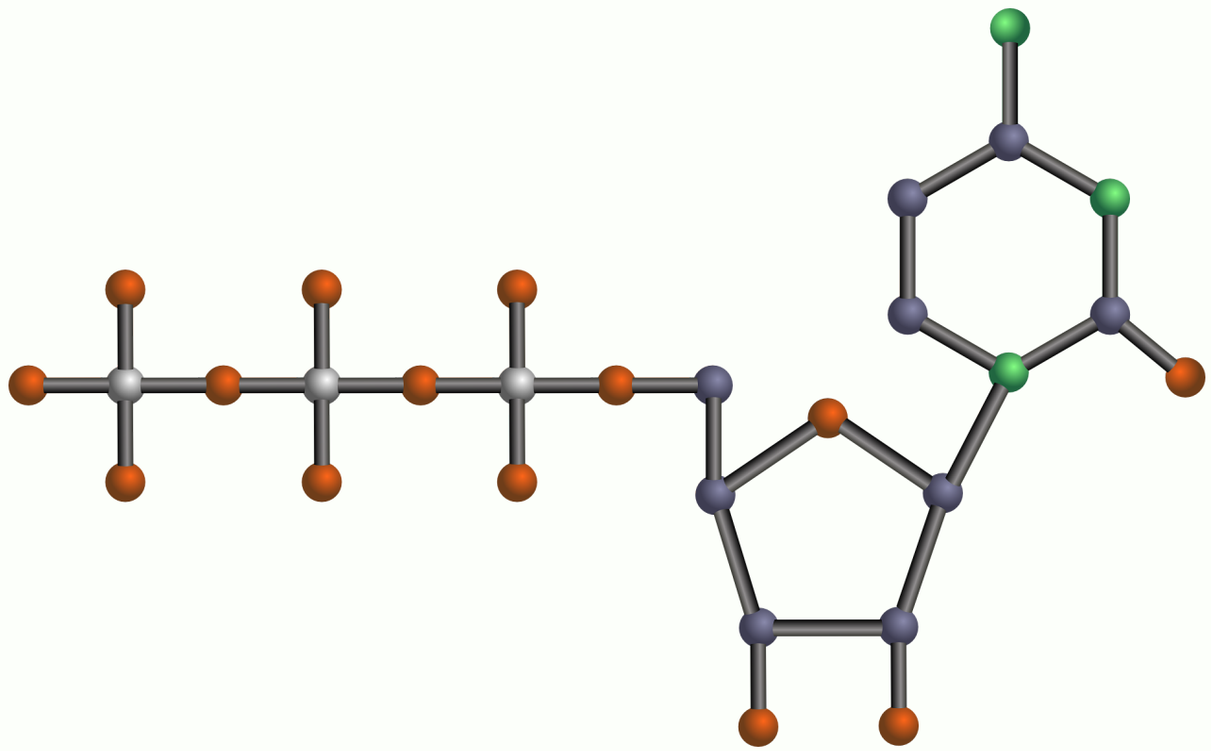



*dAMP*

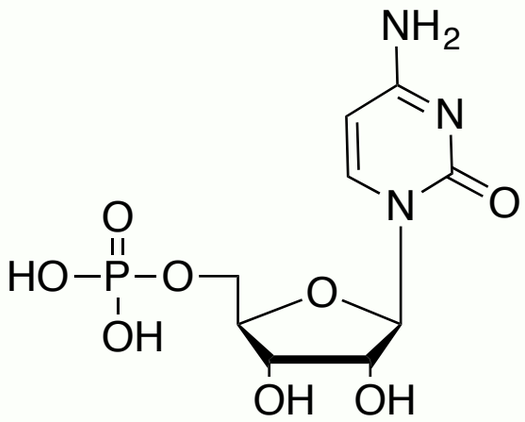



*dATP*

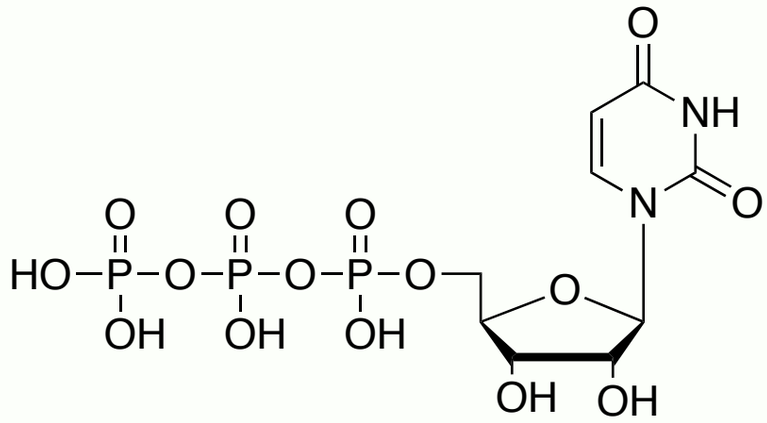
An abbreviation for the compound represented on the right is *TTP or dTTP* (same thing).



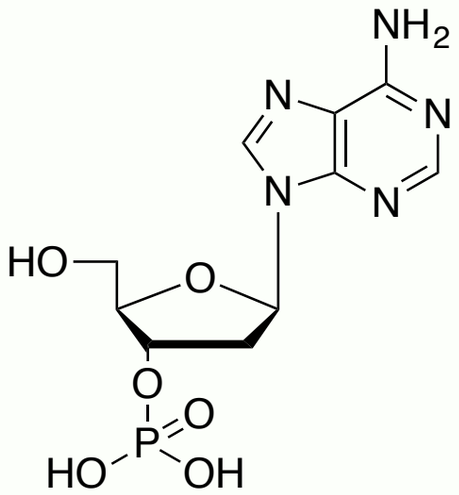
# D Phosphate Derivatives Of Nucleosides: Nucleotides



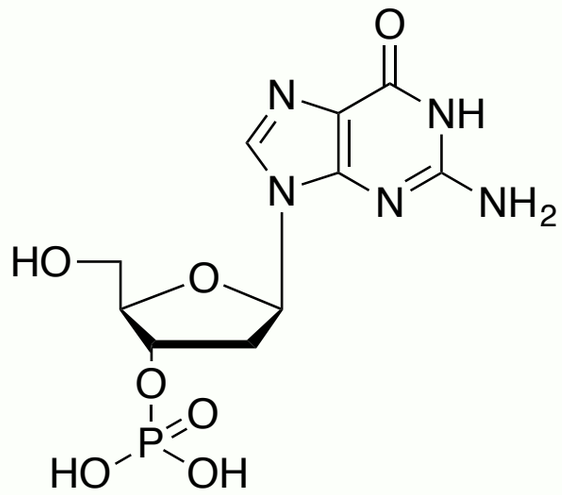
*CMP*



*UTP*

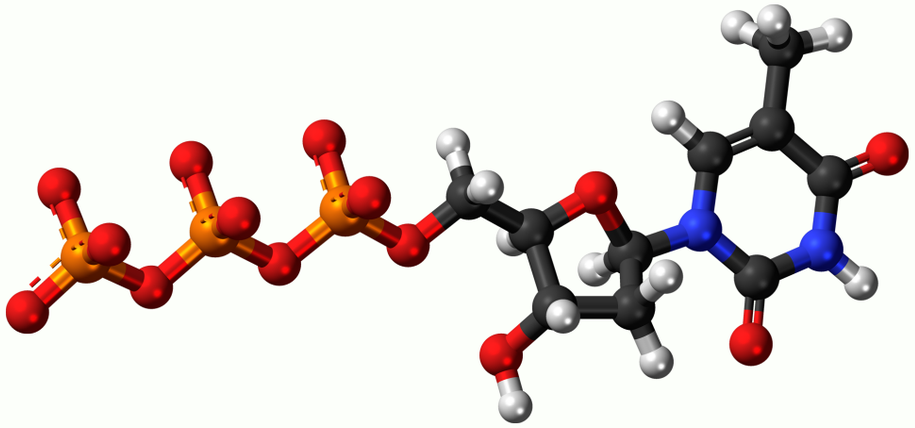


*2-deoxyadenosine 3'-monophosphate*

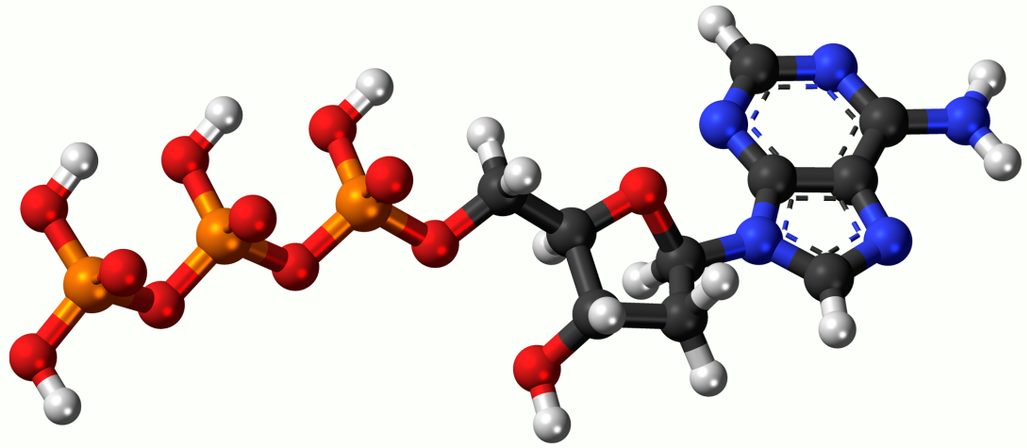


*2-deoxyguanidine 3'-monophosphate*

# E Hydrogen Bonding In Base Pairs

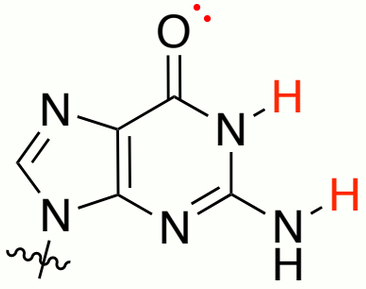


*TTP or dTTP*



*dATP*

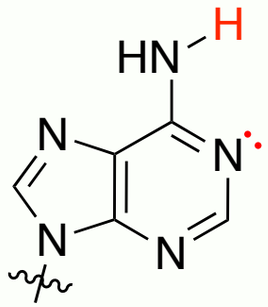
Complete the following (numbers of *H*-bond donors and acceptors refers only to the parts highlighted in red).



*G*

*# H-bond donors:*
*2*

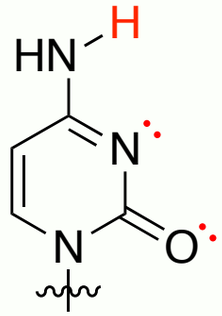
*# H-bond acc's:*
*1*



*A*

*# H-bond donors:*
*1*

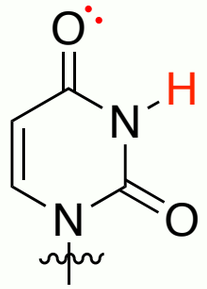
*# H-bond acc's:*
*1*



*C*

*# H-bond donors:*
*1*

*# H-bond acc's:*
*2*



*U*

*# H-bond donors:*
*1*

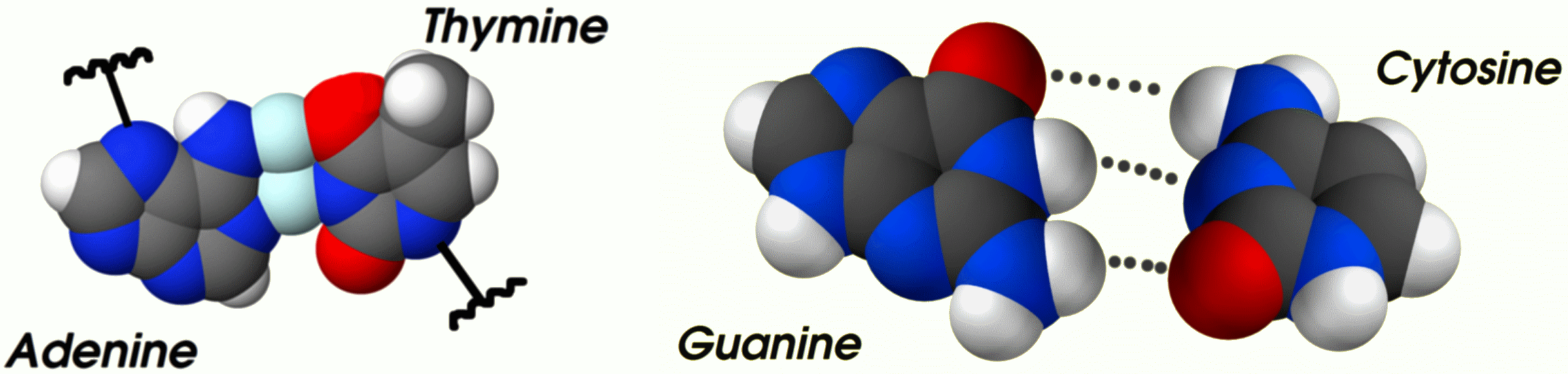
*# H-bond acc's:*
*1*

*refers to H-bond acceptors and donors, as indicated in structure*

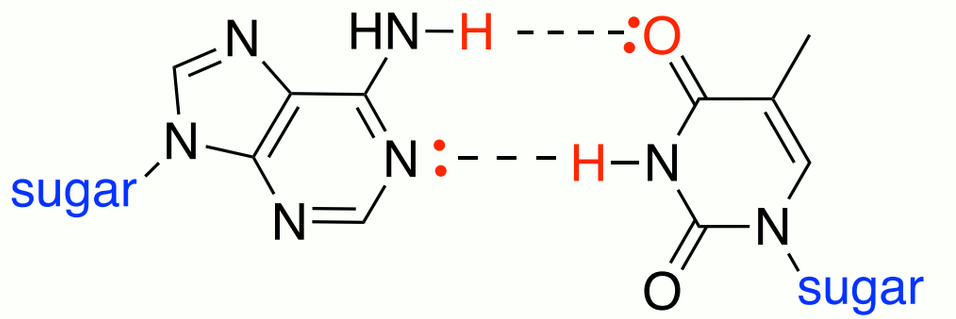
## E Hydrogen Bonding In Base Pairs

Consider only the *H*-atoms and lone pairs shown above in red. Bases with two correctly oriented *H*-bond donors and one *H*-bond acceptor should pair with others with two correctly oriented *H*-bond acceptors and donors.

