Organic structures

2

Connections

Building on

- This chapter does not depend on Chapter 1

Arriving at

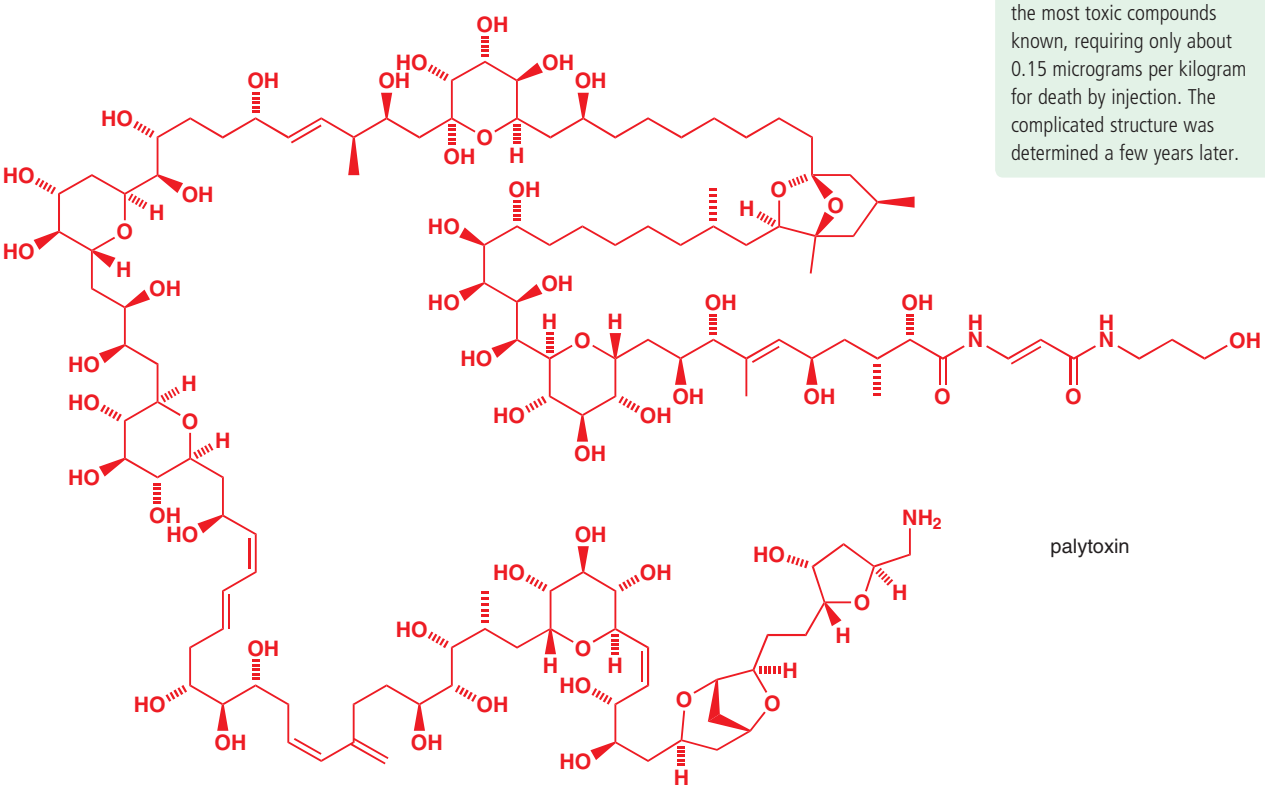
- The diagrams used in the rest of the book
- Why we use these particular diagrams
- How organic chemists name molecules in writing and in speech
- What is the skeleton of an organic molecule
- What is a functional group
- Some abbreviations used by all organic chemists
- Drawing organic molecules realistically in an easily understood style

Looking forward to

- Ascertaining molecular structure spectroscopically ch3
- What determines a molecule's structure ch4

There are over 100 elements in the periodic table. Many molecules contain well over 100 atoms—palytoxin (a naturally occurring compound with potential anticancer activity), for example, contains 129 carbon atoms, 221 hydrogen atoms, 54 oxygen atoms, and 3 nitrogen atoms. It's easy to see how chemical structures can display enormous variety, providing enough molecules to build even the most complicated living creatures.

■ Palytoxin was isolated in 1971 in Hawaii from *Limu make o Hane* ('deadly seaweed of Hana'), which had been used to poison spear points. It is one of the most toxic compounds known, requiring only about 0.15 micrograms per kilogram for death by injection. The complicated structure was determined a few years later.



palytoxin

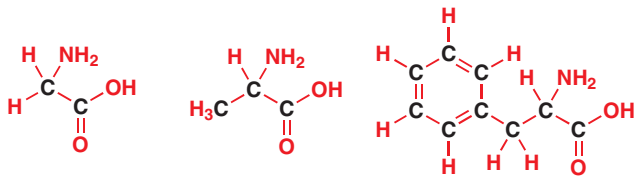
Online support. The icon in the margin indicates that accompanying interactive resources are provided online to help your understanding: just type **www.chemtube3d.com/clayden/123** into your browser, replacing **123** with the number of the page where you see the icon. For pages linking to more than one resource, type **123-1, 123-2** etc. (replacing **123** with the page number) for access to successive links.

But how can we understand what seems like a recipe for confusion? Faced with the collection of atoms we call a molecule, how can we make sense of what we see? This chapter will teach you how to interpret organic structures. It will also teach you how to draw organic molecules in a way that conveys all the necessary information and none of the superfluous.

Hydrocarbon frameworks and functional groups

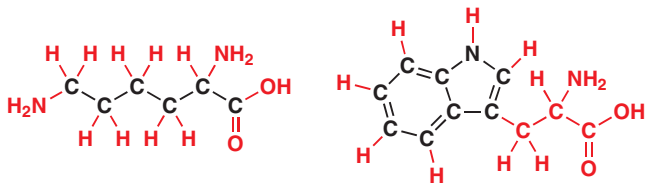
As we explained in Chapter 1, organic chemistry is the study of compounds that contain carbon. Nearly all organic compounds also contain hydrogen; most also contain oxygen, nitrogen, or other elements. Organic chemistry concerns itself with the way in which these atoms are bonded together into stable molecular structures, and the way in which these structures change in the course of chemical reactions.

Some molecular structures are shown below. These molecules are all amino acids, the constituents of proteins. Look at the number of carbon atoms in each molecule and the way they are bonded together. Even within this small class of molecules there's great variety—*glycine* and *alanine* have only two or three carbon atoms; *phenylalanine* has nine.



glycine alanine phenylalanine

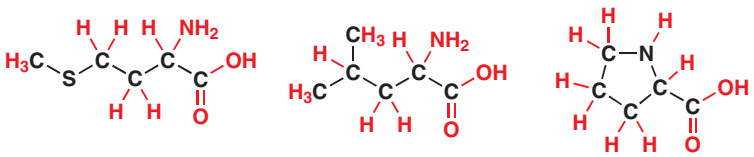
Lysine has a chain of atoms; *tryptophan* has rings.



lysine tryptophan

Interactive amino acid structures

In *methionine* the atoms are arranged in a single chain; in *leucine* the chain is branched. In *proline*, the chain bends back on itself to form a ring.

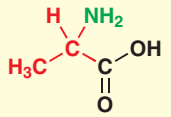


methionine leucine proline

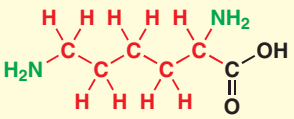
We shall return to amino acids as examples several times in this chapter but we shall leave detailed discussion of their chemistry till Chapters 23 and 42 when we look at the way they polymerize to form peptides and proteins.

Yet all of these molecules have similar properties—they are all soluble in water, they are all both acidic and basic (amphoteric), they can all be joined with other amino acids to form proteins. This is because the chemistry of organic molecules depends much less on the number or the arrangement of carbon or hydrogen atoms than on the other types of atoms (O, N, S, P, Si...) in the molecule. We call parts of molecules containing small collections of these other atoms **functional groups**, simply because they are groups of atoms that determine the way the molecule works. All amino acids contain two functional groups: an amino (NH_2 or NH) group and a carboxylic acid (CO_2H) group (some contain other functional groups as well).

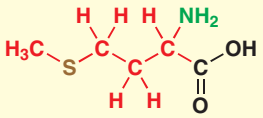
● **The functional groups determine the way the molecule works both chemically and biologically.**



alanine
contains just the amino and carboxylic acid functional groups



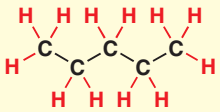
lysine
has an additional amino group



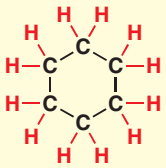
methionine
also has a sulfide functional group

That isn't to say the carbon atoms aren't important; they just play quite a different role from those of the oxygen, nitrogen, and other atoms they are attached to. We can consider the chains and rings of carbon atoms we find in molecules as their skeletons, which support the functional groups and allow them to take part in chemical interactions, much as your skeleton supports your internal organs so they can interact with one another and work properly.

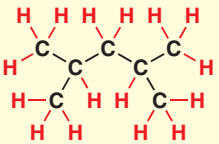
● **The hydrocarbon framework is made up of chains and rings of carbon atoms, and it acts as a support for the functional groups.**



a chain



a ring



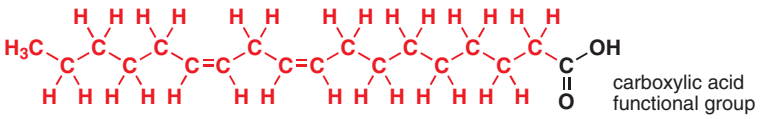
a branched chain

We will see later how the interpretation of organic structures as hydrocarbon frameworks supporting functional groups helps us to understand and rationalize the reactions of organic molecules. It also helps us to devise simple, clear ways of representing molecules on paper. You saw these structural diagrams in Chapter 1, and in the next section we shall teach you ways to draw (and ways not to draw) molecules—the handwriting of chemistry. *This section is extremely important* because it will teach you how to communicate chemistry, clearly and simply, throughout your life as a chemist.

Drawing molecules

Be realistic

Below is another organic structure—again, you may be familiar with the molecule it represents; it is a fatty acid commonly called linoleic acid.



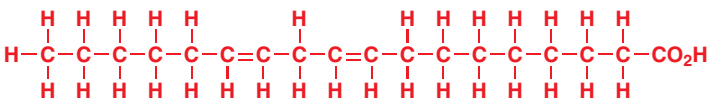
linoleic acid

We could also depict linoleic acid as

$$CH_3CH_2CH_2CH_2CH{=}CHCH_2CH{=}CHCH_2CH_2CH_2CH_2CH_2CH_2CH_2CO_2H$$

linoleic acid

or as

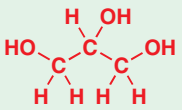


linoleic acid

Organic skeletons

Organic molecules left to decompose for millions of years in the absence of light and oxygen become literally carbon skeletons—crude oil, for example, is a mixture of molecules consisting of nothing but carbon and hydrogen, while coal consists of little else but carbon. Although the molecules in coal and oil differ widely in chemical structure, they have one thing in common: no functional groups. Many are very unreactive: about the only chemical reaction they can take part in is combustion, which, in comparison to most chemical reactions that take place in chemical laboratories, is an extremely violent process. In Chapter 5 we shall start to look at the way that functional groups direct the chemical reactions of molecules.

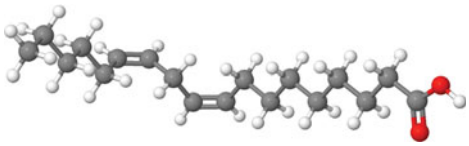
■ Three fatty acid molecules and one glycerol molecule combine to form the fats that store energy in our bodies and are used to construct the membranes around our cells. This particular fatty acid, linoleic acid, cannot be synthesized in the human body but must be an essential component of a healthy diet found, for example, in sunflower oil. Fatty acids differ in the length of their chains of carbon atoms, yet they have very similar chemical properties because they all contain the carboxylic acid functional group. We shall come back to fatty acids in Chapter 42.



glycerol

You may well have seen diagrams like these last two in older books—they used to be easy to print (in the days before computers) because all the atoms were in a line and all the angles were 90°. But are they realistic? We will consider ways of determining the shapes and structures of molecules in more detail in Chapter 3, but the picture below shows the structure of linoleic acid determined by X-ray crystallography.

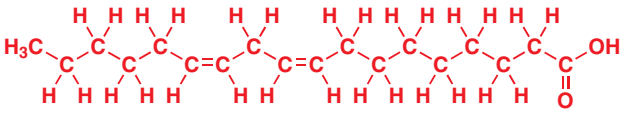
■ X-ray crystallography discovers the structures of molecules by observing the way X-rays bounce off atoms in crystalline solids. It gives clear diagrams with the atoms marked as circles and the bonds as rods joining them together.



X-ray structure of linoleic acid

Interactive linoleic acid structure

You can see that the chain of carbon atoms is not linear, but a zig-zag. Although our diagram is just a two-dimensional representation of this three-dimensional structure, it seems reasonable to draw it as a zig-zag too.



linoleic acid

This gives us our first guideline for drawing organic structures.

● Guideline 1

Draw chains of atoms as zig-zags.

Realism of course has its limits—the X-ray structure shows that the linoleic acid molecule is in fact slightly bent in the vicinity of the double bonds; we have taken the liberty of drawing it as a 'straight zig-zag'. Similarly, close inspection of crystal structures like this reveals that the angle of the zig-zag is about 109° when the carbon atom is not part of a double bond and 120° when it is. The 109° angle is the 'tetrahedral angle', the angle between two vertices of a tetrahedron when viewed from its centre. In Chapter 4 we shall look at why carbon atoms take up this particular arrangement of bonds. Our realistic drawing is a projection of a three-dimensional structure onto flat paper so we have to compromise.

Be economical

When we draw organic structures we try to be as realistic as we can be without putting in superfluous detail. Look at these three pictures.

1

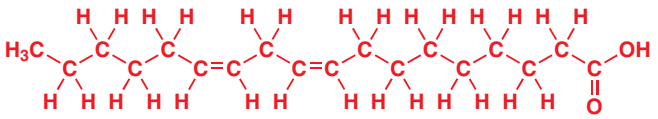
2

3

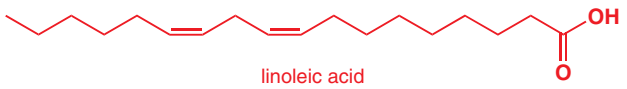
(1) is immediately recognizable as Leonardo da Vinci's Mona Lisa. You may not recognize (2)—it's also Leonardo da Vinci's Mona Lisa—this time viewed from above. The frame is very ornate, but the picture tells us as much about the painting as our rejected linear and 90° angle

diagrams did about our fatty acid. They're both correct—in their way—but sadly useless. What we need when we draw molecules is the equivalent of (3). It gets across the idea of the original, and includes all the detail necessary for us to recognize what it's a picture of, and leaves out the rest. And it was quick to draw—this picture was drawn in less than 10 minutes: we haven't got time to produce great works of art!

Because functional groups are the key to the chemistry of molecules, clear diagrams must emphasize the functional groups and let the hydrocarbon framework fade into the background. Compare the diagrams below:



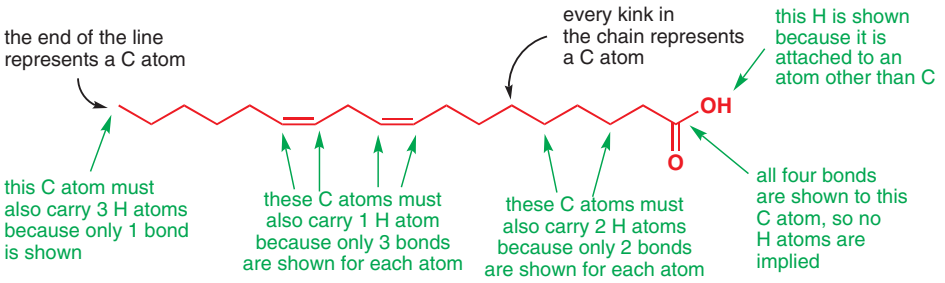
linoleic acid



linoleic acid

The second structure is the way that most organic chemists would draw linoleic acid. Notice how the important carboxylic acid functional group stands out clearly and is no longer cluttered by all those Cs and Hs. The zig-zag pattern of the chain is much clearer too. And this structure is much quicker to draw than any of the previous ones!

To get this diagram from the one above we've done two things. Firstly, we've got rid of all the hydrogen atoms attached to carbon atoms, along with the bonds joining them to the carbon atoms. Even without drawing the hydrogen atoms we know they're there—we assume that any carbon atom that doesn't appear to have its potential for four bonds satisfied is also attached to the appropriate number of hydrogen atoms. Secondly, we've rubbed out all the Cs representing carbon atoms. We're left with a zig-zag line, and we assume that every kink in the line represents a carbon atom, as does the end of the line.



We can turn these two simplifications into two more guidelines for drawing organic structures.

- **Guideline 2**

Miss out the Hs attached to carbon atoms, along with the C–H bonds (unless there is a good reason not to).

- **Guideline 3**

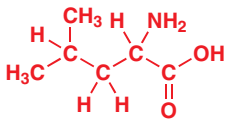
Miss out the capital Cs representing carbon atoms (unless there is a good reason not to).

■ What is 'a good reason not to'? One is if the C or H is part of a functional group. Another is if the C or H needs to be highlighted in some way, for example because it's taking part in a reaction. Don't be too rigid about these guidelines: they're not rules. It is better just to learn by example (you'll find plenty in this book): if it helps to clarify, put it in; if it clutters and confuses, leave it out. One thing you must remember, though: if you write a carbon atom as a letter C then you must add all the H atoms too. If you don't want to draw all the Hs, don't write C for carbon.

Be clear

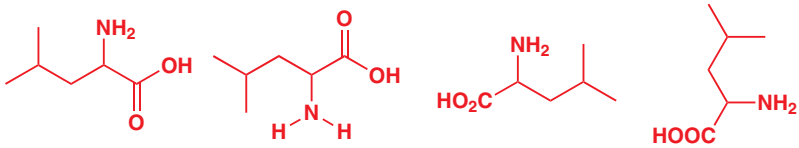
Try drawing some of the amino acids represented on p. 16 in a similar way, using the three guidelines. The bond angles at tetrahedral carbon atoms are about 109°. Make them look about 109° projected on to a plane! (120° is a good compromise, and it makes the drawings look neat.)

