# The chemistry of life

# 49

## Connections

**Building on:**

- Acidity and basicity ch8
- Carbonyl chemistry ch12 & ch14
- Stereochemistry ch16
- Conformational analysis ch18
- Enolate chemistry and synthesis ch24–ch30
- Heterocycles ch42–ch44
- Asymmetric synthesis ch45
- Sulfur chemistry ch46

**Arriving at:**

- Nucleic acids store information for the synthesis of proteins
- Modified nucleosides can be used as antiviral drugs
- Nucleotides have a role in energy storage
- Proteins catalyse reactions and provide structure
- Other amino acid derivatives act as methylating and reducing agents
- Sugars store energy, enable recognition, and protect sensitive functional groups
- How to make and manipulate sugars and their derivatives
- Lipids form the basis of membrane structures

**Looking forward to:**

- Mechanisms in biological chemistry ch50
- Natural products ch51
- Polymers ch52

Life runs on chemistry, and the chemical side of biology is fascinating for that reason alone. But from the point of view of a textbook, biological chemistry's combination of structures, mechanisms, new reactions, and synthesis is also an ideal revision aid. We shall treat this chemistry of living things in three chapters.

- Chapter 49 introduces the basic molecules of life and explains their roles along with some of their chemistry
- Chapter 50 discusses the mechanisms of biological reactions
- Chapter 51 develops the chemistry of compounds produced by life: natural products

We start with the most fundamental molecules and reactions in what is called **primary metabolism**.

## Primary metabolism

It is humbling to realize that the same molecules are present in all living things from the simplest single-cell creatures to ourselves. Nucleic acids contain the genetic information of every organism, and they control the synthesis of proteins. Proteins are partly structural—as in connective tissue—and partly functional—as in enzymes, the catalysts for biological reactions. Sugars and lipids used to be the poor relations of the other two but we now realize that, as well as having a structural role in membranes, they are closely associated with proteins and have a vital part to play in recognition and transport.

The chart overleaf shows the molecules of primary metabolism and the connections between them, and needs some explanation. It shows a simplified relationship between the key structures (emphasized in large black type). It shows their origins—from $CO_2$ in the first instance—and picks out some important intermediates. Glucose, pyruvic acid, citric acid, acetyl coenzyme A (Acetyl CoA), and ribose are players on the centre stage of our metabolism and are built into many important molecules.
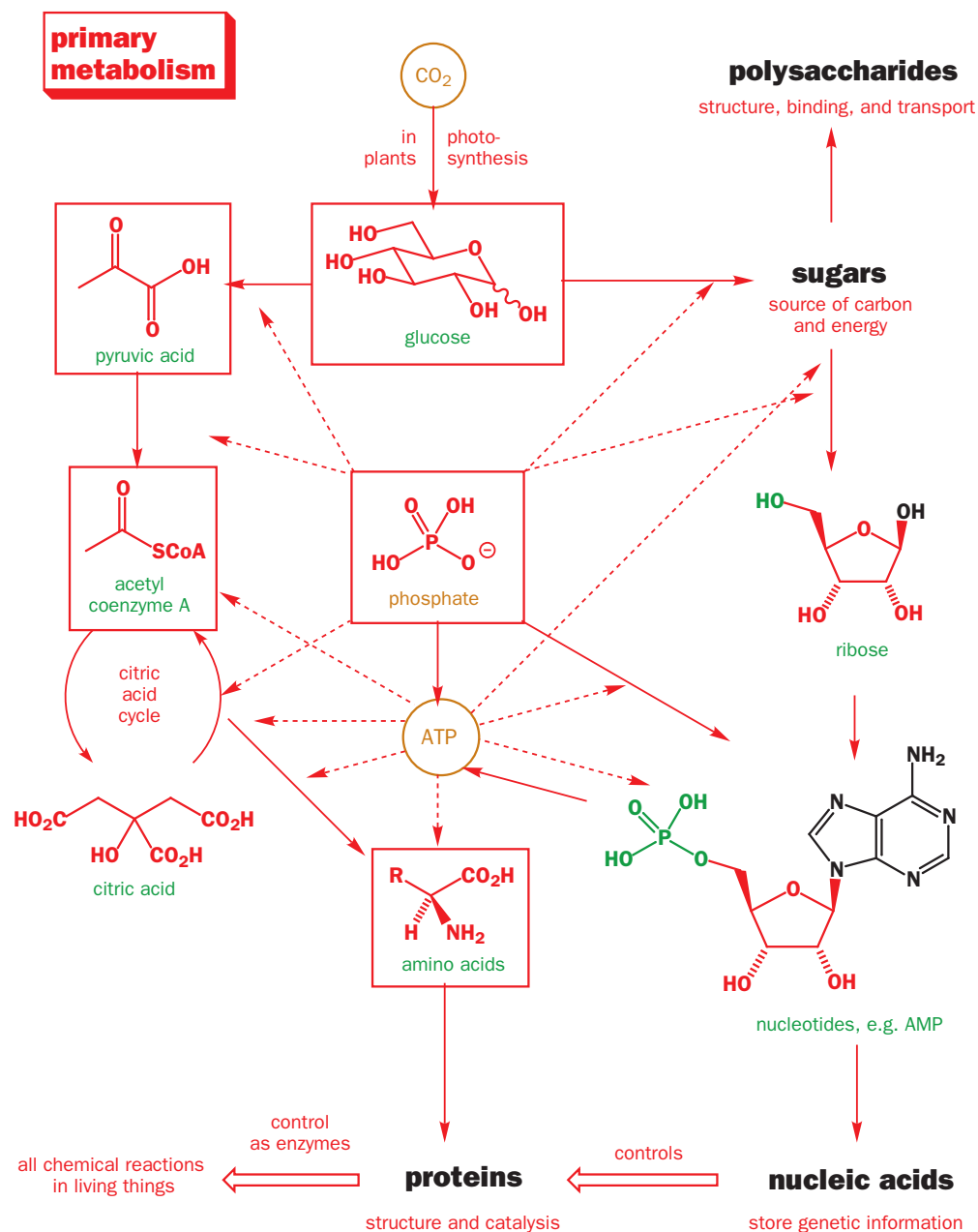
The arrows used in the chart have three functions.

→     chemical reaction in the usual sense: the starting material is incorporated into the product

----→     compound needed for the reaction but not always incorporated into the product

⇒     compound involved in controlling a reaction: *not* incorporated into the products

    We hope that this chart will allow you to keep track of the relationships between the molecules of metabolism as you develop a more detailed understanding of them. We will now look briefly at each type of molecule.

# Life begins with nucleic acids

Nucleic acids are unquestionably top level molecules because they store our genetic information. They are polymers whose building blocks (monomers) are the **nucleotides**, themselves made of three parts—a heterocyclic base, a sugar, and a phosphate ester. A **nucleo*side*** lacks the phosphate. In the example alongside, adenine is the base (black), adeno*sine* is the nucleo*side* (base and sugar), and the nucleotide is the whole molecule (base + sugar + phosphate).

This nucleotide is called **AMP**—Adenosine MonoPhosphate. Phosphates are key compounds in nature because they form useful stable linkages between molecules and can also be built up into reactive molecules by simply multiplying the number of phosphate residues. The most important of these nucleotides is also one of the most important molecules in nature—Adenosine TriPhosphate or **ATP**.



the phosphate ester group

the pyrimidine base (adenine)

the sugar (ribose)

a nucleotide (AMP)

hard (O, N)    soft (S)

Adenosine TriPhosphate—ATP

ATP is a highly reactive molecule because phosphates are stable anions and good leaving groups. It can be attacked by hard nucleophiles at a phosphate group (usually the end one) or by soft nucleophiles at the $CH_2$ group on the sugar. We shall see examples of both reactions soon. When a new reaction is initiated in nature, very often the first step is a reaction with ATP to make the compound more reactive. This is rather like our use of TsCl to make alcohols more reactive or converting acids to acid chlorides to make them more reactive.
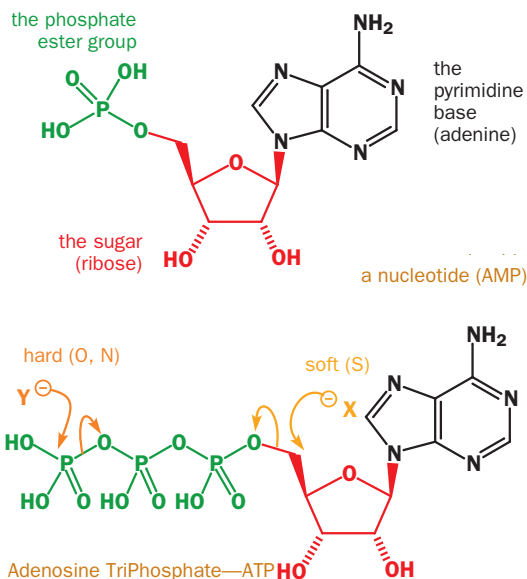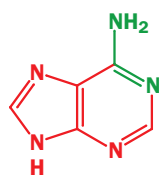
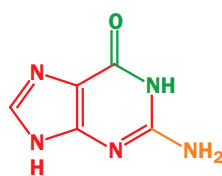## There are five heterocyclic bases in DNA and RNA

In nucleic acids there are only five bases, two sugars, and one phosphate group possible. The bases are monocyclic pyrimidines or bicyclic purines and are all aromatic.

- There are only two purine bases found in nucleic acids, adenine (A), which we have already met, and guanine (G)

- The three pyrimidine bases are the simpler and they are uracil (U), thymine (T), and cytosine (C). Cytosine is found in DNA and RNA, uracil in RNA only, and thymine in DNA only.
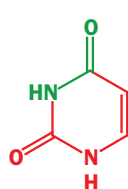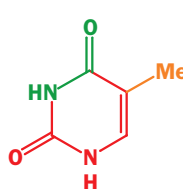
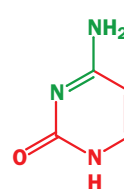■ You met pyrimidines on p. 1148 and learned how to make them on p. 1198, but the purine ring system may be new to you. It isn't always easy to find the six (or ten!) electrons in these compounds. Check for yourself that you can do this. You may need to draw delocalized structures especially for U, T, and G.

purine bases in nucleic acids

pyrimidine bases in nucleic acids



adenine     guanine     uracil     thymine     cytosine
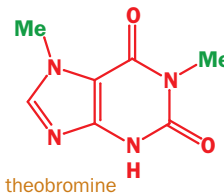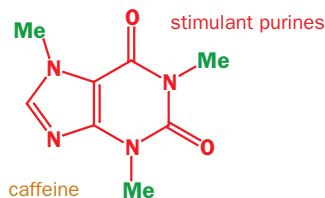
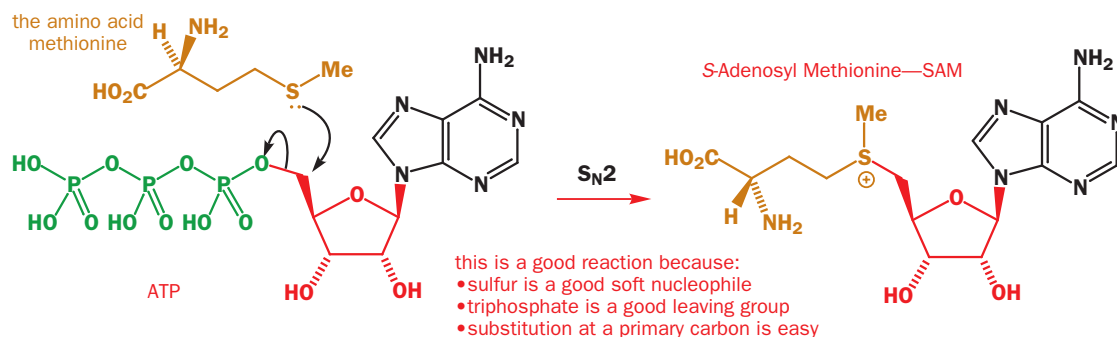## The stimulants in tea and coffee are methylated nucleic acid purines

An important natural product for most of us is a fully methylated purine present in tea and coffee—caffeine. Theobromine, the partly methylated version, is present in chocolate, and both caffeine and theobromine act as stimulants. Caffeine is a crystalline substance easily extracted from coffee or tea with organic solvents. It is extracted industrially with liquid $CO_2$ (or if you prefer 'Nature's natural effervescence') to make decaffeinated tea and coffee.

If we, as chemists, were to add those methyl groups we should use something like MeI, but Nature uses a much more complicated reagent. There is a great deal of methylating going on in living



stimulant purines

caffeine

theobromine

things—and the methyl groups are usually added by *S*-adenosyl methionine (or **SAM**), formed by reaction of methionine with ATP.



this is a good reaction because:
- sulfur is a good soft nucleophile
- triphosphate is a good leaving group
- substitution at a primary carbon is easy

The product (SAM) is a sulfonium salt and could be attacked by nucleophiles at three different carbon atoms. Two are primary centres —good for $S_N2$ reactions—but the third is the methyl group, which is even better. Many nucleophiles attack SAM in this way.

In the coffee plant, theobromine is converted into caffeine with a molecule of SAM. The methylation occurs on nitrogen partly because this preserves both the aromatic ring and the amide functionality and also because the enzyme involved brings the two molecules together in the right orientation for *N*-methylation.
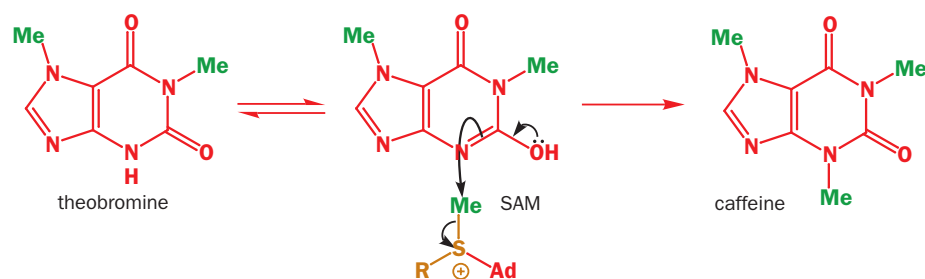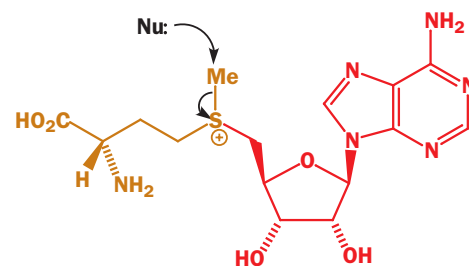


At this point we should just point out something that it's easy to forget: there is *only one chemistry*. There is no magic in biological chemistry, and Nature uses the same chemical principles as we do in the chemical laboratory. All the mechanisms that you have studied so far will help you to draw mechanisms for biological reactions and most reactions that you have met have their counterparts in nature. The difference is that Nature is very very good at chemistry, and all of us are only just learning. We still do much more sophisticated reactions *inside* our bodies without thinking about them than we can do *outside* our bodies with all the most powerful ideas available to us at the beginning of the twenty-first century.

## Nucleic acids exist in a double helix

One of the most important discoveries of modern science was the elucidation of the structures of DNA and RNA as the famous double helix by Watson and Crick in 1953. They realized that the basic structure of base–sugar–phosphate was ideal for a three-dimensional coil. The structure of a small part of DNA is shown opposite.

Notice that the 2′ (pronounced 'two prime') position on the ribose ring is vacant. There is no OH group there and that is why it is called *Deoxy*ribo-*N*ucleic *A*cid (DNA). The nucleotides link the two

remaining OH groups on the ribose ring and these are called the 3′- and 5′-positions. This piece of DNA has three nucleotides (adenine, adenine, and thymine) and so would be called –AAT– for short.
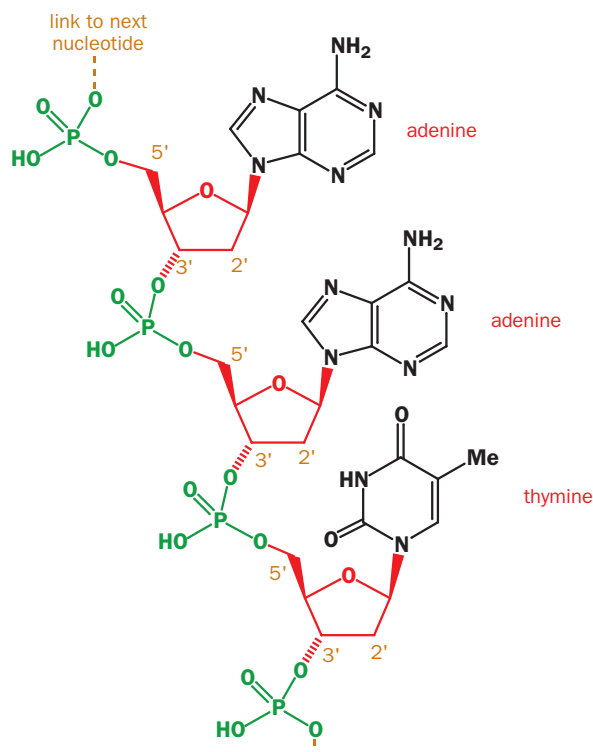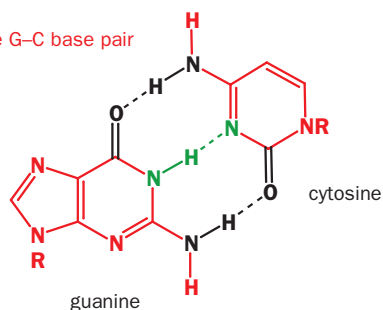
Each polymeric strand of DNA coils up into a helix and is bonded to another strand by hydrogen bonds between the bases. Each base pairs up specifically with another base —adenine with thymine (A–T) and guanine with cytosine (G–C)—like this.
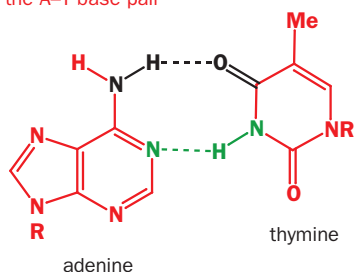
There is quite a lot to notice about these structures. Each purine (A or G) is bonded specifically to one pyrimidine (T or C) by two or by three hydrogen bonds. The hydrogen bonds are of two kinds: one links an amine to a carbonyl group (black in the diagram) and one links an amine to an imine (green in the diagram). In this way, each nucleotide reliably recognizes another and reliably pairs with its partner. The short strand of DNA above (–AAT–) would pair reliably with –TTA–.

## How the genetic information in DNA is passed to proteins

In the normal structure of DNA each strand is paired with another strand called the **complementary strand** because it has each base paired with its complementary base. When DNA replicates, the strands separate and a new strand with complementary structure grows alongside each. In this way the original double helix now becomes two identical double helices and so on.
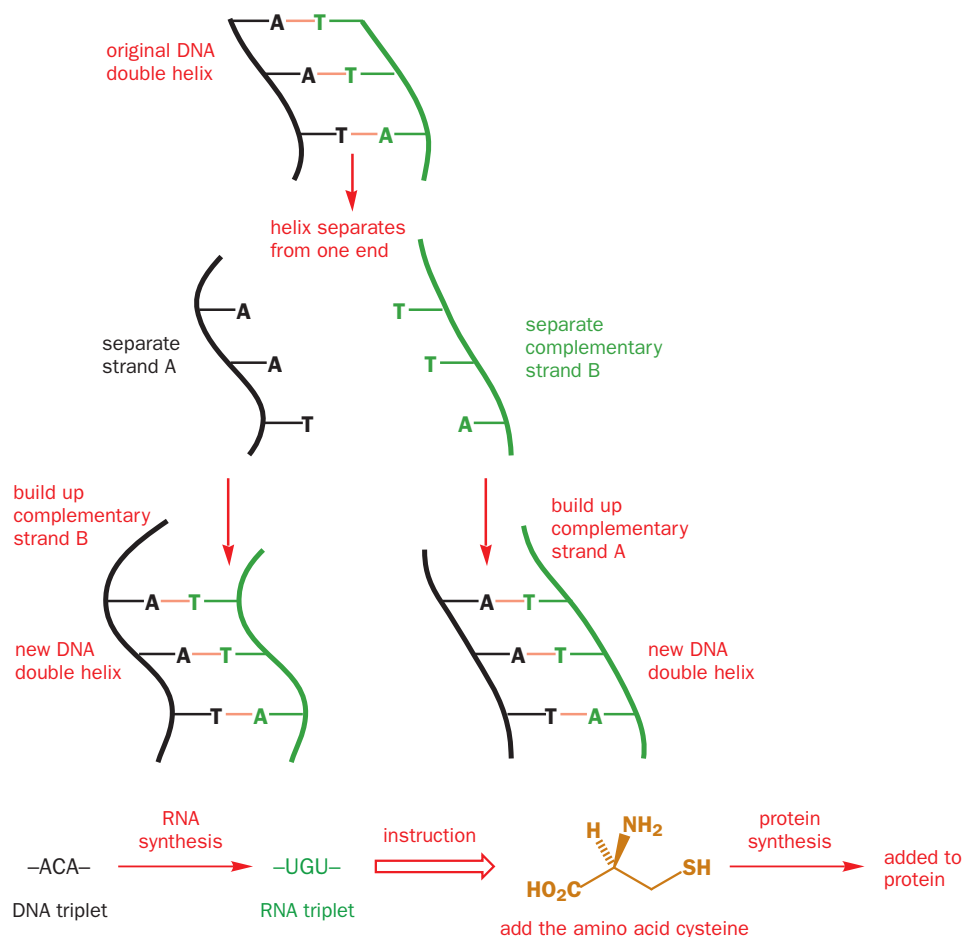
This is a crude simplification of a beautiful process and you should turn to a biochemistry text-book for more details. The actual building up of a strand of DNA obviously involves a complex series of chemical reactions. The DNA is then used to build up a complementary strand of RNA, which *does* have the 2′ hydroxyl group, and the RNA then instructs the cell on protein synthesis using three-nucleotide codes to indicate different amino acids. Again, the details of this process are beyond the scope of this book, but the code is not.