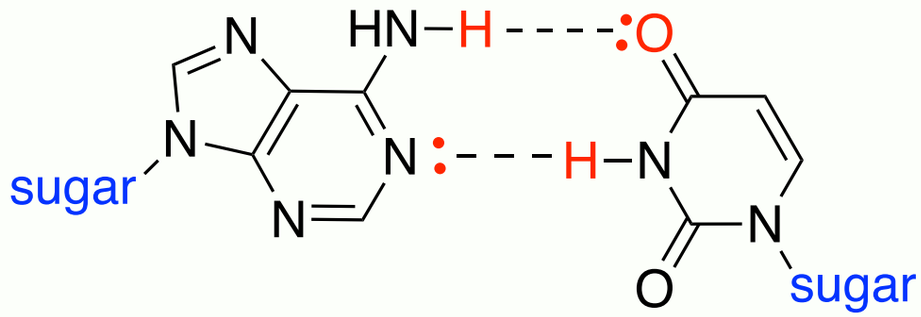
**G** with ***C*** in DNA
**G** with ***C*** in RNA
**A** with ***T*** in DNA
**A** with ***U*** in RNA
**C** with ***G*** in DNA
**C** with ***G*** in RNA
**T** with ***A*** in DNA



# E Hydrogen Bonding In Base Pairs

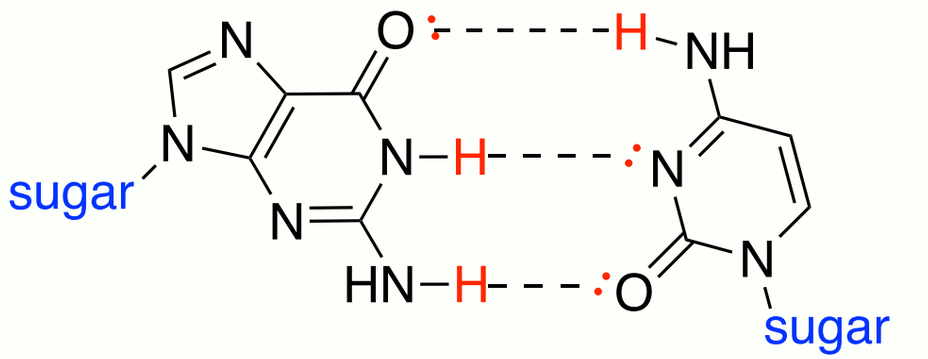



***A** H-bonded to **T***

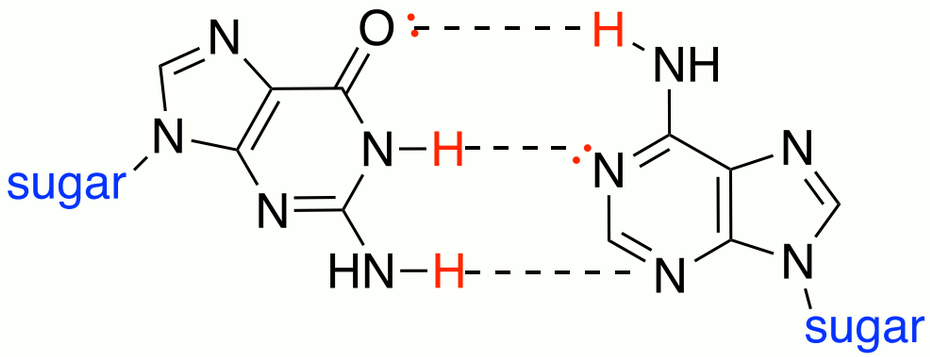



***A** H-bonded to **U***

# E Hydrogen Bonding In Base Pairs




*G H-bonded to C*



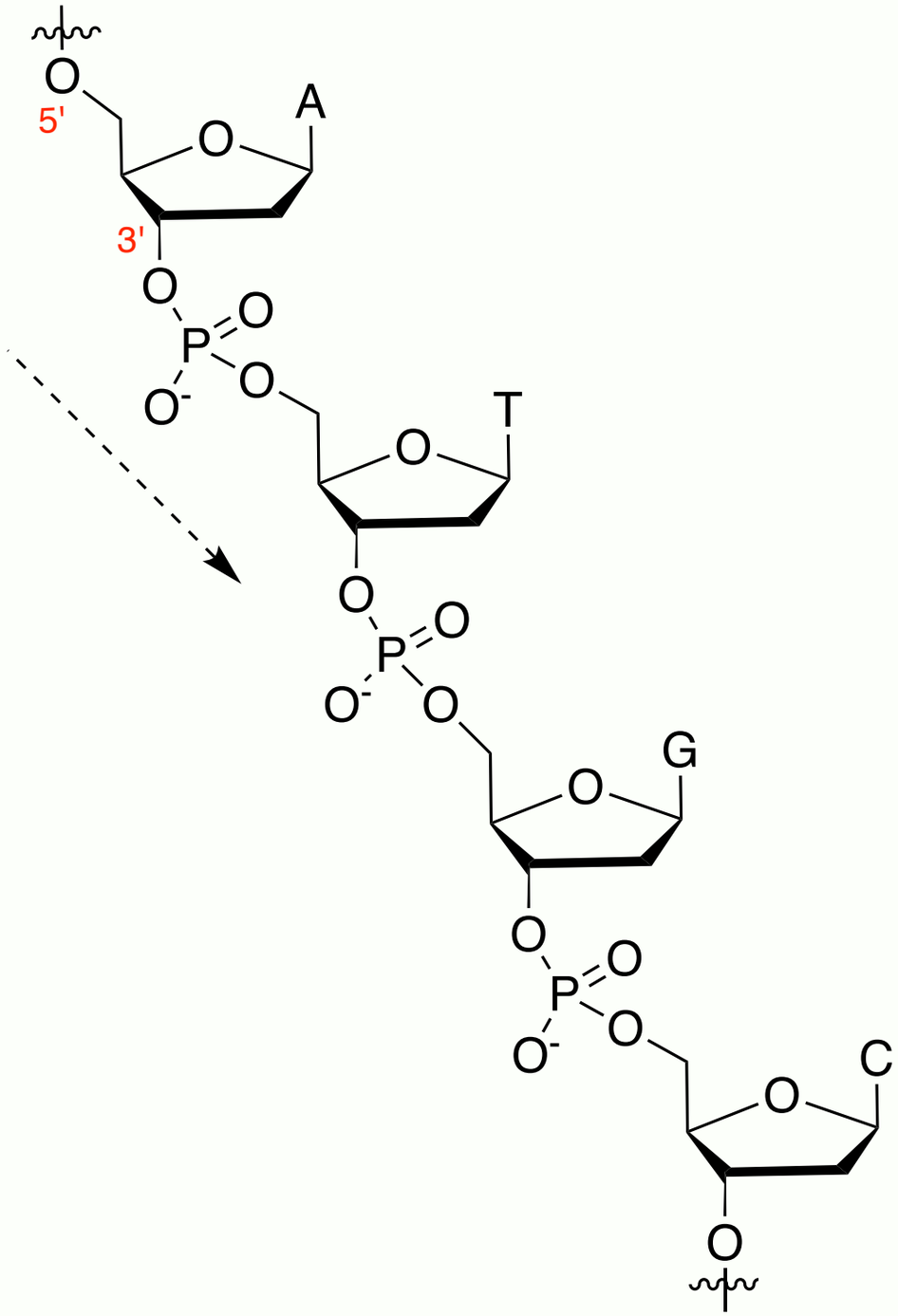

*G H-bonded to A*
*does not match well*

H-bonding between A and T is *weaker* than between G and C because there are *less* inter-residue hydrogen bonds.

*Deamination* in living organisms converts **C** into ***U***.

# F DNA Structure

**A**, **T**, **U**, **C**, and **G** symbols can be used to represent just the base in nucleosides and nucleotides, or whole nucleobase units.  Units of DNA are connected by phospho*di*esters.




RNA bases are connected by phospho*di*esters.

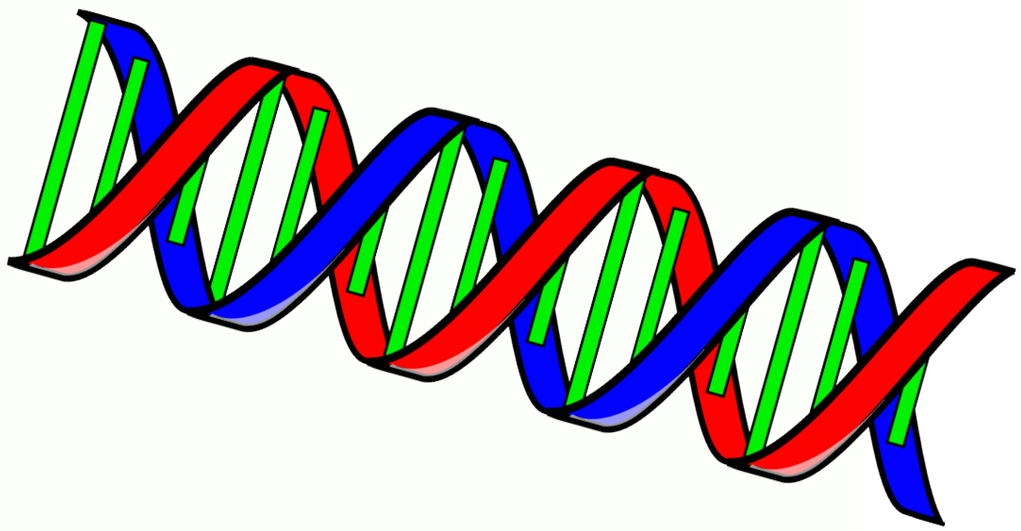
**CGTA** has the *a different* structure to **ATGC**.

When reading DNA sequences, start at the *5'* end and list towards the *3'*.

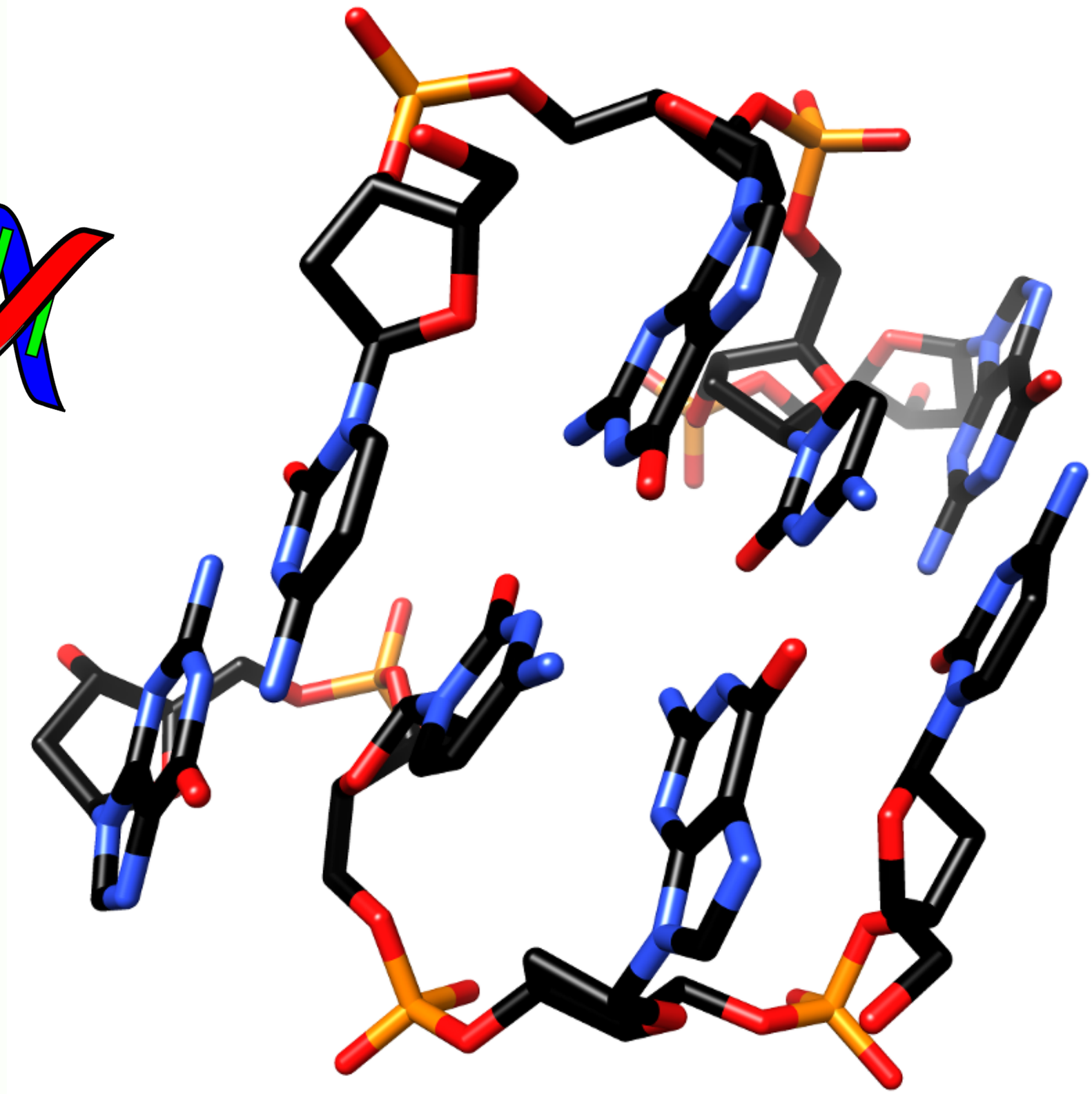
Primes in these numbers mean they refer to *the sugar part*.

Phosphorus atoms in RNA and DNA *are not* chiral in their dibasic state.

# F DNA Structure

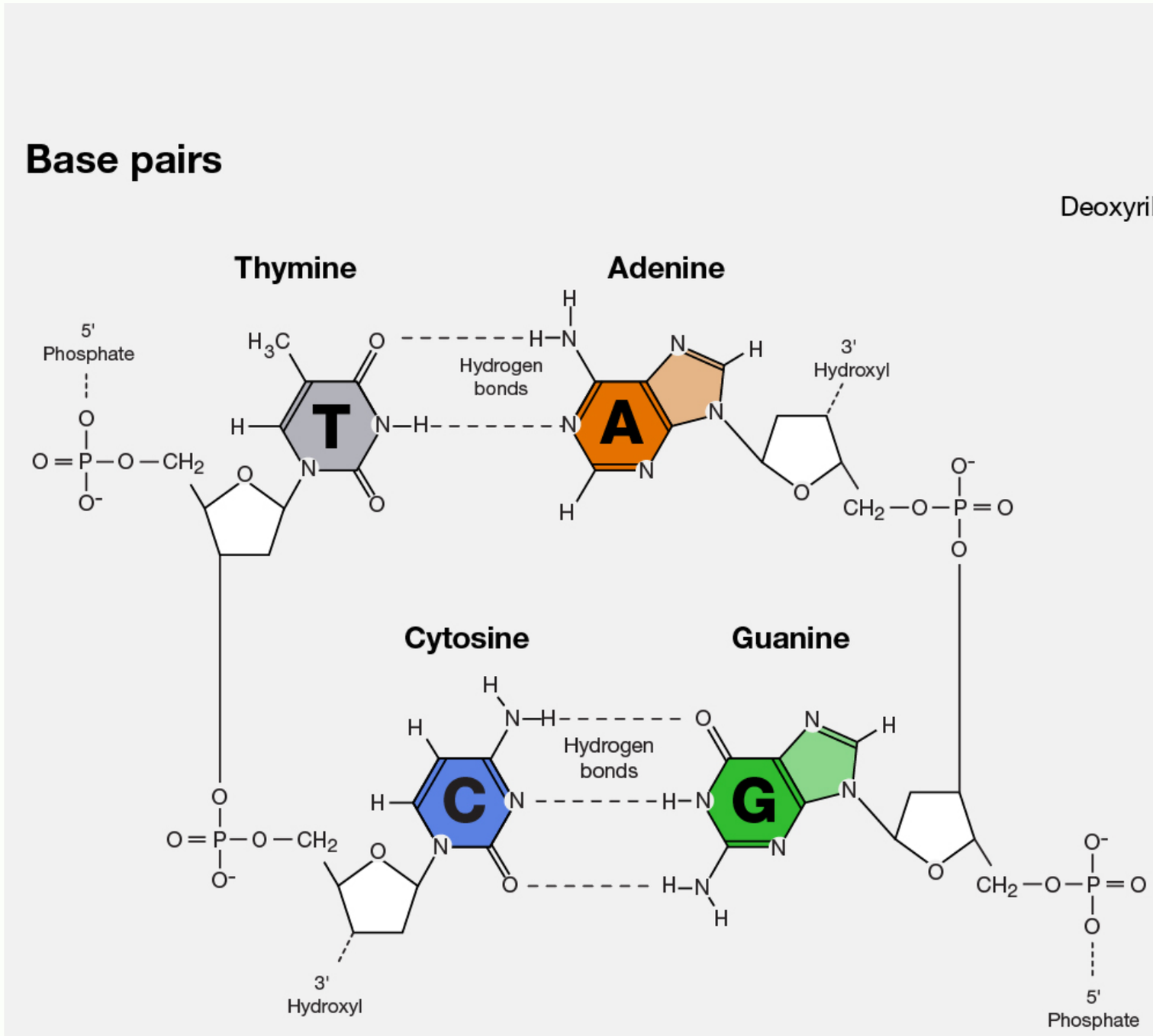


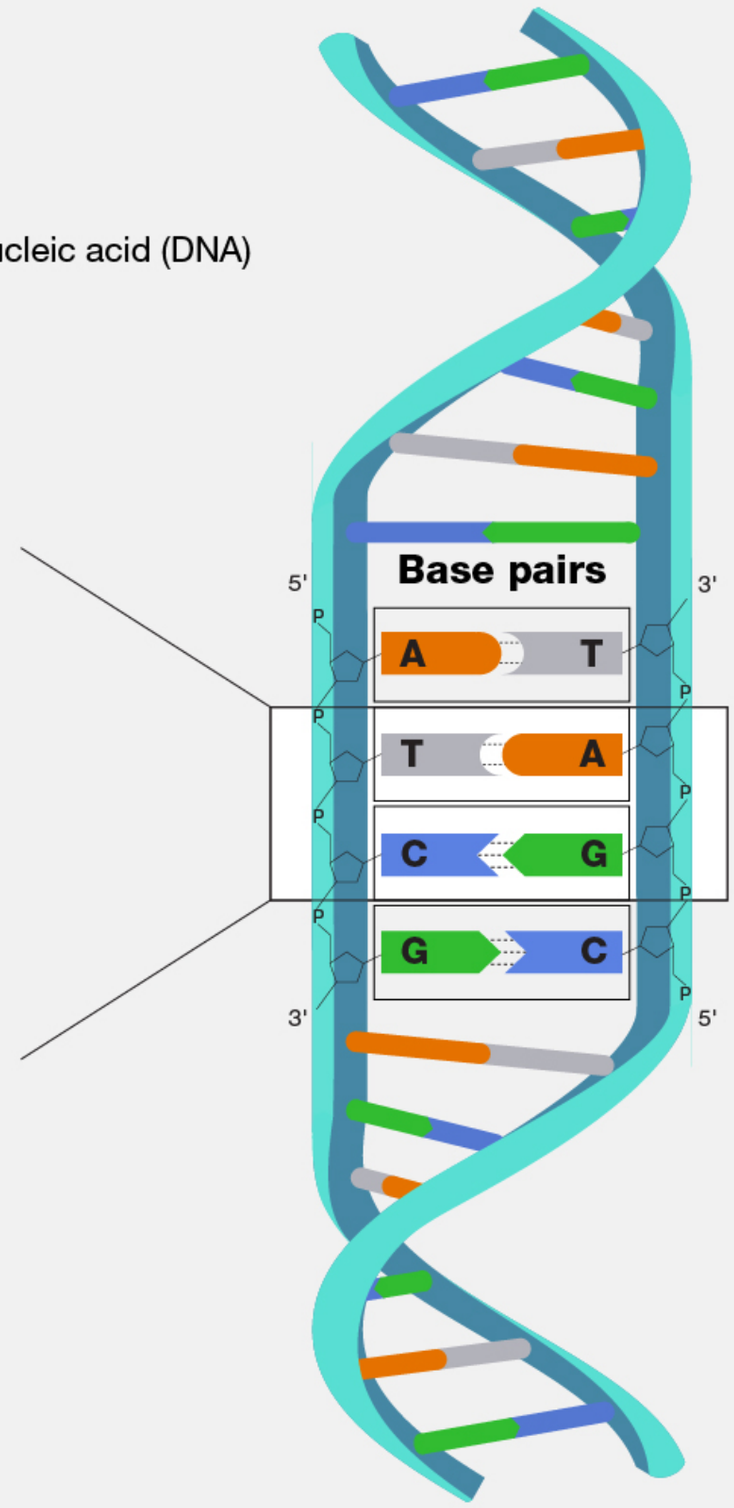
Draw the *H*-bonds between the DNA bases (diagram on right) , and identify each base by writing A, T, G, or C beside them.