Start with leucine—earlier we drew it as the structure to the right. Get a piece of paper and do it now. Once you have done this, turn the page to see how your drawing compares with our suggestions.



leucine

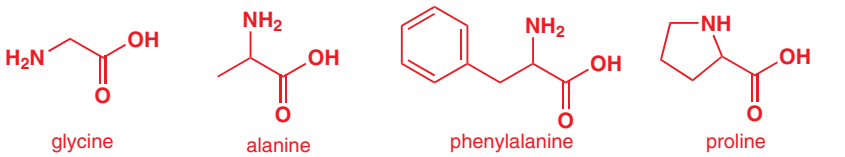
It doesn't matter which way up you've drawn it, but your diagram should look something like one of these structures below.



leucine leucine leucine leucine

The guidelines we gave were only guidelines, not rules, and it certainly does not matter which way round you draw the molecule. The aim is to keep the functional groups clear and let the skeleton fade into the background. That's why the last two structures are all right—the carbon atom shown as 'C' is part of a functional group (the carboxyl group) so it can stand out.

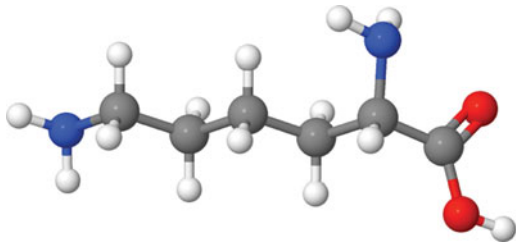
Now turn back to p. 16 and try redrawing the some of the other eight structures there using the guidelines. Don't look at our suggestions below until you've done them! Then compare your drawings with our suggestions.



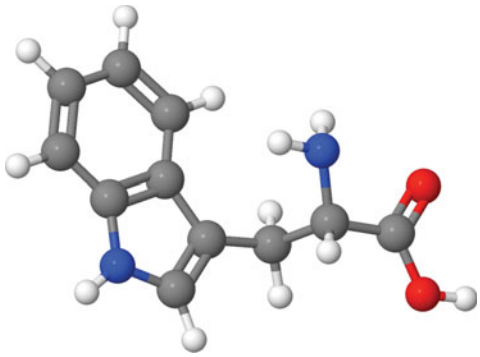
glycine alanine phenylalanine proline

tryptophan methionine lysine

Remember that these are only suggestions, but we hope you'll agree that this style of diagram looks much less cluttered and makes the functional groups much clearer than the diagrams on p. 16. Moreover, they still bear significant resemblance to the 'real thing'—compare these crystal structures of lysine and tryptophan with the structures shown above, for example.



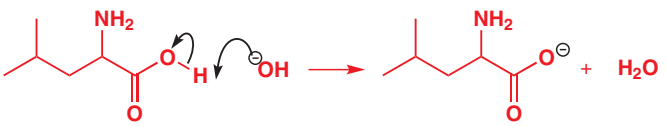
X-ray crystal structure of lysine



X-ray crystal structure of tryptophan

Structural diagrams can be modified to suit the occasion

You'll probably find that you want to draw the same molecule in different ways on different occasions to emphasize different points. Let's carry on using leucine as an example. We mentioned before that an amino acid can act as an acid or as a base. When it acts as an acid, a base (for example hydroxide, OH^-) removes H^+ from the carboxylic acid group in a reaction we can represent as:

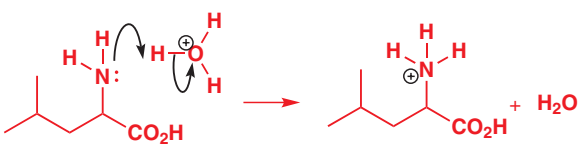


The product of this reaction has a negative charge on an oxygen atom. We have put it in a circle to make it clearer, and we suggest you do the same when you draw charges: + and – signs are easily mislaid. We shall discuss this type of reaction, the way in which reactions are drawn, and what the 'curly arrows' in the diagram mean in Chapter 5. But for now, notice that we drew out the CO_2H as the fragment on the left because we wanted to show how the O–H bond was broken when the base attacked. We modified our diagram to suit our own purposes.

■ Not all chemists put circles round their plus and minus charges—it's a matter of personal choice.

When leucine acts as a base, the amino (NH_2) group is involved. The nitrogen atom attaches itself to a proton, forming a new bond using its *lone pair*.

We can represent this reaction as:



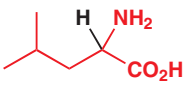
■ A lone pair is a pair of electrons that is not involved in a chemical bond. We shall discuss lone pairs in detail in Chapter 4. Again, don't worry about what the curly arrows in this diagram mean—we will cover them in detail in Chapter 5.

Notice how we drew in the lone pair this time because we wanted to show how it was involved in the reaction. The oxygen atoms of the carboxylic acid groups also have lone pairs but we didn't draw them in because they weren't relevant to what we were talking about. Neither did we feel it was necessary to draw CO_2H in full this time because none of the atoms or bonds in the carboxylic acid functional group was involved in the reaction.

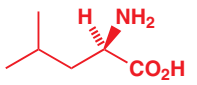
Structural diagrams can show three-dimensional information on a two-dimensional page

Of course, all the structures we have been drawing give only an idea of the real structure of the molecules. For example, the carbon atom between the NH_2 group and the CO_2H group of leucine has a tetrahedral arrangement of atoms around it, a fact which we have so far completely ignored.

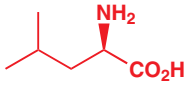
We might want to emphasize this fact by drawing in the hydrogen atom we missed out at this point, as in structure 1 (in the right-hand margin). We can then show that one of the groups attached to this carbon atom comes towards us, out of the plane of the paper, and the other one goes away from us, into the paper.

1


There are several ways of doing this. In structure 2, the bold, wedged bond suggests a perspective view of a bond coming towards you, while the hashed bond suggests a bond fading away from you. The other two 'normal' bonds are in the plane of the paper.

2


Alternatively we could miss out the hydrogen atom and draw something a bit neater, although slightly less realistic, as in structure 3. We can assume the missing hydrogen atom is behind the plane of the paper because that is where the 'missing' vertex of the tetrahedron of atoms attached to the carbon atom lies. When you draw diagrams like these to indicate the three dimensional shape of the molecule, try to keep the hydrocarbon framework in the

3


plane of the paper and allow functional groups and other branches to project forwards out of the paper or backwards into it.

These conventions allow us to give an idea of the three-dimensional shape (stereochemistry) of any organic molecule—you have already seen them in use in the diagram of the structure of palytoxin at the beginning of this chapter.

We shall look in more detail at the shapes of molecules—their *stereochemistry*—in Chapter 14.

Reminder

Organic structural drawings should be *realistic, economical*, and *clear*.

We gave you three guidelines to help you achieve this when you draw structures:

- **Guideline 1: Draw chains of atoms as zig-zags.**
- **Guideline 2: Miss out the Hs attached to the carbon atoms along with the C–H bonds.**
- **Guideline 3: Miss out the capital Cs representing carbon atoms.**

The guidelines we have given and the conventions we have illustrated in this section have grown up over decades. They are not arbitrary pronouncements by some official body but are used by organic chemists because they work! We guarantee to follow them for the rest of the book—try to follow them yourself whenever you draw an organic structure. Before you ever draw a capital C or a capital H again, ask yourself whether it's really necessary!

Now that we have considered how to draw structures, we can return to some of the structural types that we find in organic molecules. Firstly, we'll talk about hydrocarbon frameworks, then about functional groups.

Hydrocarbon frameworks

Carbon as an element is unique in the variety of structures it can form. It is unusual because it forms strong, stable bonds to the majority of elements in the periodic table, including itself. It is this ability to form bonds to itself that leads to the variety of organic structures that exist, and indeed to the possibility of life existing at all. Carbon may make up only 0.2% of the earth's crust, but it certainly deserves a whole branch of chemistry all to itself.

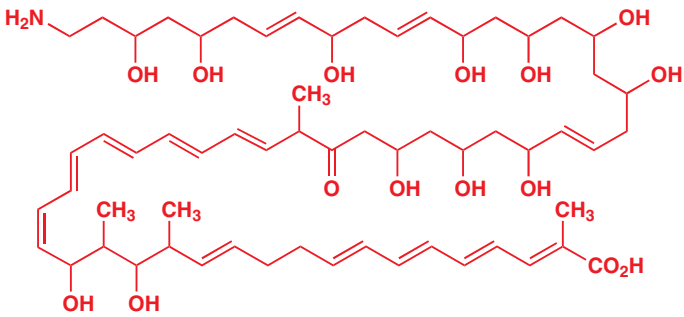
Chains

The simplest class of hydrocarbon frameworks contains just chains of atoms. The fatty acids we met earlier have hydrocarbon frameworks made of zig-zag chains of atoms, for example. Polythene is a polymer whose hydrocarbon framework consists entirely of chains of carbon atoms. The wiggly line at each end of this structure shows that we have drawn a piece in the middle of the polythene molecule. The structure continues indefinitely beyond the wiggly lines.

a section of the structure of polythene

Interactive structure of polythene

At the other end of the spectrum of complexity is this antibiotic, extracted from a fungus in 1995 and aptly named linearmycin as it has a long linear chain. The chain of this antibiotic is so long that we have to wrap it round two corners just to get it on the page. We haven't drawn whether the CH_3 and OH groups are in front of or behind the plane of the paper because, at the time of writing this book, the stereochemistry of linearmycin is unknown.



linearmycin

■ Notice we've drawn in four groups as CH_3—we did this because we didn't want them to get overlooked in such a large structure. They are the only tiny branches off this long winding trunk.

Names for carbon chains

It is often convenient to refer to a chain of carbon atoms by a name indicating its length. You have probably met some of these names before in the names of the simplest organic molecules, the alkanes. There are also commonly used abbreviations for these names: these can be very useful in both writing about chemistry and in drawing chemical structures, as we shall see shortly.

Names and abbreviations for carbon chains

Number of carbon atoms in chain	Name of group	Formula†	Abbreviation	Name of alkane (= chain + H)
1	methyl	$-CH_3$	Me	methane
2	ethyl	$-CH_2CH_3$	Et	ethane
3	propyl	$-CH_2CH_2CH_3$	Pr	propane
4	butyl	$-(CH_2)_3CH_3$	Bu	butane
5	pentyl	$-(CH_2)_4CH_3$	—‡	pentane
6	hexyl	$-(CH_2)_5CH_3$	—‡	hexane
7	heptyl	$-(CH_2)_6CH_3$	—‡	heptane
8	octyl	$-(CH_2)_7CH_3$	—‡	octane
9	nonyl	$-(CH_2)_8CH_3$	—‡	nonane
10	decyl	$-(CH_2)_9CH_3$	—‡	decane

† This representation is not recommended, except for CH_3. ‡ Names for longer chains are not commonly abbreviated.

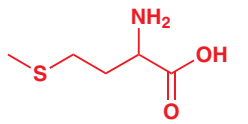
■ The names for shorter chains (which you must learn) exist for historical reasons; for chains of five or more carbon atoms, the systematic names are based on Greek number names.

Organic elements

You may notice that the abbreviations for the names of carbon chains look very much like the symbols for chemical elements: this is deliberate, and these symbols are sometimes called 'organic elements'. They can be used in chemical structures just like element symbols. It is often convenient to use the 'organic element' symbols for short carbon chains for tidiness. Here are some examples. Structure 1 to the right shows how we drew the structure of the amino acid methionine on p. 20. The stick representing the methyl group attached to the sulfur atom does, however, look a little odd. Most chemists would draw methionine as structure 2, with 'Me' representing the CH_3 (methyl) group. Tetraethyllead used to be added to petrol to prevent engines 'knocking', until it was shown to be a health hazard. Its structure (as you might easily guess from the name) is easy to write as $PbEt_4$ or Et_4Pb.

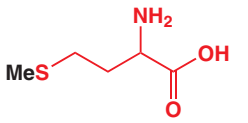
Remember that these symbols (and names) can be used only for terminal chains of atoms. We couldn't abbreviate the structure of lysine to 3, for example, because Bu represents 4 and not 5.

1

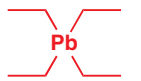


methionine

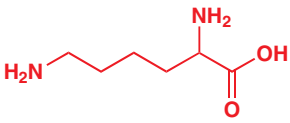
2



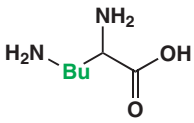
methionine



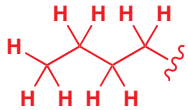
tetraethyllead



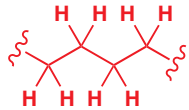
lysine



3 NOT CORRECT

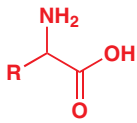


4 C_4H_9 = Bu



5 C_4H_8 NOT Bu

Before leaving carbon chains, we must mention one other very useful organic element symbol, R. R in a structure can mean *anything*—it's a sort of wild card. For example, structure 6 would indicate any amino acid, if R = H it is glycine, if R = Me it is alanine... As we've mentioned before, and you will see later, the reactivity of organic molecules is so dependent on their functional groups that the rest of the molecule can be irrelevant. In these cases, we can choose just to call it R.

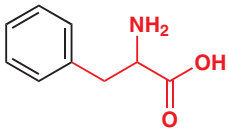


6 amino acid

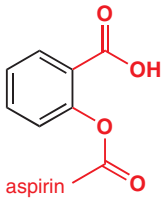
Carbon rings

benzene

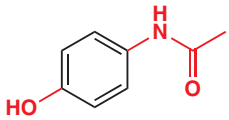
Rings of atoms are also common in organic structures. You may have heard the famous story of Auguste Kekulé first realizing that benzene has a ring structure when he dreamed of snakes biting their own tails. You have met benzene rings in phenylalanine and aspirin. Paracetamol also has a structure based on a benzene ring.



phenylalanine



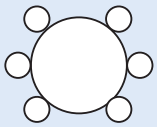
aspirin



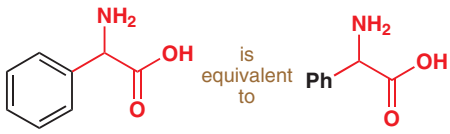
paracetamol

Benzene's ring structure

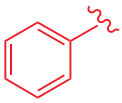
In 1865, August Kekulé presented a paper at the Academie des Sciences in Paris suggesting a cyclic structure for benzene, the inspiration for which he ascribed to a dream. However, was Kekulé the first to suggest that benzene was cyclic? Some believe not and credit an Austrian schoolteacher, Josef Loschmidt, with the first depiction of cyclic benzene structures. In 1861, 4 years before Kekulé's 'dream', Loschmidt published a book in which he represented benzene as a set of rings. It is not certain whether Loschmidt or Kekulé—or even a Scot named Archibald Couper—got it right first.

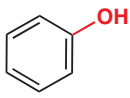

Loschmidt's structure for benzene

When a benzene ring is attached to a molecule by only *one* of its carbon atoms (as in phenylalanine, but not paracetamol or aspirin), we can call it a 'phenyl' group and give it the organic element symbol Ph.

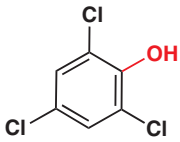


Any compound containing a benzene ring or a related (Chapter 7) ring system is known as 'aromatic', and another useful organic element symbol related to Ph is Ar (for 'aryl'). While Ph always means C_6H_5, Ar can mean any *substituted* phenyl ring, in other words phenyl with any number of the hydrogen atoms replaced by other groups. Of course Ar = argon too but there is no confusion as there are no organic compounds of argon.

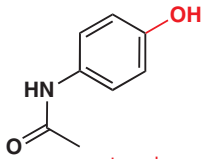

the phenyl group, Ph



PhOH = phenol



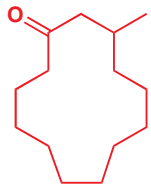
2,4,6-trichlorophenol



paracetamol

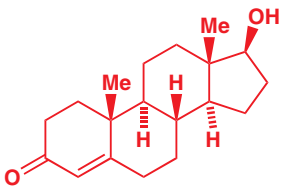
For example, while PhOH always means phenol, ArOH could mean phenol, 2,4,6-trichlorophenol (the antiseptic TCP), paracetamol, or aspirin (among many other substituted phenols). Like R, the 'wild card' alkyl group, Ar is a 'wild card' *aryl* group.

The compound known as muscone has only relatively recently been made in the laboratory. It is the pungent aroma that makes up the base-note of musk fragrances. Before chemists had determined its structure and devised a laboratory synthesis the only source of musk was the musk deer, now rare for this very reason. Muscone's skeleton is a 13-membered ring of carbon atoms.

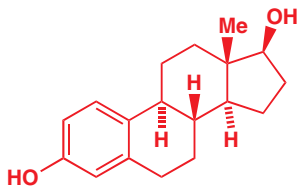


muscone

The steroid hormones have several (usually four) rings fused together. These hormones are testosterone and oestradiol, the important human male and female sex hormones.



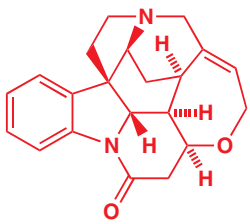
testosterone



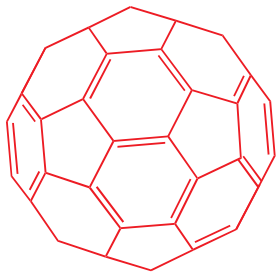
oestradiol

■ A reminder: solid wedge-shaped bonds are coming towards us out of the paper while cross-hatched bonds are going back into the page away from us.

Some ring structures are much more complicated. The potent poison strychnine is a tangle of interconnecting rings.



strychnine


buckminsterfullerene

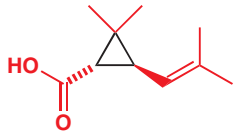
Interactive structures of testosterone, oestradiol, strychnine, and buckminsterfullerene

One of the most elegant ring structures is shown above and is known as buckminsterfullerene. It consists solely of 60 carbon atoms in rings that curve back on themselves to form a football-shaped cage. Count the number of bonds at any junction and you will see they add up to four so no hydrogens need be added. This compound is C_{60}. Note that you can't see all the atoms as some are behind the sphere.

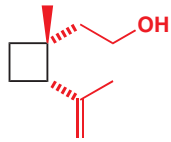
Buckminsterfullerene

Buckminsterfullerene is named after the American inventor and architect Richard Buckminster Fuller, who designed the structures known as 'geodesic domes'.