Each set of three nucleotides (called a **triplet** or **codon**) in a DNA molecule tells the cell to do something. Some triplets tell it to start work or stop work but most represent a specific amino acid. The code UGU in RNA tells the cell 'add a molecule of cysteine to the protein you are building'. The code UGA tells the cell 'stop the protein at this point'. So a bit of RNA reading UGUUGA would produce a protein with a molecule of cysteine at the end.
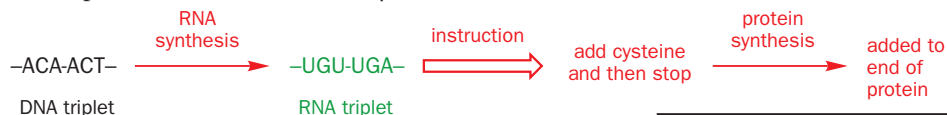


There are four bases available for DNA and so there are $4^3 = 64$ different triplet codons using three bases in each codon. There are only 20 amino acids used in proteins so that gives plenty of spare codons. In fact 61 of the 64 are used as codons for amino acids and the remaining three are 'stop' signals. Thus the code ATT in DNA would produce the complementary UAA and this is another 'stop' signal.
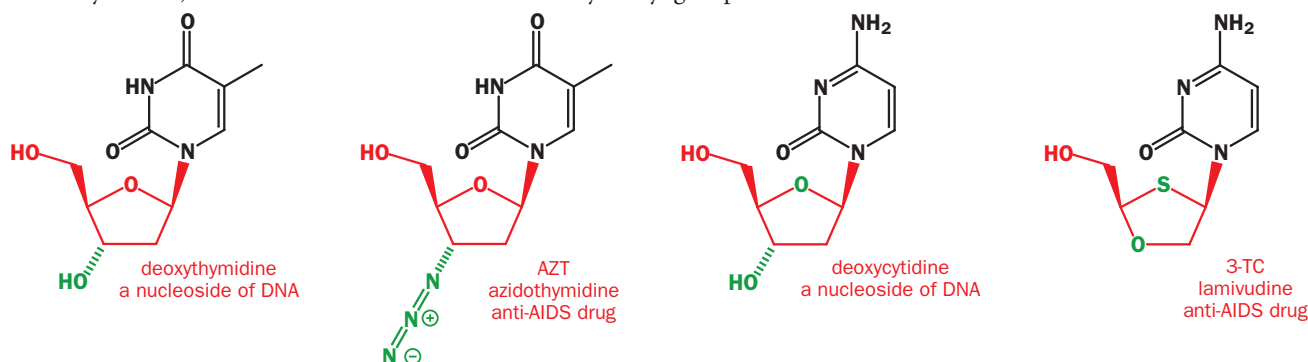
| Base | Complementary base in DNA | Complementary base in RNA |
|---|---|---|
| A | T | U |
| C | G | G |
| G | C | C |
| U | * | * |
| T | A | a |

* T occurs in DNA only and is replaced by U in RNA.

But that doesn't leave a 'start' signal! This signal is the same (TAC in DNA = AUG in RNA) as that for the amino acid methionine, which you met as a component of SAM, the biological methylating agent. In other words, all proteins start with methionine. At least, they are all made that way, though the methionine is sometimes removed by enzymes before the protein is released. These code letters are the same for all living things except for some minor variations in some microorganisms.

## AIDS is being treated with modified nucleosides

Modified nucleosides have proved to be among the best antiviral compounds. The most famous anti-AIDS drug, AZT (zidovudine from GlaxoWellcome), is a slightly modified DNA nucleoside (3′-azidothymidine). It has an azide at C3′ instead of the hydroxyl group in the natural nucleoside.



deoxythymidine
a nucleoside of DNA

AZT
azidothymidine
anti-AIDS drug

deoxycytidine
a nucleoside of DNA

3-TC
lamivudine
anti-AIDS drug

Doctors are having some spectacular success at the moment (1999) against HIV and AIDS by using a combination of AZT and a much more modified nucleoside 3-TC (lamivudine) which is active against AZT-resistant viruses. This drug is based on cytosine but the sugar has been replaced by a different heterocycle though it is recognizably similar especially in the stereochemistry.

The last drug to mention is acyclovir (Zovirax), the cold sore (herpes) treatment. Here is a modified guanosine in which only a ghost of the sugar remains. There is no ring at all and no stereochemistry.

The bottom edge of the sugar ring has been done away with so that a simple alkyl chain remains. This compound has proved amazingly successful as an antiviral agent and it is highly likely that more modified nucleosides will appear in the future as important drugs.



deoxyguanosine
a nucleoside of DNA

acyclovir
(Zovirax)
anti-herpes drug
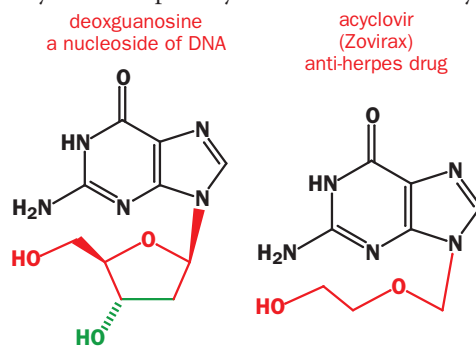
## Cyclic nucleosides and stereochemistry

We know the relative stereochemistry around the ribose ring of the nucleosides in DNA and RNA because the bases can be persuaded to cyclize on to the ring in certain reactions. Treatment of deoxythymidine with reagents that make oxygen atoms into leaving groups leads to cyclization by intramolecular $S_N2$ reaction. The amide oxygen of the base attacks the 3′-position in the sugar ring.



deoxythymidine

cyclization with inversion

cyclic nucleoside
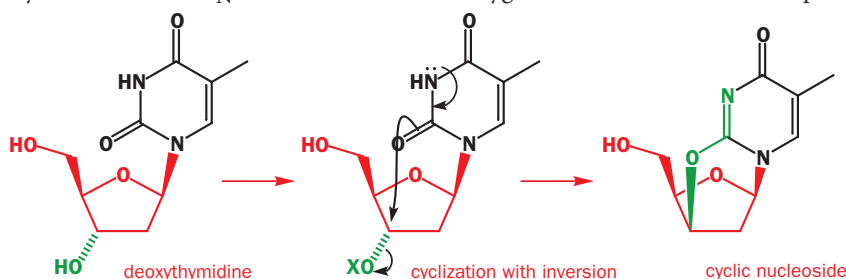
This $S_N2$ reaction has to happen with inversion, proving that the base and the 3′-OH group are on opposite sides of the ribose ring. The cyclized product is useful too. If it is reacted with azide ion the ring reopens with inversion in another $S_N2$ reaction and AZT is formed.
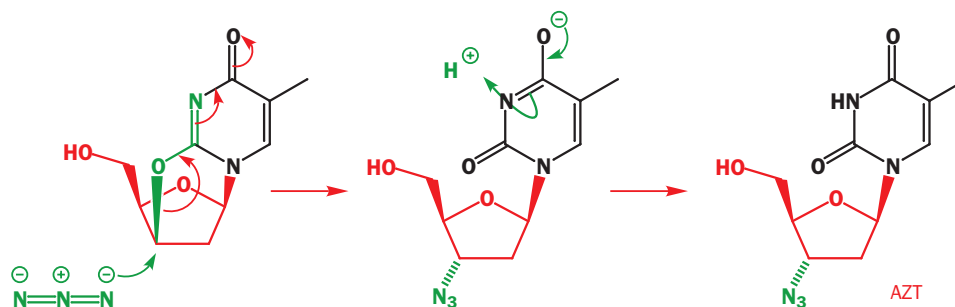
We can show that the primary alcohol is on the same side of the ring as the base by another cyclization reaction. Treatment of the related iodide with a silver(I) salt gives a new seven-membered ring. This reaction can happen only with this stereochemistry of starting material.
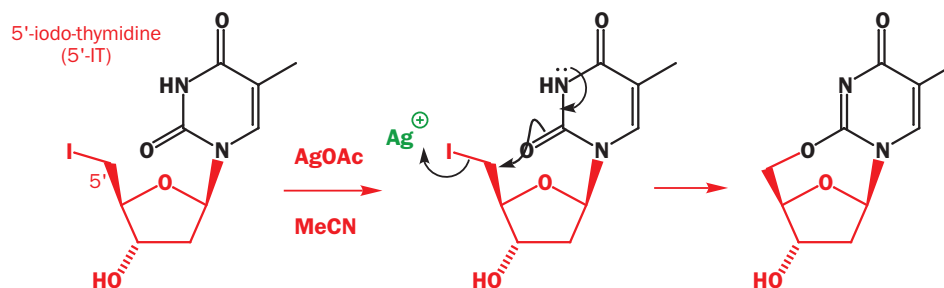
In ribonucleic acids, the fact that the 2′- and 3′-OH groups are on the same side of the ring makes alkaline hydrolysis of such dinucleotides exceptionally rapid by intramolecular nucleophilic catalysis.

■
The substituents $B^1$ and $B^2$ represent any purine or pyrimidine base.
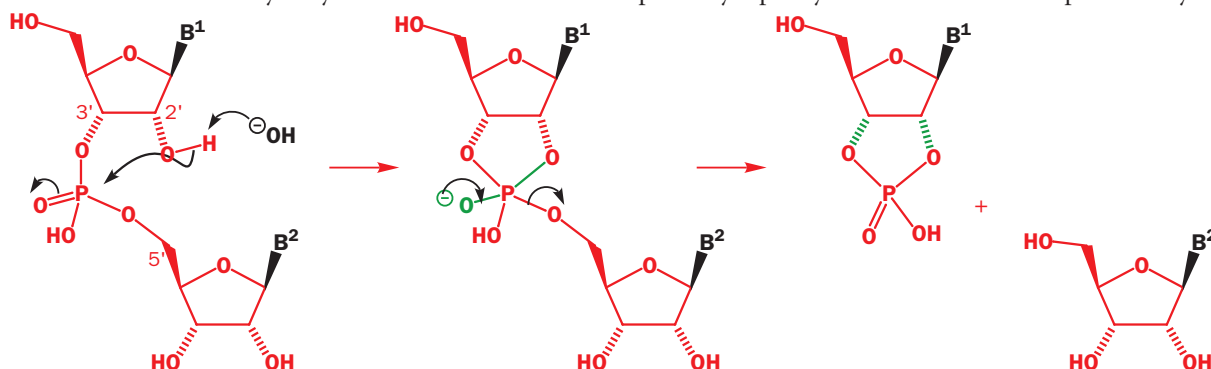
The alkali removes a proton from the 2′-OH group, which cyclizes on to the phosphate link—possible only if the ring fusion is *cis*. The next reaction involves breakdown of the pentacovalent phosphorus intermediate to give a cyclic phosphate. One nucleoside is released by this reaction and the second follows when the cyclic phosphate is itself cleaved by alkali.

The simplest cyclic phosphate that can be formed from a nucleotide is also important biologically as it is a messenger that helps to control such processes as blood clotting and acid secretion in the stomach. It is **cyclic AMP** (**cAMP**), formed enzymatically from ATP by nucleophilic displacement of pyrophosphate by the 3′-OH group.

►
Note that cAMP has a trans 6,5-fused ring junction.
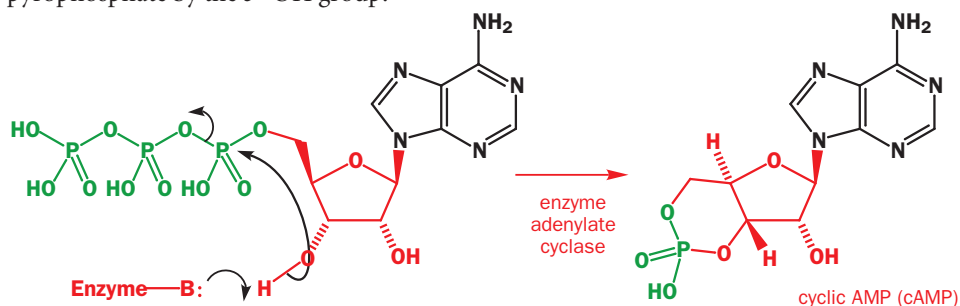
# Proteins are made of amino acids

The molecule of methionine, which we met as a component of SAM, is a typical amino acid of the kind present in proteins. It is the starter unit in all proteins and is joined to the next amino acid by an amide bond. In general, we could write:
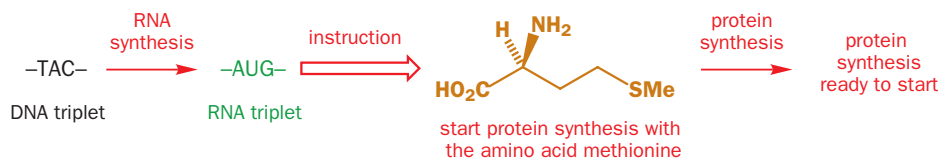


Now we can add the next amino acid using its correct codon, but we want to show the process in general so we shall use the general structure in the margin. All amino acids have the same basic structure and differ only in the group 'R'. Both structures are the same and have the same (S) stereochemistry.

*two views of the general amino acid structure*
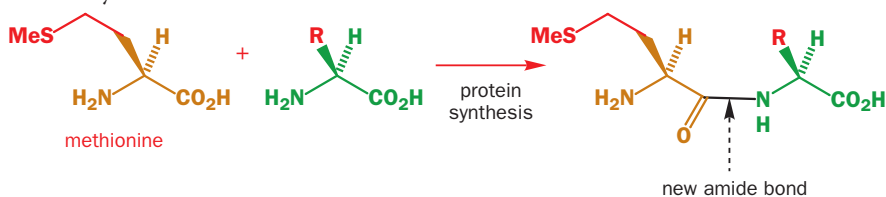




The process then continues with more amino acids added in turn to the right-hand end of the growing molecule. A section of the final protein drawn in a more realistic conformation might look like this.
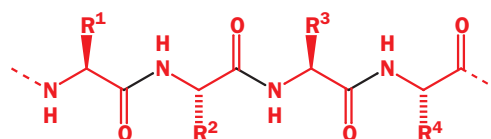


The basic skeleton of the protein zig-zags up and down in the usual way; the amide bonds (shown in black) are rigid because of the amide conjugation and are held in the shape shown. Each amino acid may have a different substituent ($R^1$, $R^2$, $R^3$, etc.) or some may be the same.

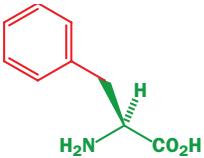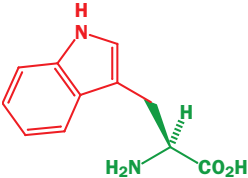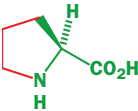## A catalogue of the amino acids

So what groups are available when proteins are being made? The simplest amino acid, glycine, has no substituents except hydrogen and is the only amino acid that is not chiral. Four other amino acids have alkyl groups without further functionality. The next table gives their structures together with two abbreviations widely used for them. The three-letter code (which has nothing to do with the codon in DNA!) is almost self-explanatory as are the one-letter codes in this group, but some of the one-letter codes for the other amino acids are not so obvious.

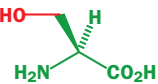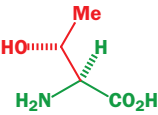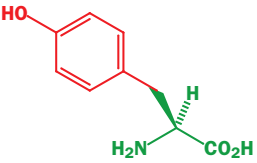| Name | Three-letter code | One-letter code | Structure |
|---|---|---|---|
| glycine | Gly | G |  |
| alanine | Ala | A |  |
| valine | Val | V |  |
| leucine | Leu | L |  |
| isoleucine | Ile | I |  |
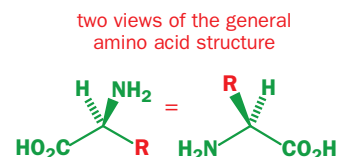
► Many of the compounds we discuss in this chapter will be salts under biological conditions. Most carboxylic acids will exist as anions, as will the phosphates you have just seen, and most amines as cations as they would be protonated at pH 7. Amino acids exist in biological systems as zwitterions. For simplicity, we will usually draw functional groups in the simplest and most familiar way, leaving the question of protonation to be addressed separately if required.

These amino acids form hydrophobic (water-repelling) nonpolar regions in proteins. There are three more of this kind with special roles. Phenylalanine and tryptophan have aromatic rings and, though they are still hydrophobic, they can form attractive π-stacking interactions with other aromatic molecules. Enzyme-catalysed hydrolysis of proteins often happens next to one of these residues. Proline is very special. It has its amino group inside a ring and has a different shape from all the other amino acids. It appears in proteins where a bend or a twist in the structure is needed.
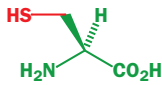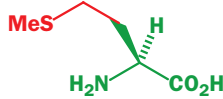
| Name | Three-letter code | One-letter code | Structure |
|------|-------------------|-----------------|-----------|
| phenylalanine | Phe | F |  |
| tryptophan | Trp | W |  |
| proline | Pro | P |  |

The rest of the amino acids have functional groups of various kinds and we shall deal with them by function. The simplest have hydroxyl groups and there are three of them—two alcohols and a phenol. Serine in particular is important as a reactive group in enzymatic reactions. It is a good nucleophile for carbonyl groups.

| Name | Three-letter code | One-letter code | Structure |
|------|-------------------|-----------------|-----------|
| serine | Ser | S |  |
| threonine | Thr | T |  |
| tyrosine | Tyr | Y |  |



disulfide cross-link
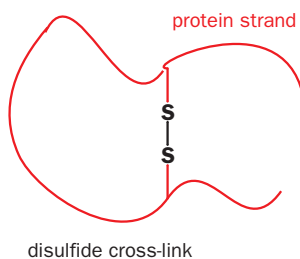
Next come the two compounds we have already met, the sulfur-containing cysteine and methionine. Cysteine has a thiol group and methionine a sulfide. These are very important in protein structure—methionine starts off the synthesis of every new protein as its *N*-terminal amino acid, while cysteine forms S–S bridges linking two parts of a protein together. These disulfide links may be important in holding the three-dimensional shape of the molecule.