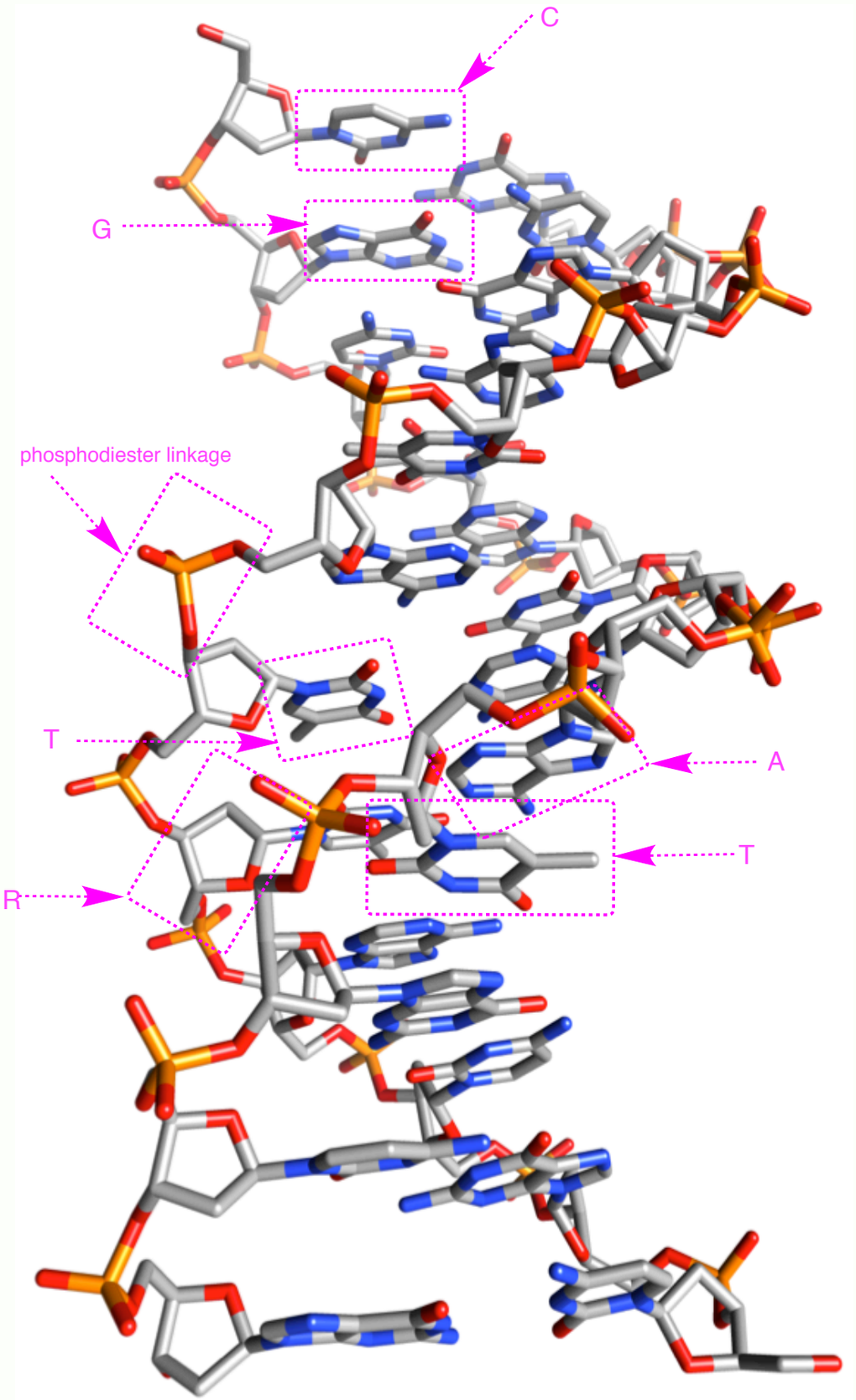
# F DNA Structure

*Phosphodiesters* are on the perimeter of double stranded DNA, and *nucleobases* form the core.



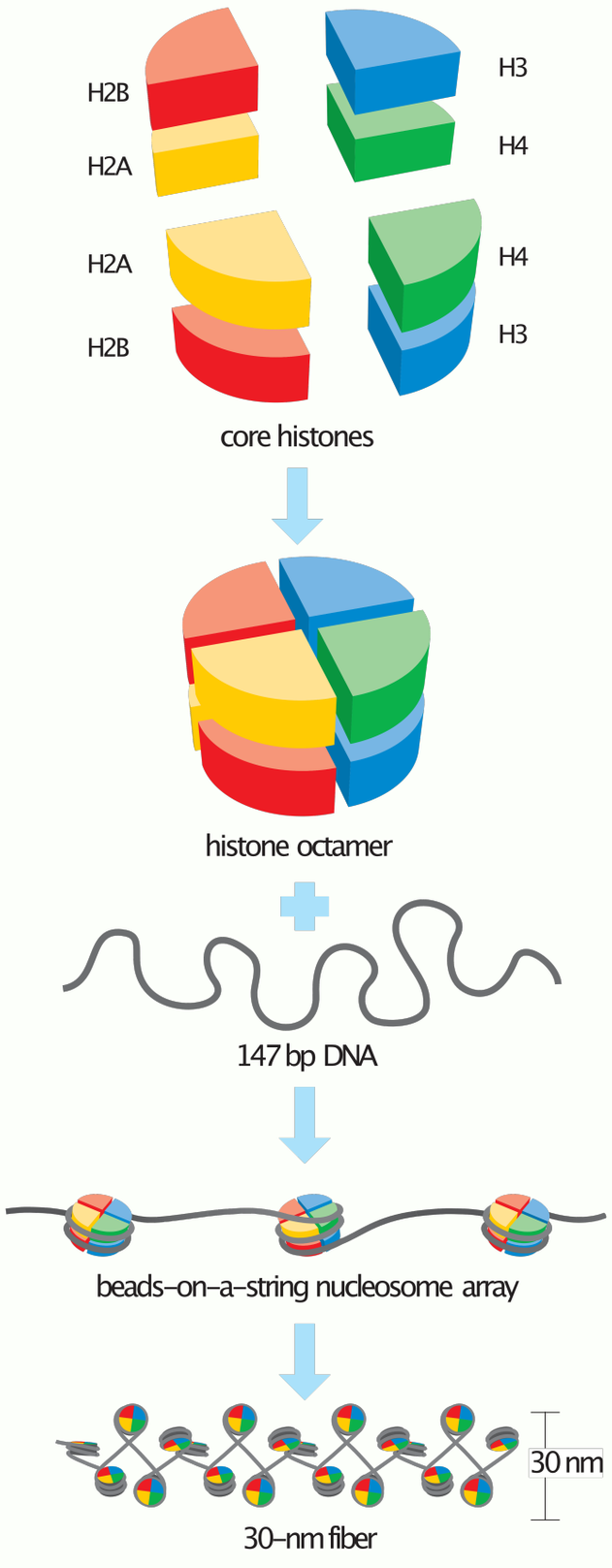



# F DNA Structure




This diagram represents double stranded DNA. A phosphodiester linkage, various **T**, **A**, **G**, **C** nucleobases and a ribose fragment are indicated (perhaps more than once).

# F DNA Structure



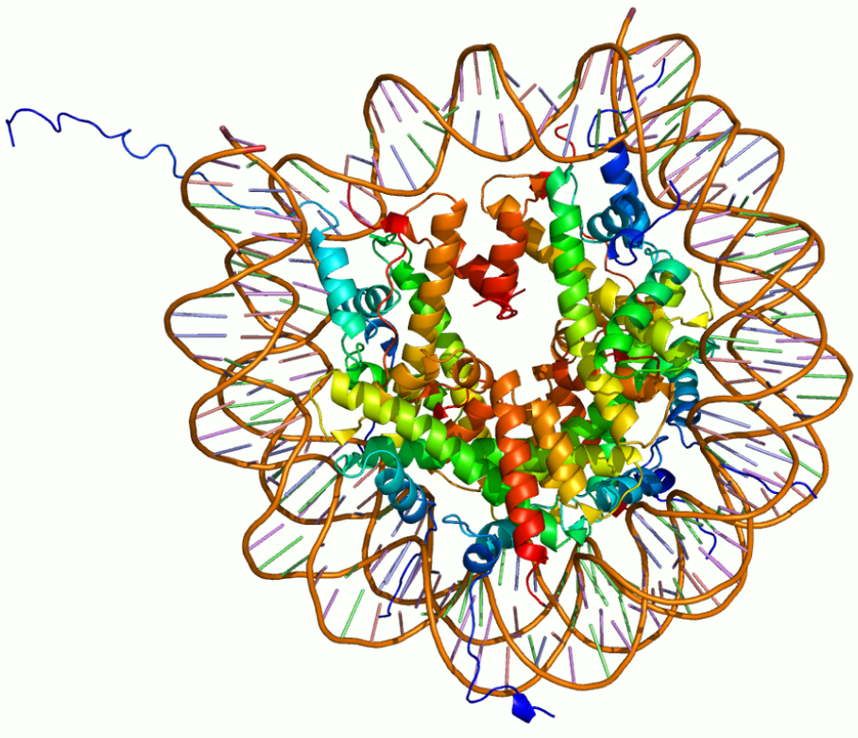

DNA tends to exist in a *right* handed helical arrangements, *the same as* protein helices.

Bases of *DNA* can be distinguished by the prefix “d” as in **dA**, **dT**, **dC**, and **dG**, relative to **A**, **U**, **C**, and **G** used for *RNA*.

DNA synthesis is mediated by enzymes called *polymerases*. These build *antiparallel* strands of DNA to complement the first by adding the complementary triphosphate (dATP, dGTP, dCTP or dTTP for DNA polymerases) to pair with the next base at *5’ / 3’*. A *di*phosphate is lost whenever a base is added by a polymerase.

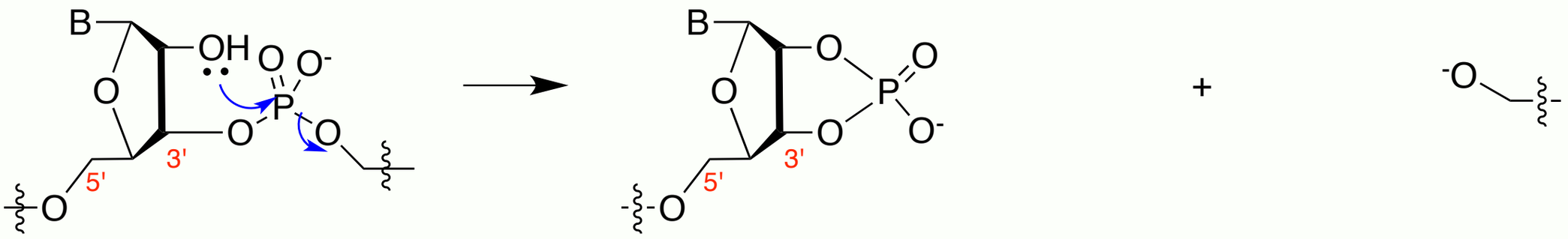
Syntheses of mRNA stops at particular three-base sequences of DNA called stop codons. Exon DNA and RNA encodes for proteins, and regions which do not are called *introns* (but some encode functional RNA).

In cells, DNA is protected in a resting state: cells wind it around proteins called histones (as left), in “bead on string” arrangements which give genes their shape. One “bead” is shown on the right. DNA is unwound when called upon to express proteins, then rewound after use to protect them from trauma in less controlled intracellular environments.



# G RNA Structure

RNA is *less* stable than DNA because it has a 2'-*OH* that can cyclize onto the phosphorus giving a 2',3'-cyclic phosphate, displacing the 5'-*O* of the next residue.




The reaction shown above cannot happen with DNA because it has no *2'-OH*.

In all organisms, excluding some viruses, genetic information is stored as deoxynucleic acid (DNA) sequences. That DNA information is *transcribed* into RNA (ribonucleic acid) when genes are switched on, then *translated* into proteins.

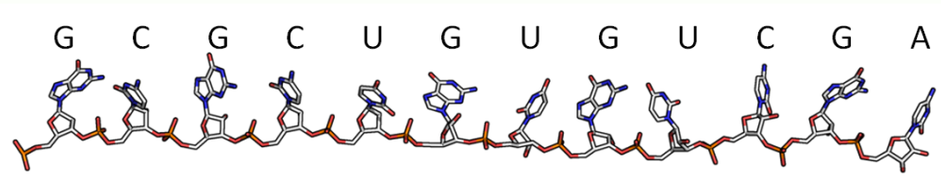
Genes are turned on by interacting with *promoter* proteins bind to *promoting* regions of DNA sequences to initiate synthesis of a complementary strand of *messenger* RNA.

Non-encoding regions in the transcribed mRNA are cut out in the process of RNA *splicing*.

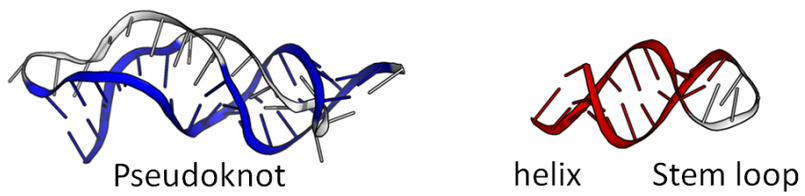
In the ribosome, *transfer* RNA reads three base codons on the *messenger* RNA and converts this sequence information into proteins.