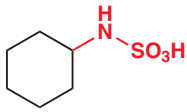
Rings of carbon atoms are given names starting with 'cyclo', followed by the name for the carbon chain with the same number of carbon atoms. Structure 1 shows chrysanthemic acid, part of the naturally occurring pesticides called pyrethrins (an example appears in Chapter 1), which contains a cyclopropane ring. Propane has three carbon atoms. Cyclopropane is a three-membered ring. Grandisol (structure 2), an insect pheromone used by male boll weevils to attract females, has a structure based on a cyclobutane ring. Butane has four carbon atoms. Cyclobutane is a four-membered ring. Cyclamate (structure 3), formerly used as an artificial sweetener, contains a cyclohexane ring. Hexane has six carbon atoms. Cyclohexane is a six-membered ring.



1 chrysanthemic acid



2 grandisol

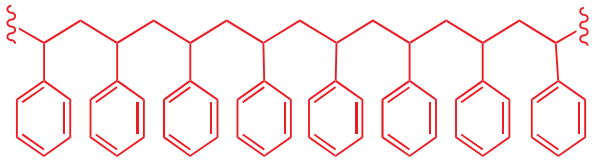


3 cyclamate

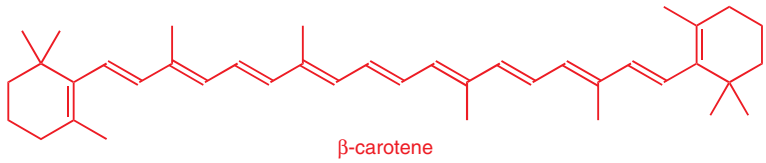
Branches

Hydrocarbon frameworks rarely consist of single rings or chains, but are often branched. Rings, chains, and branches are all combined in structures like that of the marine toxin palytoxin that we met at the beginning of the chapter, polystyrene, a polymer made of six-membered rings dangling from linear carbon chains, or of β-carotene, the compound that makes carrots orange.

Interactive structure of polystyrene



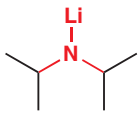
part of the structure of polystyrene



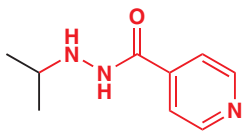
β-carotene

the isopropyl group *i*-Pr

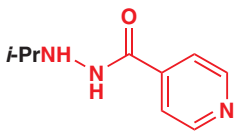
Just like some short straight carbon chains, some short branched carbon chains are given names and organic element symbols. The most common is the isopropyl group. Lithium diisopropylamide (also called LDA) is a strong base commonly used in organic synthesis.



lithium diisopropylamide (LDA)
is equivalent to LiN*i*-Pr$_2$



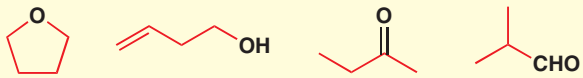
is equivalent to

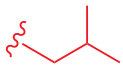


iproniazid

Notice how the 'propyl' part of 'isopropyl' still indicates three carbon atoms; they are just joined together in a different way—in other words, as an *isomer* of the straight chain propyl group. Sometimes, to avoid confusion, the straight chain alkyl groups are called '*n*-alkyl' (for example, *n*-Pr, *n*-Bu)—*n* for 'normal'—to distinguish them from their branched counterparts. Iproniazid is an antidepressant drug with *i*-Pr in both structure and name. 'Isopropyl' may be abbreviated to *i*-Pr, ^{i}Pr, or Pr^{i}. We shall use the first in this book, but you may see the others used elsewhere.

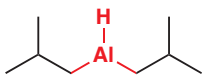
● ***Isomers*** **are molecules with the same kinds and numbers of atoms joined up in different ways. *n*-propanol, *n*-PrOH, and isopropanol, *i*-PrOH, are isomeric alcohols. Isomers need not have the same functional groups—these compounds are all isomers of C_4H_8O:**



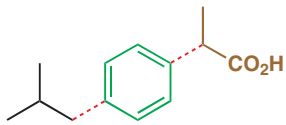


the isobutyl group *i*-Bu

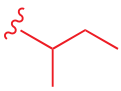
The isobutyl (*i*-Bu) group is a CH_2 group joined to an *i*-Pr group. It is *i*-PrCH_2–. Two isobutyl groups are present in the reducing agent diisobutyl aluminium hydride (DIBAL). The painkiller ibuprofen (marketed as Nurofen®) contains an isobutyl group. Notice how the invented name ibuprofen is a medley of 'ibu' (from *i*-Bu for isobutyl) + 'pro' (for propyl, the three-carbon unit shown in brown) + 'fen' (for the phenyl ring). We will talk about the way in which compounds are named later in this chapter.



diisobutyl aluminium hydride (DIBAL)
is equivalent to **HAl*i*-Bu$_2$**

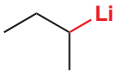


Ibuprofen



the *sec*-butyl group, *s*-Bu

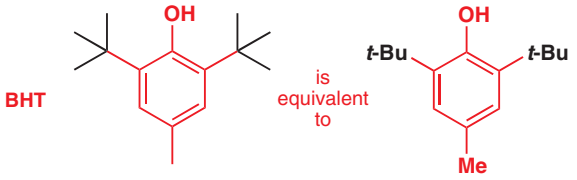
There are two more isomers of the butyl group, both of which have common names and abbreviations. The *sec*-butyl group (*s*-butyl or *s*-Bu) has a methyl and an ethyl group joined to the same carbon atom. It appears in an organolithium compound, *sec*-butyl lithium, used to introduce lithium atoms into organic molecules.



is equivalent to ***s*-BuLi**

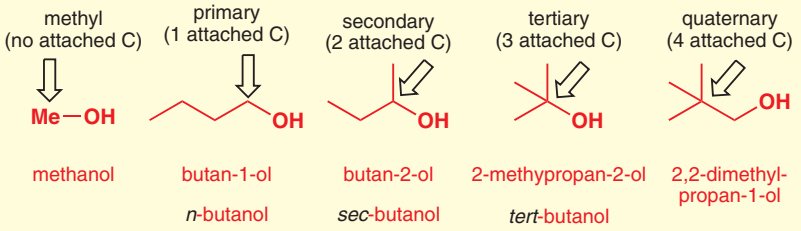
The *tert*-butyl group (*t*-butyl or *t*-Bu) group has three methyl groups joined to the same carbon atom. Two *t*-Bu groups are found in butylated hydroxy toluene (BHT E321), an antioxidant added to some processed foods.

the *tert*-butyl group *t*-Bu



● Primary, secondary, and tertiary

The prefixes *sec* and *tert* are really short for secondary and tertiary, terms that refer to the carbon atom that attaches these groups to the rest of the molecular structure.



A primary carbon atom is attached to only one other C atom, a secondary to two other C atoms, and so on. This means there are five types of carbon atom. These names for bits of hydrocarbon framework are more than just useful ways of writing or talking about chemistry. They tell us something fundamental about the molecule and we shall use them when we describe reactions.

This quick architectural tour of some of the molecular edifices built by nature and by humans serves just as an introduction to some of the hydrocarbon frameworks you will meet in the rest of this chapter and this book. Yet, fortunately for us, however complicated the hydrocarbon framework might be, it serves only as a support for the functional groups. And, by and large, a functional group in one molecule behaves in much the same way as it does in another molecule. What we now need to do, and we start in the next section, is to introduce you to some functional groups and explain why it is that their attributes are the key to understanding organic chemistry.

Functional groups

If you bubble ethane gas (CH_3CH_3, or EtH) through acids, bases, oxidizing agents, reducing agents—in fact almost any chemical you can think of—it will remain unchanged. Just about the only thing you can do with it is burn it. Yet ethanol (CH_3CH_2OH, or preferably EtOH—structure in the margin) not only burns, it reacts with acids, bases, and oxidizing agents.

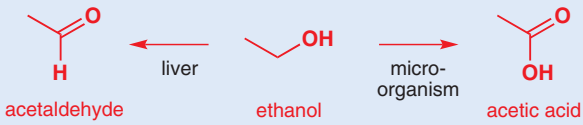
ethanol

The difference between ethanol and ethane is the functional group—the OH, or hydroxyl group. We know that these chemical properties (being able to react with acids, bases, and oxidizing agents) are properties of the hydroxyl group and not just of ethanol because other compounds containing OH groups (in other words, other alcohols) have similar properties, whatever their hydrocarbon frameworks.

Your understanding of functional groups will be the key to your understanding of organic chemistry. We shall therefore now go on to meet some of the most important functional groups. We won't say much about the properties of each group; that will come in Chapter 5

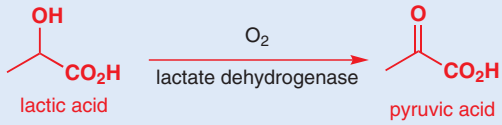
Ethanol

The reaction of ethanol with oxidizing agents makes vinegar from wine and sober people from drunk ones. In both cases, the oxidizing agent is oxygen from the air, catalysed by an enzyme in a living system. The oxidation of ethanol by micro-organisms that grow in wine left open to the air leads to acetic acid (ethanoic acid) while the oxidation of ethanol by the liver gives acetaldehyde (ethanal).



Human metabolism and oxidation

The human metabolism makes use of the oxidation of alcohols to render harmless other toxic compounds containing the OH group. For example, lactic acid, produced in muscles during intense activity, is oxidized by an enzyme called lactate dehydrogenase to the metabolically useful compound pyruvic acid.



and later. Your task at this stage is to learn to recognize them when they appear in structures, so make sure you learn their names. The classes of compound associated with some functional groups also have names, for example compounds containing the hydroxyl group are known as alcohols. Learn these names too as they are more important than the systematic names of individual compounds. We've told you a few snippets of information about each group to help you to get to know something of the group's character.

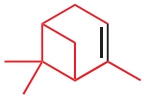
pentane

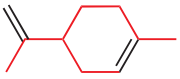
hexane

isooctane

Alkanes contain no functional groups

The alkanes are the simplest class of organic molecules because they contain no functional groups. They are extremely unreactive and therefore rather boring as far as the organic chemist is concerned. However, their unreactivity can be a bonus, and alkanes such as pentane and hexane are often used as solvents, especially for the purification of organic compounds. Just about the only thing alkanes will do is burn—methane, propane, and butane are all used as domestic fuels, and petrol is a mixture of alkanes containing largely isooctane.

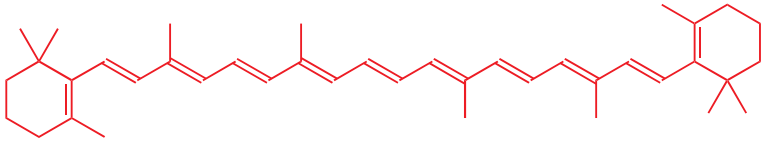

α-pinene


limonene

Alkenes (sometimes called olefins) contain C=C double bonds

It may seem strange to classify a type of bond as a functional group, but you will see later that C=C double bonds impart reactivity to an organic molecule just as functional groups consisting of, say, oxygen or nitrogen atoms do. Some of the compounds produced by plants and used by perfumers are alkenes (see Chapter 1). For example, pinene has a smell evocative of pine forests, while limonene smells of citrus fruits.

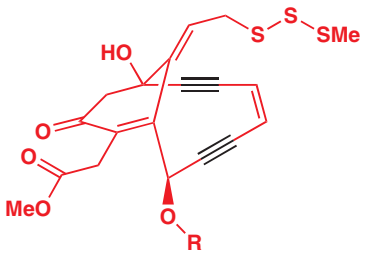
You've already met the orange pigment β-carotene. Eleven C=C double bonds make up most of its structure. Coloured organic compounds often contain chains or rings of C=C double bonds like this. In Chapter 7 you will find out why this is so.


β-carotene

Alkynes contain C≡C triple bonds

Just like C=C double bonds, C≡C triple bonds have a special type of reactivity associated with them, so it's useful to call a C≡C triple bond a functional group. Alkynes are linear so we

draw them with four carbon atoms in a straight line. Alkynes are not as widespread in nature as alkenes, but one fascinating class of compounds containing C≡C triple bonds is a group of antitumour agents discovered during the 1980s. Calicheamicin is a member of this group. The high reactivity of this combination of functional groups enables calicheamicin to attack DNA and prevent cancer cells from proliferating. For the first time we have drawn a molecule in three dimensions, with two bonds crossing one another—can you see the shape?



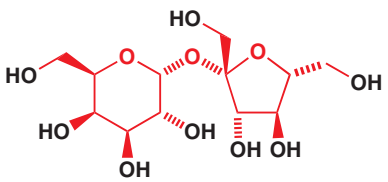
calicheamicin
(R = a string of sugar molecules)

Saturated and unsaturated

In an alkane, each carbon atom is joined to four other atoms (C or H). It has no potential for forming more bonds and is therefore **saturated**. In alkenes, the carbon atoms making up the C=C double bond are attached to only three atoms each. They still have the potential to bond with one more atom, and are therefore unsaturated. In general, carbon atoms attached to four other atoms are saturated; those attached to three, two, or one are **unsaturated**. Remember that R may mean any alkyl group.

Alcohols (R–OH) contain a hydroxyl (OH) group

We've already talked about the hydroxyl group in ethanol and other alcohols. Carbohydrates are peppered with hydroxyl groups; sucrose has eight of them, for example (a more three-dimensional picture of the sucrose molecule appears in Chapter 1, p.3).



sucrose

Interactive structure of sucrose

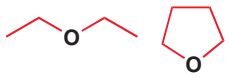
Molecules containing hydroxyl groups are often soluble in water, and living things often attach sugar groups, containing hydroxyl groups, to otherwise insoluble organic compounds to keep them in solution in the cell. Calicheamicin, a molecule we have just mentioned, contains a string of sugars for just this reason. The liver carries out its task of detoxifying unwanted organic compounds by repeatedly hydroxylating them until they are water soluble, and they are then excreted in the bile or urine.

Ethers (R^1–O–R^2) contain an alkoxy group (–OR)

The name ether refers to any compound that has two alkyl groups linked through an oxygen atom. 'Ether' is also used as an everyday name for diethyl ether, Et_2O. You might compare this use of the word 'ether' with the common use of the word 'alcohol' to mean ethanol. Diethyl ether is a highly flammable solvent that boils at only 35°C. It used to be used as an anaesthetic. Tetrahydrofuran (THF) is another commonly used solvent and is a cyclic ether.

Brevetoxin B (overleaf) is a fascinating naturally occurring compound that was synthesized in the laboratory in 1995. It is packed with ether functional groups in ring sizes from 6 to 8.

■ If we want a structure to contain more than one 'R', we give the Rs numbers and call them R^1, R^2... Thus R^1–O–R^2 means an ether with two different unspecified alkyl groups. (*Not* R_1, R_2..., which would mean 1 × R, 2 × R...)



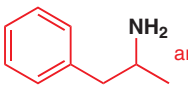
diethyl ether
'ether'

THF

Amines (R–NH_2) contain the amino (NH_2) group

We met the amino group when we were discussing the amino acids: we mentioned that it was this group that gave these compounds their basic properties. Amines often have powerful fishy smells: the smell of putrescine is particularly foul. It is formed as meat decays. Many neurologically active compounds are also amines: amphetamine is a notorious stimulant.

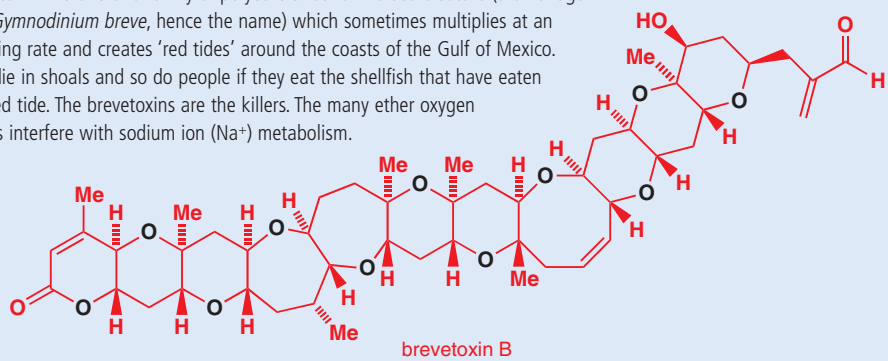
putrescine



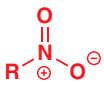
amphetamine

Brevetoxin B

Brevetoxin B is one of a family of polyethers found in a sea creature (a dinoflagellate *Gymnodinium breve*, hence the name) which sometimes multiplies at an amazing rate and creates 'red tides' around the coasts of the Gulf of Mexico. Fish die in shoals and so do people if they eat the shellfish that have eaten the red tide. The brevetoxins are the killers. The many ether oxygen atoms interfere with sodium ion (Na^+) metabolism.



brevetoxin B



the nitro group

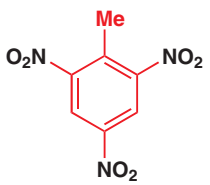
nitrogen cannot have five bonds!

incorrect structure for the nitro group

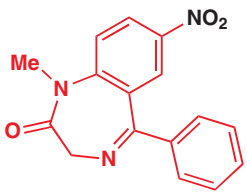
Nitro compounds (R–NO_2) contain the nitro group (NO_2)

The nitro group (NO_2) is sometimes incorrectly drawn with five bonds to nitrogen which as you will see in Chapter 4 is impossible. Make sure you draw it correctly when you need to draw it out in detail. If you write just NO_2 you are all right!

Several nitro groups in one molecule can make it quite unstable and even explosive. Three nitro groups give the most famous explosive of all, trinitrotoluene (TNT), its kick. However, functional groups refuse to be stereotyped. Nitrazepam also contains a nitro group, but this compound is marketed as Mogadon®, the sleeping pill.



TNT



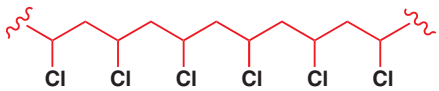
nitrazepam

Alkyl halides (fluorides R–F, chlorides R–Cl, bromides R–Br, or iodides R–I) contain the fluoro, chloro, bromo, or iodo groups

These four functional groups have similar properties, although alkyl iodides are the most reactive and alkyl fluorides the least. Polyvinyl chloride (PVC) is one of the most widely used polymers—it has a chloro group on every other carbon atom along a linear hydrocarbon framework. Methyl iodide (MeI), on the other hand, is a dangerous carcinogen since it reacts with DNA and can cause mutations in the genetic code. These compounds are also known as haloalkanes (fluoroalkanes, chloroalkanes, bromoalkanes, or iodoalkanes).