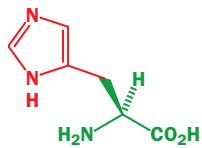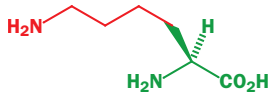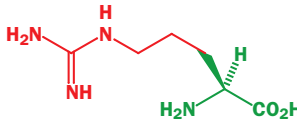| Name | Three-letter code | One-letter code | Structure |
|---|---|---|---|
| cysteine | Cys | C |  |
| methionine | Met | M |  |

## Cysteine and hairdressing

Thiols (RSH) are easily oxidized, by air, for example, to disulfides (RS–SR). This chemistry of cysteine is used by hairdressers to give 'perms' or permanent waves. The hair proteins are first reduced so that any disulfide (cysteine to cysteine) cross-links within each strand are reduced to thiols. Then the hair is styled and the final stage is the 'set' when the hair is oxidized so that disulfide cross-links are established to hold its shape for a good time. The disulfide resulting from cross-links between the thiol groups of *cysteine* is known as *cystine*—beware of confusing the names!
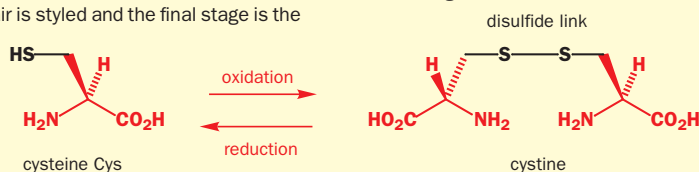


The amino acids with a second amino group are important because of their basicity and they are vital to the catalytic activity of many enzymes. Histidine has a $pK_{aH}$ very close to neutrality (6.5) and can function as an acid or a base. Lysine and arginine are much more basic, but are normally protonated in living things. An extra column in this table gives the $pK_{aH}$ of the extra amino groups.
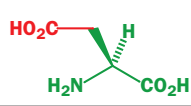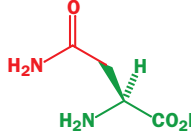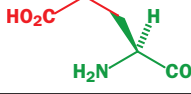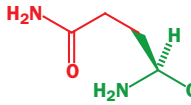
| Name | Three-letter code | One-letter code | $pK_{aH}$ | Structure |
|---|---|---|---|---|
| histidine | His | H | 6.5 |  |
| lysine | Lys | K | 10.0 |  |
| arginine | Arg | R | 12.0 |  |

## Essential amino acids

If you saw 'Jurassic Park' you may recall that the failsafe device was the 'lysine option'. The dinosaurs were genetically modified so as to need lysine in their diet. The idea was that they would die unless lysine was provided by their keepers. Lysine was a good choice as it is one of the 'essential' amino acids for humans. If we are not given it in our diet, we die. Of course, any normal diet, including the human beings eaten by the escaped dinosaurs, would also contain plenty of lysine. The other essential amino acids (for humans) are His, Ile, Leu, Met, Phe, Thr, Try, and Val.

Finally, we come to the acidic amino acids—those with an extra carboxylic acid group. We are going to include their amides too as they also occur in proteins. This group is again very much involved in the catalytic activity of enzymes. The two acids have $pK_a$s for the extra $CO_2H$ group of about 4.5.

| Name | Three-letter code | One-letter code | Structure |
|------|-------------------|-----------------|-----------|
| aspartic acid | Asp | D |  |
| asparagine | Asn | N |  |
| glutamic acid | Glu | E |  |
| glutamine | Gln | Q |  |

Sometimes it is not known whether the acids or their amides are present and sometimes they are present interchangeably. Aspartic acid or asparagine has the codes Asx and B while glutamic acid or glutamine is Glx or Z.

Now perhaps you can see that a protein is an assembly of many different kinds of group attached to a polyamide backbone. Some of the groups are purely structural, some control the shape of the protein, some help to bind other molecules, and some are active in chemical reactions.

Most amino acids are readily available to chemists. If proteins are hydrolysed with, say, concentrated HCl, they are broken down into their amino acids. This mixture is tricky to separate, but the acidic ones are easy to extract with base while the aromatic ones crystallize out easily.

### Amino acids combine to form peptides and proteins

In nature, the amino acids are combined to give proteins with hundreds or even thousands of amino acids in each one. Small assemblies of amino acids are known as **peptides** and the amide bond that links them is called a **peptide bond**. One important dipeptide is the sweetening agent aspartame, whose synthesis was discussed in Chapter 25. It is composed (and made) of the amino acid aspartic acid (Asp) and the methyl ester of phenylalanine. Only this enantiomer has a sweet taste and it is very sweet indeed—about 160 times as sweet as sucrose. Only a tiny amount is needed to sweeten drinks and so it is much less fattening than sucrose and is 'safe' because it is degraded in the body to Asp and Phe, which are there in larger amounts anyway.



aspartame



aspartic acid

phenylalanine methyl ester

An important tripeptide is **glutathione**. So important is this compound that it is present in almost all tissues of most living things. It is the 'universal thiol' that removes dangerous oxidizing agents by allowing itself to be oxidized to a disulfide.



glutathione = RSH

Glutathione is not quite a simple tripeptide. The left-hand amino acid is normal glutamic acid but it is joined to the next amino acid through its γ-CO$_2$H group instead of the more normal α-CO$_2$H group. The middle amino acid is the vital one for the function—cysteine with a free SH group. The *C*-terminal acid is glycine.



γ-Glu (glutamic acid joined through its γ-CO$_2$H group)

cysteine

glycine
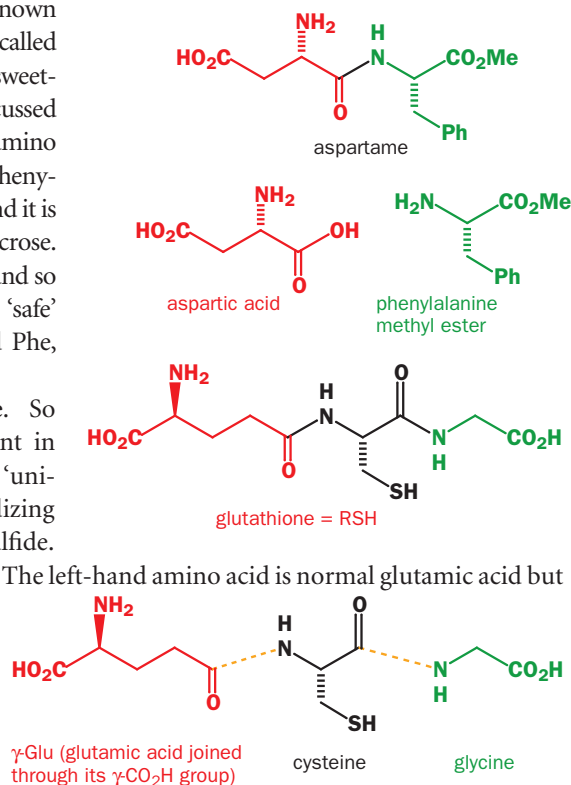
Thiols are easily oxidized to disulfides, as we have already seen in our discussion on hairdressing (though the redox chemistry of glutathione is a matter of life or death and not merely a bad hair day), and glutathione sacrifices itself if it meets an oxidizing agent. Later, the oxidized form of glutathione is reduced back to the thiol by reagents we shall meet in the next chapter (NADH, etc.).
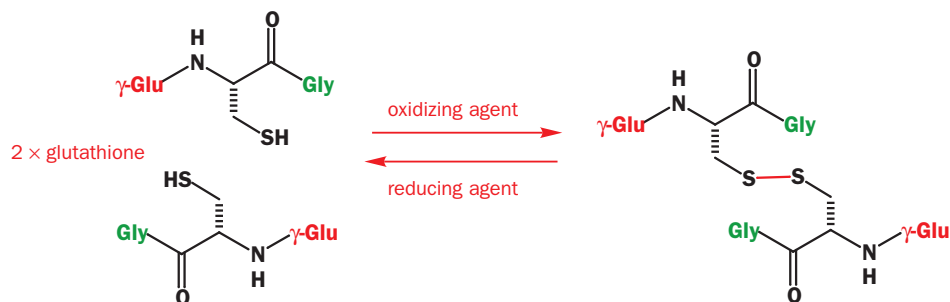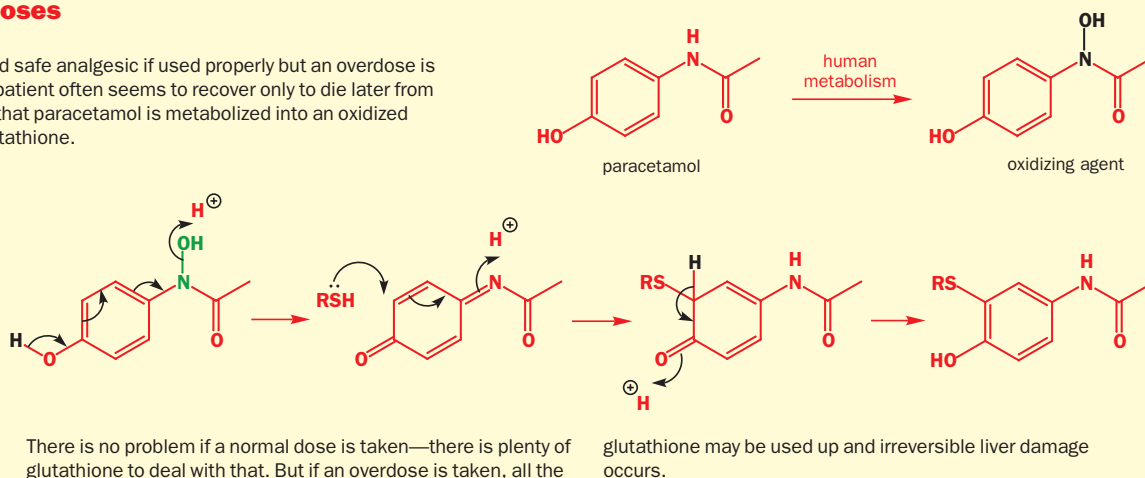


If we imagine that the stray oxidizing agent is a peroxide, say, $H_2O_2$, we can draw a mechanism to show how this can be reduced to water as glutathione (represented as RSH) is oxidized to a disulfide.
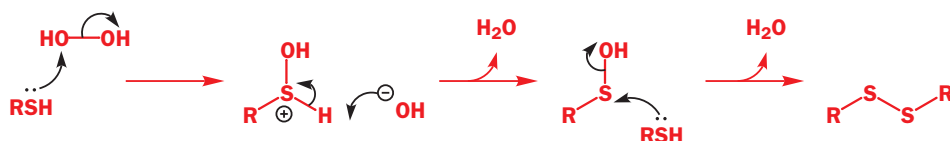


### Paracetamol overdoses

Paracetamol is a popular and safe analgesic if used properly but an overdose is insidiously dangerous. The patient often seems to recover only to die later from liver failure. The problem is that paracetamol is metabolized into an oxidized compound that destroys glutathione.



Glutathione detoxifies this oxidizing agent by a most unusual mechanism. The unstable hydroxylamine loses water to give a reactive quinone imine that is attacked by glutathione on the aromatic ring. The adduct is stable and safe but, for every molecule of paracetamol, one molecule of glutathione is consumed.



There is no problem if a normal dose is taken—there is plenty of glutathione to deal with that. But if an overdose is taken, all the glutathione may be used up and irreversible liver damage occurs.

Glutathione also detoxifies some of the compounds we have earlier described as very dangerous carcinogens such as Michael acceptors and 2,4-dinitrohalobenzenes. In both cases the thiol acts as a nucleophile for these electrophiles. Most of the time there is enough glutathione present in our cells to attack these poisons before they attack DNA or an enzyme.
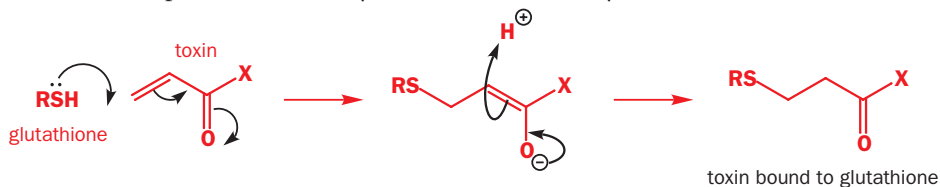


The toxin is now covalently bound to glutathione and so is no longer electrophilic. It is harmless and can be excreted. More glutathione will be synthesized from glutamic acid, cysteine, and glycine to replace that which is lost.

## Proteins are Nature's chemical laboratories

Longer peptides are called **proteins**, though where exactly the boundary occurs is difficult to say.

The structure of the hormone insulin (many diabetics lack this hormone and must inject themselves with it daily) was deduced in the 1950s by Sanger. It has two peptide chains, one of 21 amino acids and one of 30, linked by three disulfide bridges—just like the links in oxidized glutathione. This is a very small protein.

Enzymes are usually bigger. One of the smaller enzymes—ribonuclease (which hydrolyses RNA)—from cows—has a chain of 124 amino acids with four internal disulfide bridges. The abundance of the various amino acids in this enzyme is given in this table.

| Type | Amino acid (number)[a] | Total |
|---|---|---|
| structural | A (12), F (3), M (4), L (2), P (4), V (9), G (3), I (3) | 40 |
| cross-linking | C (8) | 8 |
| basic | K (10), R (4), H (4) | 18 |
| acids and amides | E (5), Q (7), D (5), N (10) | 27 |
| hydroxyl | T (10), S (15), Y (6) | 31 |

[a] See tables earlier in this section for one-letter codes of amino acids.

There are 48 structural and cross-linking amino acids concerned with the shape of the protein but over half of the amino acids have functional groups sticking out of the chain—amino, hydroxy, acid groups, and the like. In fact, the enzyme uses only a few of these functional groups in the reaction it catalyses (the hydrolysis of RNA)—probably only two histidines and one lysine—but it is typical of enzymes that they have a vast array of functional groups available for chemical reactions.

Below is part of the structure of ribonuclease surrounding one of the catalytic amino acids His12. There are seven amino acids in this sequence. Every one is different and every one has a functionalized side chain. This is part of a run of ten amino acids between Phe8 and Ala19. This strip of peptide has six different functional groups (two acids, one each of amide, guanidine, imidazole, sulfide, and alcohol) available for chemical reactions. Only the histidine is actually used.
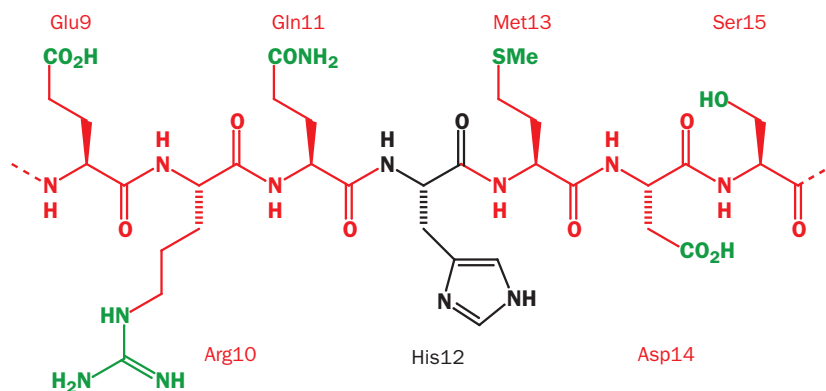
► Proteins are conventionally drawn and described with the amino (*N*) terminus to the left and the carboxyl (*C*) terminus to the right. This section of ribonuclease would be called 'glutamyl arginyl glutaminyl histidyl methionyl aspartyl seryl…' or, more briefly, –Glu–Arg–Gln–His–Met–Asp–Ser– or, more briefly still, –ERQHMDS–. The numbers on the diagram such as 'Glu9' tell us that this glutamic acid residue is number 9 from the *N*-terminus.



One reason for disease is that enzymes may become overactive and it may be necessary to design specific inhibitors for them to treat the disease. *A*ngiotensin-*C*onverting *E*nzyme (ACE) is a zinc-dependent enzyme that cleaves two amino acids off the end of angiotensin I to give angiotensin II, a protein that causes blood pressure to rise.



captopril

natural dipeptide (Ala–Pro)
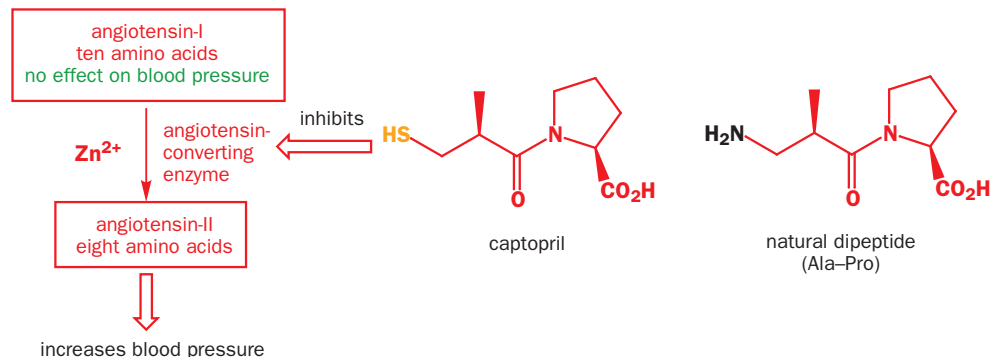
It is necessary in some situations for our blood pressure to rise (when we stand up for instance!) but too much too often is a very bad thing leading to heart attacks and strokes. Captopril is a treatment for high blood pressure called an 'ACE inhibitor' because it works by inhibiting the enzyme. It is a dipeptide mimic, having one natural amino acid and something else. The 'something else' is an SH group replacing the NH$_2$ group in the natural dipeptide. Captopril binds to the enzyme because it is *like* a natural dipeptide but it inhibits the enzyme because it is *not* a natural dipeptide. In particular, the SH group is a good ligand for Zn(II). Many people are alive today because of this simple deception practised on an enzyme.

### Structural proteins must be tough and flexible

In contrast with the functional enzymes, there are purely structural proteins such as collagen. Collagen is the tough protein of tendons and is present in skin, bone, and teeth. It contains large amounts of glycine (every third amino acid is glycine), proline, and hydroxyproline (again about a third of the amino acids are either Pro or Hyp).

In the enzyme above there were only three glycines and four prolines and no hydroxyproline at all. Hydroxyproline is a specialized amino acid that appears almost nowhere else and, along with proline, it establishes a very strong triply coiled structure for collagen. The glycine is necessary as there is no room in the inside of the triple coil for any larger amino acid. Functionalized amino acids are rare in collagen.


glycine
Gly; G


(*S*)-proline Pro; P


(2*S*,4*R*)-hydroxyproline Hyp
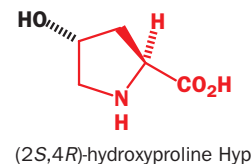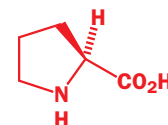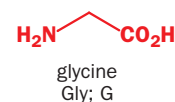
### Hydroxyproline and scurvy

Hydroxyproline is a very unusual amino acid. There is no genetic codon for the insertion of Hyp into a growing protein because collagen is not made that way. The collagen molecule is first assembled with Pro where Hyp ends up. Then some proline residues are oxidized to hydroxyproline. This oxidation requires vitamin C, and without it collagen cannot be formed. This is why vitamin C deficiency causes scurvy—the symptoms of scurvy (teeth falling out, sores, blisters) are caused by the inability to make collagen.

Proteins are enormously diverse in structure and function and we will be looking at a few of their reactions in the next chapter.

# Sugars—just energy sources?

Sugars are the building blocks of carbohydrates. They used to be thought of as essential but rather dull molecules whose only functions were the admittedly useful provision of energy and cell wall construction. We have already noted that ribose plays an intimate role in DNA and RNA structure and function. More recently, biochemists have realized that carbohydrates are much more exciting. They are often found in intimate association with proteins and are involved in recognition of one protein by another and in adhesion processes.

That may not sound very exciting, but take two examples. How does a sperm recognize the egg and penetrate its wall? The sperm actually binds to a carbohydrate on the wall of the egg in what was the first event in all of our lives. Then how does a virus get inside a cell? If it fails to do so, it has no life. Viruses depend on host cells to reproduce. Here again, the recognition process involves specific carbohydrates. One of the ways in which AIDS is being tackled with some success is by a combination of the antiviral drugs we met earlier in this chapter with HIV protease inhibitor drugs, which aim to prevent recognition and penetration of cells by HIV.

We now know that many vital activities as diverse as healing, blood clotting, infection, prevention of infection, and fertilization all involve carbohydrates. Mysterious compounds such as 'sialyl Lewis-X', unknown a few years ago, are now known to be vital to our health and happiness. Far from being dull, carbohydrates are exciting molecules and our future depends on them. It is well worthwhile to spend some time exploring their structure and chemistry.

## Sugars normally exist in cyclic forms with much stereochemistry
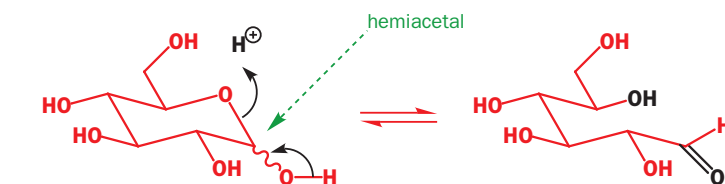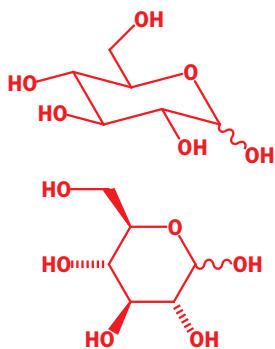
The most important sugar is glucose. It has a saturated six-membered ring containing oxygen and it is best drawn in a chair conformation with nearly all the substituents equatorial. It can also be drawn reasonably as a flat configurational diagram.