# G RNA Structure

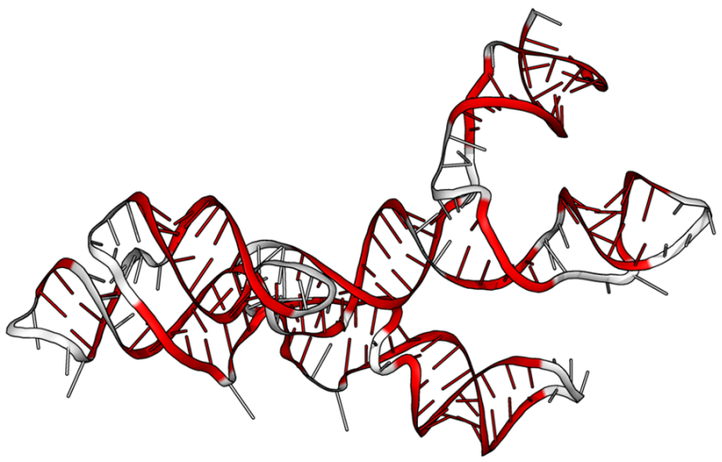
Primary



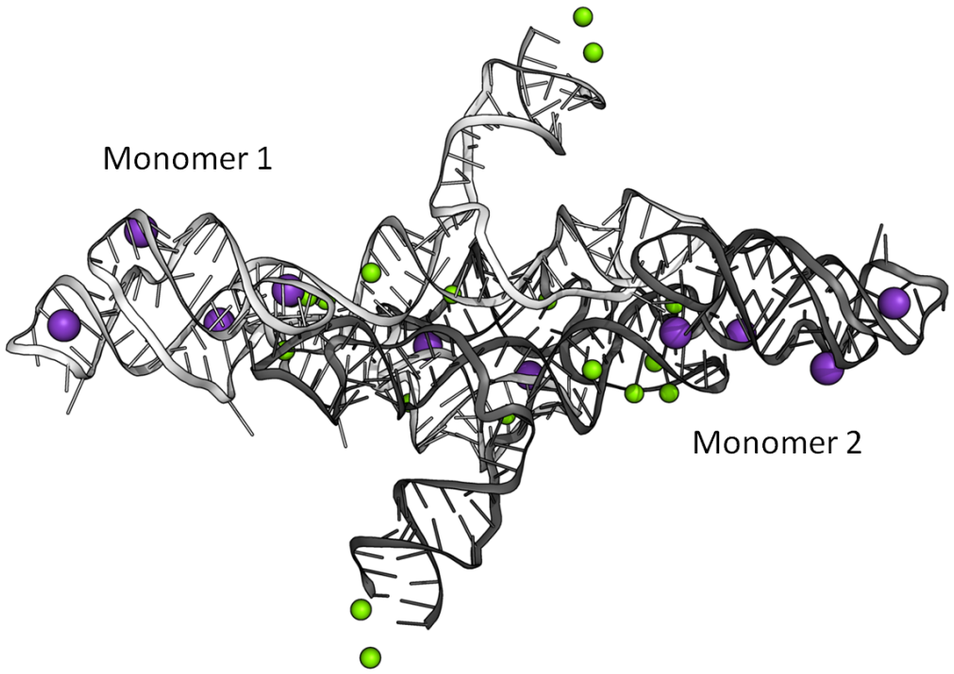

Secondary




Tertiary

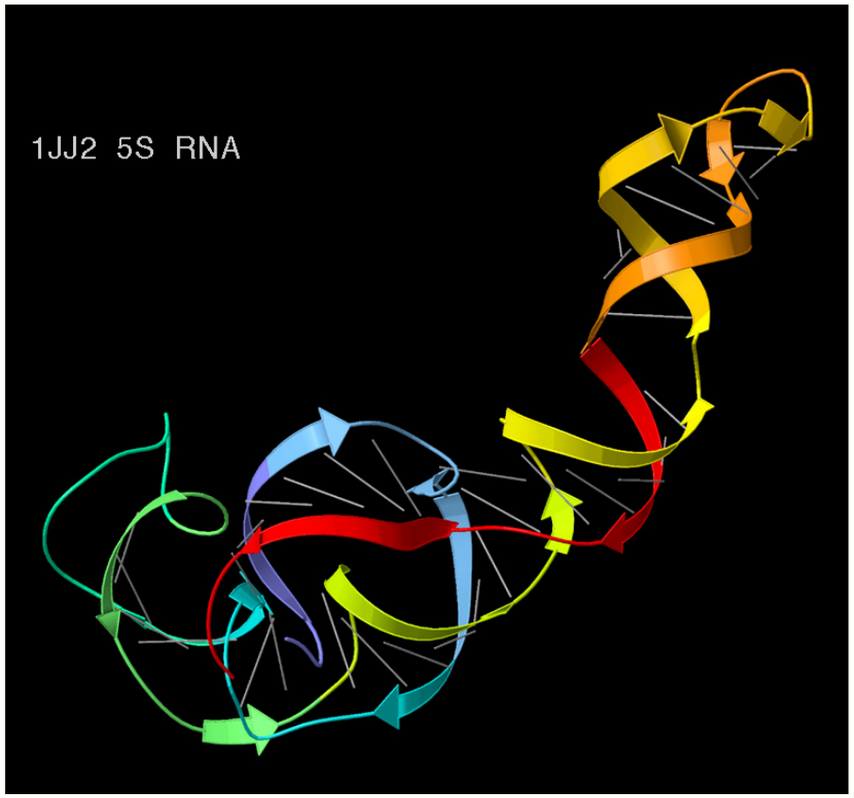


Quaternary



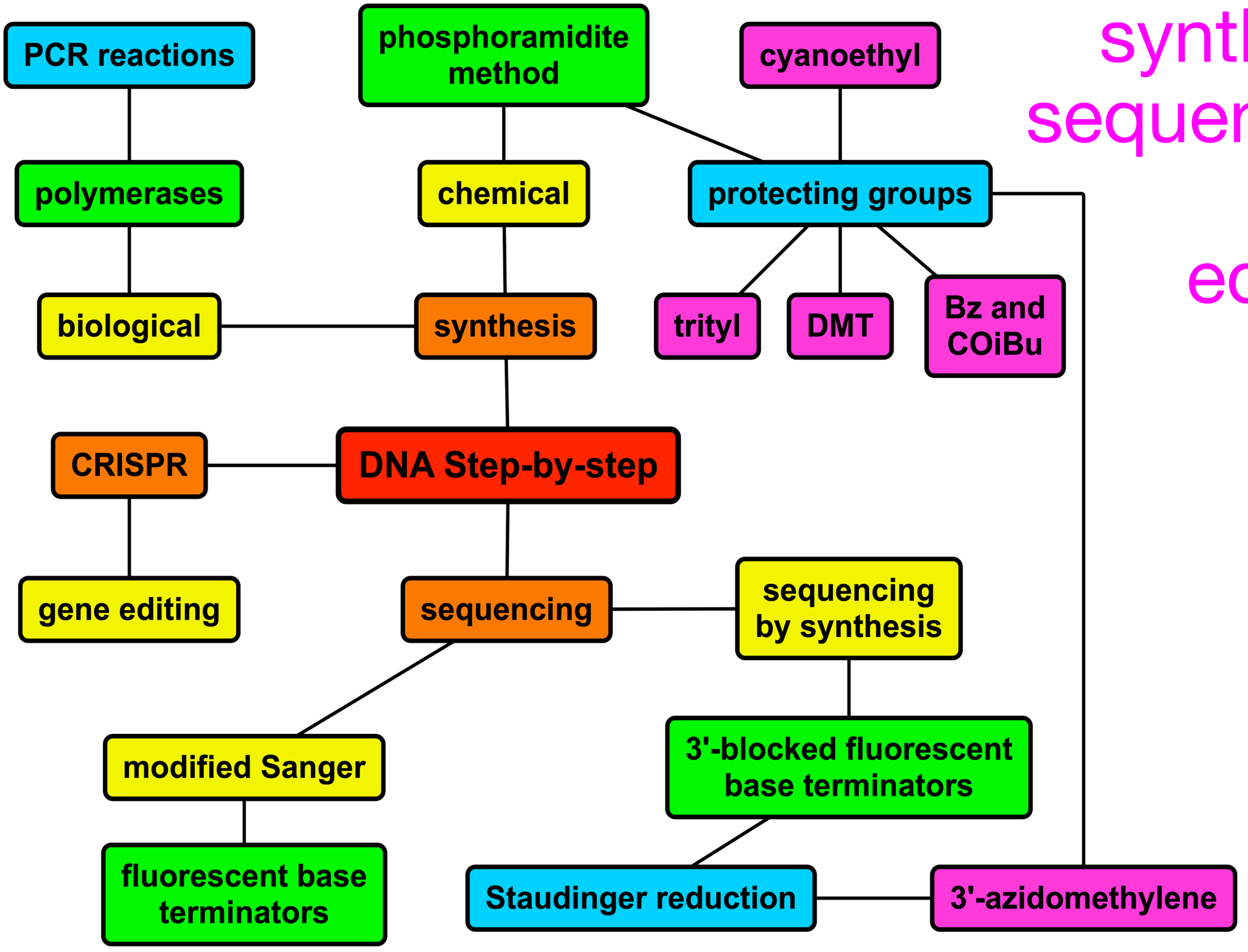

RNA has *less* regularly ordered structures than DNA. These *are* often reinforced by coordination to magnesium. Sequences without 3D structure information are *primary*. Local shapes are *secondary* structures. Whole shapes of an RNA fragments are *tertiary* structures. *Quaternary* structures refer to more than one RNA unit interacting.

Left is an outline for naming RNA structures. Below is a crystal structure of an RNA fragment which *does* contain a helix stem loop.



# 4 DNA Step-by-step

synthesis
sequencing
and
editing

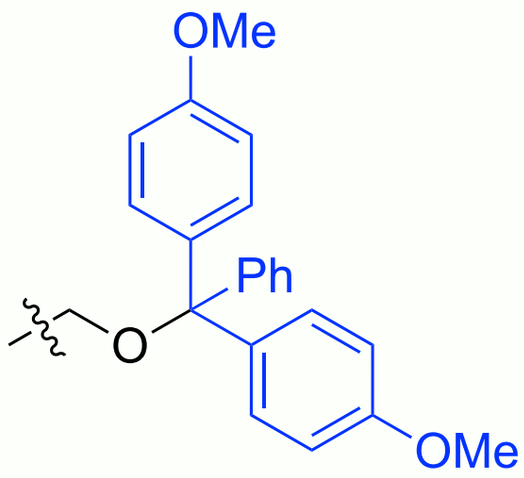


# B Chemical Synthesis Of DNA

Sequencing DNA begins by making primers which complement parts of the template which have a partially known sequence, and allowing a polymerase to extend this by making a complement of the unknown overhang region at the *3'*-end. Typically the template comes from a living organism, but the primer is made chemically.

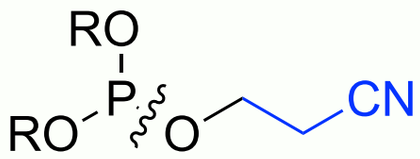
## Protecting Groups

It is necessary to introduce masked bases sequentially in chemical syntheses of oligonucleotides. They are masked because otherwise sensitive functionality may be changed in the synthetic process. Masking is achieved via *protecting groups* to cover up sensitive functionalities that might otherwise react undesirably during syntheses, then removing them; the final product is ssDNA. Here are protecting groups for DNA synthesis via phosphoramidites.



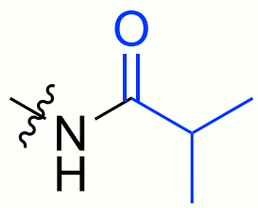

*DMT*
*dimethoxytrityl*

*protects sugar 5' O*



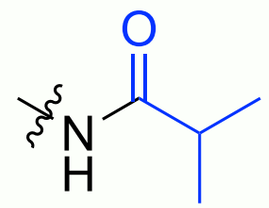

*cyanoethyl*

*protects phosphoramidite*




*CO$^{i}$Bu*
*iso-butyryl*

*protects exocyclic* ***G*** *amine*




*Bz*
*benzoyl*

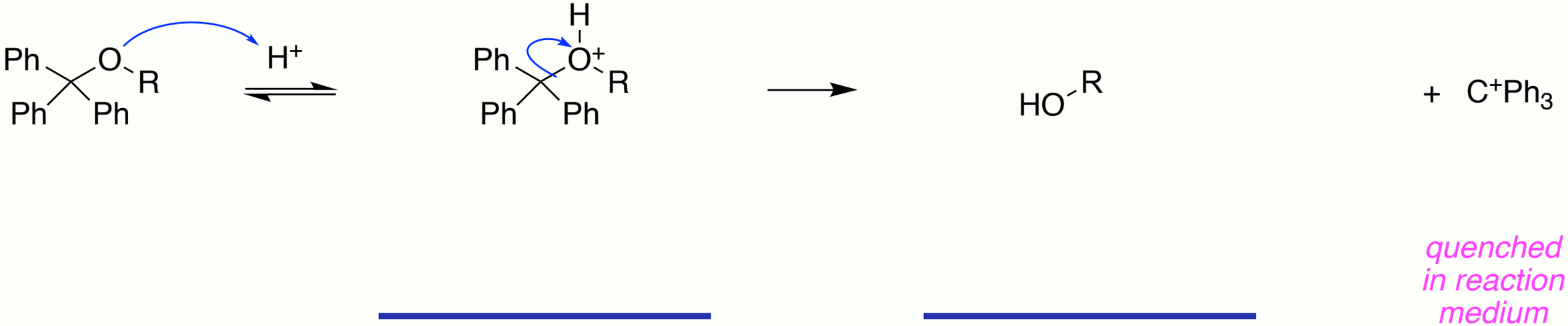
*protects exocyclic* ***A*** *and* ***C*** *amines*

# B Chemical Synthesis Of DNA

DNA sequences of 100 bases, or even longer, can be made via automated syntheses, though most research laboratories tend to buy them from commercial suppliers.

A widespread method for DNA synthesis involves protection of the 5'-O with a trityl ($CPh_3$) derivative, *P* – *O* bonds with *O*-cyanoethyl, *G* exocyclic amine with $CO^{i}Bu$, and exocyclic amines of C and *A* with Bz.

## Trityl Groups

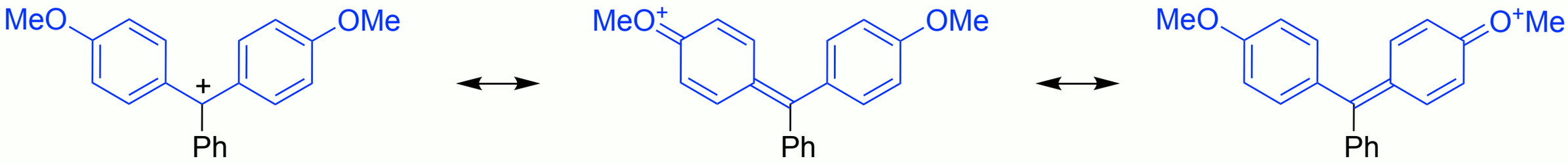



This is a $S_N1$ reaction, "seen through the eyes of the protecting group".

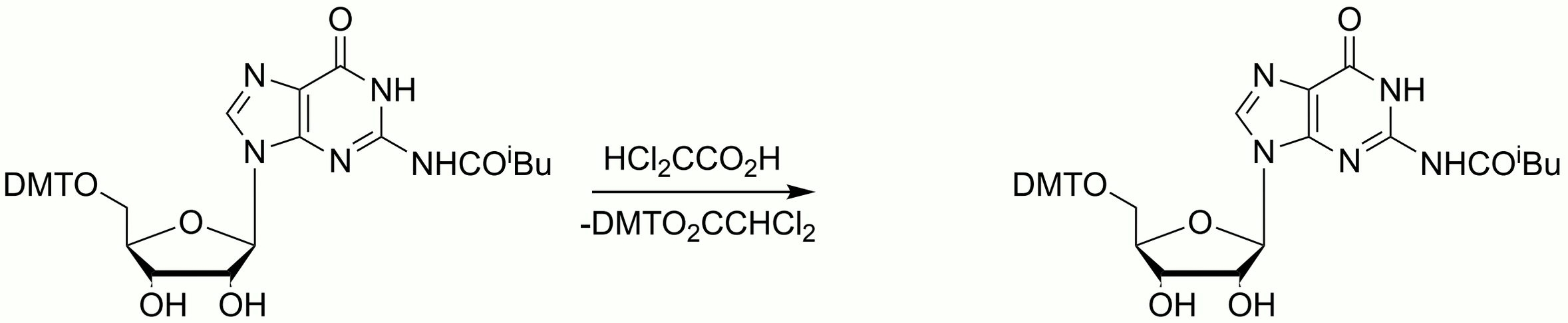
# B Chemical Synthesis Of DNA

## DMT

Substitution of two phenyl groups in trityl with 4-methoxybenzene gives dimethoxytrityl, DMT. Removal of DMT using protons (like above) would be *easier* than for trityl because cations are *less* stabilized by resonance. Show some of the resonance structures which do this stabilization by completing the following (add bonds and curly arrows).

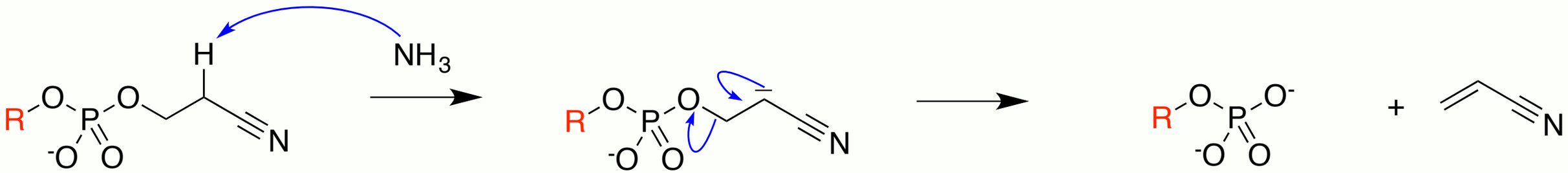



Show the product of the following.