Interactive structure of PVC

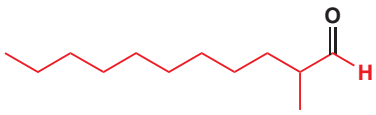
Because alkyl halides have similar properties, chemists use yet another wild card organic element symbol, X, as a convenient substitute for Cl, Br, or I and sometimes F: R–X is any alkyl halide.



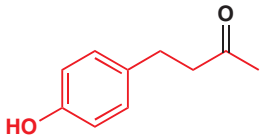
a section of the structure of PVC

Aldehydes (R–CHO) and ketones (R^1–CO–R^2) contain the carbonyl group C=O

Aldehydes can be formed by oxidizing alcohols—in fact the liver detoxifies ethanol in the bloodstream by oxidizing it first to acetaldehyde (ethanal, CH_3CHO) (see p. 28). Acetaldehyde in the blood is the cause of hangovers. Aldehydes often have pleasant smells—2-methylundecanal is a key component of the fragrance of Chanel No. 5, and 'raspberry ketone' is the major component of the flavour and smell of raspberries.



2-methylundecanal



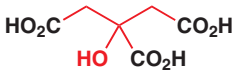
'raspberry ketone'

■ –CHO represents

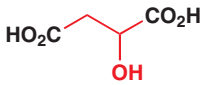
When we write aldehydes as R–CHO, we have no choice but to write in the C and H (because they're part of the functional group)—one important instance where you should ignore Guideline 3 for drawing structures. Another point: always write R–CHO and never R–COH, which looks too much like an alcohol.

Carboxylic acids (R–CO_2H) contain the carboxyl group CO_2H

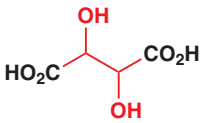
As their name implies, compounds containing the carboxylic acid (CO_2H) group can react with bases, losing a proton to form carboxylate salts. Edible carboxylic acids have sharp flavours and several are found in fruits—citric, malic, and tartaric acids are found in lemons, apples, and grapes, respectively.



citric acid



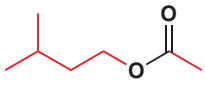
malic acid



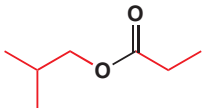
tartaric acid

Esters (R^1–CO_2R^2) contain a carboxyl group with an extra alkyl group (CO_2R)

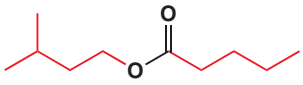
Fats are esters; in fact they contain three ester groups. They are formed in the body by condensing glycerol, a compound with three hydroxyl groups, with three fatty acid molecules. Other, more volatile, esters have pleasant, fruity smells and flavours. These three are components of the flavours of bananas, rum, and apples:



isopentyl acetate (bananas)



isobutyl propionate (rum)

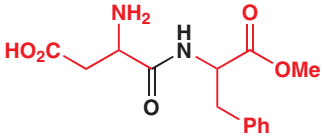


isopentyl valerate (apples)

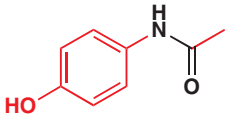
■ The terms 'saturated fats' and 'unsaturated fats' are familiar—they refer to whether the R groups are saturated (no C=C double bonds) or unsaturated (contain C=C double bonds)—see the box on p. 29. Fats containing R groups with several double bonds (for example those that are esters formed from linoleic acid, which we met at the beginning of this chapter) are known as 'polyunsaturated'.

Amides (R–$CONH_2$, R^1–$CONHR^2$, or R^1–$CONR^2R^3$)

Proteins are amides: they are formed when the carboxylic acid group of one amino acid condenses with the amino group of another to form an amide linkage (also known as a peptide bond). One protein molecule can contain hundreds of amide bonds. Aspartame, the artificial sweetener marketed as NutraSweet®, on the other hand, contains just two amino acids, aspartic acid and phenylalanine, joined through one amide bond. Paracetamol is also an amide.



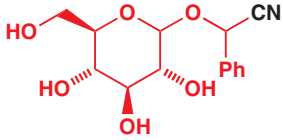
aspartame



paracetamol

Nitriles or cyanides (R–CN) contain the cyano group –C≡N

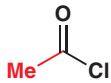
Nitrile groups can be introduced into molecules by reacting potassium cyanide with alkyl halides. The organic nitrile group has quite different properties from those associated with lethal inorganic cyanide: laetrile, for example, is extracted from apricot kernels, and was once developed as an anticancer drug.



laetrile

Acyl chlorides (acid chlorides, R–COCl)

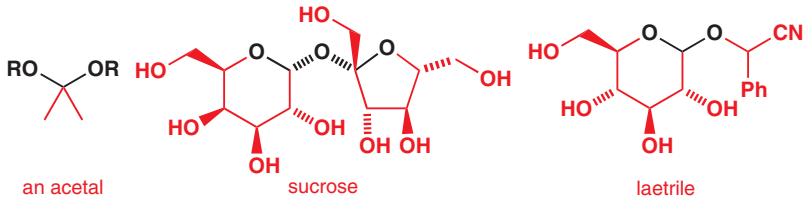
Acyl chlorides are reactive compounds used to make esters and amides. They are derivatives of carboxylic acids with the –OH replaced by –Cl, and are too reactive to be found in nature.



acetyl chloride

Acetals

Acetals are compounds with two single-bonded oxygen atoms attached to the same carbon atom. Many sugars are acetals, as is laetrile, which you have just met.



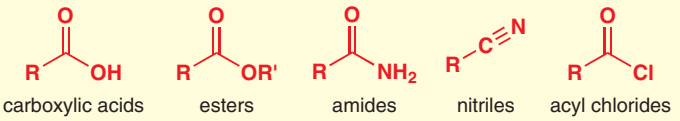
Carbon atoms carrying functional groups can be classified by oxidation level

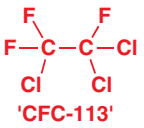
A heteroatom is an atom that is not C or H

You've seen that a functional group is essentially any deviation from an alkane structure, either because the molecule has fewer hydrogen atoms than an alkane (alkenes, alkynes) or because it contains a collection of atoms that are not C and not H. There is a useful term for these 'different' atoms: heteroatoms. A heteroatom is any atom in an organic molecule other than C or H.

All functional groups are different, but some are more different than others. For example, the structures of a carboxylic acid, an ester, and an amide are all very similar: in each case the carbon atom carrying the functional group is bonded to two **heteroatoms**, one of the bonds being a double bond. You will see in Chapter 10 that this similarity in structure is mirrored in the reactions of these three types of compounds and in the ways in which they can be interconverted. Carboxylic acids, esters, and amides can be changed one into another by reaction with simple reagents such as water, alcohols, or amines plus appropriate catalysts. To change them into aldehydes or alcohols requires a different type or reagent, a reducing agent (a reagent which adds hydrogen atoms). We say that the carbon atoms carrying functional groups that can be interconverted without the need for reducing agents (or oxidizing agents) have the same oxidation level—in this case, we call it the 'carboxylic acid oxidation level'.

● The carboxylic oxidation level

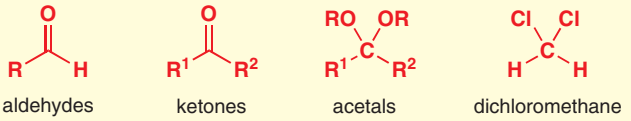




In fact, amides can quite easily be converted into nitriles just by dehydration (removal of water), so we must give nitrile carbon atoms the same oxidation level as carboxylic acids, esters, and amides. Maybe you're beginning to see the structural similarity between these four functional groups that you could have used to assign their oxidation level? In all four cases, the carbon atom has *three* bonds to heteroatoms, and only one to C or H. It doesn't matter how many heteroatoms there are, just how many bonds to them. Having noticed this, we can also assign both carbon atoms in 'CFC-113', one of the environmentally unfriendly aerosol propellants/refrigerants that have caused damage to the earth's ozone layer, to the carboxylic acid oxidation level.

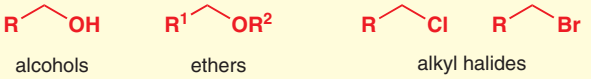
■ Don't confuse oxidation level with oxidation state. Oxidation level is determined by the number of heteroatoms bonded to carbon while oxidation state is determined by the number of bonds to carbon, including those to C and H. In all of these compounds, carbon has four bonds and is in oxidation state +4.

● The aldehyde oxidation level



Aldehydes and ketones contain a carbon atom with *two* bonds to heteroatoms; they are at the 'aldehyde oxidation level'. The common laboratory solvent dichloromethane CH_2Cl_2 also has two bonds to heteroatoms, so it too contains a carbon atom at the aldehyde oxidation level, as do acetals.

● The alcohol oxidation level



Alcohols, ethers, and alkyl halides have a carbon atom with only *one* single bond to a heteroatom. We assign these the 'alcohol oxidation level', and they are all easily made from alcohols without oxidation or reduction.

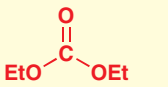
● The alkane oxidation level

methane

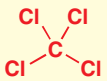
We must include simple alkanes, which have no bonds to heteroatoms, as an 'alkane oxidation level'.

● The carbon dioxide oxidation level

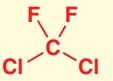
carbon dioxide



diethyl carbonate useful reagent for adding ester groups



carbon tetrachloride formerly used as a dry cleaning fluid

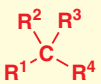
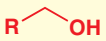
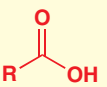
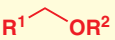
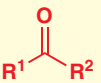
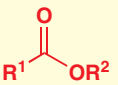
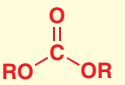
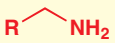
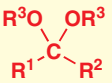
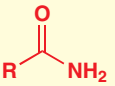
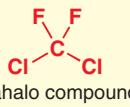
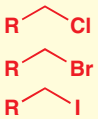
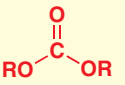


'CFC-12' one of the aerosol propellants causing damage to the ozone layer

The small class of compounds that have a carbon atom with four bonds to heteroatoms is related to CO_2 and best described as at the carbon dioxide oxidation level.

Alkenes and alkynes obviously don't fit easily into these categories as they have no bonds to heteroatoms. Alkenes can be made from alcohols by dehydration without any oxidation or reduction so it seems sensible to put them in the alcohol column. Similarly, alkynes and aldehydes are related by hydration/dehydration without oxidation or reduction.

● Summary: Important functional groups and oxidation level

Zero bonds to heteroatoms: alkane oxidation level	One bond to heteroatoms: alcohol oxidation level	Two bonds to heteroatoms: aldehyde oxidation level	Three bonds to heteroatoms: carboxylic acid oxidation level	Four bonds to heteroatoms: carbon dioxide oxidation level
 alkanes	 alcohols	aldehydes	 carboxylic acids	O=C=O carbon dioxide
	 ethers	 ketones	 esters	 carbonates
	 amines	 acetals	 amides	 tetrahalo compounds
	 alkyl halides	alkynes	nitriles	 ureas
	alkenes		acyl chlorides	

Naming compounds

So far, we have talked a lot about compounds by name. Many of the names we've used (palytoxin, muscone, brevetoxin) are simple names given to complicated molecules without regard for the actual structure or function of the molecule—these three names, for example, are all derived from the name of the organism from which the compound was first extracted.

They are known as **trivial names**, not because they are unimportant, but because they are used in everyday scientific conversation.

Names like this are fine for familiar compounds that are widely used and referred to by chemists, biologists, doctors, nurses, and perfumers alike. But there are over 16 million known organic compounds. They can't all have simple names, and no-one would remember them if they did. For this reason, the International Union of Pure and Applied Chemistry (IUPAC) have developed **systematic nomenclature**, a set of rules that allows any compound to be given a unique name that can be deduced directly from its chemical structure. Conversely, a chemical structure can be deduced from its systematic name.

The problem with systematic names is that they tend to be grotesquely unpronounceable for anything but the most simple molecules. In everyday speech and writing, chemists therefore do tend to disregard them, and use a mixture of systematic and trivial names. Nonetheless, it's important to know how the rules work. We shall look next at systematic nomenclature, before going on to look at the real language of chemistry.

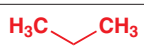
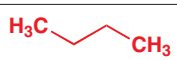
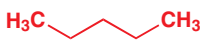
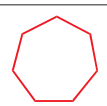
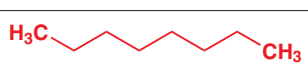
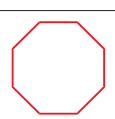
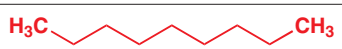
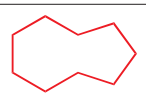
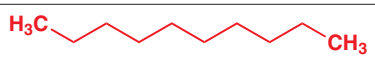
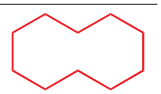
Systematic nomenclature

There isn't space here to explain all the rules for giving systematic names for compounds—they fill several desperately dull volumes, and there's no point knowing them anyway since computers will do the naming for you. What we will do is to explain the principles underlying systematic nomenclature. You should understand these principles because they provide the basis for the names used by chemists for the vast majority of compounds that do not have their own trivial names.

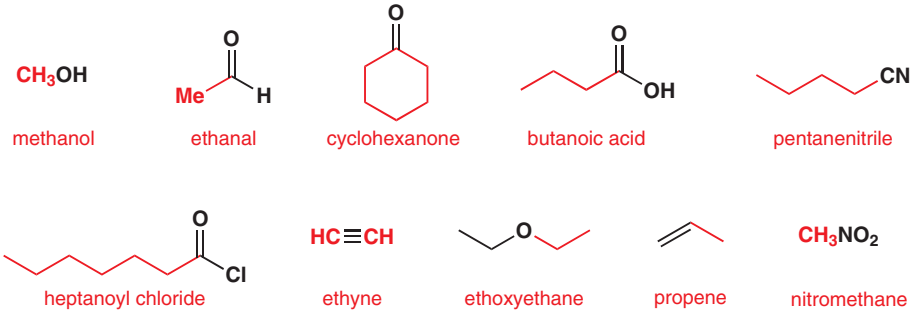
Systematic names can be divided into three parts: one describes the hydrocarbon framework, one describes the functional groups, and one indicates where the functional groups are attached to the skeleton.

You have already met the names for some simple fragments of hydrocarbon framework (methyl, ethyl, propyl). Adding a hydrogen atom to these alkyl fragments and changing -yl to -ane makes the alkanes and their names. You should hardly need reminding of their structures:

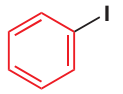
Names for the hydrocarbon framework

one carbon	methane	CH_4		
two carbons	ethane	$H_3C—CH_3$		
three carbons	propane		cyclopropane	
four carbons	butane		cyclobutane	
five carbons	pentane		cyclopentane	
six carbons	hexane		cyclohexane	
seven carbons	heptane		cycloheptane	
eight carbons	octane		cyclooctane	
nine carbons	nonane		cyclononane	
ten carbons	decane		cyclodecane	

The name of a functional group can be added to the name of a hydrocarbon framework either as a suffix or as a prefix. Some examples follow. It is important to count all of the carbon atoms in the chain, even if one of them is part of a functional group: pentanenitrile is actually BuCN.



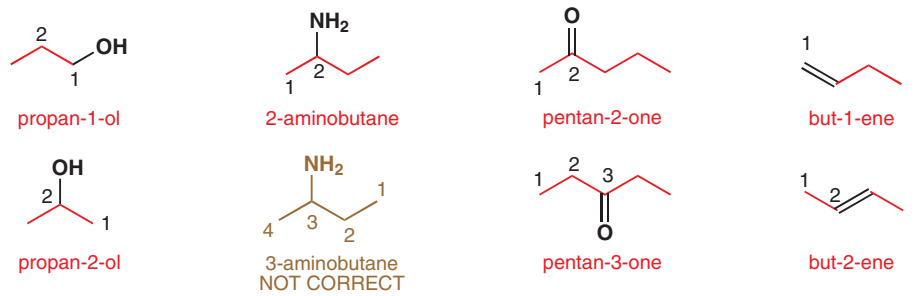
Compounds with functional groups attached to a benzene ring are named in a similar way.



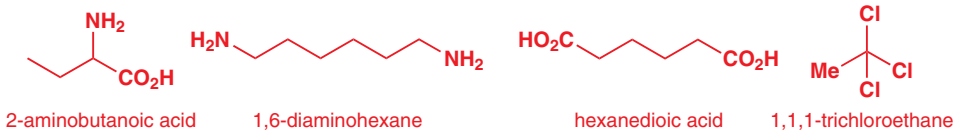
iodobenzene

Numbers are used to locate functional groups

Sometimes a number can be included in the name to indicate which carbon atom the functional group is attached to. None of the above list needed a number—check that you can see why not for each one. When numbers are used, the carbon atoms are counted from one end. In most cases, either of two numbers could be used (depending on which end you count from); the one chosen is always the lower of the two. Again, some examples will illustrate this point. Notice again that some functional groups are named by prefixes, some by suffixes, and that the number always goes directly before the functional group name.



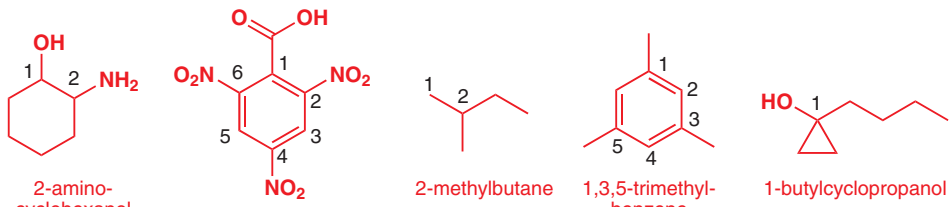
One carbon atom can have as many as four functional groups: this limit is reached with tetrabromomethane, CBr_4. Here are some other examples of compounds with more than one functional group.



Again, the numbers indicate how far the functional groups are from the end of the carbon chain. Counting must always be from the same end for each functional group. Notice how we use di-, tri-, and tetra- if there is more than one of the same functional group.