We have already met one sugar in this chapter, ribose, because it was part of the structure of nucleic acids. This sugar is a five-membered saturated oxygen heterocycle with many OH groups. Indeed, you can define a sugar as an oxygen heterocycle with every carbon atom bearing an oxygen-based functional group—usually OH, but alternatively C=O.

Both our drawings of glucose and ribose show a number of stereogenic centres and one centre undefined—the OH group is marked with a wavy line. This is because one centre in both sugars is a hemiacetal and therefore the molecule is in equilibrium with an open-chain hydroxy-aldehyde. For glucose, the open-chain form is this.



When the ring closes again, any of the OH groups could cyclize on to the aldehyde but there is no real competition—the six-membered ring is more stable than any of the alternatives (which could have three-, four-, five-, or seven-membered rings—check for yourself). However, with ribose there is a reasonable alternative.
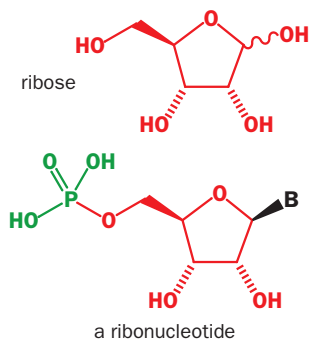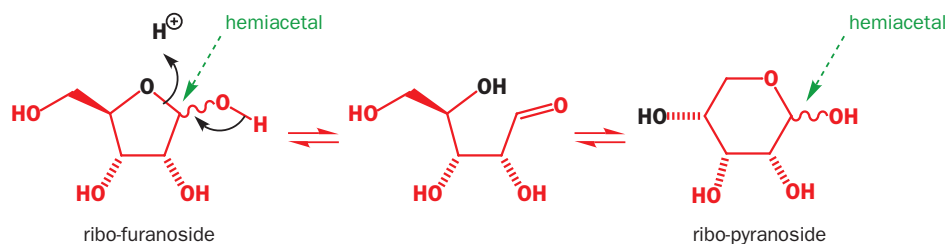


ribo-furanoside                    ribo-pyranoside

The most important sugars may exist in an open-chain form, as a five-membered oxygen heterocycle (called a furanoside after the aromatic furan) or a six-membered oxygen heterocycle (called a pyranoside after the compound pyran).

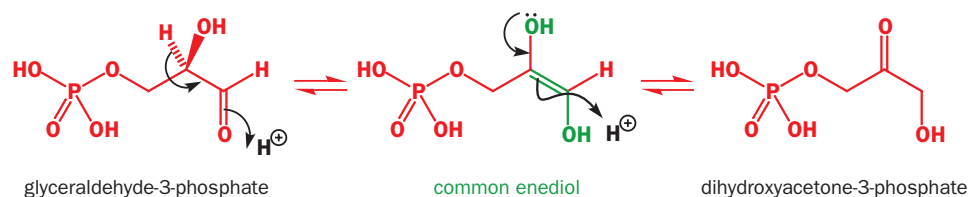## From triose to glucose requires doubling the number of carbon atoms

We will return to that in a moment, but let us start from the beginning. The simplest possible sugar is glyceraldehyde, a three-carbon sugar that cannot form a cyclic hemiacetal.

Glyceraldehyde is present in cells as its phosphate which is in equilibrium with dihydroxyacetone phosphate. This looks like a complicated rearrangement but it is actually very simple—the two compounds have a common enol through which they interconvert.



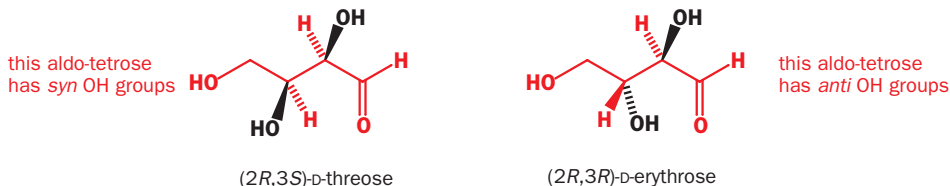glyceraldehyde-3-phosphate          common enediol          dihydroxyacetone-3-phosphate

Glyceraldehyde is an aldehyde sugar or **aldose** and dihydroxyacetone is a keto-sugar or **ketose**. That ending '-ose' just refers to a sugar. These two molecules combine to form the six-carbon sugar,

two representations of glucose



ribose

a ribonucleotide

furan          pyran

D-glyceraldehyde
(*R*)–(+)–glyceraldehyde

**fructose**, in living things and this reaction is a key step in the synthesis of organic compounds from $CO_2$ in plants.

When we come to the four-carbon sugars, or **tetroses**, two are important. They are diastereoisomers called erythrose and threose. You can see from this series that each aldose has $n-2$ stereogenic centres in its carbon chain where $n$ is the total number of carbon atoms in that chain.



this aldo-tetrose has *syn* OH groups

(2*R*,3*S*)-D-threose

this aldo-tetrose has *anti* OH groups

(2*R*,3*R*)-D-erythrose

We shall take a longer look at the stereochemistry and reactions of glucose and the important keto-hexose, fructose. These two are often found together in cells and are combined in the same molecule as **sucrose**—ordinary sugar. In this molecule, glucose appears as a pyranose (six-membered ring) and fructose as a furanose (five-membered ring). They are joined through an acetal at what were hemiacetal positions, and sucrose is a single diastereoisomer.



sucrose

acetal

acetal

fructose
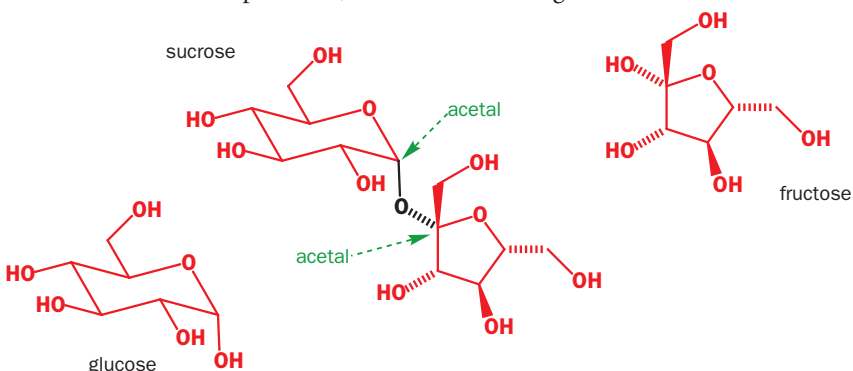
glucose

## Sugars can be fixed in one shape by acetal formation

This is the simplest way to fix glucose in the pyranose form—any alcohol, methanol, for example, gives an acetal and, remarkably, the acetal has an *axial* OR group.



glucose

MeOH

$H^\oplus$

OMe
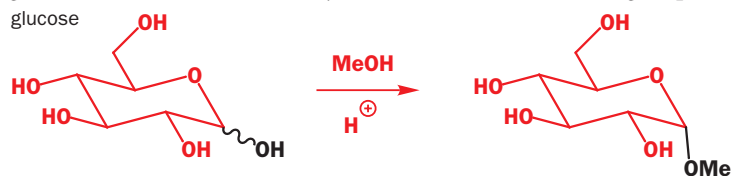
Acetal formation is under thermodynamic control (Chapter 14) so the axial compound must be the more stable. This is because of the **anomeric effect**—so called because this C atom is called the anomeric position and the acetal diastereoisomers are called anomers. The effect is a bonding interaction between the axial lone pair on the oxygen atom in the ring and the σ* orbital of the OMe group.

the anomeric effect

axial lone pair

bonding interaction

C–O σ*

axial anomer

OMe

equatorial anomer
no stabilization from anomeric effect
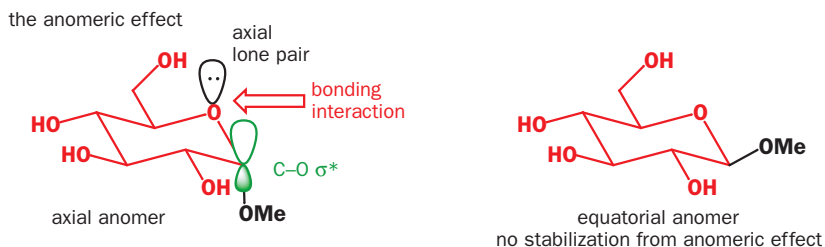
The formation of acetals allows a remarkable degree of control over the chemistry of sugars. Apart from the simple glucoside acetal we have just seen, there are three important acetals worth understanding because of the way in which they illustrate stereoelectronic effects—the interplay of

stereochemistry and mechanism. If we make an acetal from methyl glucoside, we get a single compound as a single stereoisomer.
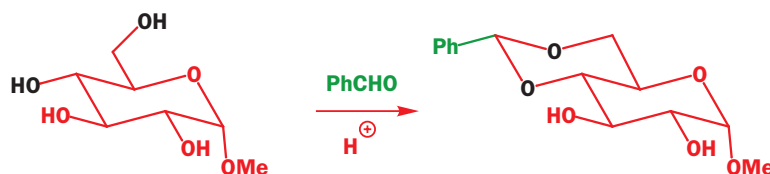


The new acetal could have been formed between any of the adjacent OH groups in the starting material but it chose the only pair (the black OH groups) to give a six-membered ring. The stereochemistry of glucose is such that the new six-membered ring is *trans*-fused on the old so that a beautifully stable all-chair bicyclic structure results, with the phenyl group in an equatorial position in the new chair acetal ring. It does not matter which OH group adds to benzaldehyde first because acetal formation is under thermodynamic control and this product is the most stable possible acetal.

Acetals formed from sugars and acetone have a quite different selectivity. For a start, cyclic acetals of acetone prefer to be five- rather than six-membered rings. In a six-membered ring, one of the acetone's methyl groups would have to be axial, so the five-membered ring is preferred. A 5/5 or 5/6 ring fusion is more stable if it is *cis*, and so acetone acetals ('acetonides') form preferentially from *cis* 1,2-diols. Glucose has no neighbouring *cis* hydroxyls in the pyranose form, but in the furanose form it can have two pairs. Formation of an acetal with acetone fixes glucose in the furanose form. This is all summarized in the scheme below.
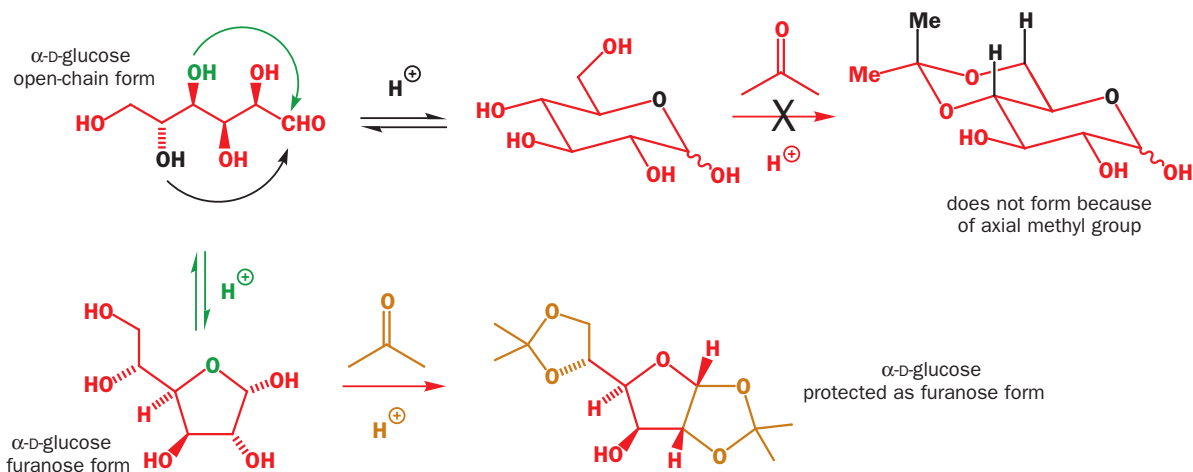


The open-chain form of glucose is in equilibrium with both pyranose and furanose forms by hemiacetal formation with the black and green OH groups, respectively. Normally, the pyranose form is preferred, but the furanose form can form a double acetal with acetone, one acetal having *cis*-fused 5/5 rings and the other being on the side chain. This is the product.

If we want to fix glucose in the open-chain form, we must make an 'acetal' of quite a different kind using a thiol (RSH) instead of an alcohol, an aldehyde, or a ketone.
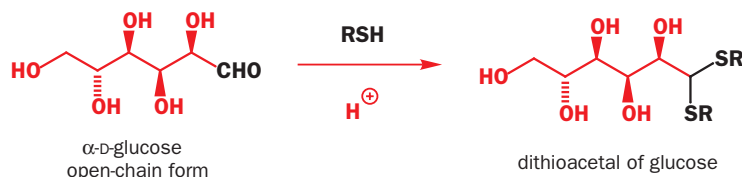


The thiol combines with the aldehyde group of the open-chain form to give a stable dithioacetal. The dithioacetal is evidently more stable than the alternative hemiacetals or monothioacetals that could be formed from the pyranose or furanose forms.

hemithioacetal    dithioacetal
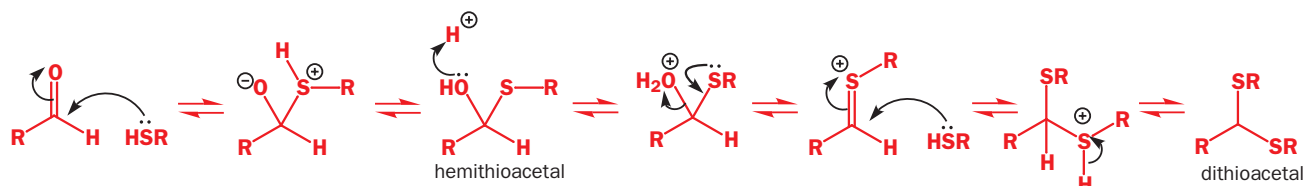
## Sugar alcohols are important in food chemistry

Another reaction of the open-chain form of sugars is reduction of the aldehyde group. This leads to a series of polyols having an OH group on each carbon atom. We will use **mannose** as an example. Mannose is a diastereoisomer of glucose having one axial OH group (marked in black) and, like glucose, is in equilibrium with the open-chain form.

If we redraw the open-chain form in a more realistic way, and then reduce it with $NaBH_4$, the product is mannitol whose symmetry is interesting. It has $C_2$ symmetry with the $C_2$ axis at right angles to the chain and marked with the orange dot.



mannose:
open-chain structure

mannitol



mannose:
pyranose structure

mannose:
open-chain structure
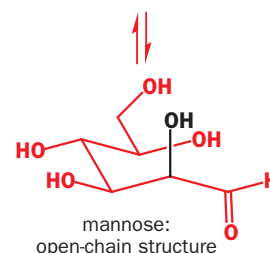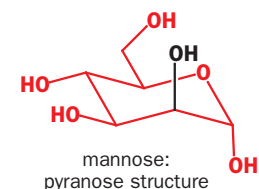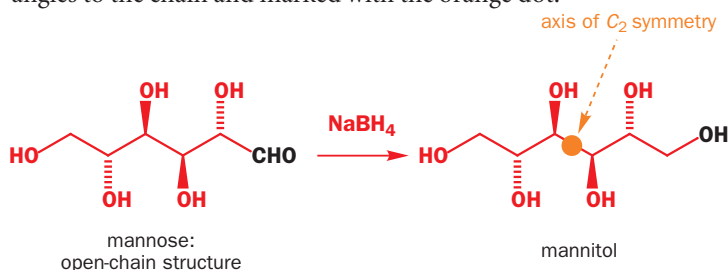
■ $C_2$ means that the axis of symmetry is twofold: rotating 180° gives the same structure.

The simplification of stereochemistry results because the two ends of the sugar both now have $CH_2OH$ groups so that the possibility of $C_2$ and planar symmetry arises. If we look at the two four-carbon sugars we can establish some important stereochemical correlations. Threose is reduced to threitol which has a $C_2$ axis like that of mannitol.



(2R,3S)-D-threose
furanose form

(2R,3S)-D-threose
open-chain form

(2R,3R)-threitol
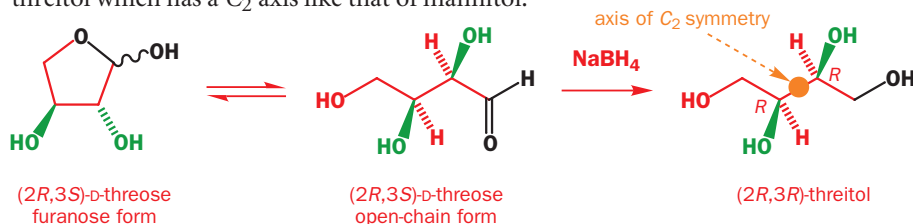
Erythrose on the other hand reduces to erythritol, which is not chiral.



(2R,3R)-D-erythrose
furanose form

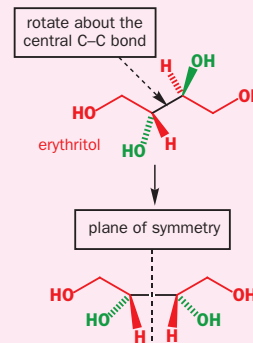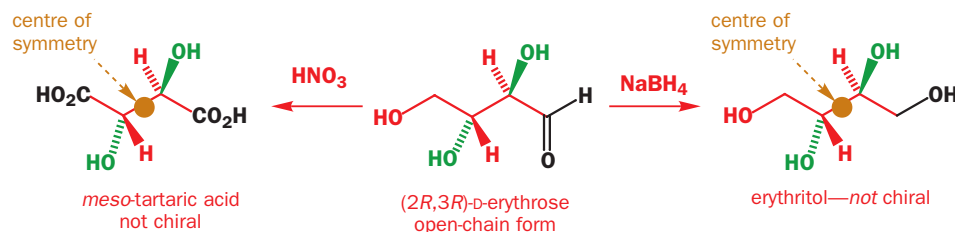(2R,3R)-D–erythrose
open-chain form

(2R,3S)-erythritol
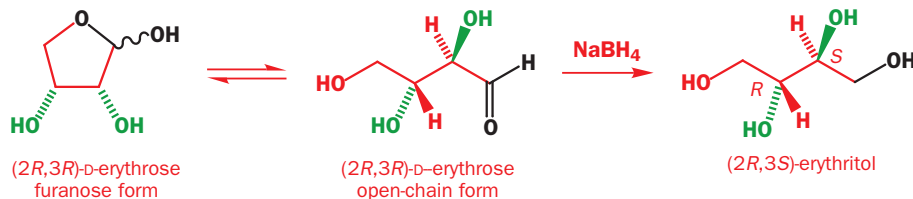
The important correlation is that threose is reduced or oxidized to chiral compounds—the oxidation product is tartaric acid—while erythrose is reduced or oxidized to *meso* compounds. This may help you to remember the labels **erythro-** and **threo-** should you need to.