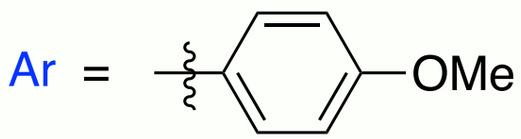
# B Chemical Synthesis Of DNA

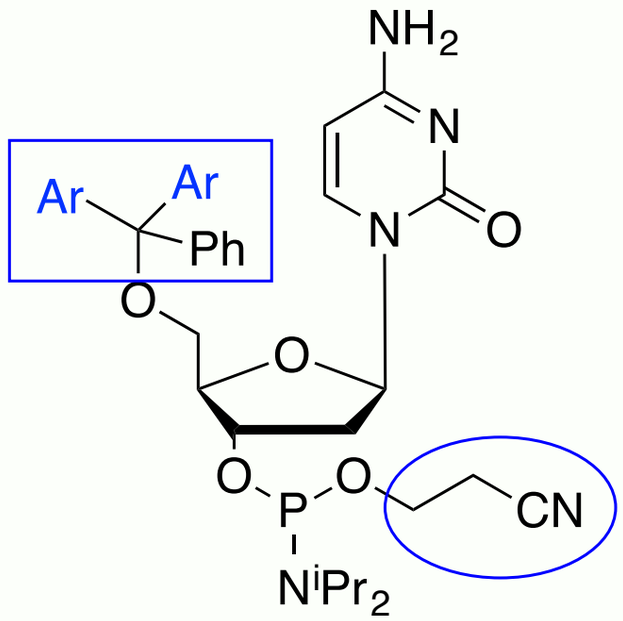
## Cyanoethyl




This is an *E1cb* reaction, also “seen through the eyes of the leaving group”.

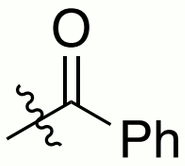
# B Chemical Synthesis Of DNA



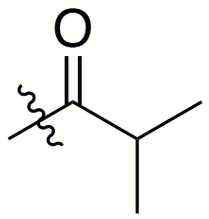



# B Chemical Synthesis Of DNA

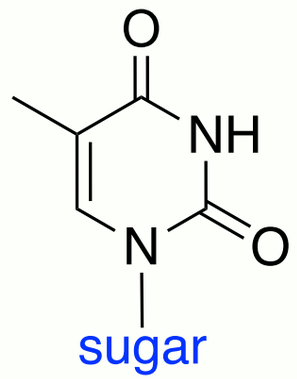
Exocyclic amine groups in **G**, **A**, and **C** must be protected to withstand DNA synthesis.



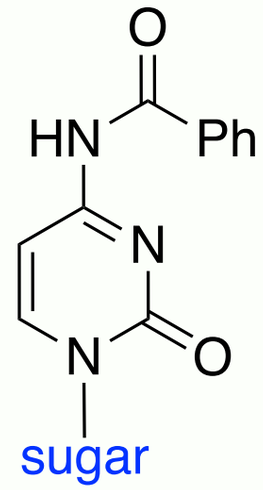

*benzoyl*
*Bz*



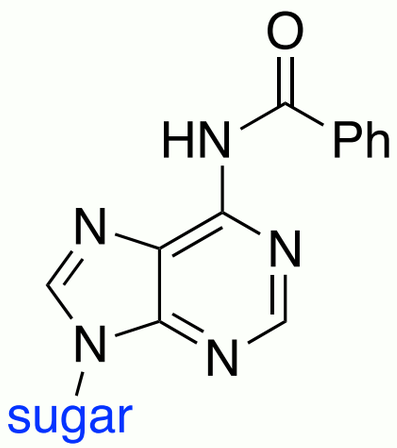

*iso-butyryl*
*CO$^{i}$Bu*



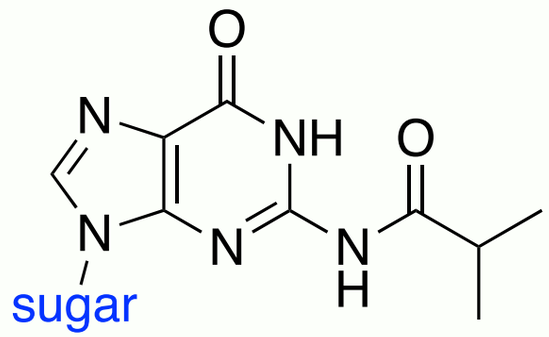

*no protection required*




*N-Bz **C***




*N-Bz **A***

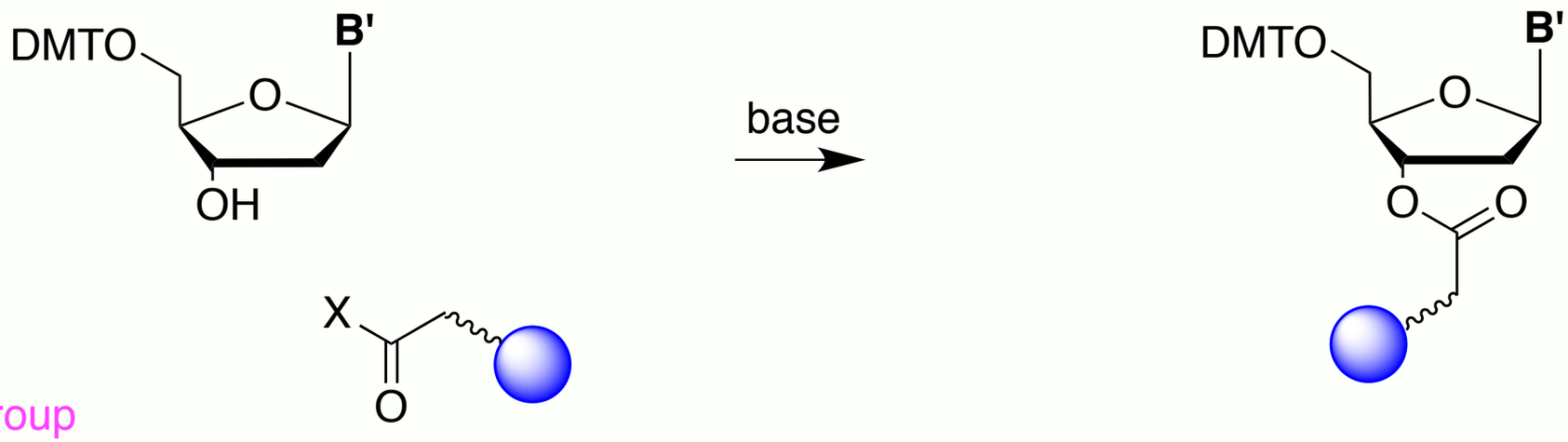



*N-iso-butyryl **G***

# B Chemical Synthesis Of DNA

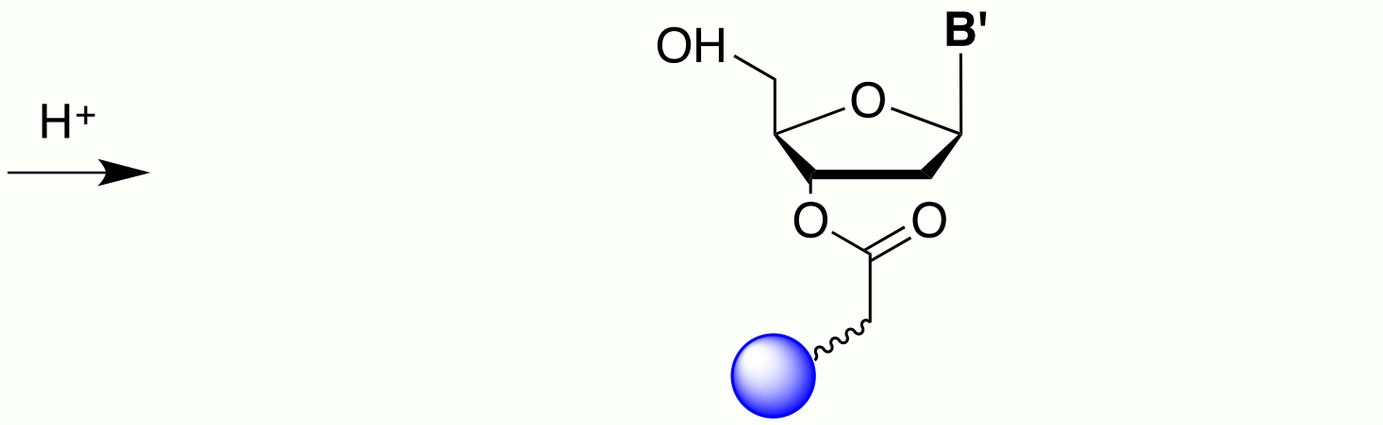
## Phosphoramidite Method

In the *phosphoramidite method* the first base is attached via the 3'-OH to a support, typically a glass bead.



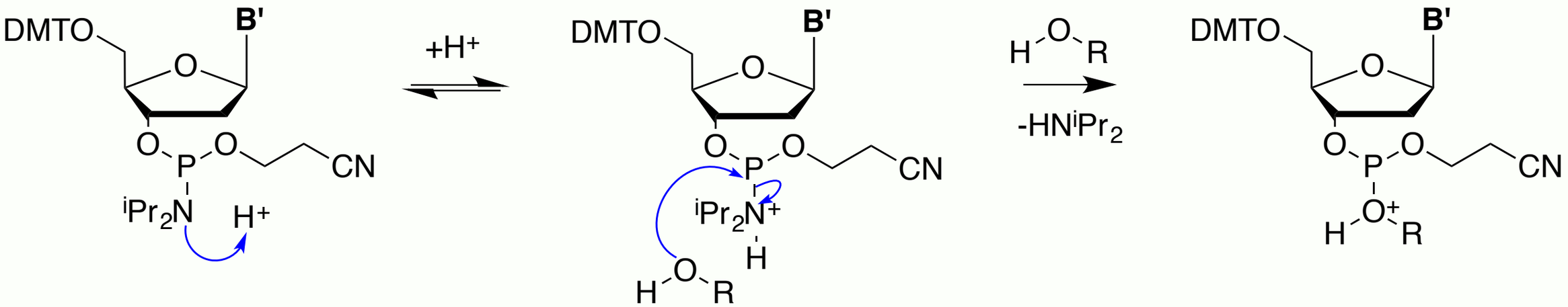

X = leaving group

**B'** = protected nucleobase



# B Chemical Synthesis Of DNA

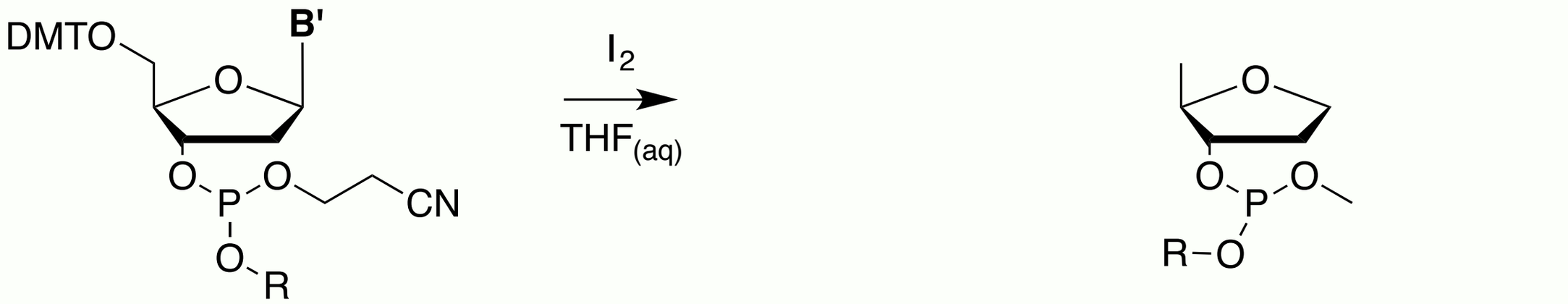
Coupling steps involve solution activation of the next phosphoramidite via protonation at P – $NR_2$.




The acid used must be mild enough to promote activation without removing the DMT group. *Tetrazole* (only one of these is an acid) is preferred. A *stronger* acid, dichloroacetic, is used to remove the DMT group later in the process.

# B Chemical Synthesis Of DNA

Coupling, as shown above, is the first step in each cycle of DNA syntheses, and the last is DMT removal. Between these comes *oxidation* of the *P*-atom in the phosphoramidite to the phosphorus(5+) form.

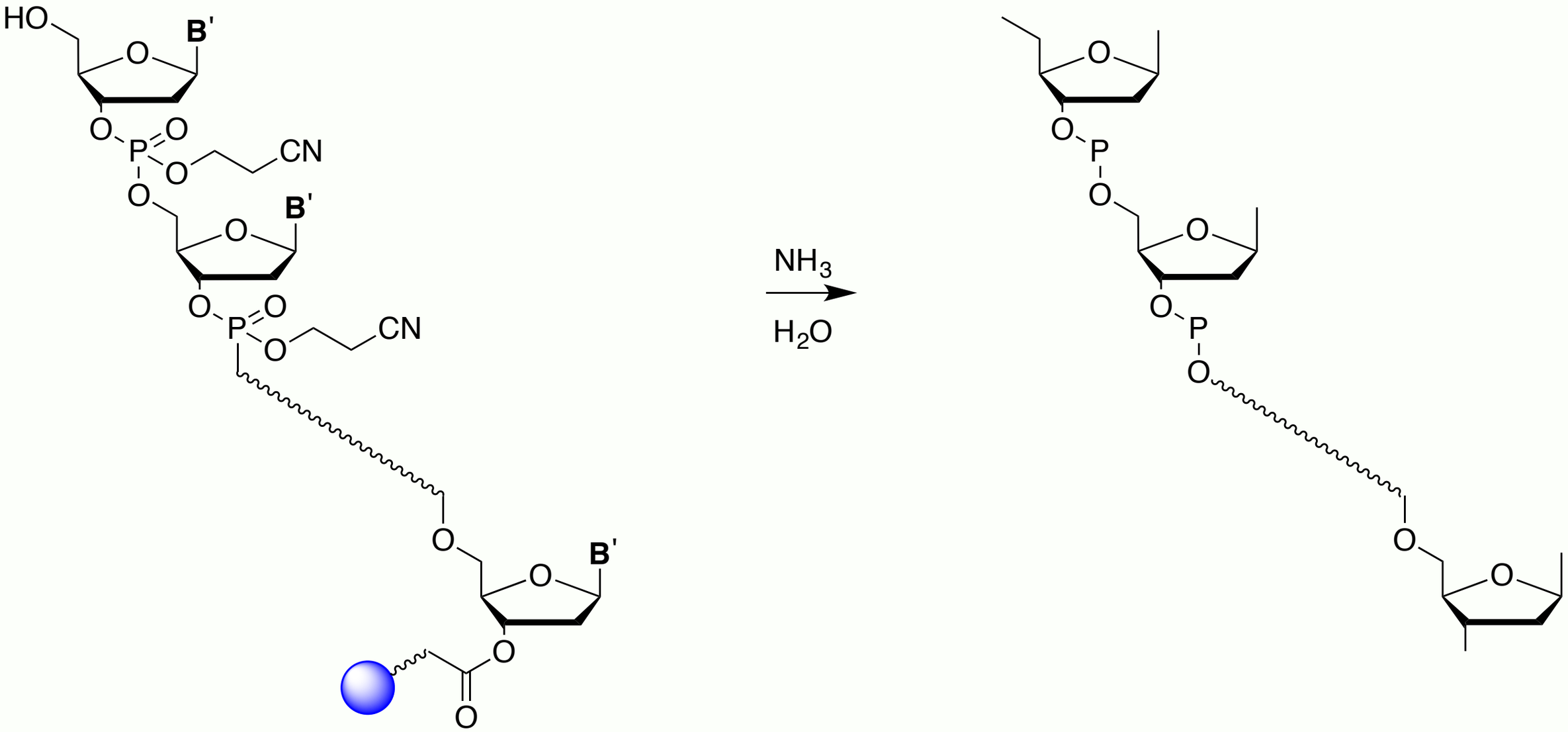



*phosphotriester*

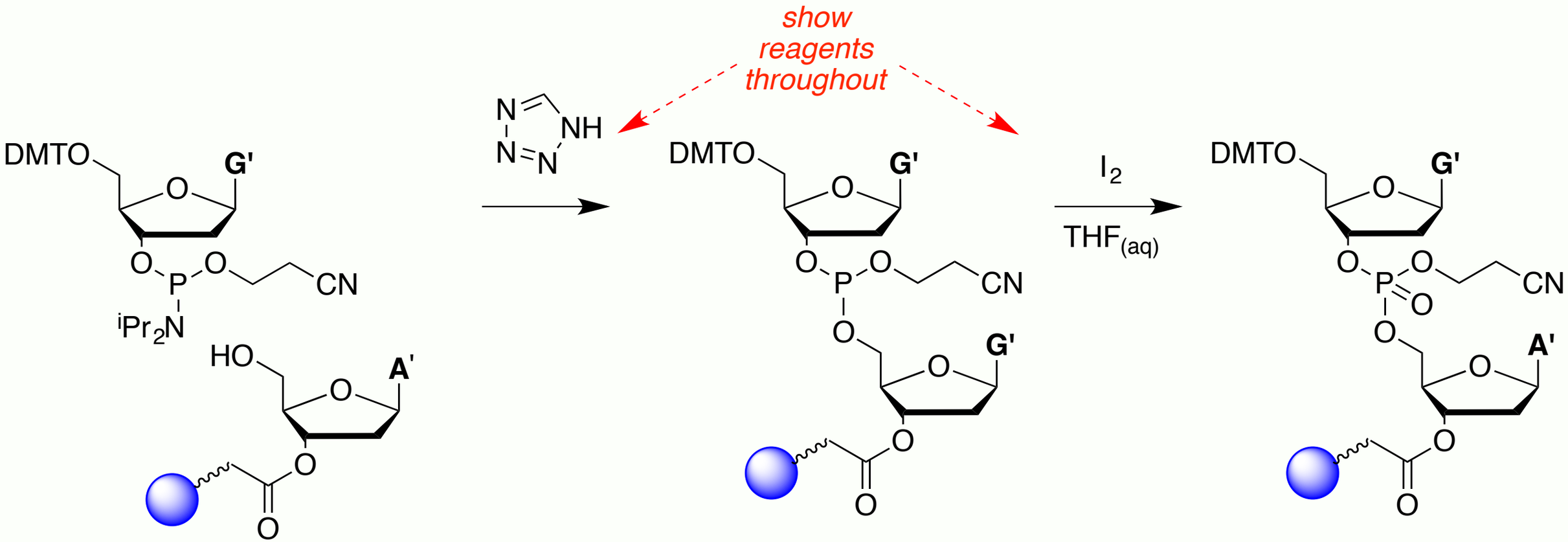
The *P* oxidation state of a phosphoramidite is *3*, while phosphodiesters are *5*.