With cyclic compounds, there isn't an end to the chain, but we can use numbers to show the distance between the two groups—start from the carbon atom carrying one of the functional groups, then count round. These rules work for hydrocarbon frameworks that are

chains or rings, but many skeletons are branched. We can name these by treating the branch as though it were a functional group.

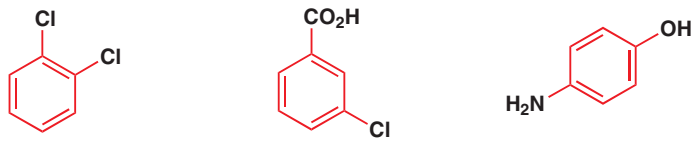


2-amino-cyclohexanol | 2,4,6-trinitrobenzoic acid | 2-methylbutane | 1,3,5-trimethyl-benzene | 1-butylcyclopropanol

Ortho, *meta*, and *para*

■ *ortho*, *meta*, and *para* are often abbreviated to *o*, *m*, and *p*.

With substituted benzene rings, an alternative way of identifying the positions of the substituents is to use the terms *ortho*, *meta*, and *para*. *Ortho* compounds are 1,2-disubstituted, *meta* compounds are 1,3-disubstituted, and *para* compounds are 1,4-disubstituted. Some examples should make this clear.



1,2-dichlorobenzene or *ortho*-dichlorobenzene or *o*-dichlorobenzene | 3-chlorobenzoic acid or *meta*-chlorobenzoic acid or *m*-chlorobenzoic acid | 4-aminophenol or *para*-aminophenol or *p*-aminophenol

Beware! *Ortho*, *meta*, and *para* are used in chemistry to mean other things too: you may come across orthophosphoric acid, metastable states, and paraformaldehyde—these have nothing to do with the substitution patterns of benzene rings.

The terms *ortho*, *meta*, and *para* are used by chemists because they're easier to remember than numbers, and the words carry with them chemical meaning. *Ortho* shows that two groups are next to each other on the ring even though the atoms may not happen to be numbered 1 and 2. They are one example of the way in which chemists don't always use systematic nomenclature but revert to more convenient 'trivial' terms. We consider trivial names in the next section.

What do chemists really call compounds?

The point of naming a compound is to be able to communicate with other chemists. Most chemists are happiest communicating chemistry by means of structural diagrams, and structural drawings are far more important than any sort of chemical nomenclature. That's why we explained in detail how to draw structures, but only gave an outline of how to name compounds. Good diagrams are easy to understand, quick to draw, and difficult to misinterpret.

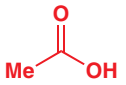
● **Always give a diagram alongside a name unless it really is something very simple, such as ethanol.**

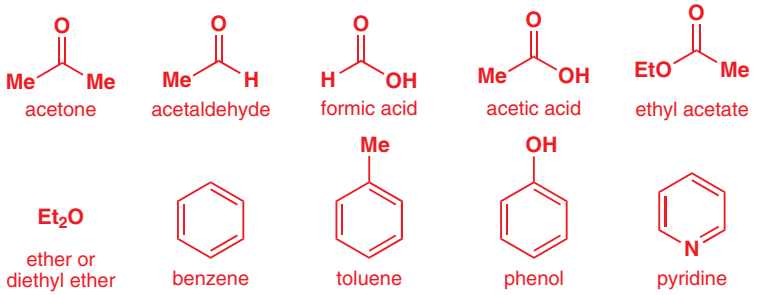
But we do need to be able to communicate by speech and by writing as well. In principle we could do this by using systematic names. In practice, however, the full systematic names of anything but the simplest molecules are far too clumsy for use in everyday chemical speech. There are several alternatives, mostly based on a mixture of trivial and systematic names.

Names for well-known and widely used simple compounds

A few simple compounds are called by trivial names not because the systematic names are complicated, but just out of habit. We know them so well that we use their familiar names.

You may have met the compound on the right before and perhaps called it ethanoic acid, its systematic name. But in a chemical laboratory everyone would refer to this acid as acetic acid, its trivial name. The same is true for all these common substances.

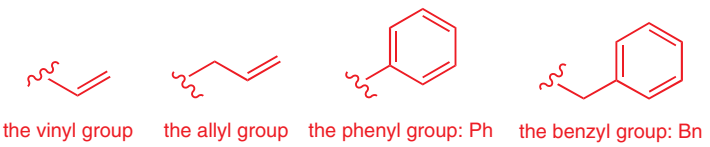




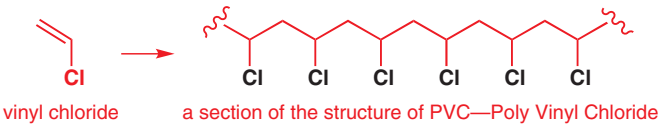
■ We haven't asked you to remember any trivial names of molecules yet, but these 10 compounds are so important you must be able to remember them. Learn them now.

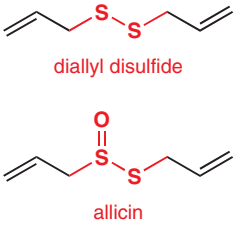
Trivial names like this are often long-lasting, well-understood historical names that are less easy to confuse than their systematic counterparts. 'Acetaldehyde' is easier to distinguish from 'ethanol' than is 'ethanal'.

Trivial names also extend to fragments of structures containing functional groups. Acetone, acetaldehyde, and acetic acid all contain the acetyl group (MeCO-, ethanoyl) abbreviated Ac and chemists often use this organic element symbol in writing AcOH for acetic acid or EtOAc for ethyl acetate. Chemists use special names for four fragments because they have mechanistic as well as structural significance. These are vinyl and allyl, phenyl and benzyl.



Giving the vinyl group a name allows chemists to use simple trivial names for compounds like vinyl chloride, the material that polymerizes to give PVC (polyvinyl chloride) but the importance of the name lies more in the difference in reactivity (Chapter 15) between the vinyl and allyl groups.

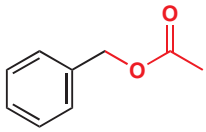


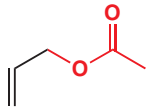


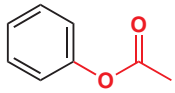
The allyl group gets its name from garlic (*Allium* sp.) because it makes up part of the structure of the compounds on the right responsible for the taste and smell of garlic.

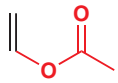
Allyl and vinyl are different in that the vinyl group is attached directly to a double-bonded C=C carbon atom, while the allyl group is attached to a carbon atom *adjacent* to the C=C double bond. The difference is extremely important chemically: allyl compounds are typically quite reactive, while vinyl compounds are fairly unreactive.

For some reason, the allyl and vinyl groups have never acquired organic element symbols, but the benzyl group has and it is Bn. It is again important not to confuse the benzyl group with the phenyl group: the phenyl group is joined through a carbon atom in the ring, while the benzyl group is joined through a carbon atom attached to the ring. Phenyl compounds are typically unreactive but benzyl compounds are often reactive. Phenyl is like vinyl, and benzyl is like allyl. We shall review all the organic element symbols you have met at the end of the chapter.


benzyl acetate

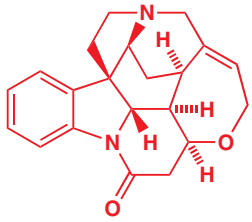

allyl acetate


phenyl acetate

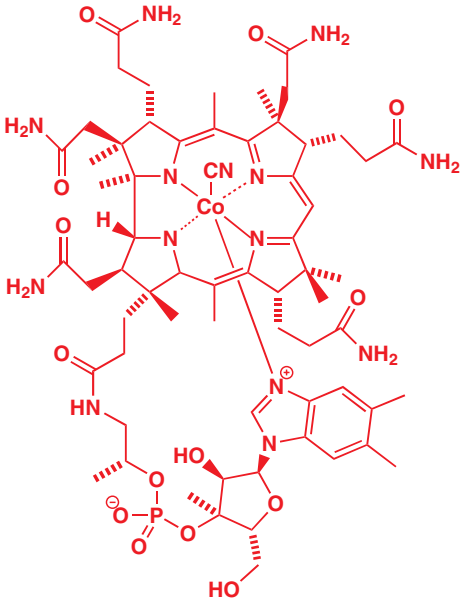

vinyl acetate

Names for more complicated but still well-known molecules

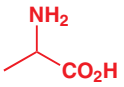
Complicated molecules that have been isolated from natural sources are always given trivial names because in these cases the systematic names really are impossible! Strychnine is a famous poison featured in many detective stories and a molecule with a beautiful structure. All chemists refer to it as strychnine as the systematic name is virtually unpronounceable. Two groups of experts at IUPAC and *Chemical Abstracts* also have different ideas on the systematic name for strychnine. Others like this are penicillin, DNA, and folic acid.


strychnine, or
(1R,11R,18S,20S,21S,22S)-12-oxa-8.17-diazaheptacyclo[15.5.0^{1,8}.0^{2,7}.0^{15,20}]tetracosa-2,4,6,14-tetraene-9-one (IUPAC)
or
4aR-[4aα,5aα,8aR*,15aα,15bα,15cβ]-2,4a,5,5a,7,8,15,15a,15b,15c-decahydro-4,6-methano-6H,14H-indolo[3,2,1-*ij*]oxepino[2,3,4-*de*]pyrrolo[2,3-*h*]quinolone
(*Chemical Abstracts*)

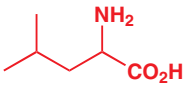
But the champion is vitamin B_{12}, a complicated cobalt complex with a three-dimensional structure of great intricacy. No chemist would learn this structure but would look it up in an advanced textbook of organic chemistry. You will find it in such books in the index under vitamin B_{12} and not under its systematic name. We do not even know what its systematic name might be and we are not very interested.


vitamin B_{12}, or...

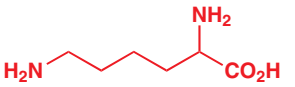
Even fairly simple but important molecules, the amino acids for example, which have systematic names that are relatively easy to understand, are normally referred to by their trivial names, which are, with a bit of practice, easy to remember and hard to muddle up. They are given in full in Chapter 23.



alanine, or
2-aminopropanoic acid

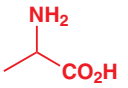


leucine, or
2-amino-4-methylpentanoic acid

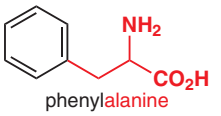


lysine, or
2,6-diaminohexanoic acid

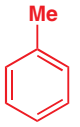
A very flexible way of getting new, simple names for compounds can be to combine a bit of systematic nomenclature with trivial nomenclature. Alanine is a simple amino acid that occurs in proteins. Add a phenyl group and you have phenylalanine, which is a more complex amino acid also in proteins. Toluene, the common name for methylbenzene, can be combined (both chemically and in making names for compounds!) with three nitro groups to give the famous explosive trinitrotoluene or TNT.



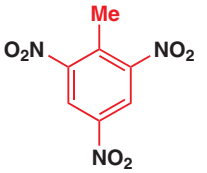
alanine



phenylalanine



toluene

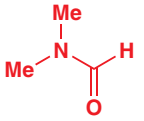


2,4,6-trinitrotoluene

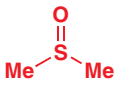
Compounds named as acronyms

Some compounds are referred to by acronyms, shortened versions of either their systematic or their trivial name. We just saw TNT as an abbreviation for TriNitroToluene but the more common use for acronyms is to define solvents and reagents in use all the time. Later in the book you will meet these solvents:

THF
(TetraHydroFuran)



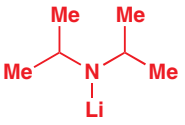
DMF
(DiMethylFormamide)



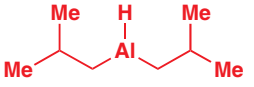
DMSO
(DiMethylSulfOxide)

■ The names and structures of these common solvents need learning too.

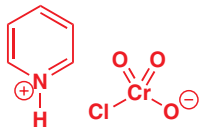
The following reagents are usually referred to by acronym and their functions will be introduced in other chapters so you do not need to learn them now. You may notice that some acronyms refer to trivial and some to systematic names.



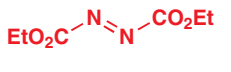
LDA
Lithium Di-isopropylAmide



DIBAL
Di-IsoButylALuminium hydride



PCC
Pyridinium ChloroChromate

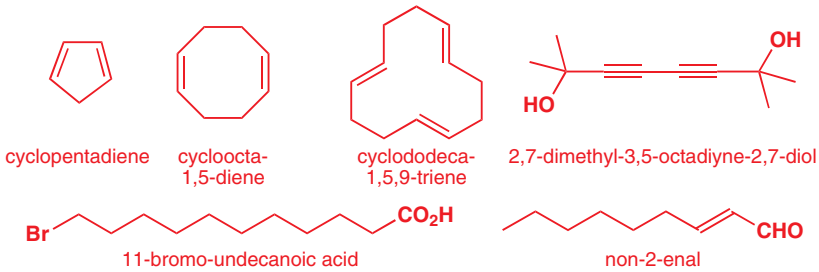


DEAD
DiEthyl AzoDicarboxylate

Compounds for which chemists use systematic names

You may be surprised to hear that practising organic chemists use systematic names at all in view of what we have just described, but they do! Systematic names really begin with derivatives of

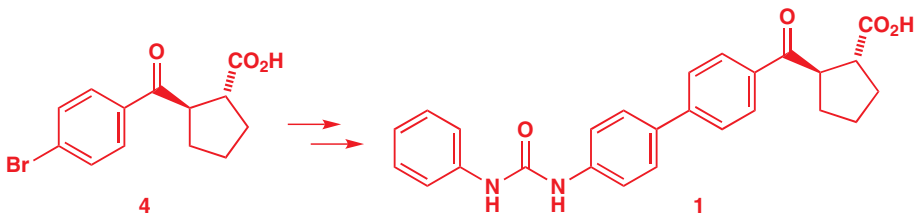
pentane (C_5H_{12}) since the prefix *pent-* means five, whereas *but-* does not mean four. Chemists refer to simple derivatives of open-chain and cyclic compounds with 5 to about 20 carbon atoms by their systematic names, providing that there is no common name in use. Here are some examples.



These names contain a syllable that tells you the framework size: penta- for C_5, octa- for C_8, nona for C_9, undeca- for C_{11}, and dodeca- for C_{12}. These names are easily worked out from the structures and, what is more important, you get a clear idea of the structure from the name. One of them might make you stop and think a bit (which one?), but the others are clear even when heard without a diagram to look at.

Complicated molecules with no trivial names

When chemists make complex new compounds in the laboratory, they publish the method for making them in a chemical journal, giving their full systematic names in the experimental account, however long and clumsy those names may be. But in the text of the paper, and while talking in the laboratory about the compounds they have made, they will just call them 'the amine' or 'the alkene'. Everyone knows which amine or alkene is meant because at some point they remember seeing a chemical structure of the compound. This is the best strategy for talking about almost any molecule: draw a structure, then give the compound a 'tag' name like 'the amine' or 'the acid'. In written chemistry it's often easiest to give every chemical structure a 'tag' number as well. To illustrate what we mean, let's talk about a recent drug synthesis.



This potential anti-obesity drug 1, which might overcome insulin resistance in diabetics, was recently made at Abbott laboratories from a simpler intermediate 4. In the published work the drug is called 'a selective DGAT-1 inhibitor' but that doesn't mean much to us. In the text of the paper they refer to it by its compound number 1. How much more sensible than using its systematic name: *trans*-(1*R*,2*R*)-2-(4′-(3-phenylureido)biphenylcarbonyl) cyclopentanecarboxylic acid. The simpler intermediate they call 'the ketoacid 4' or 'the aryl bromide 4' or 'the free acid 4' depending on what aspect of its structure they want to emphasize. Notice that in both cases a clear diagram of the structure appears with its number.

How should you name compounds?

So what should you call a compound? It really depends on circumstances, but you won't go far wrong if you follow the example of this book. We shall use the names for compounds

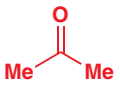
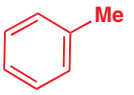
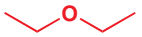
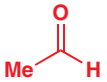
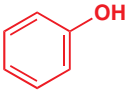
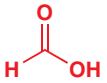
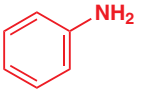
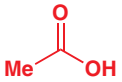
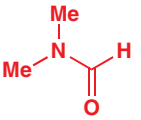
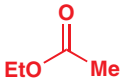
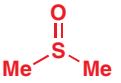
that real chemists use. There's no need to learn all the commonly used names for compounds now, but you should log them in your memory as you come across them. Never allow yourself to pass a compound name by unless you are sure you know what chemical structure it refers to.

Our advice on chemical names—six points in order of importance

- **Draw a structure first and worry about the name afterwards.**
- **Learn the names of the *functional groups* (ester, nitrile, etc.).**
- **Learn and use the names of a few simple compounds used by all chemists.**
- **In speech, refer to compounds as 'that acid' (or whatever) while pointing to a diagram.**
- **Grasp the principles of systematic (IUPAC) nomenclature and use it for compounds of medium size.**
- **Keep a notebook to record acronyms, trivial names, structures, etc. that you might need later.**

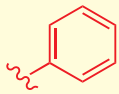
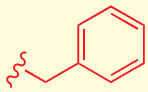
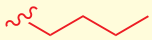
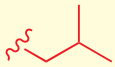
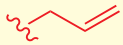
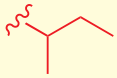
We've met a great many molecules in this chapter. Most of them were just there to illustrate points so don't learn their structures! Instead, learn to recognize the names of the functional groups they contain. However, there were 10 names for simple compounds and three for common solvents that we advised you to learn. Cover up the right-hand part of each column and draw the structures for these 14 compounds.

Important structures to learn

acetone		toluene	
ether or diethyl-ether		pyridine	
acetaldehyde		phenol	
formic acid		aniline	
acetic acid or AcOH		THF or tetrahydrofuran	
benzene		DMF, Me_2NCHO, or dimethylformamide	
ethyl acetate or EtOAc		DMSO	

That's all we'll say on the subject of nomenclature—you'll find that as you practise using these names and start hearing other people referring to compounds by name you'll soon pick up the most important ones. But, to reiterate, make sure you never pass a compound name by without being absolutely sure what it refers to—draw a structure to check.