(+)-tartaric acid

(2R,3S)-D-threose
open-chain form

threitol—chiral
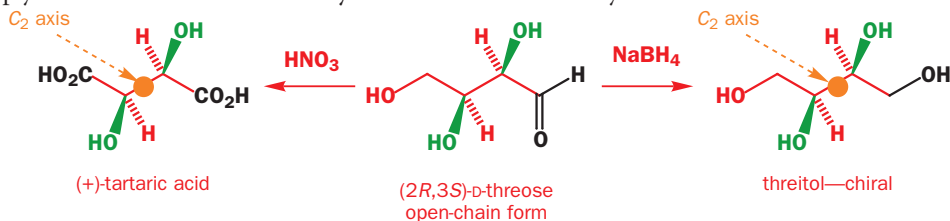
(continued overleaf)

▶ This may not be obvious in the normal drawing (which has a centre of symmetry), but rotation around the central C–C bond clearly shows the plane of symmetry. Neither plane nor centre of symmetry may be present in a chiral molecule, but a $C_2$ axis is allowed (Chapter 16).



rotate about the central C–C bond

erythritol

plane of symmetry

meso-tartaric acid
not chiral

(2R,3R)-D-erythrose
open-chain form

erythritol—*not* chiral

In the pentoses and hexoses there are again sugars that are reduced to *meso* alcohols and some that are reduced to $C_2$ symmetric alcohols. The $C_5$ sugar xylose has the same stereochemistry as glucose from C2 to C4 but lacks the $CH_2OH$ group at C6.



xylose:
pyranose structure

xylose:
open chain-structure
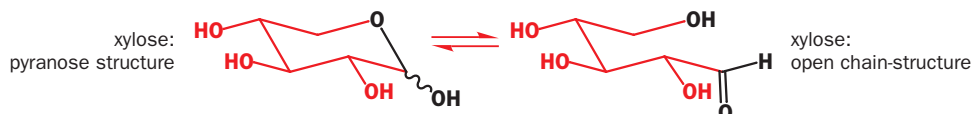
Xylose is reduced to the *meso* alcohol xylitol. This alcohol is more or less as sweet as sugar and, as xylose (which is not sweet) can be extracted in large quantities from waste products such as sawdust or corncobs, xylitol is used as a sweetener in foods. There is an advantage in this. Though we can digest xylitol (so it is fattening), the bacteria on teeth cannot so that xylitol does not cause tooth decay.



xylose:
open-chain structure

xylitol:
achiral
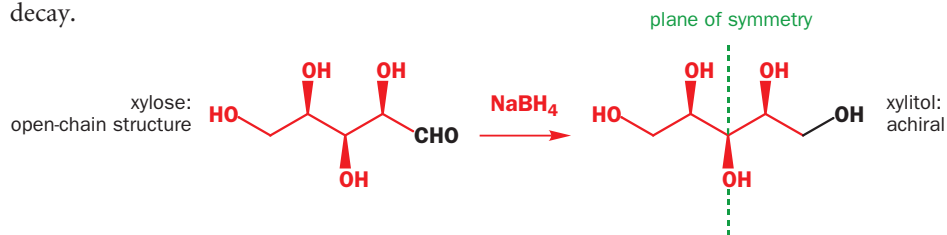
By careful manipulation of protecting groups such as acetals and reactions such as reduction and oxidation, it is possible to transform sugars into many different organic compounds retaining the natural optical activity of the sugars themselves. As some sugars are also very cheap, they are ideal starting points for the synthesis of other compounds and are widely used in this way (Chapter 45). Sucrose and glucose are very cheap indeed—probably the cheapest optically active compounds available. Here are the relative (to glucose = 1) prices of some other cheap sugars.

| Sugar | Price[a] | Sugar | Price[a] |
|---|---|---|---|
| glucose | 1 | sorbitol | 2 |
| mannose | 75 | mannitol | 4 |
| galactose | 8 | dulcitol[b] | 70 |
| xylose | 20 | xylitol | 15 |
| ribose | 100 | sucrose | 1 |

[a] Prices relative to glucose = 1.
[b] Dulcitol is the reduction product of galactose.

## Chemistry of ribose—from sugars to nucleotides

We have said little about selective reactions of pentoses so we shall turn now to the synthesis of nucleotides such as AMP. In nature, ribose is phosphorylated on the primary alcohol to give ribose-5-phosphate. This is, of course, an enzyme-catalysed reaction but it shows straightforward chemoselectivity such as we should expect from a chemical reaction.



ribose

ribose-5-phosphate

5-phosphoribosyl-1-pyrophosphate (PRPP)
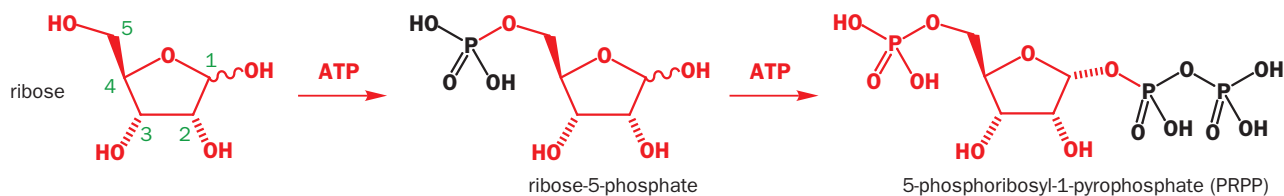
The second step is a pyrophosphorylation at the anomeric position to give PRPP. Only one diastereoisomer is produced so presumably the two anomers interconvert rapidly and only the one isomer reacts under control by the enzyme. This selectivity would be very difficult to achieve chemically.
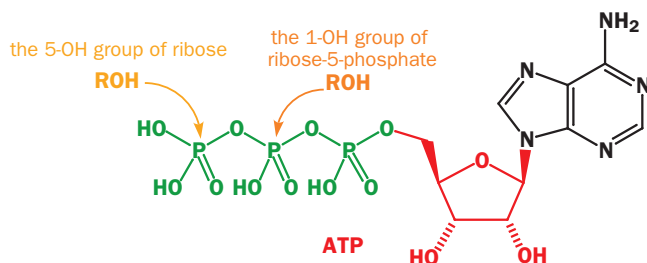
Now the stage is set for an $S_N2$ reaction. The nucleophile is actually the amide group of glutamine but the amide is hydrolysed by the same enzyme in the same reaction and the result is as if a molecule of ammonia had done an $S_N2$ reaction displacing the pyrophosphate from the anomeric position. An $NH_2$ group is introduced, which is then built into the purine ring-system in a series of reactions involving simple amino acids. These reactions are too complex to describe here.
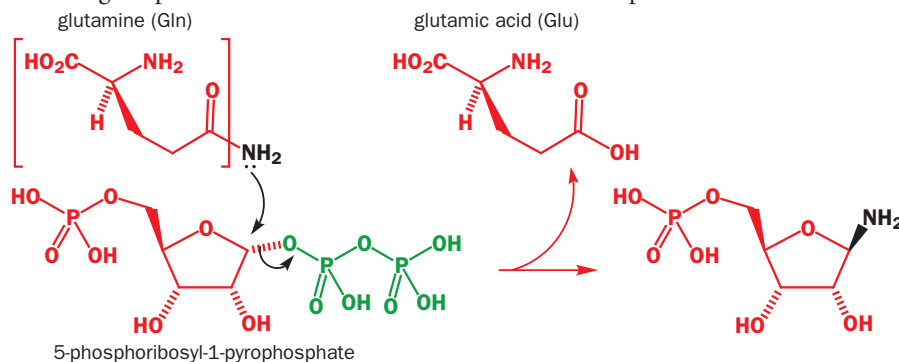


By contrast, if a pyrimidine is to be made, Nature assembles a general pyrimidine structure first and adds it in one step to the PRPP molecule, again in an $S_N2$ reaction using a nitrogen nucleophile. This general nucleotide, orotidylic acid, can be converted into the other pyrimidine nucleotides by simple chemistry.
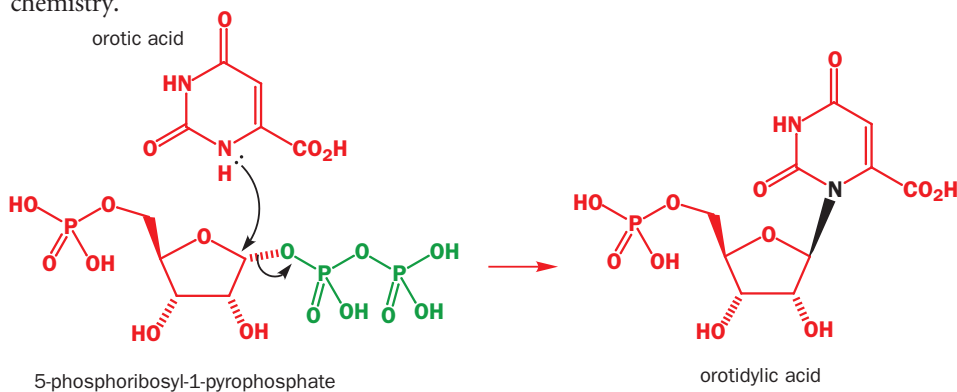


## The chemical version—protection all the way

In a chemical synthesis (work that led to Alexander (Lord) Todd's Nobel prize) there are rather different problems. We cannot achieve the remarkable selectivity between the different OH groups achieved in Nature so we have to protect any OH group that is not supposed to react. We also prefer to add pre-formed purines and pyrimidines to a general electrophile derived from ribose. The first step is to form acetate esters from all the OH groups. Since ribose is rather unstable to acetylation conditions, the methyl glycoside (which is formed under very mild conditions) is used. This fixes the sugar in the furanose form. Now the tetraacetate can be made using acetic anhydride in acidic solution. All of the OH groups react by nucleophilic attack on the carbonyl group of the anhydride with retention of configuration except for the anomeric OH, which esterifies by an $S_N1$ mechanism. This, of course, epimerizes the anomeric centre but the crystalline diastereoisomer shown can be isolated easily.
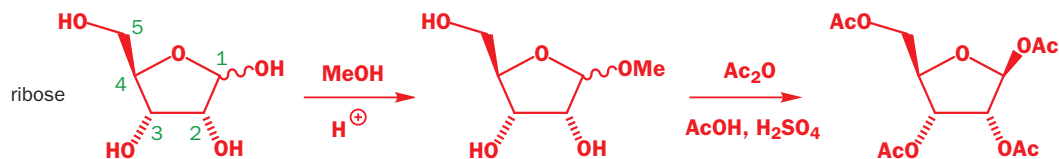
Now the anomeric centre can be activated towards nucleophilic attack by replacement of acetate by chloride. This is again an $S_N1$ reaction and produces a mixture of chlorides. The other esters are stable to these conditions.
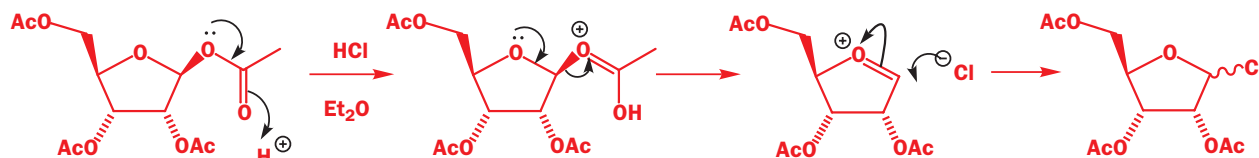


Replacement of the chlorine by the purine or pyrimidine base is sometimes quite tricky and silver or silyl derivatives are often used. Lewis acid catalysis is necessary to help the chloride ion leave in this $S_N1$ reaction. We shall avoid detailed technical discussion and simply draw the adenosine product from a general reaction.
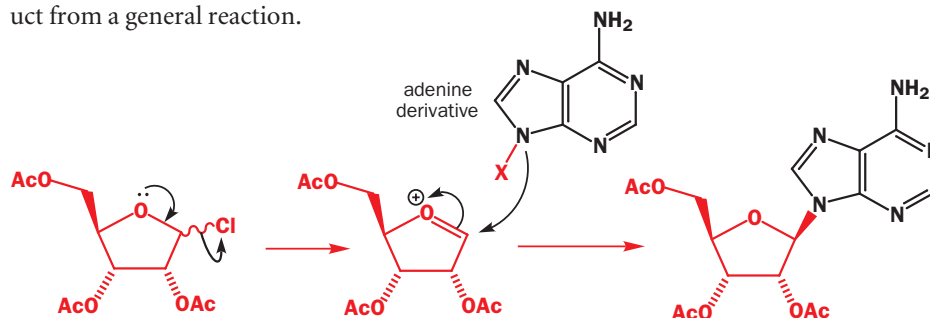


Now we need to remove the acetates and put a phosphate specifically on the 5-position. The acetates can be removed with retention by ester hydrolysis and we already know how to protect the 2-OH and 3-OH groups. They are *cis* to each other so they will form an acetal with acetone leaving the 5-OH group free.
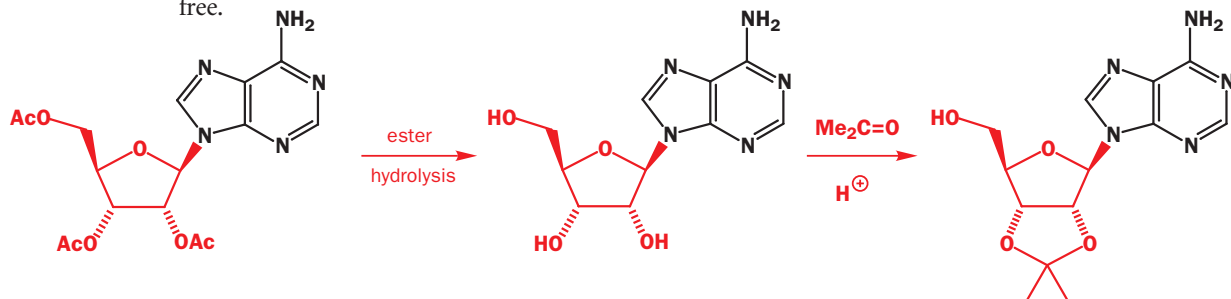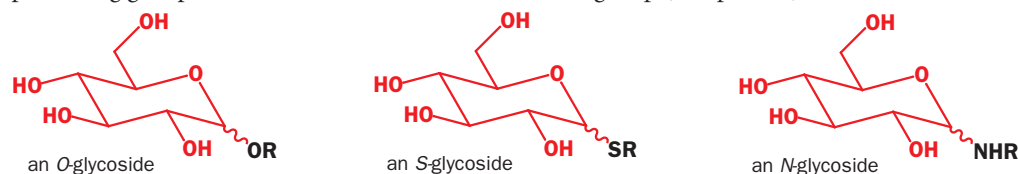


Putting on the phosphate is tricky too and more protection is necessary. This phosphorus compound with one chloride as leaving group and two benzyl esters as protecting groups proved ideal. The benzyl esters can be removed by hydrogenation (Chapter 24) and the acetal by treatment with dilute acid to give AMP.
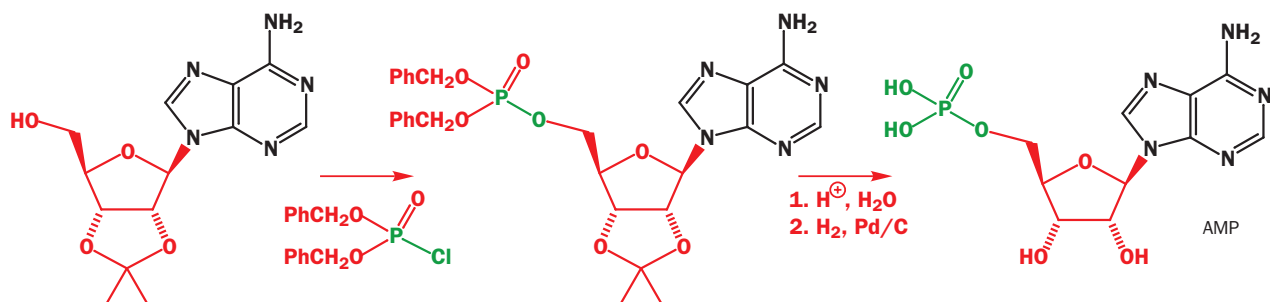
The chemical synthesis involves a lot more selective manipulation of functional groups, particularly by protection, than is necessary in the biological synthesis. However, this synthesis paved the way to the simple syntheses of nucleotides and polynucleotides carried out routinely nowadays. The usual method is to build short runs of nucleotides and then let the enzymes copy them—a real partnership between biology and chemistry.
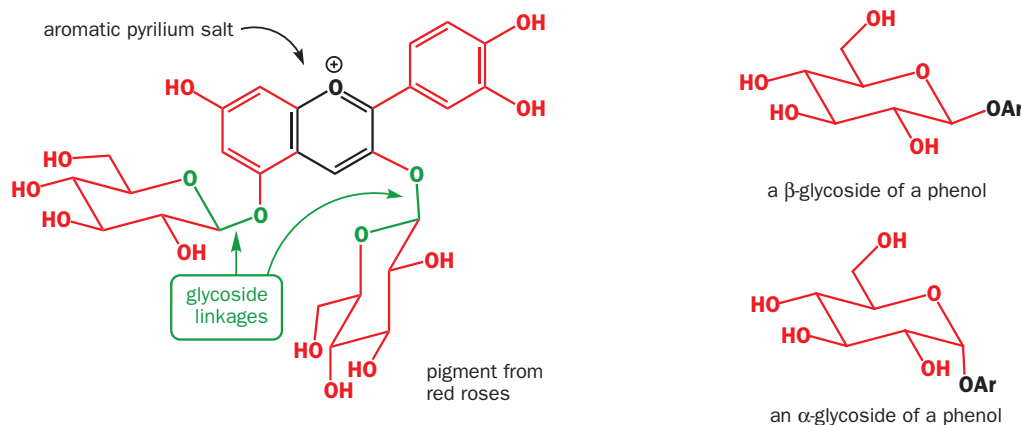
# Glycosides are everywhere in nature

Many alcohols, thiols, and amines occur in nature as **glycosides**, that is as *O*-, *S*-, or *N*-acetals at the anomeric position of glucose. The purpose of attaching these compounds to glucose is often to improve solubility or transport across membranes—to expel a toxin from the cell, for example. Sometimes glucose is attached in order to stabilize the compound so that glucose appears as Nature's protecting group, rather as a chemist would use a THP group (Chapter 24).

■ The most important *N*-glycosides are, of course, the nucleotides and we have already described them in some detail.



an *O*-glycoside     an *S*-glycoside     an *N*-glycoside

*O*-Glycosides occur in immense variety with glucose and other sugars being joined to the OH groups of alcohols and phenols to form acetals. The stereochemistry of these compounds is usually described by the Greek letters α and β. If the OR bond is down, we have an α-glycoside; if up, a β-glycoside.
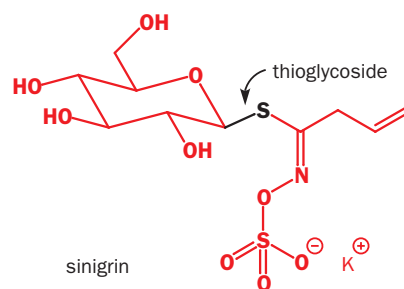
An attractive example is the pigment of red roses, which is an interesting aromatic oxygen heterocycle (an anthocyanidin). Two of the phenolic OH groups are present as β-glycosides.

■ We saw an example in Chapter 6 where acetone cyanohydrin is found in the cassava plant as a glucoside and suitable precautions must be taken when eating cassava to avoid poisoning by HCN.



aromatic pyrilium salt

glycoside linkages

pigment from red roses

a β-glycoside of a phenol

an α-glycoside of a phenol

▶ It is easy to remember which is which. People who devise nomenclature are maliciously foolish and, just as *E* means *trans* and *Z* means *cis* (each letter has the shape of the *wrong* isomer), so α means *below* and β means *above*—each word begins with the *wrong* letter.

## Protect yourself from cancer with green vegetables: *S*-glycosides

We will take an important series of *S*-glycosides for further chemical discussion in this chapter. It is clear that there are special benefits to health in eating broccoli and brussels sprouts because of their potent sulfur-containing anticancer compounds. These compounds are unstable isothiocyanates and are not, in fact, present in the plant but are released on damage by, for example, cutting or cooking when a glycosidase (an enzyme which hydrolyses glycosides) releases the sulfur compound from its glucose protection. A simple example is sinigrin.
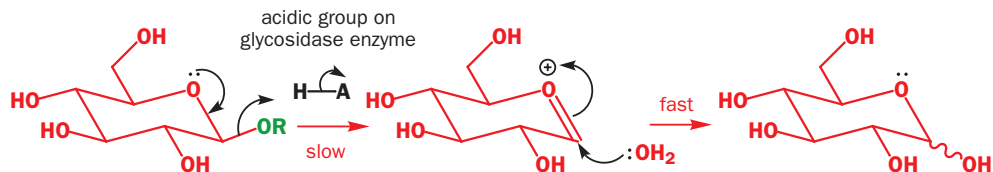


thioglycoside

sinigrin

When a glycosidase enzyme cleaves an *O*-glycoside, we should expect a simple general acid-catalysed first step followed by fast addition of water to the intermediate oxonium ion, essentially the same mechanism as is shown by the chemical reaction (Chapter 13).
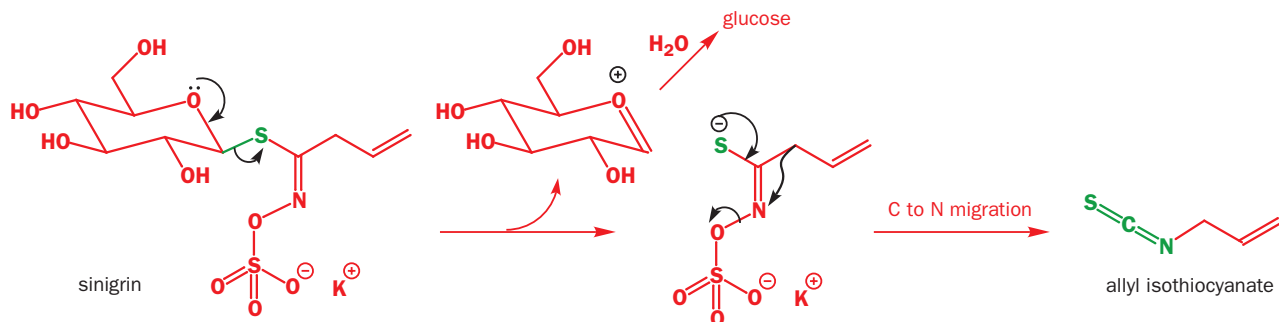
The *S*-glycosides of the sinigrin group start to hydrolyse in the same way. The sulfur atom is the better leaving group when it leaves as an anion (though worse than oxygen when the hydrolysis occurs in acidic conditions—see p. 1255) and these anions are additionally stabilized by conjugation.