# B Chemical Synthesis Of DNA

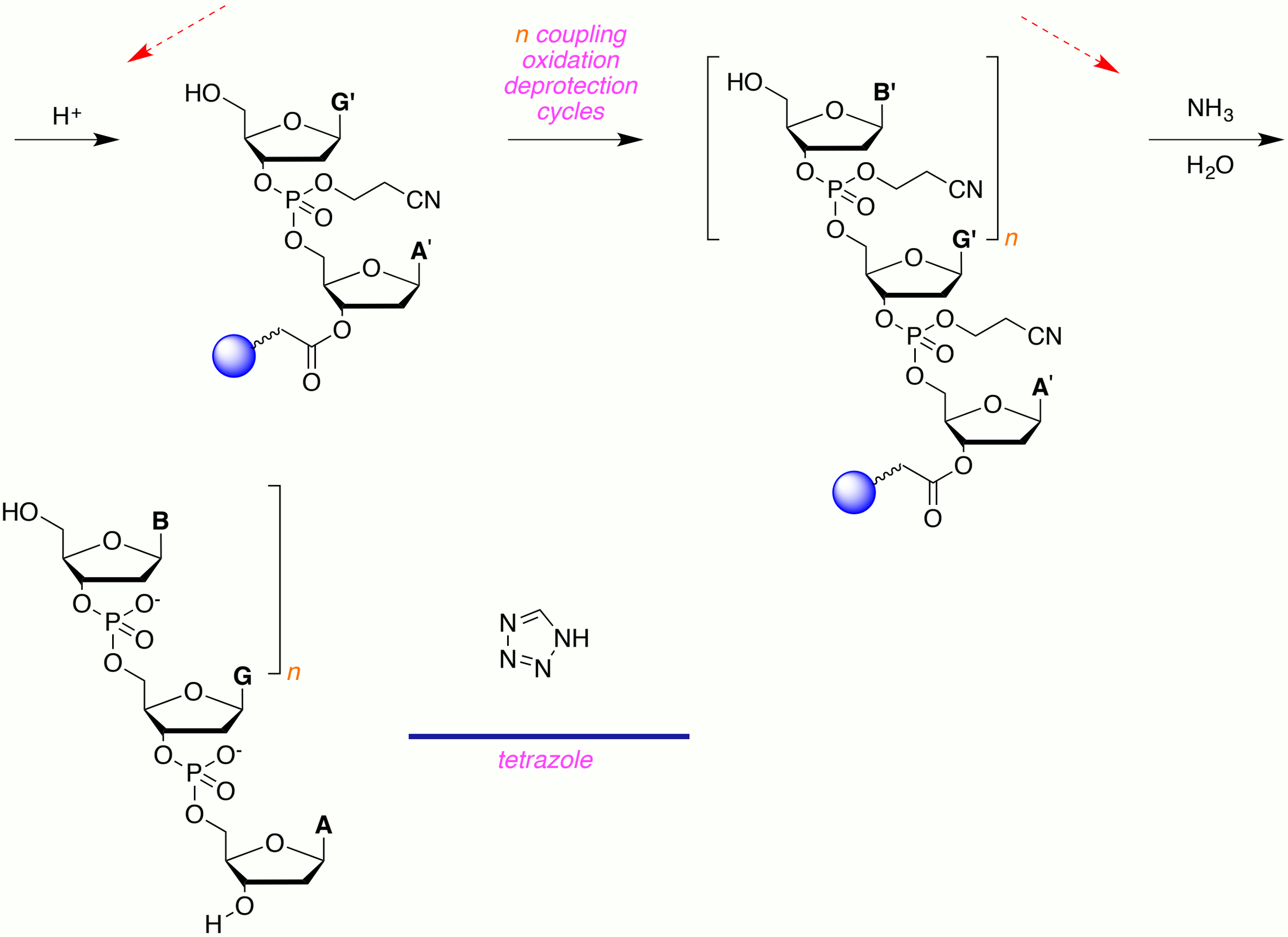
After repetitive couplings of phosphoramidite, oxidation, then DMT removal, DNA is cleaved from the support by treatment with ammonia; this *also removes* the *iso*-butyryl, benzoyl, and cyanoethyl protecting groups.



# B Chemical Synthesis Of DNA



# B Chemical Synthesis Of DNA

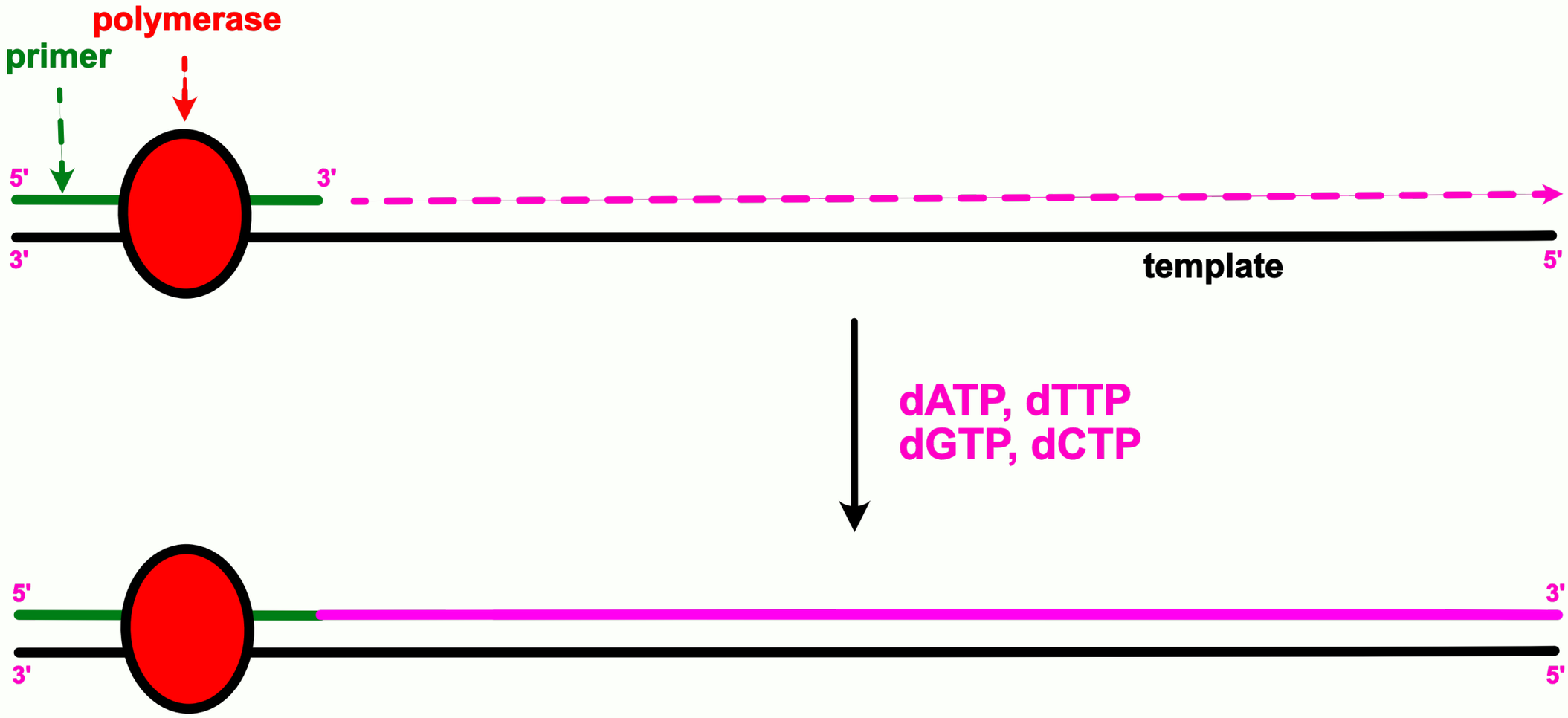


# C Polymerases

## Biological And Biochemical Syntheses Of Complementary Strands

Refer to the diagram below. Polymerases evolved to make complementary DNA begin by forming a complexes to *dsDNA* regions. In biotechnology, this dsDNA is created from primers, prepared as above, specifically for this purpose. The long DNA strand is called the *template.*

Polymerases sequentially accept dATP, dTTP, dCTP, or dGTP from solution as required to make the *complement*. They proceed down the template quickly and dissociate once it is all copied. Imagine a cartoon train in on a double track, then encountering only one rail ahead. It responds by quickly laying the second track in front it. Polymerases work like that except using four different kinds of sleepers (wooden supports under the rails).



# C Polymerases

## PCR Reactions

Refer to the diagram. The first step gives *1* complement of the template.

The third gives *4* strands total.

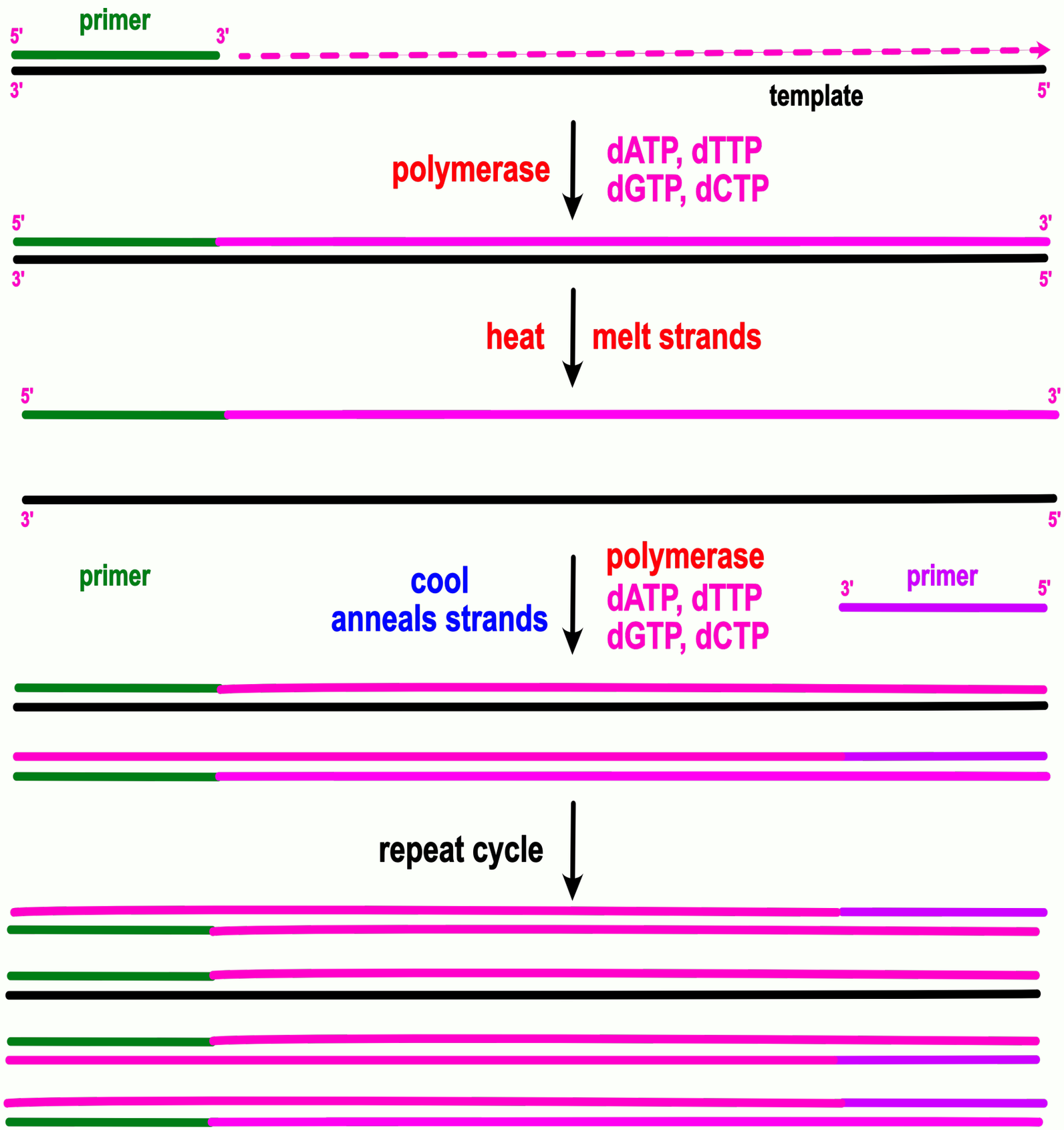
The next cycle gives *8* total.

The number of strands in PCR is $2^n$ where *n* is the number of cycles.

Polymerases used in PCR are based on *thermostable* ones, originally found in hot springs.

PCR *does* enable tiny amounts of DNA to be replicated many times giving larger amounts necessary for DNA sequencing experiments.

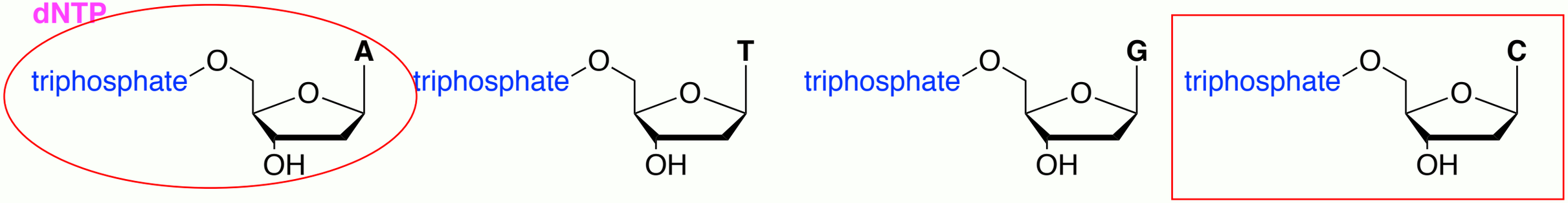
The diagram shows primers at the end of the template strands; they *do not* have to be there. Primers are designed to bracket the sequence to be amplified.



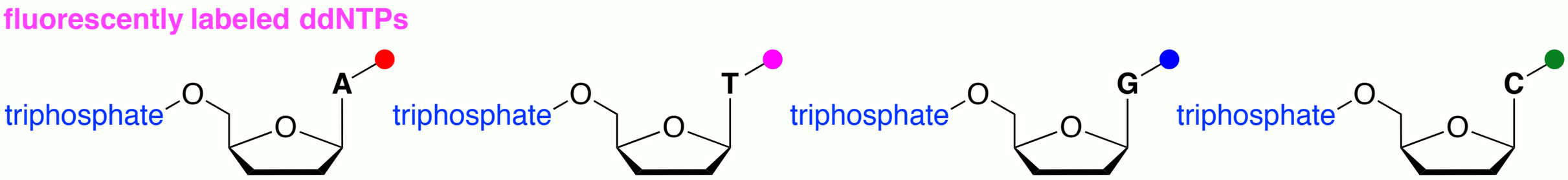
# G DNA Sequencing

## Modified Sanger

Polymerase enzymes recognize double strand (ds) DNA having long strand 5'-ends extending beyond the short strand 3' termini. In this case the short strand is called a *primer* and the long strand is called the *template*.




Imagine an imperfectly selective polymerase which sometimes incorporates differentially fluorescently-labeled, 2',3'-didexoynucleotide triphosphates (**ddNTP**s, as shown below) *even* if the complementary dNTP is present.



# G DNA Sequencing

If the polymerase is supplied all the dNTPs and labeled-ddNTP in carefully determined ratios then it may occasionally incorporate a ddNTP instead of a dNTP, *eg* dCTP opposite *G*; if it does the synthesis *cannot* progress.