● Review box: Table of fragment names and 'organic elements'

R	alkyl		***t*-Bu**	*tert*-butyl	
Me	methyl		**Ar**	aryl	any aromatic ring
Et	ethyl		**Ph**	phenyl	
Pr (*n*-Pr)	propyl		**Bn**	benzyl	
Bu (*n*-Bu)	butyl		**Ac**	acetyl	
***i*-Pr**	isopropyl			vinyl	
***i*-Bu**	isobutyl			allyl	
s-Bu	*sec*-butyl		**X**	halide	F, Cl, Br or I

Further reading

All the big American textbooks have early chapters on structure, shape, and the drawing of molecules but they tend to use Lewis structures with all atoms and electrons in bonds shown and often right angles between bonds.

A short and sensible introduction is in the Oxford Primer *Foundations of Organic Chemistry* by M. Hornby and J. Peach, OUP, Oxford, 1996.

For more on palytoxin: E. M. Suh and Y. Kishi, *J. Am. Chem. Soc.*, 1994, **116**, 11205–11206.

For an account of the competing claims to the first proposal of a cyclic structure of benzene, see Alfred Bader's article 'Out of the Shadow' in the 17 May 1993 issue of *Chemistry and Industry*.

Check your understanding

To check that you have mastered the concepts presented in this chapter, attempt the problems that are available in the book's Online Resource Centre at **http://www.oxfordtextbooks.co.uk/orc/clayden2e/**

42 Organic chemistry of life

Connections

Building on

- Acidity and basicity ch8
- Carbonyl chemistry ch10 & ch11
- Stereochemistry ch14
- Conformational analysis ch16
- Enolate chemistry and synthesis ch25 & ch26
- Sulfur chemistry ch27
- Heterocycles ch29 & ch30
- Asymmetric synthesis ch41

Arriving at

- Nucleic acids store information for the synthesis of proteins
- Modified nucleosides can be used as antiviral drugs
- 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
- The main sorts of natural products are alkaloids, polyketides, terpenes, and steroids
- Alkaloids are amines made from amino acids
- Fatty acids are built up from acetyl CoA and malonyl CoA subunits

Looking forward to

- Three more comprehensive web chapters:
  - The chemistry of life
  - Mechanisms in biological chemistry
  - Natural products
- Organic chemistry today ch43

Primary metabolism

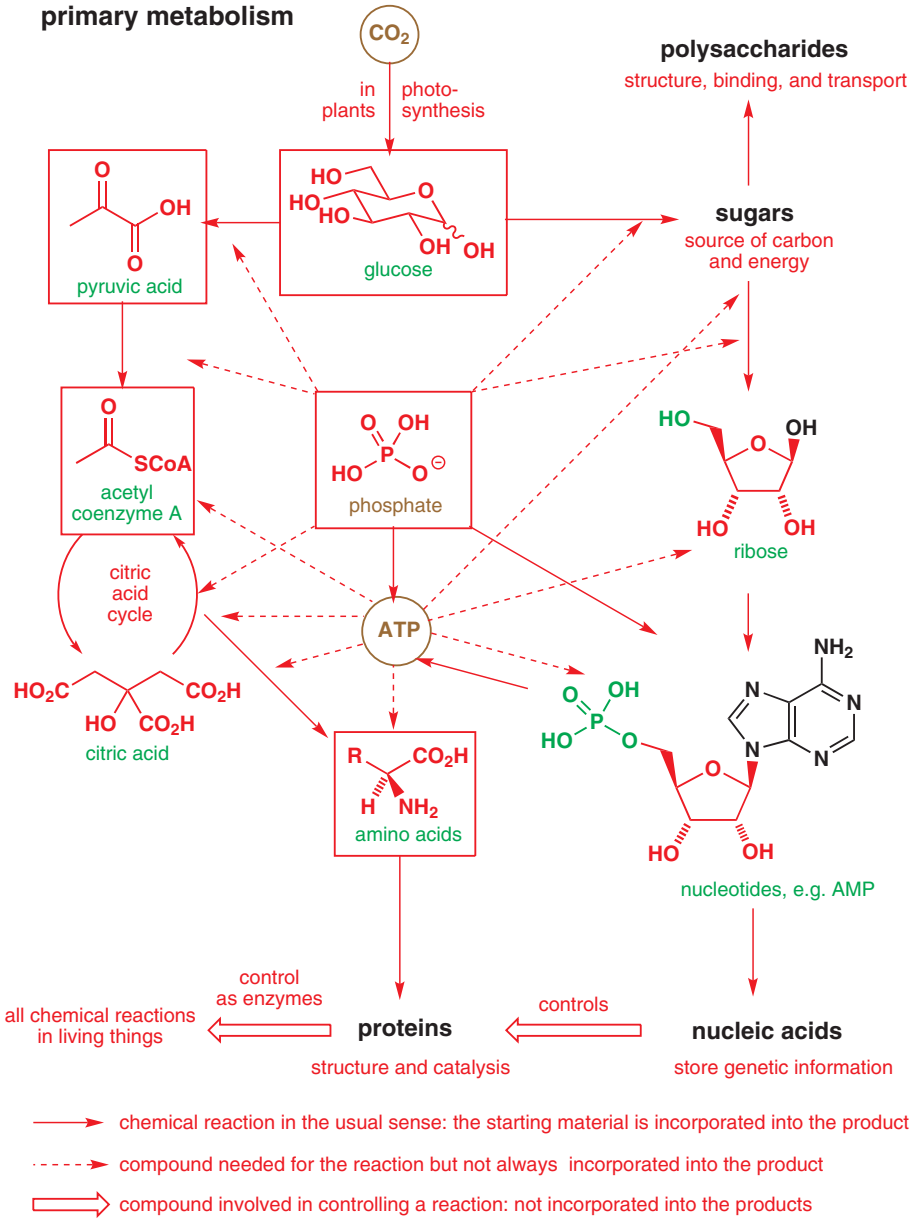
Secondary metabolism is, by contrast, chemistry less fundamental to the workings of life and restricted to smaller groups of organisms. Later in this chapter you will meet alkaloids produced by some plants and terpenes produced by others. Humans produce neither of these, but we do make steroids, as do other animals (and a few plants). All of these molecules are the products of secondary metabolism.

Life runs on chemistry, and the chemical side of biology is fascinating for that reason alone. 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 thought of as the poor relations of the other two, storing energy and building membranes, but it is now clear that they also have a vital part to play in recognition and transport.

The chemistry common to all living things is known as *primary metabolism* and 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 metabolism and are built into many

Online support. The icon in the margin indicates that accompanying interactive resources are provided online to help your understanding: just type **www.chemtube3d.com/clayden/123** into your browser, replacing **123** with the number of the page where you see the icon. For pages linking to more than one resource, type **123-1, 123-2** etc. (replacing **123** with the page number) for access to successive links.

important biological molecules. Use this chart to keep track of the relationships between the molecules of metabolism as you develop a more detailed understanding of them. We start with nucleic acids.

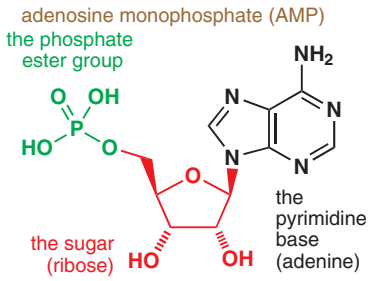


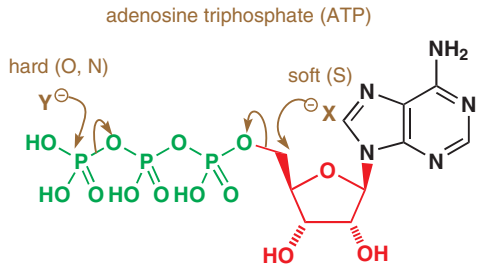
Life begins with nucleic acids

Nucleic acids store 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. In the example below, adenine is the base (shown in black), adenosine is the nucleoside (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.

Nucleosides and nucleotides

A nucleoside differs from a nucleotide in lacking the phosphate—a nucleoside is just a base and a sugar.





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. 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.

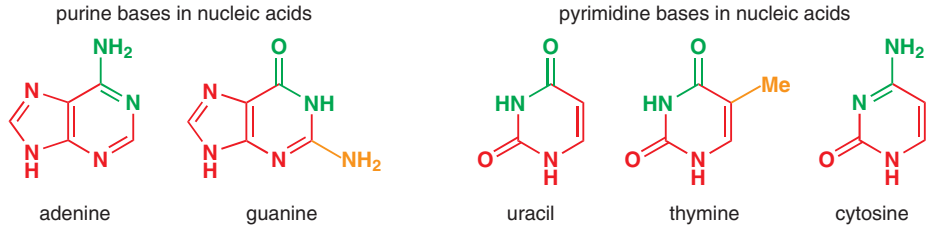
There are five heterocyclic bases in DNA and RNA

Nucleic acids are made up of a selection of five bases, two sugars, and the phosphate group. The bases are monocyclic pyrimidines or bicyclic purines and are all aromatic.

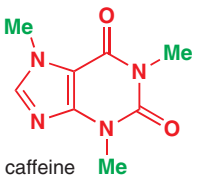
- There are only two purine bases found in nucleic acids: adenine (A), which we have already met, and guanine (G)
- The three pyrimidine bases are simpler: uracil (U), thymine (T), and cytosine (C). Cytosine is found in DNA and RNA, uracil in RNA only, and thymine in DNA only.

The coloured parts of the molecules below emphasize the characteristic features of the bases.

■ You met pyrimidines in Chapter 29 and learned how to make them in Chapter 30, but the purine ring system may be new to you. Make sure you can find the six (or ten) electrons making these compounds aromatic. You may need to draw delocalized structures, especially for U, T, and G.

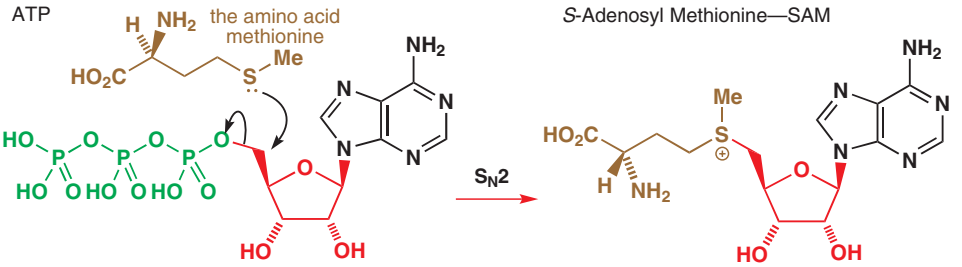


The stimulants in tea and coffee are methylated purines

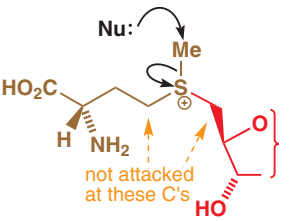


An important stimulant for many is a fully methylated purine present in tea and coffee—caffeine. Caffeine is a crystalline substance easily extracted from coffee or tea with organic solvents. It is extracted industrially with supercritical CO_2 (or, if you prefer, 'nature's effervescence') to make decaffeinated tea and coffee.

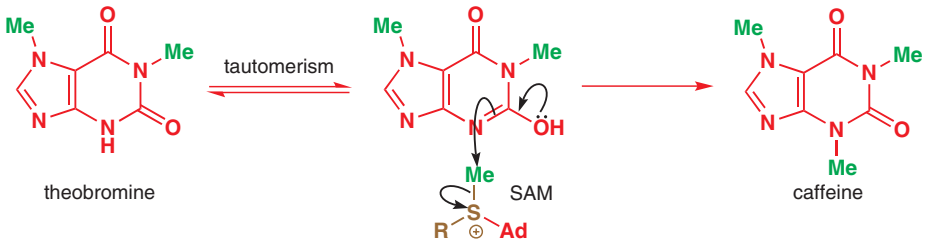
If we, as chemists, were to add those methyl groups we would choose to use a reagent such as methyl iodide, but nature uses a much more complicated molecule. There is a great deal of methylating going on in living things—and the methyl groups are usually added by (*S*)-adenosyl methionine (or SAM), formed by reaction of methionine with ATP. This is a good reaction because sulfur is a good soft nucleophile, triphosphate is a good leaving group, and substitution at primary carbon is easy.



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 (a purine also found in cocoa) 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.



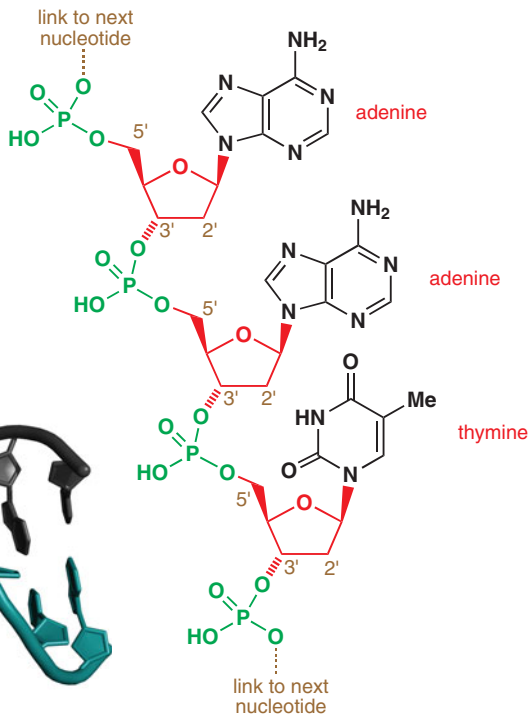
nucleophilic attack on SAM

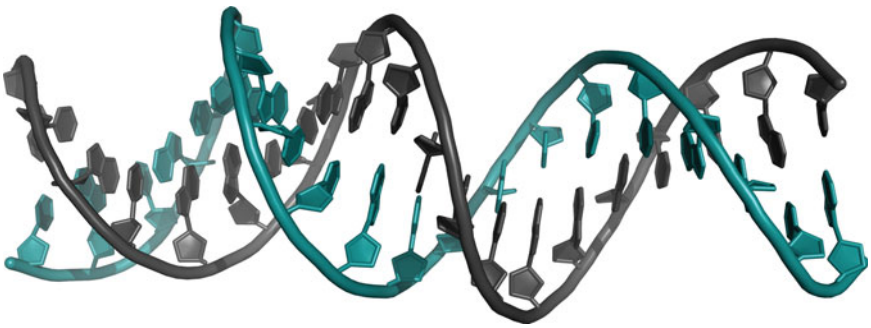


At this point we should just point out something that it's easy to forget: there is *only one chemistry*. There is no magic in biological chemistry, and nature uses the same chemical principles as we do in the chemical laboratory. All the mechanisms that you have studied so far will help you to draw 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 we humans 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 in the 21st century.

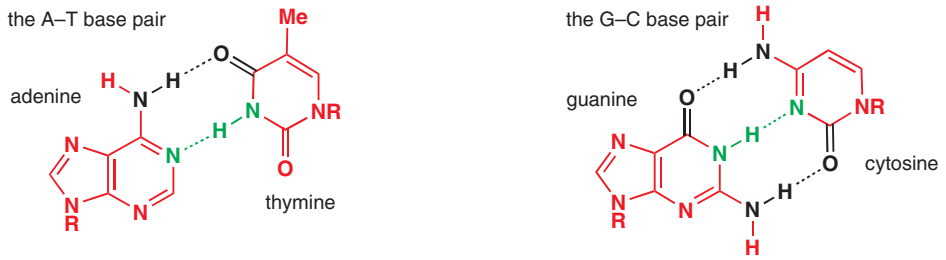
Nucleic acids exist in a double helix

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 on the right. Notice that the 2′ (pronounced 'two prime') position on the ribose ring is vacant. There is no hydroxyl group there: that is why it is called *deoxy*ribonucleic acid. 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.





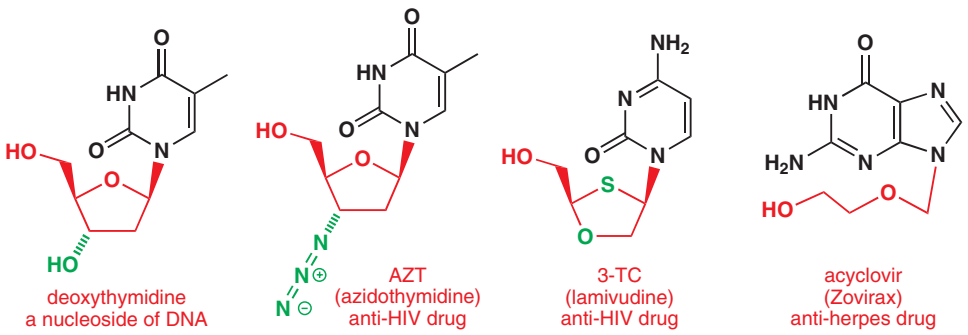
Each polymeric strand of DNA coils up into a helix and is bonded to another strand by hydrogen bonds between the bases. Each base pairs up specifically with another base—adenine with thymine (A–T) and guanine with cytosine (G–C)—like this.



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). A purine has to pair with a pyrimidine because only the combination of larger purine and smaller pyrimidine bridges the gap between the nucleic acid coils. Look back at the green and orange parts of the structures on p. 1136 and you will see that only one hydrogen bond pairing pattern can work. 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–.

HIV and AIDS are treated with modified nucleosides

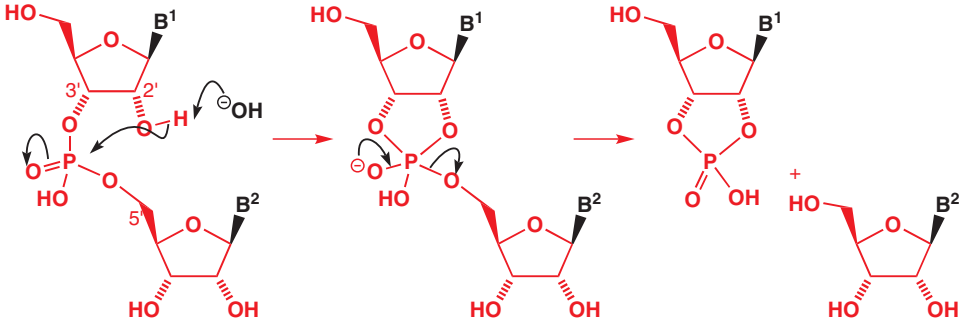
Modified nucleosides are among the best antiviral compounds. The anti-HIV drug AZT (zidovudine) is a slightly modified DNA nucleoside (3′-azidothymidine). It has an azide at C3′ instead of the hydroxyl group in the natural nucleoside. A more radically modified nucleoside 3-TC (lamivudine) is active against AZT-resistant viruses. This drug is based on cytosine with the sugar replaced by a different heterocycle, although it is recognizably similar, especially in the stereochemistry. Acyclovir (Zovirax), the cold sore (herpes) treatment, is a modified guanosine in which only a ghost of the sugar remains. There is no ring at all and no stereochemistry.