The next step is very surprising. A rearrangement occurs, rather similar to the Beckmann rearrangement (Chapter 37), in which the alkyl group migrates from carbon to nitrogen and an isothiocyanate (R–N=C=S) is formed. Sinigrin occurs in mustard and horseradish and it is the release of the allyl isothiocyanate that gives them their 'hot' taste. When mustard powder is mixed with water, the hot taste develops over some minutes as sinigrin is hydrolysed to the isothiocyanate.

The *S*-glycoside in broccoli and brussels sprouts that protects from cancer is somewhat similar but has one more carbon atom in the chain and contains a sulfoxide group as well. Hydrolysis of the *S*-glycoside is followed by the same rearrangement, producing a molecule called sulforaphane. Sulforaphane protects against cancer-causing oxidants by inducing the formation of a reduction enzyme.
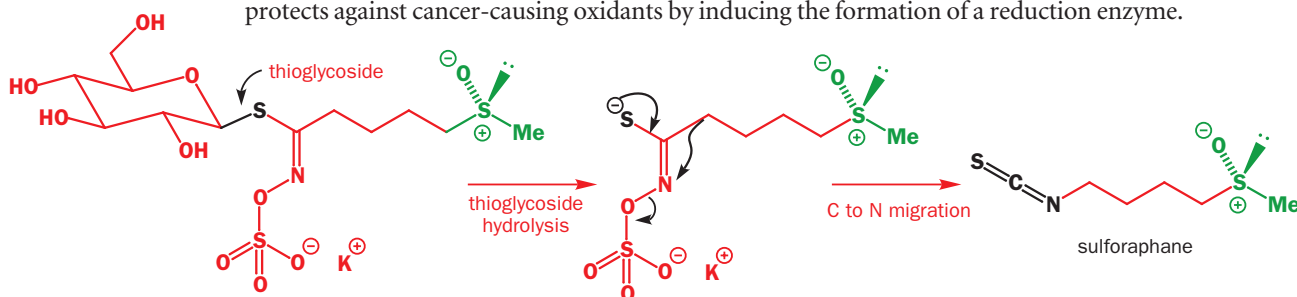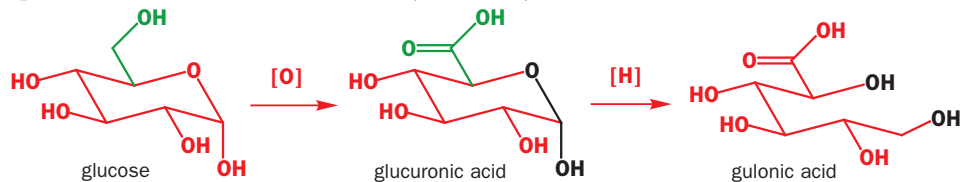


## Compounds derived from sugars

### Vitamin C

Nature makes some important compounds from simple sugars. Vitamin C—ascorbic acid—is one of these. Like glutathione, it protects us from stray oxidants as well as being involved in primary redox pathways (we mentioned earlier its role in collagen synthesis). Its reduced and oxidized forms are these.



ascorbic acid
reduced form of vitamin C                                                          oxidized form of vitamin C
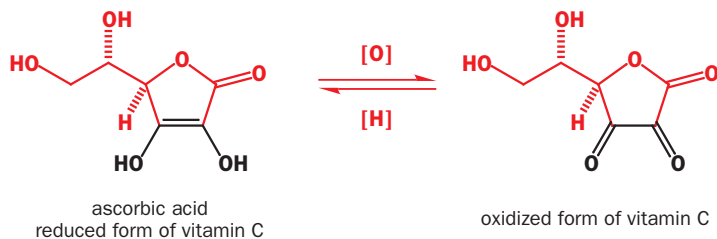
Vitamin C looks very like a sugar as it has six carbon atoms, each having an oxygen atom as substituent as well as an oxygen heterocycle, and it is no surprise that it is made in nature from glucose. We shall give just an outline of the process, which appropriately involves a lot of oxidation and reduction. The first step takes the primary alcohol of glucose to a carboxylic acid known as glucuronic acid. Next comes a reduction of the masked aldehyde to give 'gulonic acid'. Both reactions are quite reasonable in terms of laboratory chemistry.



glucose        [O]        glucuronic acid        [H]        gulonic acid

> ►
> We have given names for these relatively well-known sugar derivatives, but you do not need to learn them.

It is pretty obvious what will happen to this compound as it is an open-chain carboxylic acid with five OH groups. One of the OH groups will cyclize on to the acid to form a lactone. Kinetically, the most favourable cyclization will give a five-membered ring, and that is what happens. Now we are getting quite close to ascorbic acid and it is clear that oxidation must be the next step so that the double bond can be inserted between C2 and C3.



gulonic acid        cyclization (lactonization)        [O]
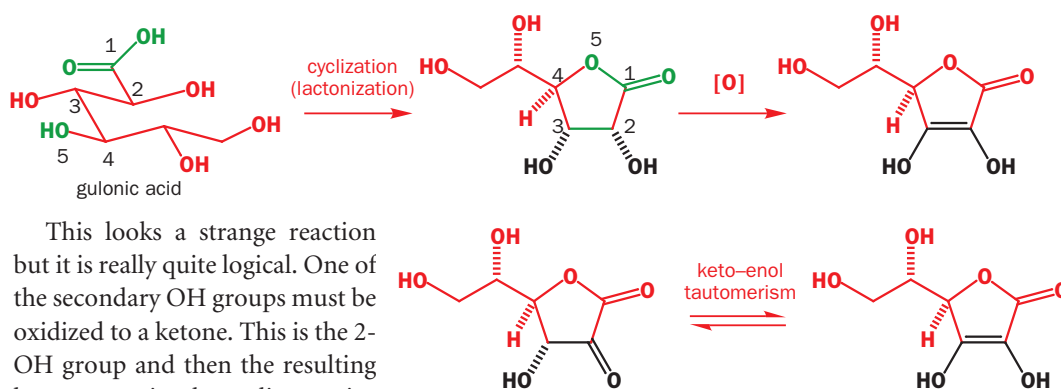
This looks a strange reaction but it is really quite logical. One of the secondary OH groups must be oxidized to a ketone. This is the 2-OH group and then the resulting ketone can simply enolize to give ascorbic acid.



keto–enol tautomerism

## Inositols

We have already discussed the widespread sugar alcohols such as mannitol but more important compounds are cyclic sugar alcohols having a carbocyclic ring (**cyclitols**). The most important is **inositol** which controls many aspects of our chemistry that require communication between the inside and the outside of a cell. Inositol-1,4,5-triphosphate (IP$_3$) can open calcium channels in cell membranes to allow calcium ions to escape from the cell.



inositol-1,4,5-triphosphate



inositol

> ■
> This inositol is known as 'myo-inositol' and is just one of many possible stereoisomers. Inositol was mentioned in Chapter 16.

Inositol is made in nature from glucose-6-phosphate by an aldol reaction that requires preliminary ring opening and selective oxidation (this would be tricky in the lab without protecting groups!).



glucose-6-phosphate pyranose form        glucose-6-phosphate open-chain form        [O]

The resulting ketone can be enolized on the phosphate side and added to the free aldehyde group to form the cyclohexane ring. We can draw the mechanism for the aldol reaction easily if we first change the conformation.
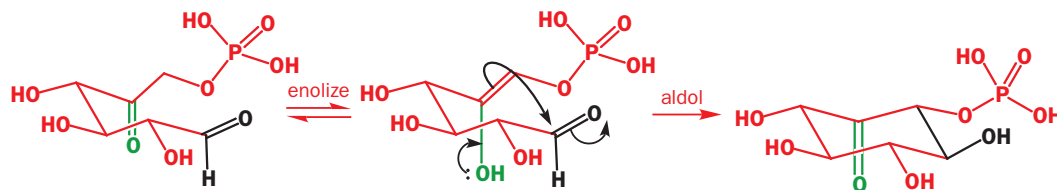


Finally, a stereochemically controlled reduction to give the axial alcohol (this would be the stereoselectivity expected with NaBH$_4$ for example: see Chapter 18) gives myo-inositol. The number and position of the phosphate esters can be controlled biochemically. This control is vital in the biological activity and would be difficult in the laboratory.
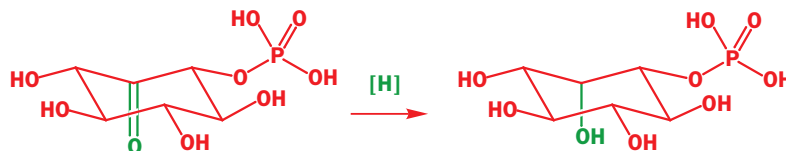


## Learning from Nature—the synthesis of inositols

If we wish to devise a chemical version of the biosynthesis of inositols, we need to use cleverly devised protecting groups to make sure that the right OH group is oxidized to a ketone. We can start with glucose trapped in its furanose form by a double acetone acetal as we discussed above. The one remaining OH group is first blocked as a benzyl ether.
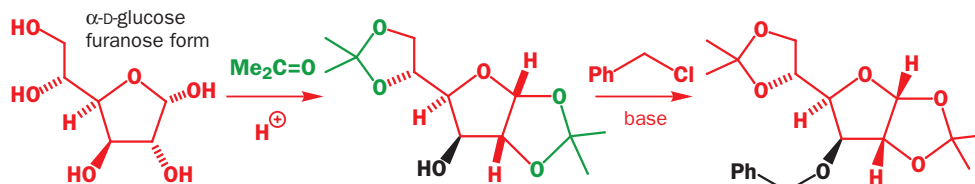


Next, one of the acetals is hydrolysed under very mild conditions, and the primary alcohol is protected as a trityl ether. This is an S$_N$1 reaction with an enormous electrophile—so big that it goes on primary alcohols only.
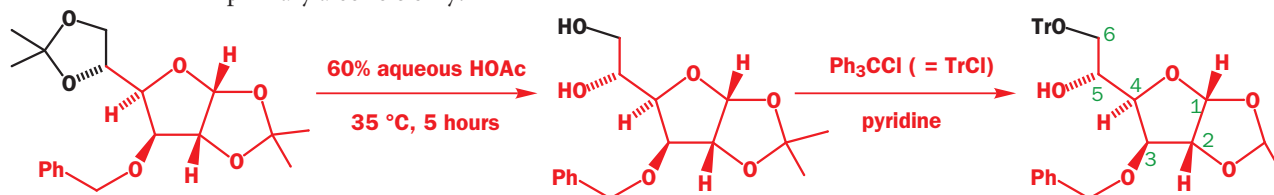


Notice that each oxygen atom in this molecule of protected glucose is now different. Only the OH at C5 is free, and its time has come: it can now be oxidized using a Swern procedure with dimethylsulfoxide as the oxidant (Chapter 46).
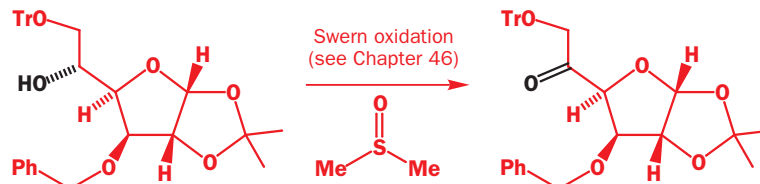


Now we can strip away the protecting groups one by one and it is instructive to see how selective these methods are. The trityl group comes off in aqueous acetic acid by another S$_N$1 reaction in which water captures the triphenylmethyl cation, and the benzyl group is removed by hydrogenolysis—hydrogen gas over a 10% palladium on charcoal catalyst in ethanol.
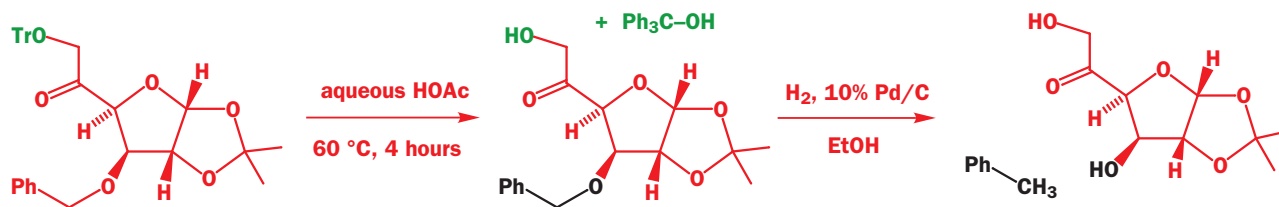
Finally, the acetone acetal is removed by acid hydrolysis. Because free sugars are difficult to isolate it is convenient to use an acidic resin known as 'Dowex'. The resin (whose polymeric structure is discussed in Chapter 52) can simply be filtered off at the end of the reaction and the solid product isolated by **lyophilization**—evaporation of water at low pressure below freezing point. The yield is quantitative.
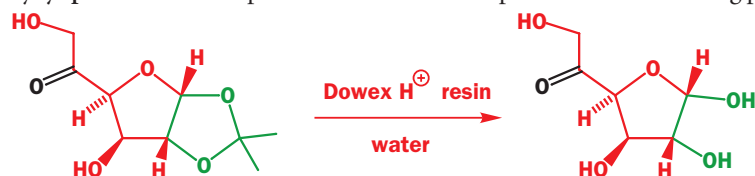


All of the hydroxyl groups are now free except the one tied up in the hemiacetal and that, of course, is in equilibrium with the open-chain hydroxy-aldehyde as we have already seen. Treatment of this free 'glucose ketone' with aqueous NaOH gives the ketone of myo-inositol as the major product together with some of the other diastereoisomers.
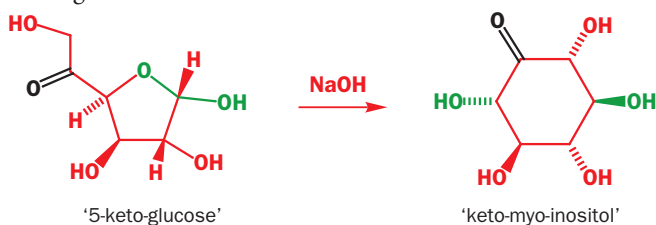


The simplest explanation of this result is that the chemical reaction has followed essentially the same course as the biological one. First, the hemiacetal is opened by the base to give the open-chain keto-aldehyde. Rotation about a C–C bond allows a simple aldol condensation between the enolate of the ketone as nucleophile and the aldehyde as electrophile.
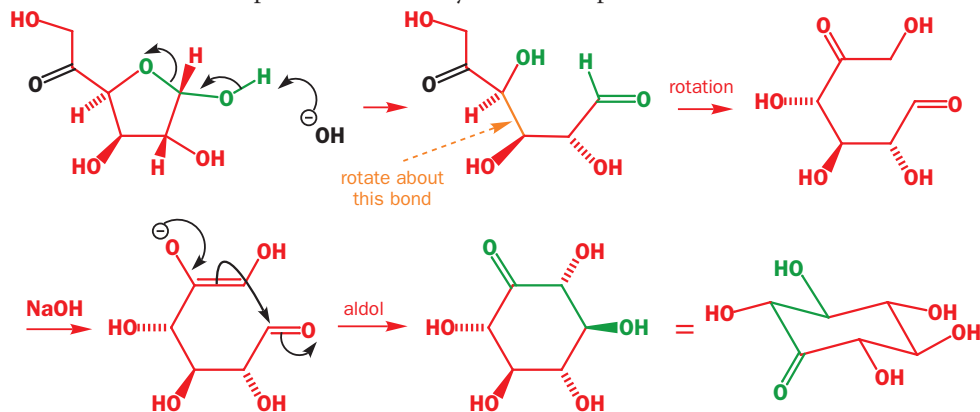


The enolate must prefer to attack the aldehyde in the same way as in the biological reaction to give the all-equatorial product as the conformational drawing shows. The arrangement of the enolate in the aldol reaction itself will be the same as in the cyclization of the phosphate above.
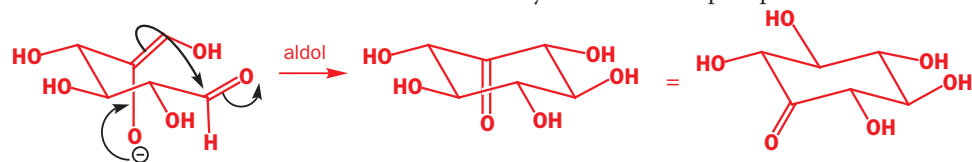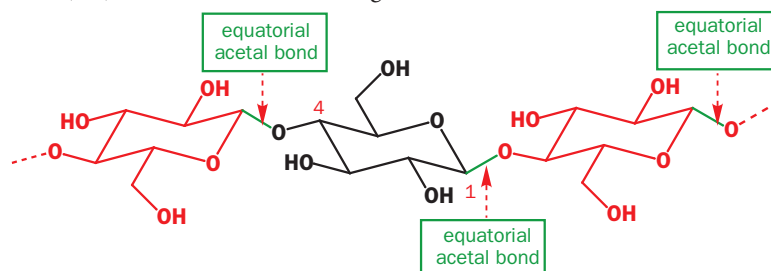
As in many other cases, by improving the rate and perfecting the stereoselectivity, the enzyme makes much better a reaction that already works.

# Most sugars are embedded in carbohydrates

Before we leave the sugars we should say a little about the compounds formed when sugars combine together. These are the **saccharides** and they have the same relationship to sugars as peptides and proteins have to amino acids. We have met one simple disaccharide, sucrose, but we need to meet some more important molecules.

One of the most abundant compounds in nature is cellulose, the structural material of plants. It is a glucose polymer and is produced in simply enormous quantities (about $10^{15}$ kg per year). Each glucose molecule is joined to the next through the anomeric bond (C1) and the other end of the molecule (C4). Here is that basic arrangement.

▶
This is literally an astronomical amount: it's about the mass of one of the moons of Mars, Deimos. Our moon weighs $10^{22}$ kg.



Notice that the anomeric bonds are all equatorial. This means that the cellulose molecule is linear in general outline. It is made rigid by extra hydrogen bonds between the 3-OH groups and the ring oxygen atoms—like this.
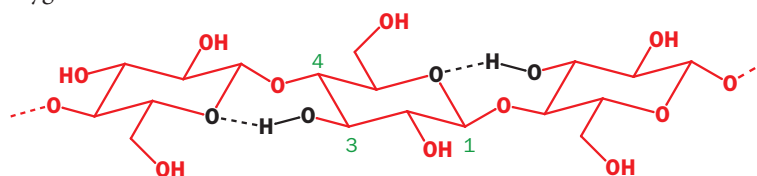


The polymer is also coiled to increase stability still further. All this makes cellulose very difficult to hydrolyse, and humans cannot digest cellulose as we do not have the necessary enzymes. Only ruminants, such as cows, whose many stomachs harbour some helpful bacteria, can manage it.