When the polymerase replicates primed DNA, each terminated complementary strand contains *all* the fluors.

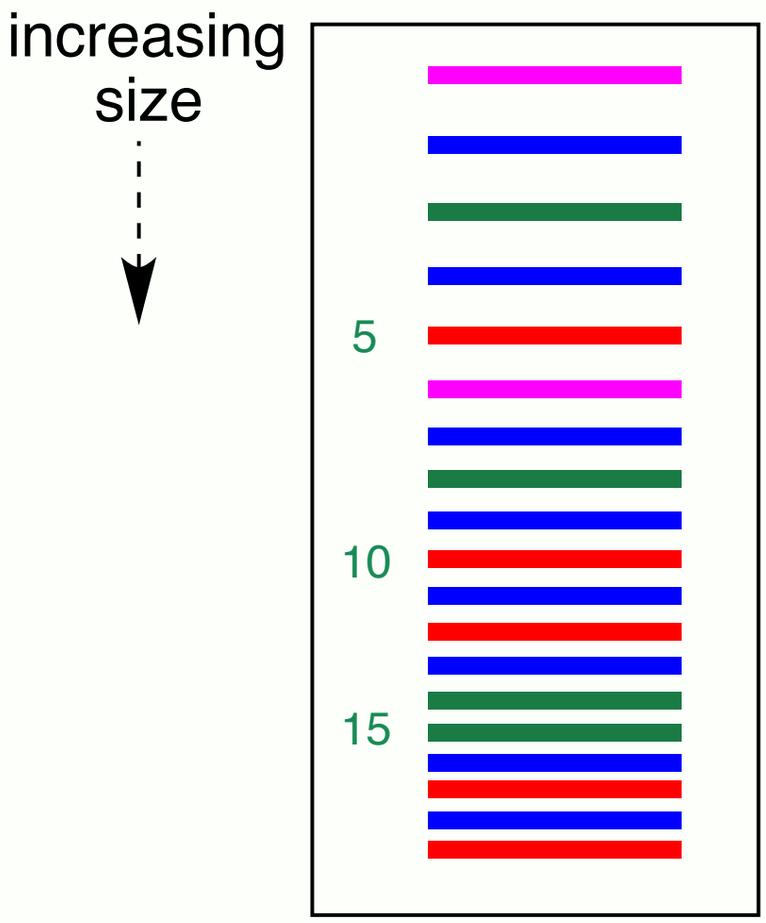
If the terminated DNA strands are separated according to number of nucleobases incorporated, then the order of incorporation *can* be read by fluorescence detection, and the sequence of the complement *can* be deduced.

The sequence of 19 bases incorporated as revealed by the gel to the left that separates according to size (colors on the gel represent bases added as coded on the previous page) is:

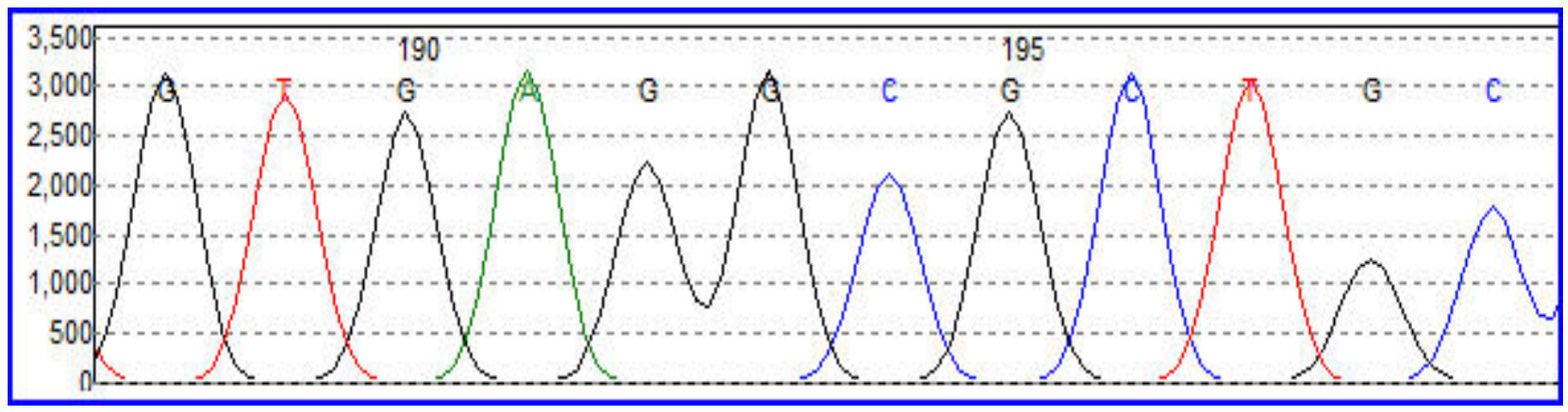
5'-TGCGATGCGAGAGCCGAGA-3'

The sequence of the template DNA after the primer is:

5'-TCTCGGCTCTCGCATCGCA-3'




Capillary electrophoresis (Wiki it) can be used instead of gels for separations, then fluorescence of progressively longer fragments are read as they emerge from the column giving traces like the one on the right.



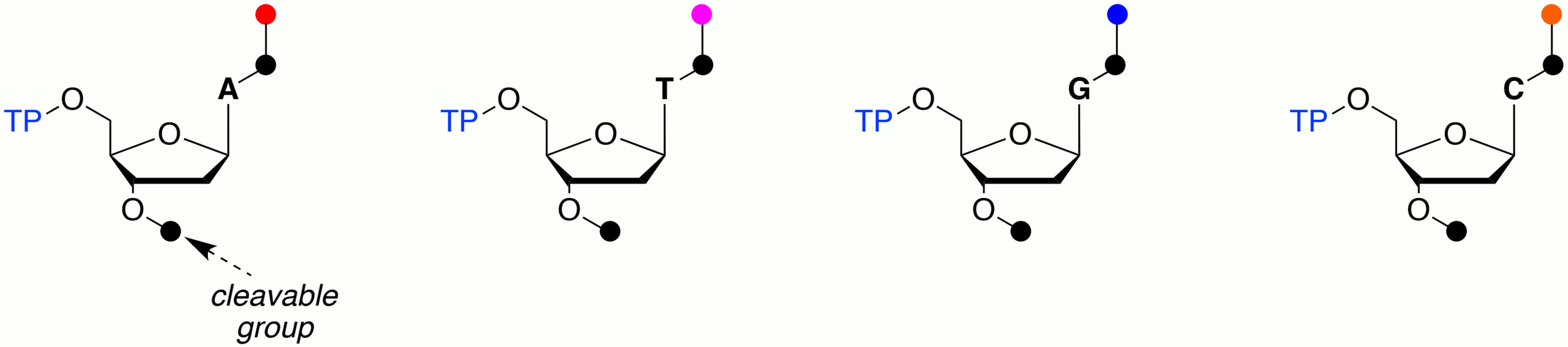
# G DNA Sequencing

## Sequencing By Synthesis (SBS)

Genomic sequencing is now performed via several different *next generation* methods, of which sequencing by synthesis (SBS) is one.

Envisage a polymerase, a primed DNA template (to be sequenced), and *only* the four derivatives below (the colored circles represent fluorescent dyes attached to the bases via a cleavable linker shown as a black circle; no dNTPs).

**fluorescently labeled, blocked, dNTPs**




Further synthesis *would not* occur after addition of one base, just as in the Sanger method.

The identity of that base and its complement *could* be deduced from fluorescence.

# G DNA Sequencing

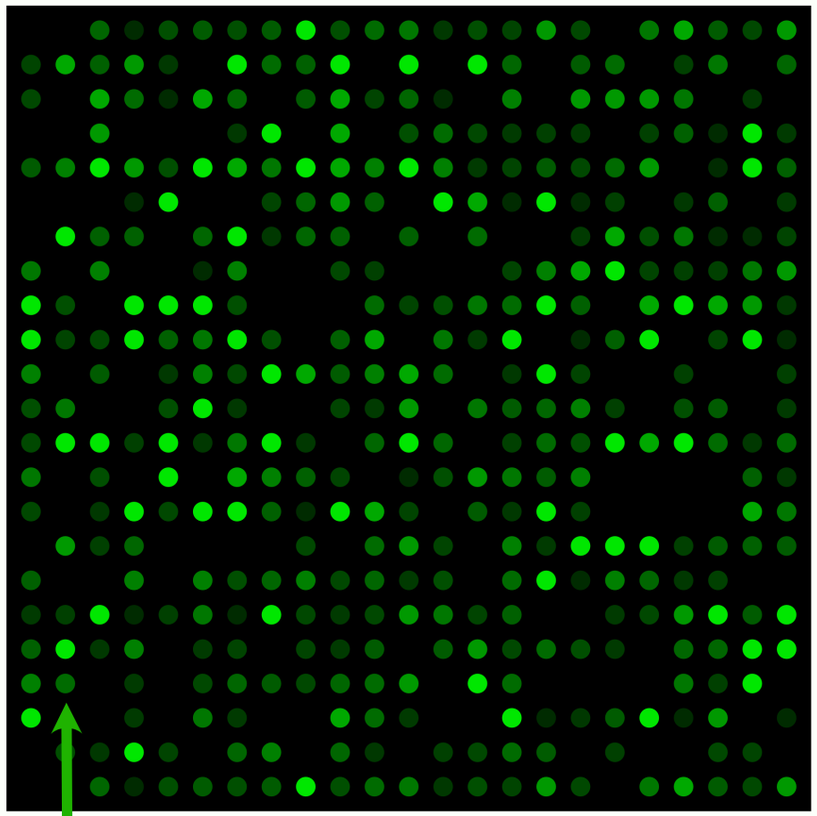
If the cleavable group were cleaved, the fluor would be lost, and further synthesis *would / would not* be possible because the 3'-OH was also liberated.

Imagine different primers anchored to a solid surface, so groups of these molecules, or even single molecules, had constant locations which could be tracked and uniquely addressed. These could then be treated with template mixtures which would anneal to *different* spots on the plate, as on the microarray chip below left.

If polymerase and the four fluorescently-labeled, blocked dNTPs were added, then SBS could be performed to give *different* sequences on all the primed templates. Consequently, *y* different DNA regions could be sequenced at the same time.

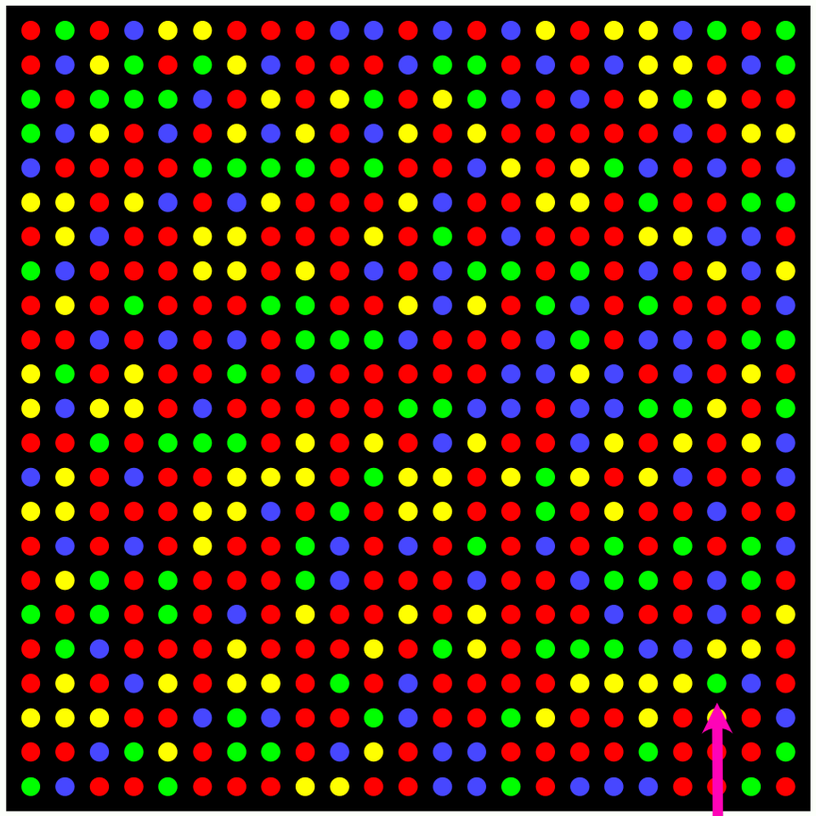
If *n* bases could be sequenced for *y* locations then the efficiency would be *y* x *n*. In practice, *y* can be >>*n* because huge numbers of spatially addressable DNA templates can be supported. Consequently, SBS can be made to be much *more* efficient than Sanger sequencing.

microarray chip



each spot is a supported primer annealed with a template strand

sequencing flow cell

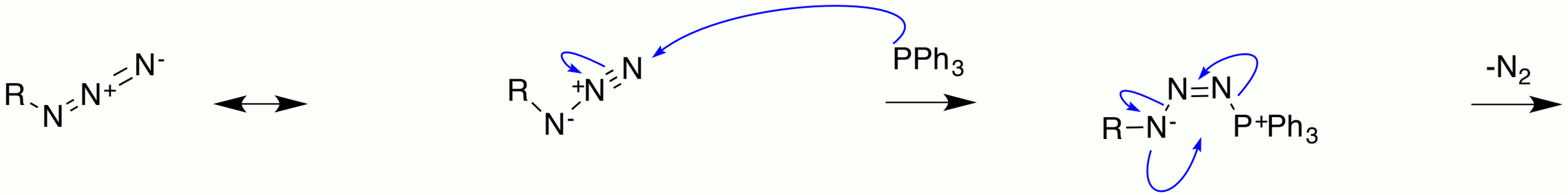


after incorporation of differently labeled, 3'-blocked nucleotides

# G DNA Sequencing

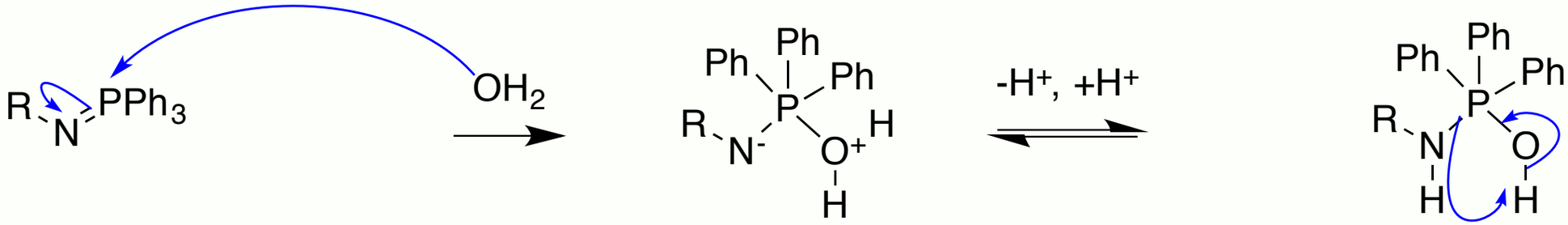
## A Cleavable Group For (SBS) Designed Around The Staudinger Reaction

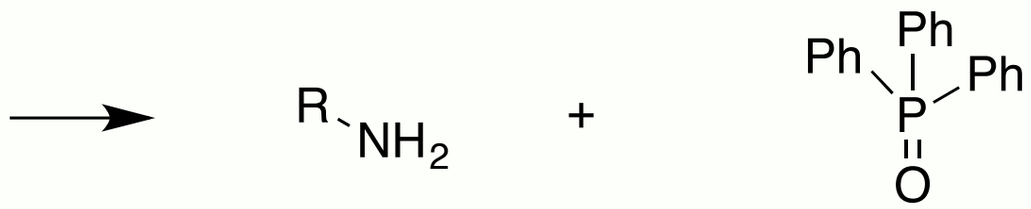
3'-Protection of fluorescently labeled-ddNTPs as described above has to be compatible with the polymerase. It is also important the dsDNA *remains annealed* when this group is removed; consequently the conditions must be *gentle*. One protection strategy for this is based on reduction of azides with phosphines, *ie* the *Staudinger* reduction.




