Skeletal formula

The **skeletal formula**, **line-angle formula**, **bond-line formula** or **shorthand formula** of an organic compound is a type of molecular structural formula that serves as a shorthand representation of a molecule's bonding and some details of its molecular geometry. A skeletal formula shows the **skeletal structure** or skeleton of a molecule, which is composed of the **skeletal atoms** that make up the molecule.[1] It is represented in two dimensions, as on a piece of paper. It employs certain conventions to represent carbon and hydrogen atoms, which are the most common in organic chemistry.

An early form of this representation was first developed by organic chemist August Kekulé, while the modern form is closely related to and influenced by the Lewis structure of molecules and their valence electrons. Hence they are sometimes termed **Kekulé structures**[a] or **Lewis–Kekulé structures**. Skeletal formulae have become ubiquitous in organic chemistry, partly because they are relatively quick and simple to draw, and also because the curved arrow notation used for discussions of reaction mechanisms and electron delocalization can be readily superimposed.

The skeletal formula of the antidepressant drug escitalopram, featuring skeletal representations of heteroatoms, a triple bond, phenyl groups and stereochemistry
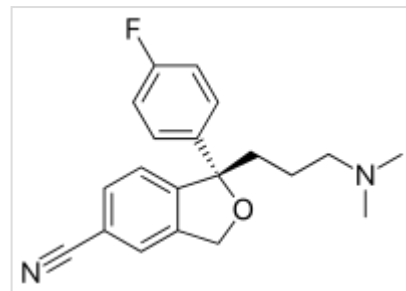
Several other ways of depicting chemical structures are also commonly used in organic chemistry (though less frequently than skeletal formulae). For example, conformational structures look similar to skeletal formulae and are used to depict the approximate positions of atoms in 3D space, as a perspective drawing. Other types of representation, such as Newman projection, Haworth projection or Fischer projection, also look somewhat similar to skeletal formulae. However, there are slight differences in the conventions used, and the reader needs to be aware of them in order to understand the structural details encoded in the depiction. While skeletal and conformational structures are also used in organometallic and inorganic chemistry, the conventions employed also differ somewhat.

The skeleton

Terminology

The skeletal structure of an organic compound is the series of atoms bonded together that form the essential structure of the compound. The skeleton can consist of chains, branches and/or rings of bonded atoms. Skeletal atoms other than carbon or hydrogen are called heteroatoms.[2]

The skeleton has hydrogen and/or various substituents bonded to its atoms. Hydrogen is the most common non-carbon atom that is bonded to carbon and, for simplicity, is not explicitly drawn. In addition, carbon atoms are not generally labelled as such directly (i.e. with "C"), whereas heteroatoms are always explicitly noted as such ("N" for nitrogen, "O" for oxygen, etc.)

Heteroatoms and other groups of atoms that give rise to relatively high rates of chemical reactivity, or introduce specific and interesting characteristics in the spectra of compounds are called underlined functional groups, as they give the molecule a function. Heteroatoms and functional groups are collectively called "substituents", as they are considered to be a substitute for the hydrogen atom that would be present in the parent hydrocarbon of the organic compound.

Basic structure

As in Lewis structures, covalent bonds are indicated by line segments, with a doubled or tripled line segment indicating double or triple bonding, respectively. Likewise, skeletal formulae indicate formal charges associated with each atom (although lone pairs are usually optional, *see below*). In fact, skeletal formulae can be thought of as abbreviated Lewis structures that observe the following simplifications:

- Carbon atoms are represented by the vertices (intersections or termini) of line segments. For clarity, methyl groups are often explicitly written out as Me or CH_3, while (hetero)cumulene carbons are frequently represented by a heavy center dot.
- Hydrogen atoms attached to carbon are implied. An unlabeled vertex is understood to represent a carbon attached to the number of hydrogens required to satisfy the octet rule, while a vertex labeled with a formal charge and/or nonbonding electron(s) is understood to have the number of hydrogen atoms required to give the carbon atom these indicated properties. Optionally, acetylenic and formyl hydrogens can be shown explicitly for the sake of clarity.
- Hydrogen atoms attached to a heteroatom are shown explicitly. The heteroatom and hydrogen atoms attached thereto are usually shown as a single group (e.g., OH, NH_2) without explicitly showing the hydrogen–heteroatom bond. Heteroatoms with simple alkyl or aryl substituents, like methoxy (OMe) or dimethylamino (NMe_2), are sometimes shown in the same way, by analogy.
- Lone pairs on carbene carbons must be indicated explicitly while lone pairs in other cases are optional and are shown only for emphasis. In contrast, formal charges and unpaired electrons on main-group elements are always explicitly shown.