## Amino sugars add versatility to saccharides

To go further in understanding the structural chemistry of life we need to know about **amino sugars**. These molecules allow proteins and sugars to combine and produce structures of remarkable variety and beauty. The most common amino sugars are *N*-acetyl-glucosamine and *N*-acetyl-galactosamine, which differ only in stereochemistry.
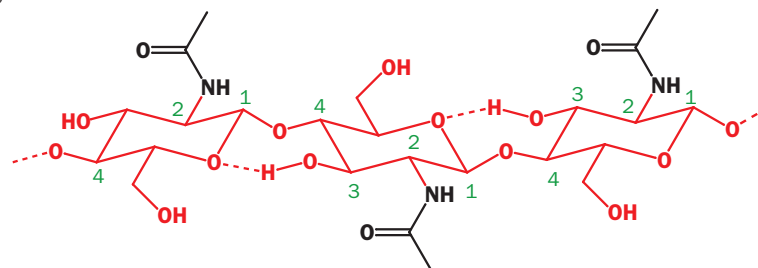
The hard outer skeleton of insects and shellfish contains chitin, a polymer very like cellulose but made of acetyl glucosamine instead of glucose itself. It coils up in a similar way and provides the toughness of crab shells and beetle cases.
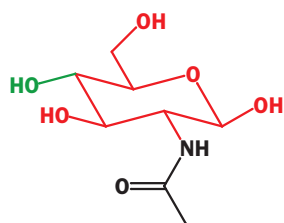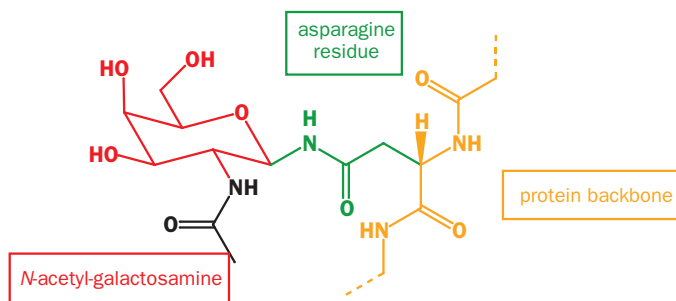


*N*-acetyl-glucosamine
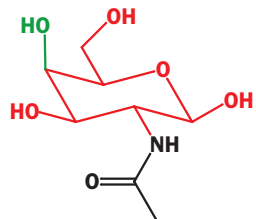


*N*-acetyl-galactosamine



Ordinary cell membranes must not be so tough as they need to allow the passage of water and complex molecules through channels that can be opened by molecules such as inositol phosphates.

These membranes contain **glycoproteins**—proteins with amino-sugar residues attached to asparagine, serine, or threonine in the protein. The attachment is at the anomeric position so that these compounds are *O*- or *N*-glycosides of the amino sugars. Here is *N*-acetyl-galactosamine attached to an asparagine residue as an *N*-glycoside.

The cell membrane normally contains less than 10% of sugars but these are vital to life. Because the sugars (*N*-acetyl-glucosamine and *N*-acetyl-galactosamine) are covered with very polar groups (OH and amide) they prefer to sit outside the membrane in the aqueous extracellular fluid rather than within the nonpolar membrane itself. When two cells meet, the sugars are the first things they see. We cannot go into the details of the biological processes here, but even the structures of these saccharides dangling from the cell are very interesting. They contain amino sugars, again particularly *N*-acetyl-glucosamine and *N*-acetyl-galactosamine, and they are rich in mannose.

In addition, they are usually branched at one of the mannose residues that is joined to two other mannoses on one side and to one glucosamine on the other. The glucosamine leads back eventually to the protein through a link to asparagine like the one we have just seen. The two mannoses are linked to more sugars at positions marked by the green arrows and provide the recognition site. The structure below is a typical branchpoint.

► Mannose is another glucose diastereoisomer and has one axial OH group at C2.
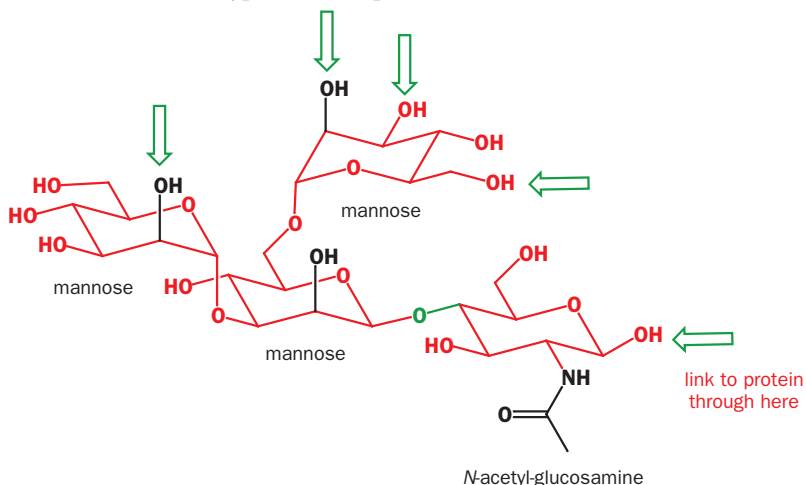
You should begin to see from structures like these just how versatile sugar molecules can be. From just four sugars we have constructed a complex molecule with up to 13 possible link sites. With more sugars added, the possibilities become enormous. It is too early to say what medical discoveries will emerge from these molecules, but one that is likely to be important is **sialyl Lewis X**. This tetrasaccharide is also branched but it contains a different type of molecule—a $C_9$ sugar with a $CO_2H$ group, called sialic acid.

Sialic acid has the $CO_2H$ group at the anomeric position, a typical *N*-acetyl group, and a unique side chain (in green) with three more OH groups. Sialyl Lewis X has sialic acid at the end of a branched sugar chain. The branchpoint is the familiar *N*-acetyl-glucosamine through which the molecule is eventually linked to the glycoprotein. The remaining sugars are galactose, a diastereoisomer of glucose, and a sugar we have not seen before, fucose. Fucose often appears in saccharides of this kind and is a six-carbon sugar without a primary OH group. It is like galactose with Me instead of $CH_2OH$.

sialyl Lewis X

fucose

Me

CO$_2$H

sialic acid

galactose

N-acetyl-glucosamine

joined to protein
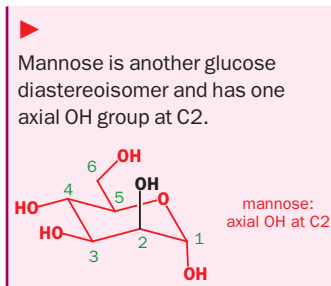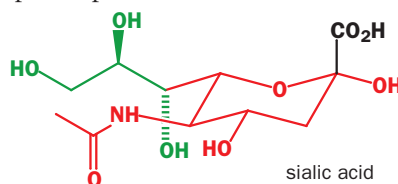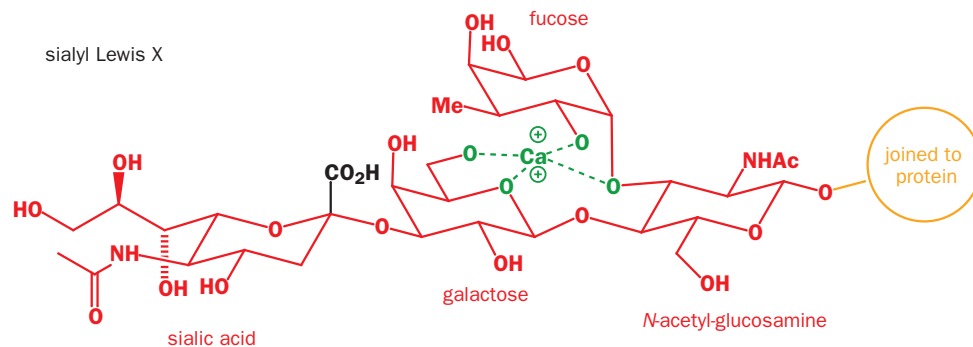
Sialyl Lewis X can also form a stable complex with calcium ions as the diagram shows and this may be vital to its activity. It is certainly involved in leukocyte adhesion to cells and is therefore vital in the prevention of infection.
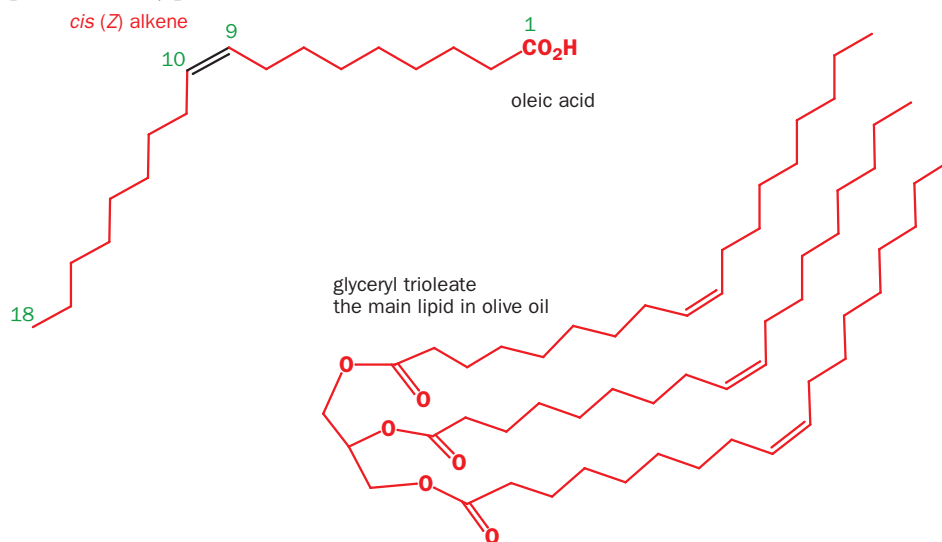
# Lipids

glycerol

glycerol monoester

glycerol 3-phosphate
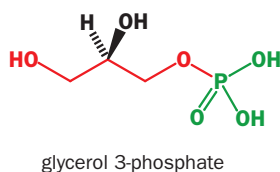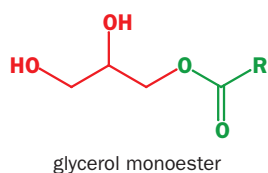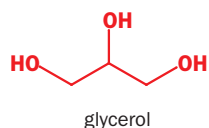
Lipids (fats) are the other important components of cell membranes. Along with cholesterol, also a component of the cell membrane, they have acquired a bad name, but they are nonetheless essential to the function of membranes as selective barriers to the movement of molecules.

The most common types of lipids are esters of glycerol. Glycerol is just propane-1,2,3-triol but it has interesting stereochemistry. It is not chiral as it has a plane of symmetry, but the two primary OH groups are enantiotopic (Chapter 16). If one of them is changed—by esterification, for example—the molecule becomes chiral. Natural glycerol phosphate is such an ester and it is optically active.

A typical lipid in foodstuffs is the triester formed from glycerol and oleic acid, which is the most abundant lipid in olive oil. Oleic acid is a 'mono-unsaturated fatty acid'—it has one *Z* double bond in the middle of the C$_{18}$ chain. This bond gives the molecule a marked kink in the middle. The compound actually present in olive oil is the triester, also kinked.

*cis* (*Z*) alkene

CO$_2$H

oleic acid

glyceryl trioleate
the main lipid in olive oil

## Oil and water do not mix

The lipid has, more or less, the conformation shown in the diagram with all the polar ester groups at one end and the hydrocarbon chains bunched together in a nonpolar region. Oil and water do not mix, it is said, but triglyceride lipids associate with water in a special way. A drop of oil spreads out on water in a very thin layer. It does so because the ester groups sit inside the water and the hydrocarbon side chains stick out of the water and associate with each other.

■
You may have done the 'Langmuir trough' experiment in a physical chemistry practical class. This involves measuring the size of a molecule by allowing an oil to spread on the surface of water in a unimolecular layer.
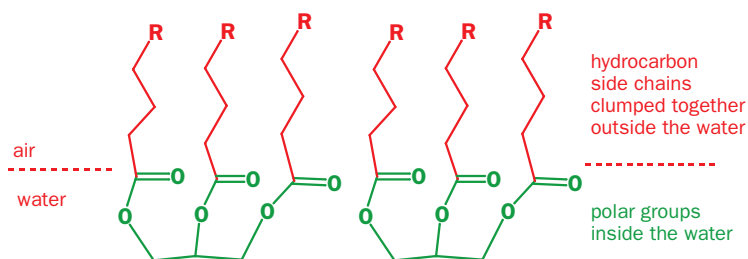
When triglycerides are boiled up with alkali, the esters are hydrolysed and a mixture of carboxylate salts and glycerol is formed. This was how soap was made—hard soap was the sodium salt and soft soap the potassium salt.
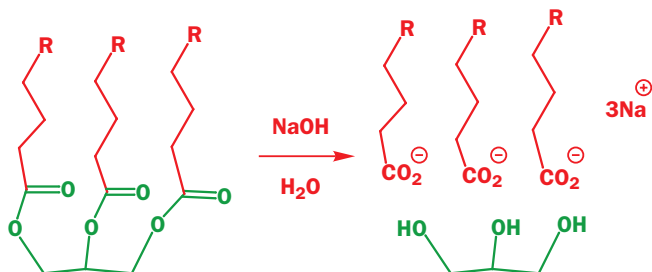


When a soap is suspended in water, the carboxylate groups have a strong affinity for the water and so oily globules or **micelles** are formed with the hydrocarbon side chain inside. It is these globules that remove greasy dirt from you or your clothes.
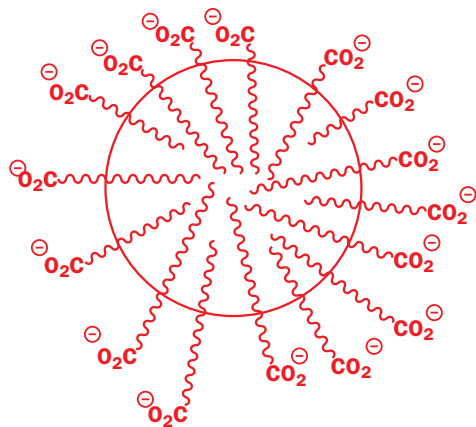


## Nature uses thiol esters to make lipids

The repulsion between molecules having oily or aqueous properties is the basis for membrane construction. The lipids found in membranes are mostly based on glyceryl phosphate and normally contain three different side chains—one saturated, one unsaturated, and one very polar.
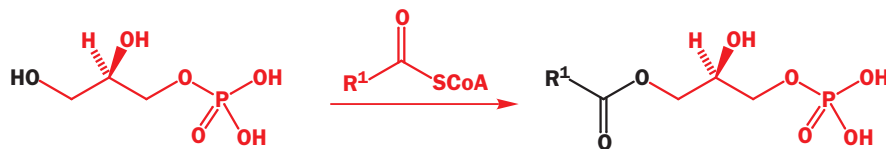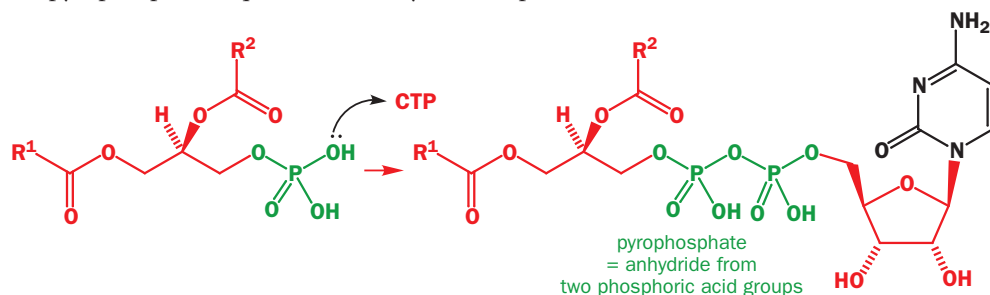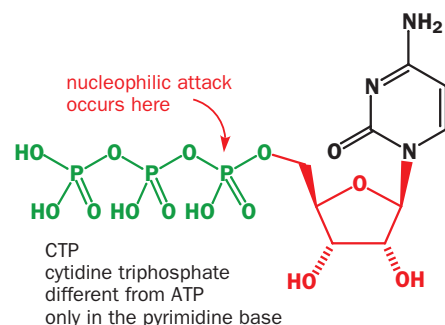


The saturated chain is added first, at C1 of glyceryl phosphate. The reagent is a thiol ester called acyl coenzyme A, whose full structure you will see in the next chapter. This reaction occurs by simple nucleophilic attack on the carbonyl group of the thiol ester followed by loss of the better leaving group, the thiolate anion. Then the process is repeated at the second OH group where an unsaturated fatty acid, perhaps oleic acid, is added by the same mechanism.

■
We discussed acylation by thioesters, the laboratory version of this reaction, in Chapter 27.
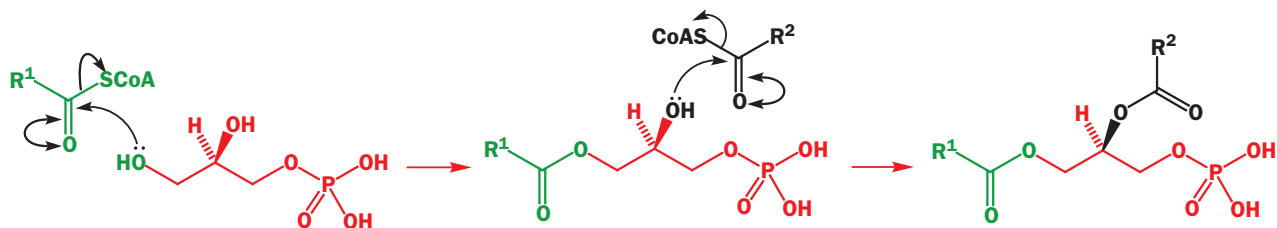
The third acylation requires the phosphate to act as the acylating agent and a polar alcohol to be introduced to form a phosphate ester. This reaction actually occurs by the activation of the phosphate as a pyrophosphate. Pyrophosphates are really acid anhydrides so it is not surprising that they act as acylating agents. The first step is a reaction with cytidine triphosphate (CTP) doing a job we might expect from ATP.

Nucleophilic attack by the phosphate group of the phosphoglyceride at the point indicated on CTP gives the pyrophosphate required for the acylation step.



CTP
cytidine triphosphate
different from ATP
only in the pyrimidine base



pyrophosphate
= anhydride from
two phosphoric acid groups

The anhydride is now attacked by an alcohol acting as a nucleophile. The attack occurs only at the electrophilic phosphorus centre further from the nucleotide. This is an impressive piece of regioselectivity and is presumably controlled by the enzyme.





inositol



the amino acid
serine (Ser)
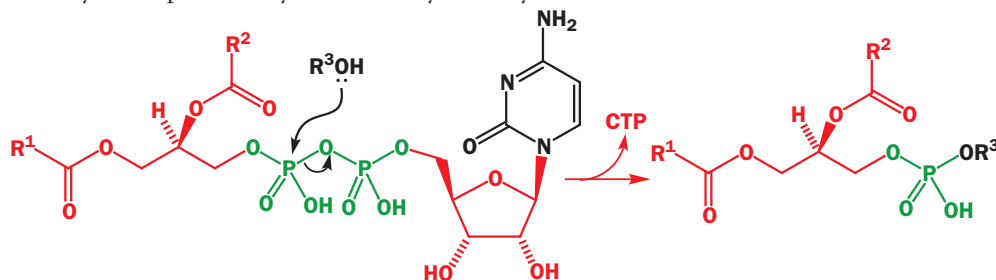
This third chain is rather different from the other two—it's a phosphate diester, and the alcohol portion can be inositol joined through the OH group at C1 or it can be the amino acid serine, joined through its OH group.

The compound formed from serine is particularly important as it can be transformed into the most dramatically contrasted of these phospholipids. A decarboxylation using a coenzyme (we shall look at the mechanism of this reaction in Chapter 51) gives a very simple molecule, phosphatidyl ethanolamine.



phosphatidyl serine

phosphatidyl ethanolamine
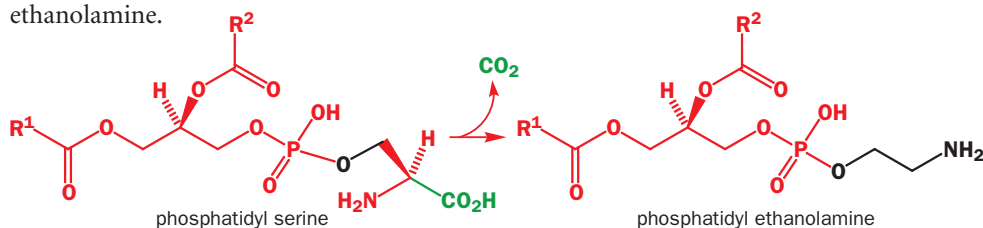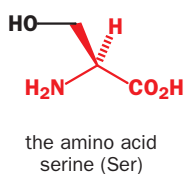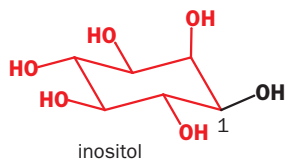
Finally, three methylations on the nitrogen atom by SAM (see p. 1348) gives the zwitterion phosphatidyl choline.
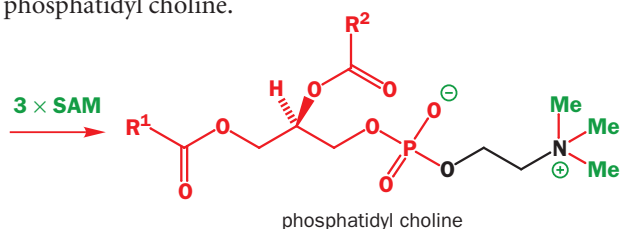


phosphatidyl choline

choline
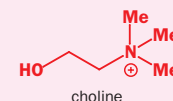
## Phospholipids form membranes spontaneously

The choline terminus of the molecule is very polar indeed. Phosphatidyl choline adopts a shape with the nonpolar chains ($R^1$ and $R^2$) close together, and it should be clear that this is an ideal molecule for the construction of membranes.



hydrophobic region · glycerol link · hydrophilic region
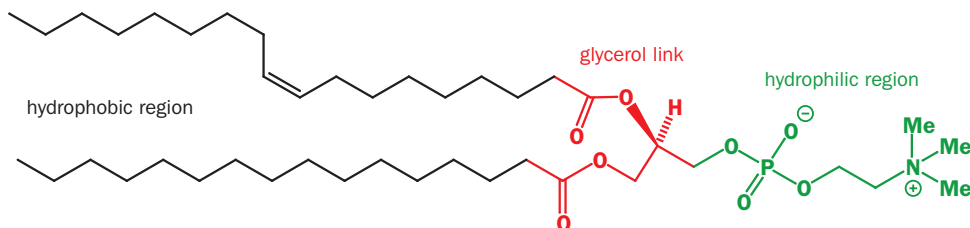
We have already seen how oils such as glyceryl trioleate form thin layers on water while soaps from the alkaline hydrolysis of glycerides form micelles. Phosphatidyl choline forms yet another structure—it spontaneously forms a membrane in water. The hydrophobic hydrocarbon chains line up together on the inside of the membrane with the hydrophilic choline residues on the outside.