Cyclic nucleosides and stereochemistry

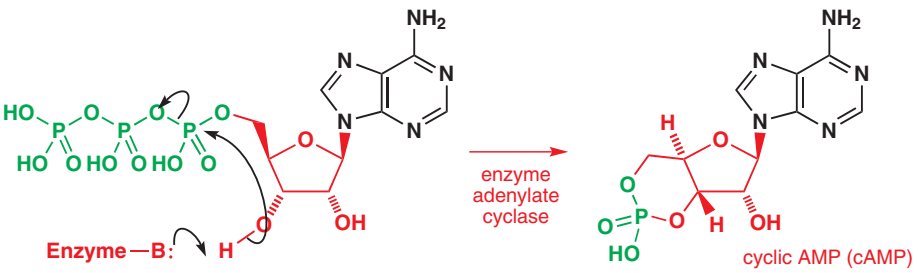
DNA is more stable than RNA because its sugars lack the 2′ hydroxyl groups. In ribonucleic acids, the fact that the 2′- and 3′-OH groups are on the same side of the ring makes alkaline hydrolysis exceptionally rapid by intramolecular nucleophilic catalysis.

■ The substituents B^1 and B^2 represent any purine or pyrimidine base.



The base 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 base.

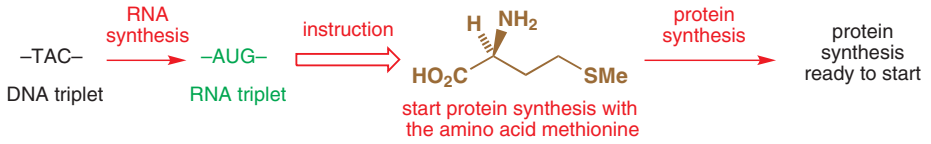
Another cyclic phosphate that can be formed from a nucleotide is important as a biological messenger that helps to control such processes as blood clotting and acid secretion in the stomach. It is cyclic AMP (cAMP), formed enzymatically from ATP by nucleophilic displacement of pyrophosphate by the 3′-OH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



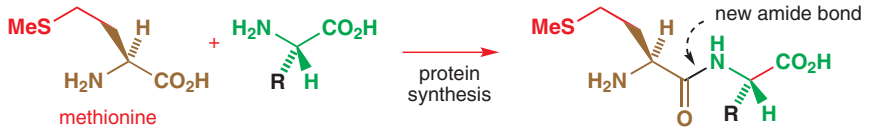
■ Note that cAMP has a *trans* 6,5-fused ring junction.

Proteins are made of amino acids

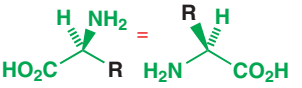
DNA encodes the information needed to make proteins in the form of triplets of bases (codons), for example thymine–adenine–cytosine (TAC) in the diagram below. As RNA is synthesized from DNA, these are turned into complementary codons (in the example below, AUG) by pairing up the bases as shown on p. 1138. This RNA forms the instructions for protein synthesis by the ribosome—perhaps the most elaborate molecular structure in the known universe. Each codon of the RNA chain tells the ribosome to add a specific amino acid to the growing protein. For example, the codon AUG indicates methionine, which we met as a component of SAM. Methionine is a typical amino acid of the kind present in proteins, but is also the starter unit of all proteins.



The next codon of RNA directs the ribosome to add the next amino acid, linked to the previous one in the chain by an amide bond. Amino acids used to make proteins have the same basic structure and stereochemistry, shown in the margin, and differ only in the group R.

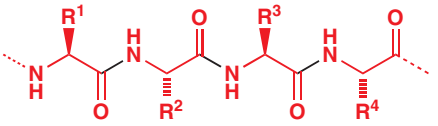


two views of the general amino acid structure



➡ There is a list of the naturally occurring amino acids in Chapter 23 (p. 554), where we discussed the *laboratory* synthesis of peptides.

The process continues as more amino acids are added in turn to the right-hand end of the growing molecule. A section of the final protein might look like the structure below. The 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.

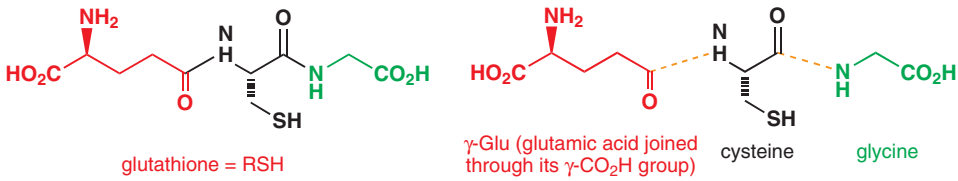


■ Much of the function of enzymes and other proteins derives from their detailed folded conformation, discussion of which is beyond the scope of this book.

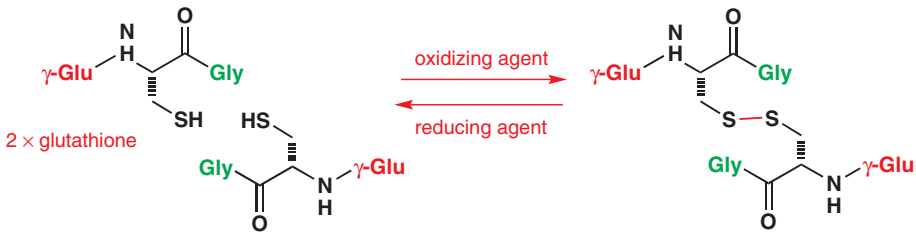
Amino acids combine to form peptides and proteins

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**.

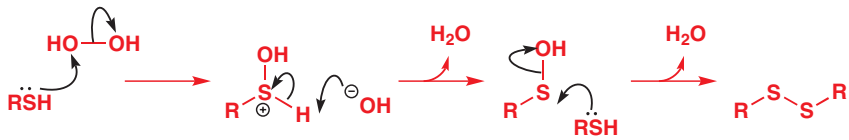
An important tripeptide is **glutathione**, present in the tissues of both animals and plants. Glutathione is the 'universal thiol' that removes dangerous oxidizing agents by allowing itself to be oxidized to a disulfide. Glutathione is, however, not quite a typical tripeptide. The left-hand amino acid is normal glutamic acid but it is joined to the next amino acid through its γ-CO_2H group instead of the more normal α-CO_2H group. The middle amino acid is the vital one for the function—cysteine with a free SH group. The *C*-terminal acid is glycine.



Thiols are easily oxidized to disulfides and glutathione sacrifices itself if it meets an oxidizing agent. The oxidized form of glutathione can later be converted back to the thiol by reduction with NADH, which you will meet later in this chapter.

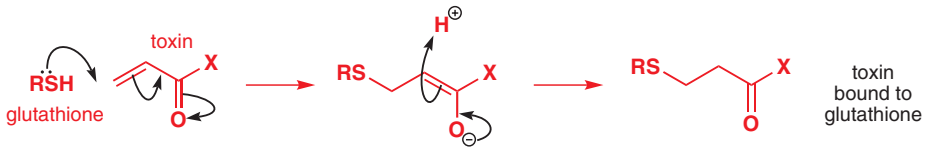


With a stray oxidizing agent such as a peroxide, say H_2O_2, the mechanism below shows how this can be reduced to water as glutathione (represented as RSH) is oxidized to a disulfide.

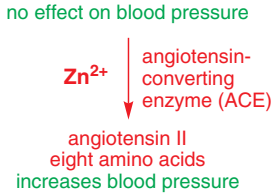


Glutathione also detoxifies some of the compounds we described earlier in this book as dangerous carcinogens such as Michael acceptors and 2,4-dinitrohalobenzenes. The thiol acts as a nucleophile, inactivating the electrophiles. Covalently bound to glutathione they are harmless and can be excreted. More glutathione will be synthesized from glutamic acid, cysteine, and glycine to replace that which is lost.

See, for example, the discussion in Chapter 22, p. 516.



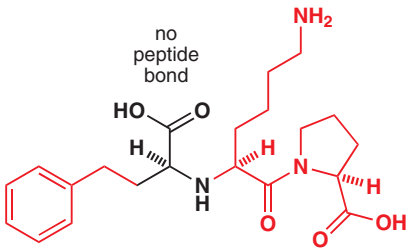
You saw some examples in Chapter 23.



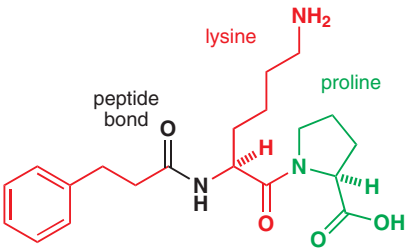
Some short peptides, of around ten amino acids, are hormones. Angiotensin II, for example, is a peptide that causes blood pressure to rise—a very necessary thing in some situations but too much and too often leads to heart attacks and strokes.

Angiotensin-converting enzyme (ACE) is the zinc-dependent enzyme that cleaves two amino acids from the end of angiotensin I to give angiotensin II, and ACE inhibitors are used as treatment for high blood pressure because they inhibit this enzyme. Lisinopril is an example: it is a dipeptide mimic, having two natural amino acids and something else. The 'something else' is the left-hand part of the molecule, linked to the dipeptide (Lys–Pro) through an amine and not

by an amide bond. This stops enzymes from hydrolysing the molecule. Lisinopril binds to ACE because it is *like* a natural dipeptide but it inhibits it because it is *not* a natural dipeptide. Many people are alive today because of this simple deception practised on an enzyme.



lisinopril



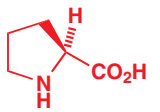
'natural peptide'

Structural proteins must be tough and flexible

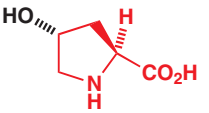
In contrast with the functional enzymes, proteins such as collagen are purely structural. Collagen is the tough protein of tendons and connective tissue, and is present in skin, bone, and teeth. It contains large amounts of glycine (every third amino acid is glycine), proline, and hydroxyproline (again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

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 triple coil for any larger amino acid. Functionalized amino acids are rare in collagen.

glycine Gly



(*S*)-proline Pro



(2*S*,4*R*)-hydroxyproline Hyp

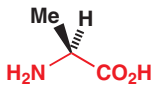
Hydroxyproline and scurvy

Hydroxyproline is a very unusual amino acid. It is not incorporated into the growing protein chain when collagen is synthesized—instead the collagen molecule is assembled with Pro where Hyp is need. Once the protein is complete, some of the 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 suffered by 18th-century sailors (loose teeth, sores, and blisters) were caused by the inability to make collagen.

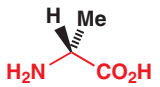
Antiobiotics exploit the special chemistry of bacteria

We have repeatedly emphasized that all life has very similar chemistry. From the biochemical point of view the most important division is that separating *prokaryotes* from *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 then, 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.

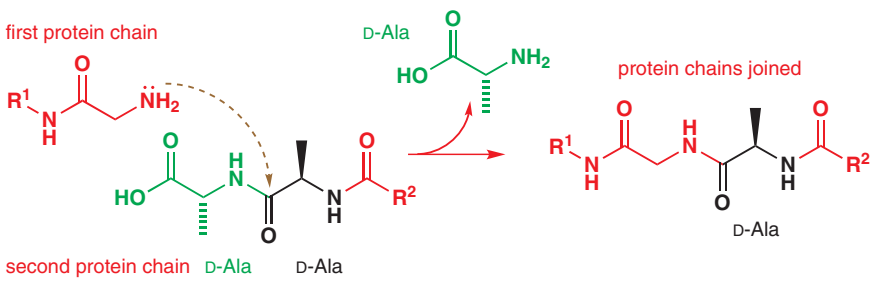


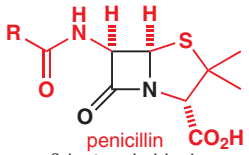
(*S*)-Ala from normal proteins



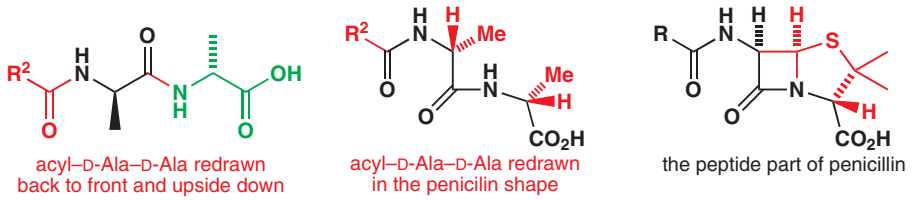
(*R*)-Ala from bacterial cell walls

■ The reason bacteria have evolved to 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.



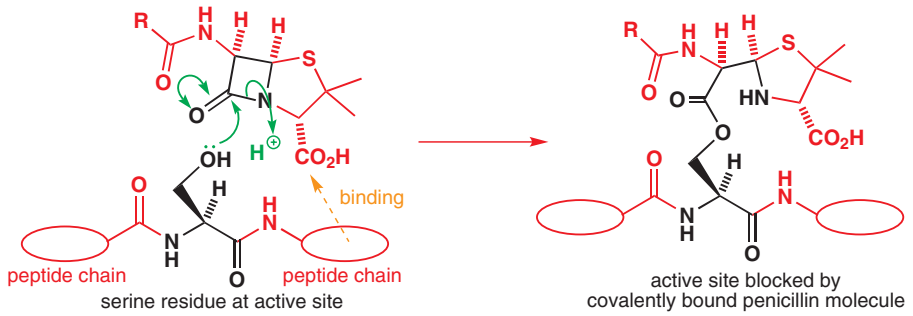


The antibiotic penicillin works by interfering with this step—although this was not even suspected when penicillin was discovered. Penicillin inhibits the enzyme that catalyses the D-Ala transfer in a very specific way. It first binds specifically to the enzyme (so it must be a mimic of the natural substrate) 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 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.



Sugars—just energy sources?

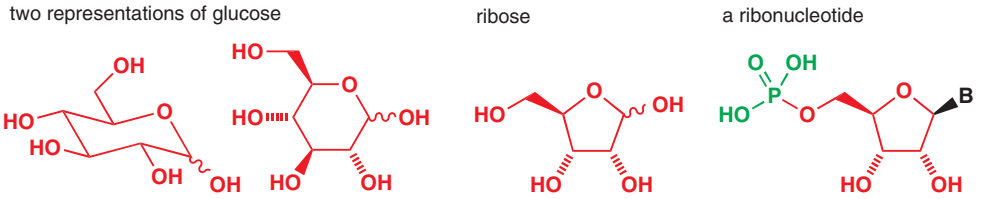
Sugars are the building blocks of carbohydrates. They used to be thought of as essential but rather dull molecules whose function was principally the (admittedly useful) storage of energy. In fact they have much more interesting and varied roles than that. We have already noted that ribose plays an intimate role in DNA and RNA structure and function. Sugars are also often found in intimate association with proteins and are involved in recognition and adhesion processes.

There is more about the development of drugs for treating HIV in Chapter 43.

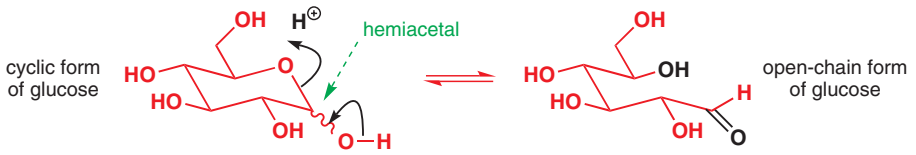
Here are two examples. How does a sperm recognize the egg and penetrate its wall? Recognition of a carbohydrate attached to the membrane of the egg was the first event in all of our lives. And how does a virus get inside a cell? Here again, the recognition process involves specific carbohydrates. One 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.

Sugars normally exist in cyclic forms with much stereochemistry

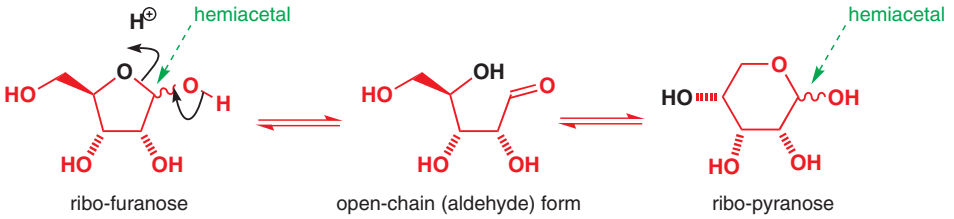
The most important sugar is glucose. It has a saturated six-membered ring containing oxygen and it is best drawn in a chair conformation with nearly all the substituents equatorial. It can also be drawn as a flat configurational diagram. We have already met one sugar in this chapter, ribose, because 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.



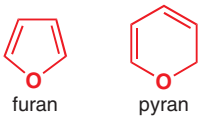
The drawings of glucose and ribose show a number of stereogenic centres, with one centre undefined—an OH group shown with a wavy bond. This is because one centre in both sugars is a hemiacetal and therefore the molecule is in equilibrium with an open-chain hydroxy-aldehyde. For glucose, the open-chain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



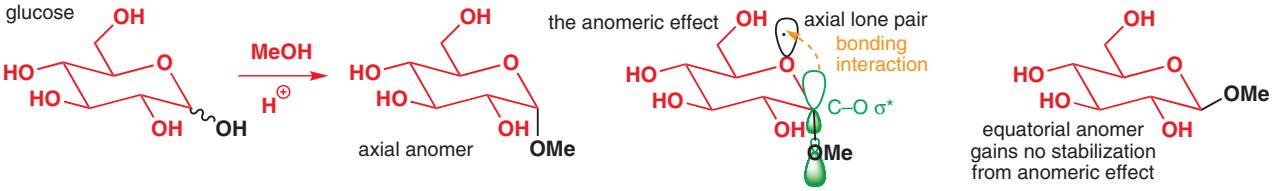
The most important sugars may exist in an open-chain form, as a five-membered oxygen heterocycle (called a *furanose*, after the five-membered aromatic compound furan) or a six-membered oxygen heterocycle (called a *pyranose*, after the six-membered pyran). Glucose prefers the pyranose structure; ribose prefers the furanose structure.