*cont. …..*

# G DNA Sequencing

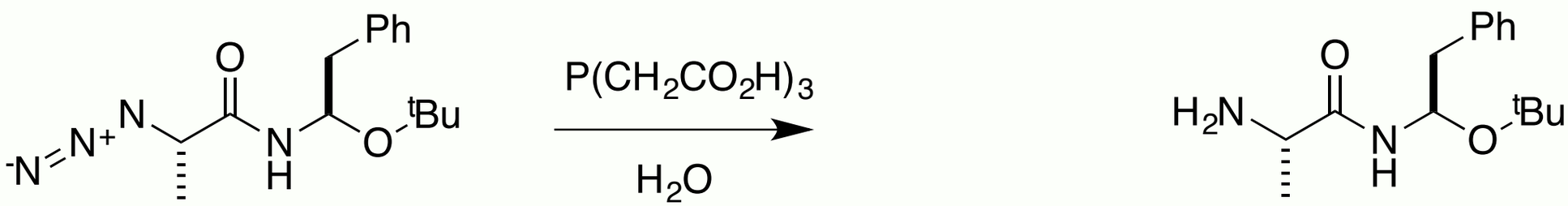


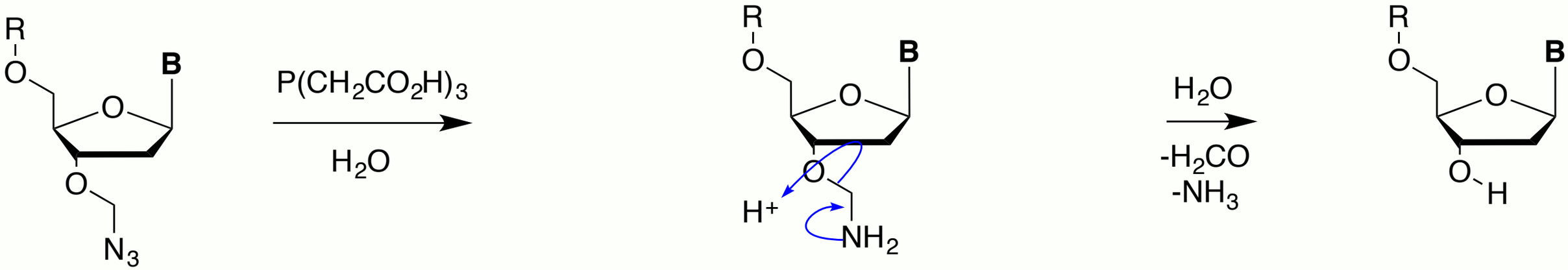

At the beginning of this reaction, the *P*-atom of triphenylphosphine is at the *+3* oxidation state, but it ends up *+ 5*, *ie* it has been *oxidized*.

The azide in this reaction is *reduced.*

# G DNA Sequencing

## 3’-Azidomethylene Protection Of Nucleotides




The alcohol protecting group in this reaction, *azidomethylene,* is the one frequently used in sequencing by synthesis. A similar functionality is used to cleave fluorescent reporter groups attached to the bases under the same conditions.

# H Restriction Enzymes

Imagine our task is to sequence a gene using a next generation method. Small parts of the sequence must be known to enable primers to be designed. One way to achieve this is to cut the DNA into fragments using *restriction enzymes*.

Restriction enzymes cut at characteristic sequences so their sequence specificities can be give the information needed to design primers. Some restriction enzymes cut at defined *palendromic sequences*.

## Palindromic DNA

Points in the following ssDNAs enable them to fold back onto themselves to form dsDNA if it is geometrically possible. Identify those points by drawing arrows on the ssDNA sequences below.

TAGC•GCTA GGCCT•AGGCC TTACGGG•CCCGTAA GCGCTATATAGC•GCTATATAGCGC

These *are* palindromic, just as words like CIVIC, KAYAK, and ROTATOR. Palindromic DNA sequences *do not* have a mirror plane of symmetry.

Circle palindromes in the following.

TT**TATCGATA** TTTAGC**TAGCGCTA**TGGCT TTTAGCT**AGCGCT**CATGGCT

Palindromic sequences are special because they *cannot* form dsDNA with themselves.

## H Restriction Enzymes

Draw dsDNA formed from the following by showing complementary strands below.

5’-TATAT•ATATA-3’ 5’-TATAGCTAGC•GCTACGTATA-3’ 5’-TATAGCTAGC•GCTAGCTATA-3’

3’-ATATA•TATAT-5’ 3’-ATATCGATCG•CGATGCATAT-5’ 3’-ATATCGATCG•CGATCGATAT-5’

The complementary strands above *are* the same as the ones above them.

Imagine an enzyme capable of specifically cutting between bases 2 and 3 residues in the 5’ direction from the center of the palindromic sequence in the following dsDNA. That enzyme would cut the dsDNA in *two* places:

5’-…….TAATTTATTATGCCGGG AA•TTCTTATTATGCCG……-3’
3’-…….ATTAAATAATACGGCCCTT•AA GAATAATACGGC……-5”

The *restriction* enzyme that cuts as shown above is *EcoRI* (on Wiki). dsDNA produced *would* have an overhanging four base ssDNA region, which might be called a *sticky* end.

The *restriction* enzyme *SmaI* (investigate using Wiki) cleaves dsDNA in exactly the center of the palindromic sequence in the following dsDNA, *ie* in *one* place:

5’-…….TAATTTATTATGCCGGCCC GGGTTATTATGCCG……-3’
3’-…….ATTAAATAATACGGCCGGG CCCAATAATACGGC……-5’

DNA produced has a *blunt* end (in fact, it has a one base overhang).

# I CRISPR

Cas9 is a helicase and an endonuclease found in *S. pyogenes*. Its role is to cleave the DNA of invading viruses. When a new viral DNA fragment is detected, the bacteria stores information about that sequence in its own DNA. Specifically, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) store information about palindromes corresponding to critical proteins in viruses which invade bacterial cells. Spacers between CRISPR repeats encode variable regions of that particular virus strain's DNA.

When a new strain of virus invades a cell, a new CRISPR repeat is created. Alternatively, if invasion occurs featuring DNA encoded before, the bacteria responds by expressing a sequence of RNA corresponding to the CRISPR and more detailed information about the DNA for that strain; this is called crRNA. Trans-activating, or tracer RNA (trRNA) is also expressed that complements part of this sequence. Together, Cas9, crRNA, and trRNA form a complex with the invading dsDNA, and the enzyme Cas9 is activated.

Stored information about the viruses also encapsulates a short segment of code (a protospacer adjacent motif, PAM, *eg* 5′-NGG-3′) incorporated to ensure CRISPR•CAS9 only works on viral DNA and not on the bacteria's own. Cas9 cuts both strands, destroying DNA of the invading virus.

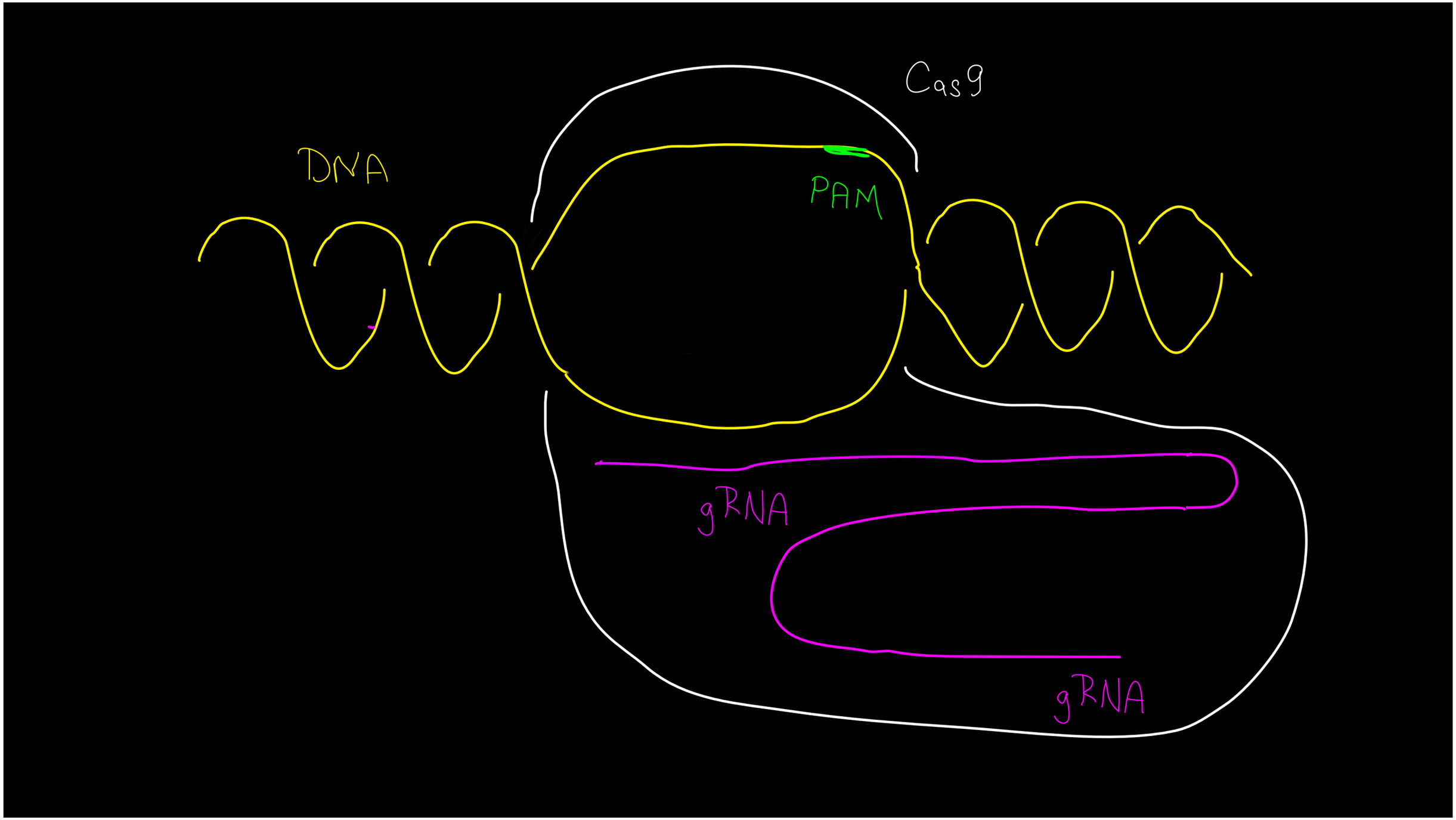
## Gene Editing

The Nobel Prize winning idea that enabled CRISPR•CAS9 to be adapted for editing genomes was to use crRNA and trRNA combined into *guide RNA*, or gRNA. gRNA contains genomic information corresponding to the genome of interest, in place of the viral DNA in the bacterial defense system. CAS9 complexes this RNA and cuts both strands of the unwound DNA where it is attached.

In this way, introduction of gRNA and CAS9 enables genomic DNA to be cut at almost any positions near a PAM code, *unlike* restriction enzymes which have relatively inflexible selectivities. These cuts can be used to prevent the targeted DNA expressing proteins (loss of function). Moreover, transfection of alien DNA fragments into the cell facilitates their introduction of other DNA into the gene (gain of function). Thus, CRISPR/CAS9 genomic editing enables target genomes to be *inactivated and functionally modified*.

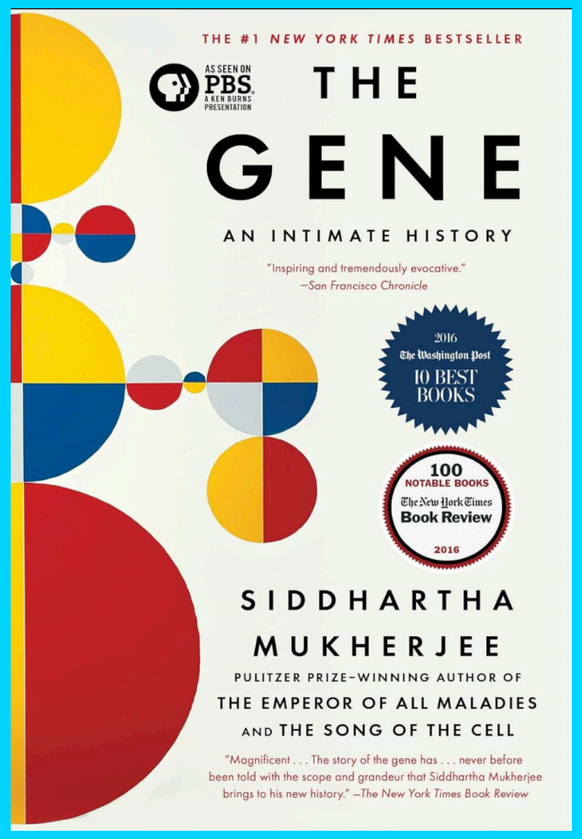
# I CRISPR

Identify components of the genomic editing CRISPR Cas9 system from the choices indicated.

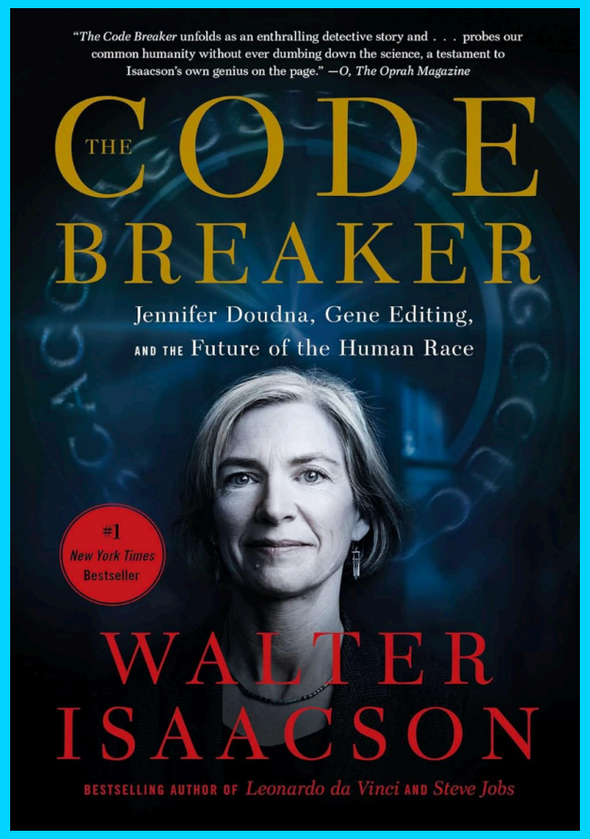


# Book Recommendations scientists who have made history

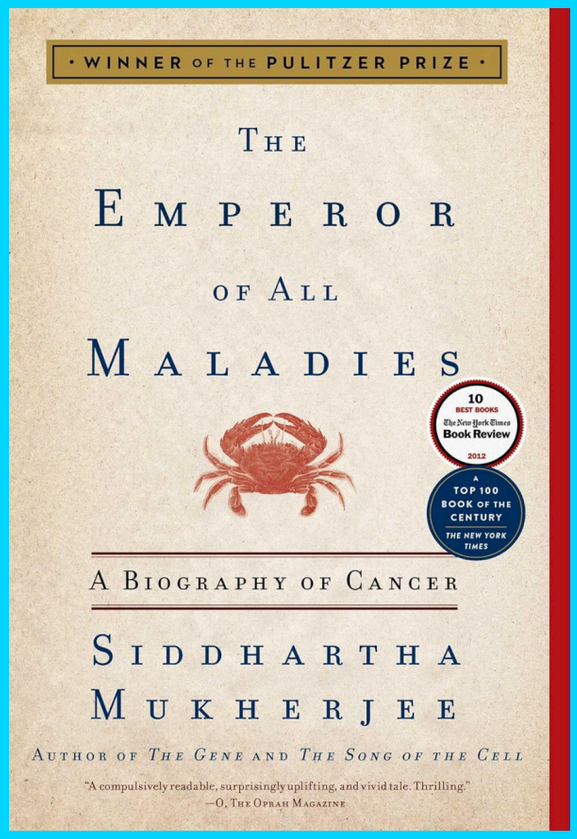
for lay people, but easier after mastering this eworkbook



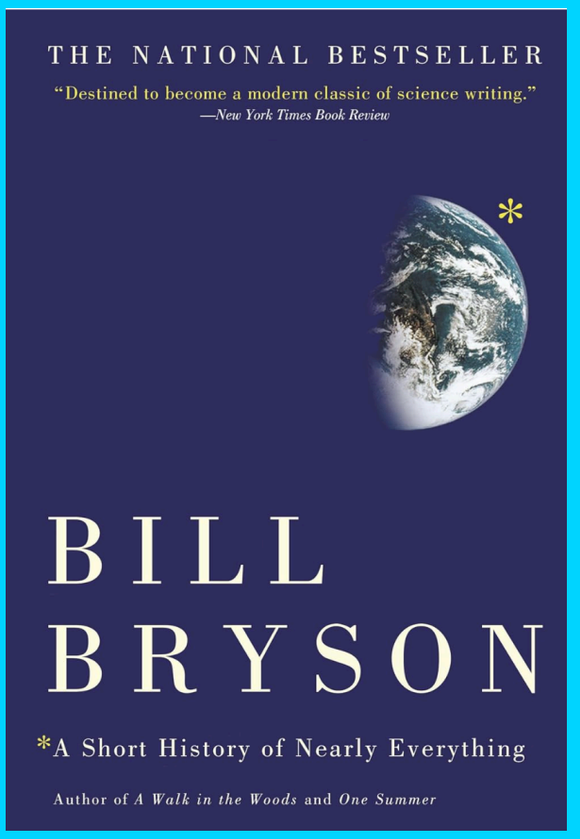

affiliate link




affiliate link




affiliate link




affiliate link