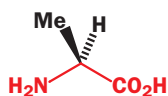hydrophilic region · glycerol link · hydrophobic region · glycerol link · hydrophilic region
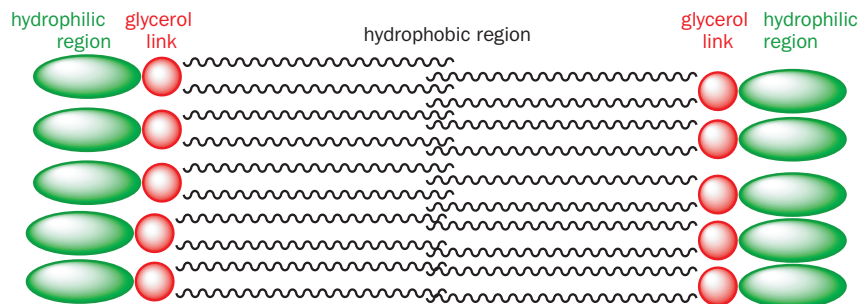
This is just a small piece of a cross-section of the membrane. These membranes are called **lipid bilayers** because two rows of molecules line up to form two layers back-to-back. The charged, hydrophilic region on the outside is solvated by the water and the hydrocarbon tails are repelled by the water and attracted to each other by weak forces such as van der Waals attractions.
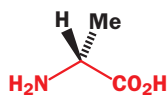
Full structural analysis of a real cell membrane reveals a chemically diverse thin sheet composed of phospholipid bilayers penetrated by glycoproteins containing the amino sugars we discussed earlier. The amount of each component varies but there is usually about 50:50 phospholipid:protein, with the protein containing about 10% sugar residues. The phospholipids' main role is as a barrier while the glycoproteins have the roles of recognition and transport.

# Bacteria and people have slightly different chemistry

We have many times emphasized that all life has very similar chemistry. Indeed, in terms of biochemistry there is little need for the classifications of mammals, plants, and so on. There is only one important division—into **prokaryotes** and **eukaryotes**. Prokaryotes, which include bacteria, evolved first and have simple cells with no nucleus. Eukaryotes, which include plants, mammals, and all
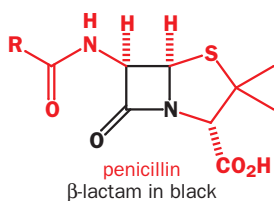
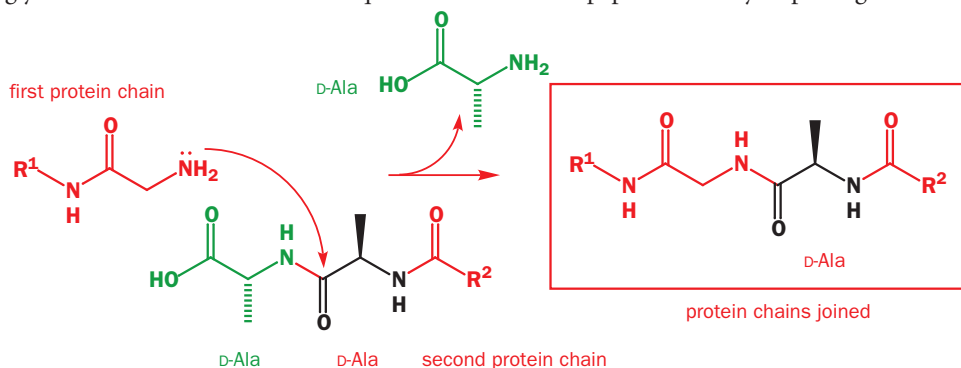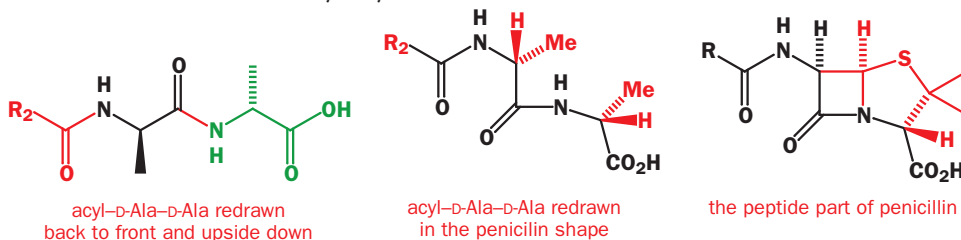other multicellular creatures, evolved later and have more complex cells including nuclei. Even so, much of the biochemistry on both sides of the divide is the same.

When medicinal chemists are looking for ways to attack bacteria, one approach is to interfere with chemistry carried out by prokaryotes but not by us. The most famous of these attacks is aimed at the construction of the cell walls of some bacteria that contain 'unnatural' ($R$)- (or D-) amino acids. Bacterial cell walls are made from glycopeptides of an unusual kind. Polysaccharide chains are cross-linked with short peptides containing ($R$)-alanine (D-Ala). Before they are linked up, one chain ends with a glycine molecule and the other with D-Ala–D-Ala. In the final step in the cell wall synthesis, the glycine attacks the D-Ala–D-Ala sequence to form a new peptide bond by displacing one D-Ala residue.



(S)-Ala from normal proteins

(R)-Ala from bacterial cell walls

> The reason bacteria use these 'unnatural' D-amino acids in their cell walls is to protect them against the enzymes in animals and plants, which cannot digest proteins containing D-amino acids.

The famous molecule that interferes with this step is penicillin, though this was not even suspected when penicillin was discovered. We now know how penicillin works. It inhibits the enzyme that catalyses the D-Ala transfer in a very specific way. It first binds specifically to the enzyme, so it must be a mimic of the natural substrate, and it then reacts with the enzyme and inactivates it by blocking a vital OH group at the active site. If we emphasize the peptide nature of penicillin and compare it with D-Ala–D-Ala, the mimicry may become clearer.



penicillin
β-lactam in black

acyl–D-Ala–D-Ala redrawn back to front and upside down          acyl–D-Ala–D-Ala redrawn in the penicilin shape          the peptide part of penicillin

Penicillin imitates D-Ala and binds to the active site of the enzyme, encouraging the OH group of a serine residue to attack the reactive, strained β-lactam. This same OH group of the same serine residue would normally be the catalyst for the D-Ala–D-Ala cleavage used in the building of the bacterial cell wall. The reaction with penicillin 'protects' the serine and irreversibly inhibits the enzyme. The bacterial cell walls cannot be completed, and the bacterial cells literally burst under the pressure of their contents. Penicillin does not kill bacteria whose cell walls are already complete but it does prevent new bacteria being formed.
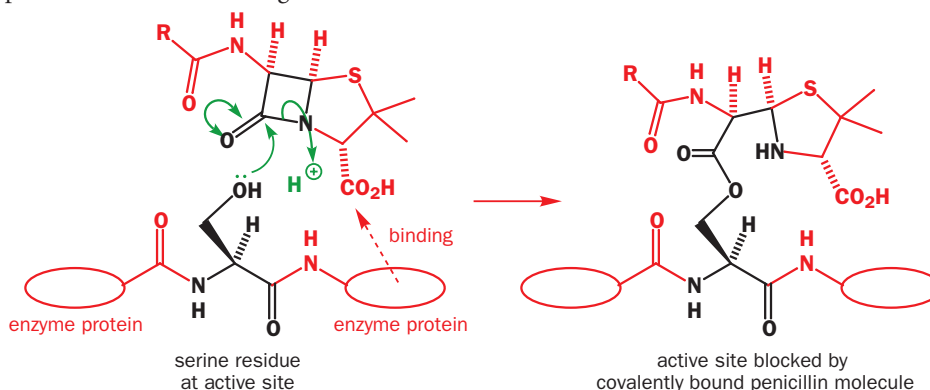


> Our current last line of defence against bacteria resistant to penicillin, and other antibiotics, is vancomycin. Vancomycin works by binding to the D-Ala–D-Ala sequences of the bacterial cell wall.

serine residue at active site          active site blocked by covalently bound penicillin molecule

You have seen many instances in this chapter of the importance of a good understanding of both the chemistry and the biochemistry of living things if medicine is to advance: it is at the frontier of chemistry and biology that many of the most important medical advances are being made.

## Problems

**1.** Do you consider that thymine and caffeine are aromatic compounds? Explain.



thymine

caffeine
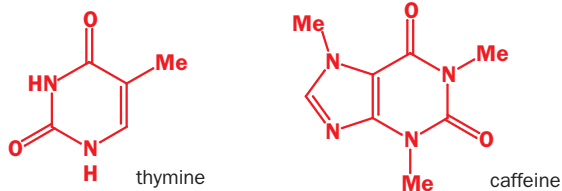
**2.** It is important that we draw certain of the purine and pyrimidine bases in their preferred tautomeric forms. The correct pairings are given early in the chapter. What alternative pairings would be possible with these (minor) tautomers of thymine and guanine? Suggest reasons (referring to Chapter 43 if necessary) why the major tautomers are preferred.



thymine tautomer

guanine tautomer
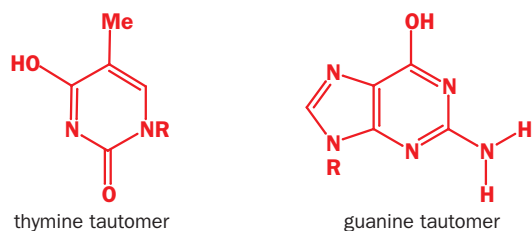
**3.** Dialkyl phosphates are generally hydrolysed quite slowly at near-neutral pHs but this example hydrolyses much more rapidly. What is the mechanism and what relevance has it to RNA chemistry?
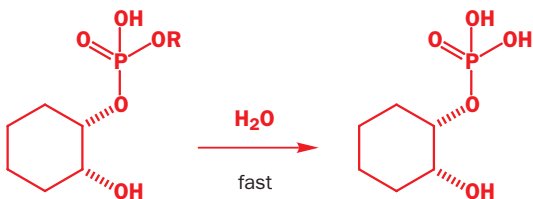


Revision of Chapter 41. This reaction is subject to general base catalysis. Explain.

**4.** Primary amines are not usually made by displacement reactions on halides with ammonia. Why not? The natural amino acids can be made by this means in quite good yield. Here is an example.



67% yield
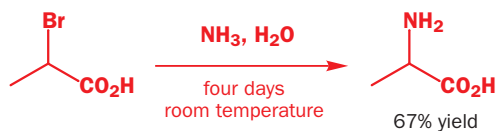
Why does this example work? Comment on the state of the reagents and products under the reaction conditions. What is the product and how does it differ from the natural amino acid?

**5.** Human hair is a good source of cystine, the disulfide dimer of cysteine. The hair is boiled with aqueous HCl and $HCO_2H$ for a day, the solution concentrated, and a large amount of sodium acetate added. About 5% of the hair by weight crystallizes out as pure cystine $[\alpha]_D$ –216. How does the process work? Why is such a high proportion of hair cystine? Why is no cysteine isolated by this process? What is the stereochemistry of cystine? Make a good drawing of cystine to show its symmetry. How would you convert the cystine to cysteine?



(S)–cysteine
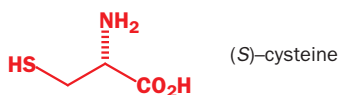
**6.** A simple preparation of a dipeptide is given below. Explain the reactions, drawing mechanisms for the interesting steps. Which steps are protection, activation, coupling, and deprotection? Explain the reasons for protection and the nature of the activation. Why is the glycine added to the coupling step as its hydrochloride? What reagent(s) would you use for the final deprotection step?
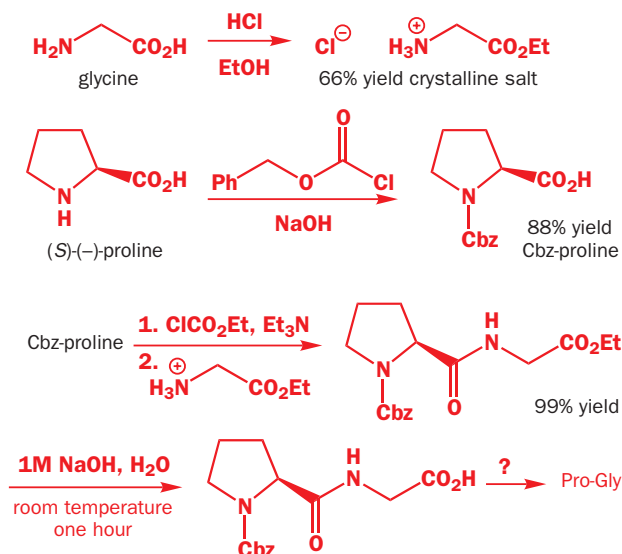


**7.** Suggest how glutathione might detoxify these dangerous chemicals in living things. Why are they still toxic in spite of this protection?
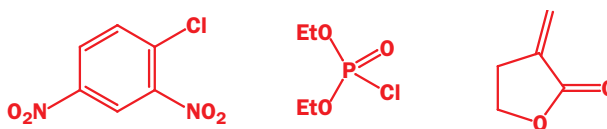
**8.** Alanine can be resolved by the following method, using a pig kidney acylase. Draw a mechanism for the acylation step. Which isomer of alanine acylates faster? In the enzyme-catalysed reaction, which isomer of the amide hydrolyses faster? In the separation, why is the mixture heated in acid solution, and what is filtered off? How does the separation of the free alanine by dissolution in ethanol work?


racemic alanine
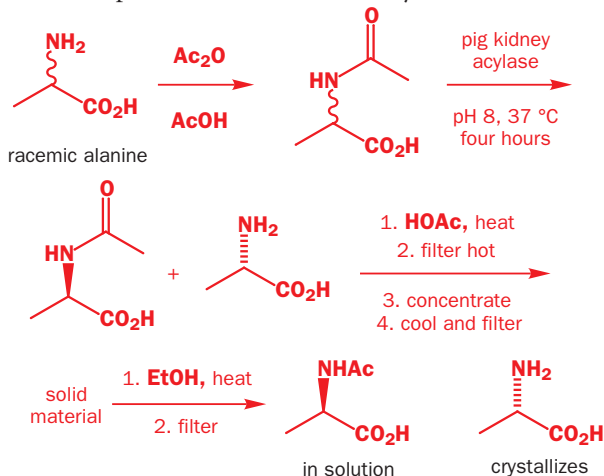
If the acylation is carried out carelessly, particularly if the heating is too long or too strong, a by-product may form that is not hydrolysed by the enzyme. How does this happen?


racemic alanine
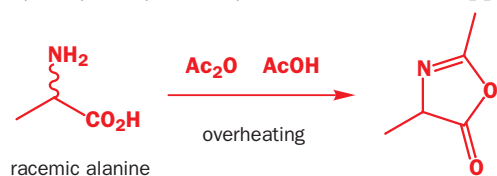
**9.** A patent discloses this method of making the anti-AIDS drug d4T. The first few stages involve differentiating the three hydroxyl groups of 5-methyluridine as shown below. Explain the reactions, especially the stereochemistry at the position of the bromine atom.
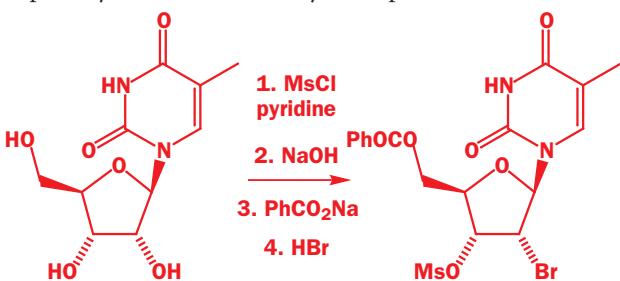


Suggest how the synthesis might be completed.
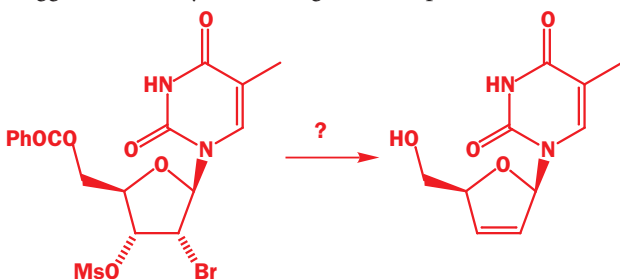


**10.** Mannose usually exists as the pyranoside shown below. This is in equilibrium with the furanoside. What is the conformation of the pyranoside and what is the stereochemistry of the furanoside? What other stereochemical change will occur more quickly than this isomerization?


manno-pyranoside          manno-furanoside
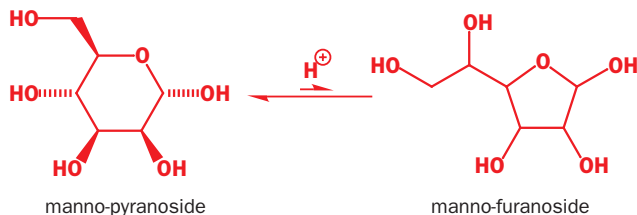
Treatment of mannose with acetone and HCl gives the acetal shown. Explain the selectivity.


manno-pyranoside
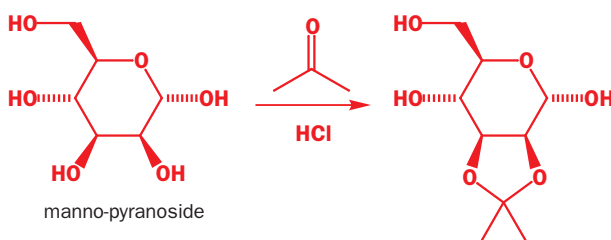
**11.** How are glycosides formed from phenols (in Nature or in the laboratory)? Why is the stereochemistry of the glycoside not related to that of the original sugar?
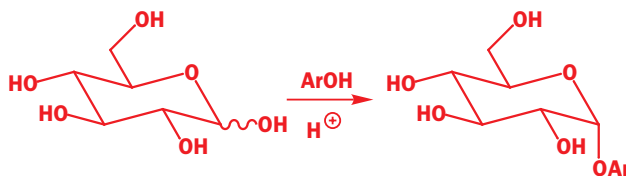


**12.** Draw all the keto and enol forms of ascorbic acid (vitamin C). Why is the one shown the most stable?


ascorbic acid
reduced form of vitamin C
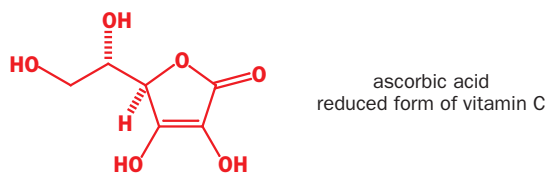
**13.** 'Caustic soda' (NaOH) was used to clean ovens and clear blocked drains. Many commercial products for these jobs with fancy names still contain NaOH. Even concentrated sodium carbonate ($Na_2CO_3$) does quite a good job. How do these cleaners work? Why is NaOH so dangerous to humans, particularly if it gets in the eye?

**14.** Bacterial cell walls contain the unnatural amino acid D-alanine. If you wanted to prepare a sample of D-ala, how would you go about it? (*Hint.* There is not enough in bacteria to make that a worthwhile source, but have you done Problem 8 yet?)