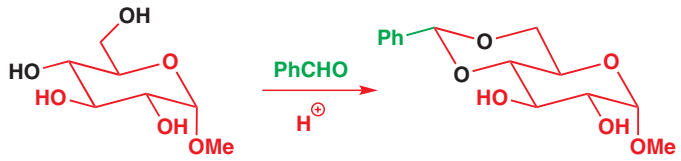
Sugars can be fixed in one shape by acetal formation

The simplest way to fix glucose in the pyranose form is to trap it as an acetal. Acid-catalysed condensation with an alcohol, methanol, for example, gives an acetal and, remarkably, the acetal has an *axial* OR group. Acetal formation is under thermodynamic control (Chapter 11) 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 was discussed in Chapter 31, and you should check that you can still write down the mechanism of acetal formation you learned in Chapter 11.

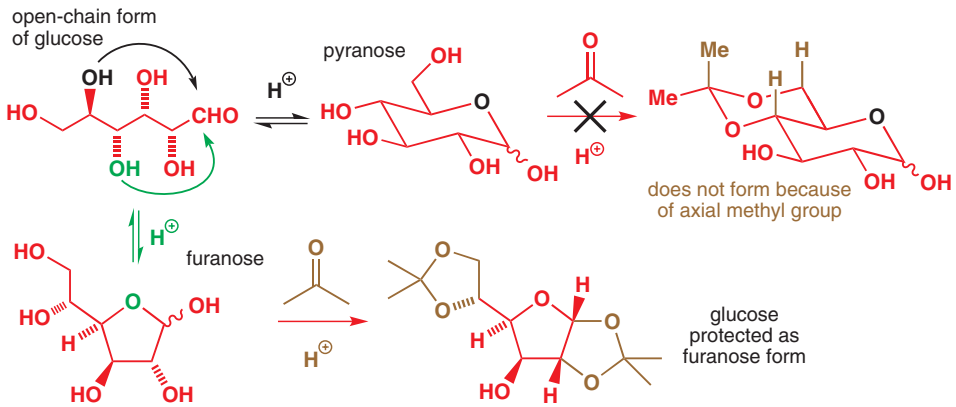


The formation of acetals allows a remarkable degree of control over the chemistry of sugars. Apart from the simple glucoside acetal we have just seen, there are three important acetals worth understanding because of the way in which they illustrate stereoelectronic effects—the interplay of stereochemistry and mechanism. If we make an acetal from methyl glucoside and benzaldehyde, we get a single compound as a single stereoisomer.



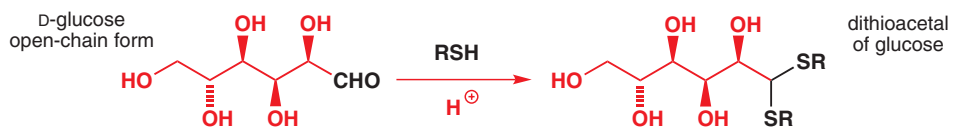
The new acetal could have been formed between any of the adjacent OH groups in the starting material but it chose the only pair (the black OH groups) which give a six-membered ring. The stereochemistry of glucose is such that the new six-membered ring is *trans*-fused to the old so that a beautifully stable all-chair bicyclic structure results, with the phenyl group in an equatorial position in the new chair acetal ring. Acetal formation is under thermodynamic control and this product is the most stable possible acetal.

Acetal formation from sugars and acetone shows 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 the pyranose and the furanose forms through reversible hemiacetal formation using 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 two *cis*-fused five-membered rings and the other being on the side chain. This double acetal is the product isolated from the reaction.

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.

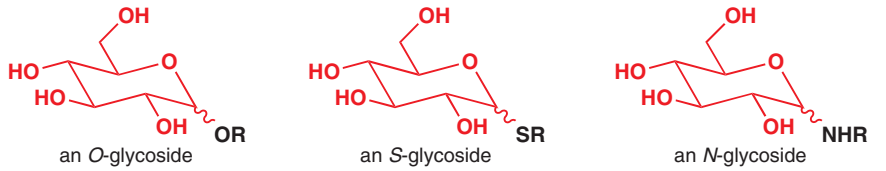


■ The most important *N*-glycosides are, of course, the nucleotides, which we have already described in some detail.

Glycosides 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 might use a THP group (Chapter 23).

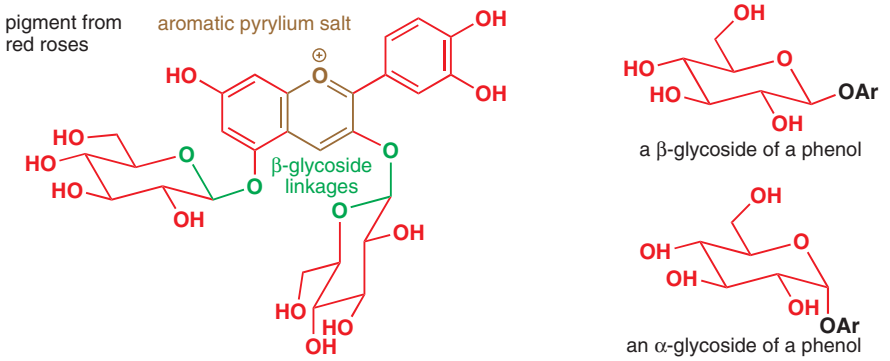
➡ You saw an example in Chapter 6 (p. 129) where acetone cyanohydrin is found in the cassava plant as a glucoside.



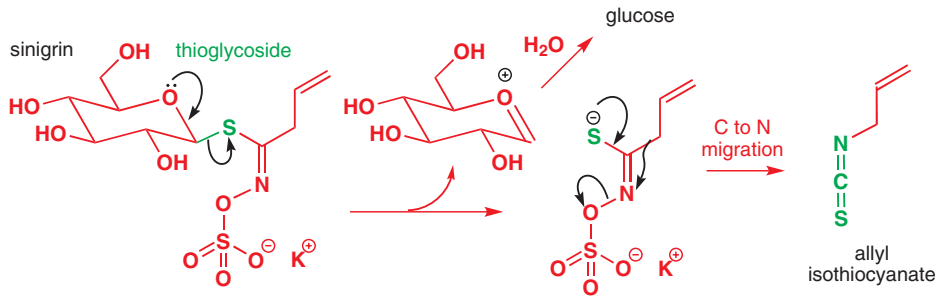
O-Glycosides occur in immense variety with glucose and other sugars being joined to the OH groups of alcohols and phenols to form acetals. The stereochemistry of these compounds is usually described by the Greek letters α and β. If the OR bond is down, it's an α-glycoside; if up, a β-glycoside. An attractive 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.

α- and β-glycosides

It is easy to remember which is which, as long as you accept that people who devise nomenclature must be maliciously foolish. 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.



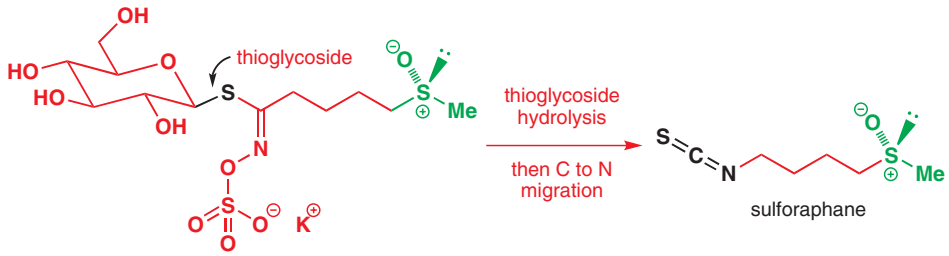
It's often proposed that there are special benefits to health in eating broccoli and brussels sprouts because of the sulfur-containing antioxidants they contain. These compounds are unstable isothiocyanates. They are not usually present in the plant; damage—by cutting or cooking, for example—induces a glycosidase (an enzyme which hydrolyses glycosides) to releases the sulfur compound from its glucose protection. A simple example is sinigrin. 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) and the anion is additionally stabilized by conjugation.



The next step is surprising. A rearrangement occurs, rather similar to the Beckmann rearrangement, 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 these their 'hot' taste. When mustard powder is mixed with water, the hot taste develops over some minutes as sinigrin is hydrolysed to the isothiocyanate.

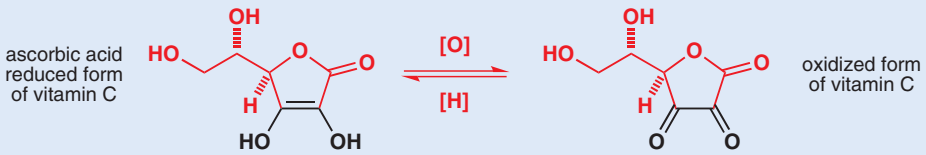
➡ The Beckmann rearrangement is described in Chapter 36, p. 958.

The *S*-glycoside in broccoli and brussels sprouts that is proposed to offer protection 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 reductive enzyme.



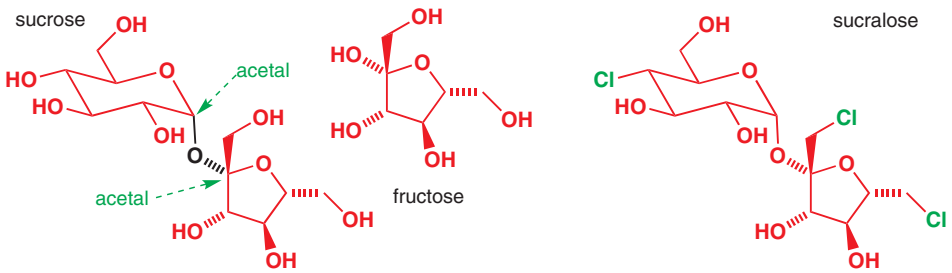
Vitamin C is a derivative of glucose

Nature makes some important compounds from simple sugars. Vitamin C—ascorbic acid—is one of these. It certainly looks very like a sugar as it has six carbon atoms, each having an oxygen atom as substituent as well as an oxygen heterocycle. Like glutathione, it protects cells 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 shown below.



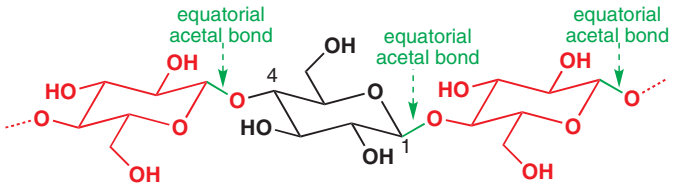
Most sugars are embedded in complex carbohydrates

The most familiar of all sugars is sucrose—the mixed acetal formed from glucose and fructose. Sucrose is of course sweet, and is easily metabolized into fats. But if three of the OH groups in sucrose are replaced by chlorine atoms, a compound 600 times as sweet is produced: less of it is needed to get the same sweet taste and the chlorines reduce the rate of metabolism so that much less fat is made. This is the compound sucralose, discovered by chemists at Tate & Lyle and now used to sweeten soft drinks.



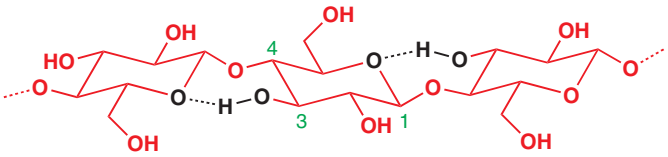
10^{15} kg per year of cellulose is literally an astronomical amount: it's about the mass of one of the moons of Mars, Deimos. Our moon weighs 10^{22} kg.

Sucrose is a disaccharide—two simple sugars linked by an acetal. In general, saccharides have the same relationship to sugars as peptides and proteins have to amino acids. One of the most abundant compounds in nature is a saccharide: 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 an acetal formed by attack of the C4 hydroxyl group of one glucose molecule on the anomeric carbon atom of the next. Here is that basic arrangement.



glucose monomers as part of the structure of cellulose

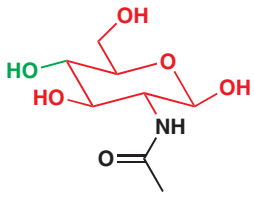
Notice that the anomeric bonds are all equatorial. This means that the cellulose molecule is linear in general outline. It is made rigid by extra hydrogen bonds between the 3-OH groups and the ring oxygen atoms—like this.



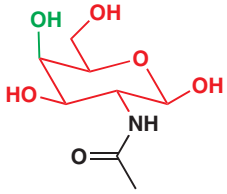
The polymer is also coiled to increase stability still further. All this makes cellulose very difficult to hydrolyse, and humans cannot digest cellulose as we do not have the necessary enzymes. Other mammals have evolved devices such as multiple stomachs (in ruminants, such as cattle) to enable them to degrade cellulose.

Amino sugars add versatility to saccharides

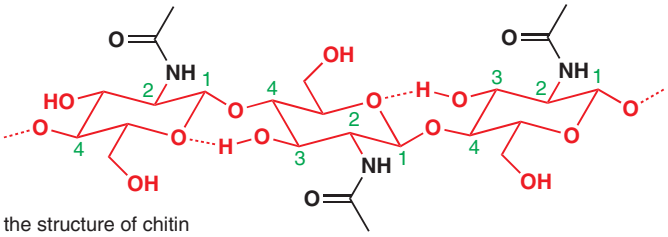
Amino sugars are carbohydrates into which nitrogen is incorporated. These molecules allow proteins and sugars to combine and produce structures of remarkable variety and beauty. The most common amino sugars are *N*-acetyl glucosamine and *N*-acetyl galactosamine, which differ only in stereochemistry. The hard outer skeletons of insects and crustaceans contain 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

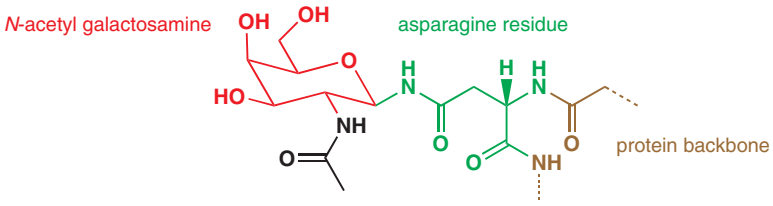


N-acetyl galactosamine

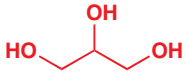


the structure of chitin

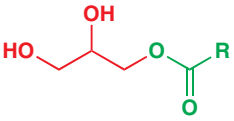
Cell membranes must not be so impermeable as they need to allow the passage of water and complex molecules. These membranes contain *glycoproteins*—proteins with amino sugar residues attached to asparagine, serine, or threonine in the protein. The attachment is at the anomeric position so that these compounds are *O*- or *N*-glycosides of the amino sugars. The structure below shows *N*-acetyl galactosamine attached to an asparagine residue as an *N*-glycoside.



Lipids

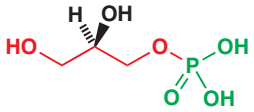


glycerol



glycerol monoester

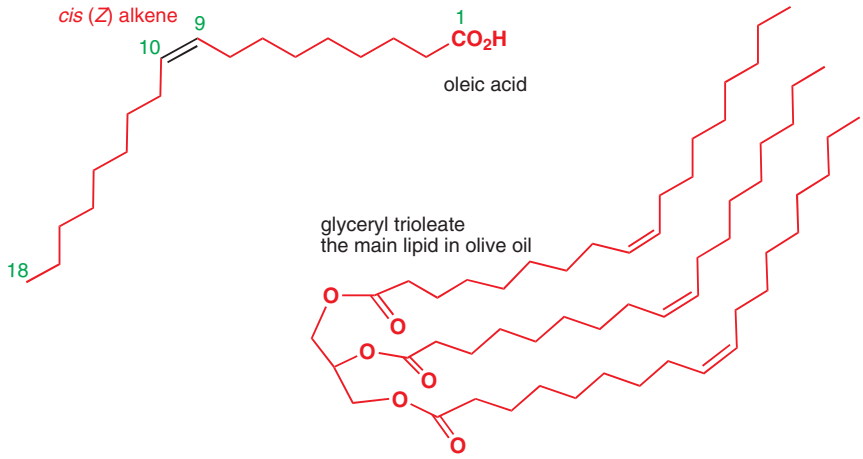
Lipids (fats) are the principal components of cell membranes. Along with cholesterol, also a component of the cell membrane, they have acquired a bad name, but they are nonetheless essential to the function of membranes as selective barriers to the movement of molecules. The most common types of lipids 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. If one of them is modified—by esterification, for



glycerol 3-phosphate

example—the molecule becomes chiral. Natural glycerol 3-phosphate is such an ester and it is optically active.

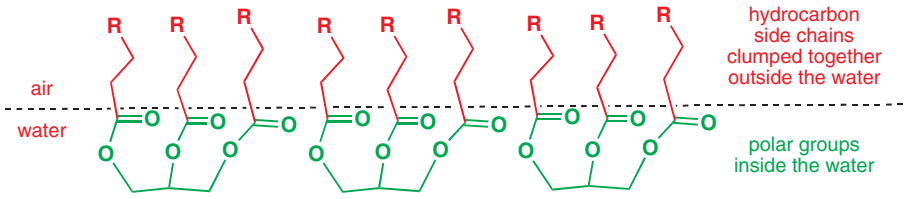
A typical lipid in foodstuffs is the triester formed from glycerol and oleic acid, which is the most abundant lipid in olive oil. Oleic acid is a mono-unsaturated fatty acid—it has one *Z* double bond in the middle of the C_{18} chain. This bond gives the molecule a marked kink in the middle. The compound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



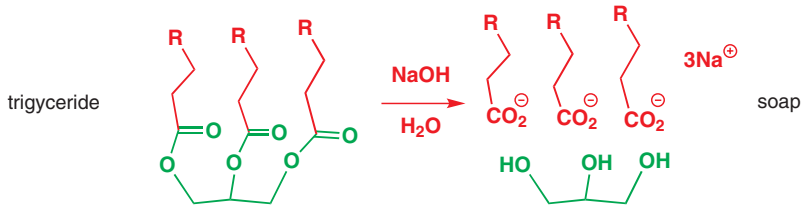
Oil and water do not mix

■ 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.

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 non-polar 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.



When triglycerides are boiled with alkali, the esters are hydrolysed and a mixture of carboxylate salts and glycerol is formed. This is how soap is made—hard soap is 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.