In the standard depiction of a molecule, the canonical form (resonance structure) with the greatest contribution is drawn. However, the skeletal formula is understood to represent the "real molecule" – that is, the weighted average of all contributing canonical forms. Thus, in cases where two or more canonical forms contribute with equal weight (e.g., in benzene, or a carboxylate anion) and one of the canonical forms is selected arbitrarily, the skeletal formula is understood to depict the true structure, containing equivalent bonds of fractional order, even though the delocalized bonds are depicted as nonequivalent single and double bonds.

Contemporary graphical conventions

Since skeletal structures were introduced in the latter half of the 19th century, their appearance has undergone considerable evolution. The graphical conventions in use today date to the 1980s. Thanks to the adoption of the ChemDraw software package as a *de facto* industry standard (by American Chemical Society, Royal Society of Chemistry, and Gesellschaft Deutscher Chemiker publications, for instance), these conventions have been nearly universal in the chemical literature since the late 1990s. A few minor conventional variations, especially with respect to the use of stereobonds, continue to exist as a result of differing US, UK and European practice, or as a matter of personal preference.[3] As another minor

variation between authors, formal charges can be shown with the plus or minus sign in a circle (⊕, ⊖) or without the circle. The set of conventions that are followed by most authors is given below, along with illustrative examples.

1. Bonds between sp^2 or sp^3 hybridized carbon or heteroatoms are conventionally represented using 120° angles whenever possible, with the longest chain of atoms following a zigzag pattern unless interrupted by a *cis* double bond. Unless all four substituents are explicit, this is true even when stereochemistry is being depicted using wedged or dashed bonds (*see below*).[b]
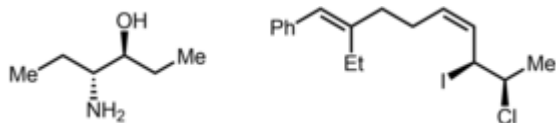
2. If all four substituents to a tetrahedral carbon are explicitly shown, bonds to the two in-plane substituents still meet at 120°; the other two substituents, however, are usually shown with wedged and dashed bonds (to depict stereochemistry) and subtend a smaller angle of 60–90°.
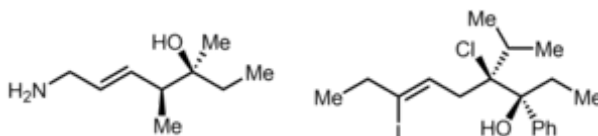
3. The linear geometry at sp hybridized atoms is normally depicted by line segments meeting at 180°. Where this involves two double bonds meeting (an allene or cumulene), the bonds are separated by a dot.
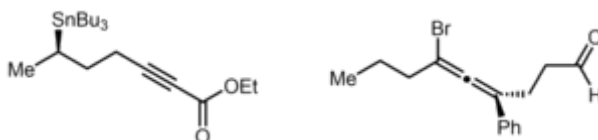
4. Carbo- and heterocycles (3- to 8-membered) are generally represented as regular polygons; larger ring sizes tend to be represented by concave polygons.[c]
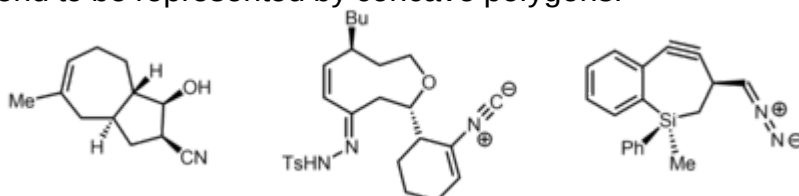
5. Atoms in a group are ordered so that the bond emanates from the atom that is directly attached to the skeleton. For example, the nitro group NO_2 is denoted —NO_2 or O_2N—, depending on the placement of the bond. In contrast, the isomeric nitrite group is denoted —ONO or ONO—.[d]
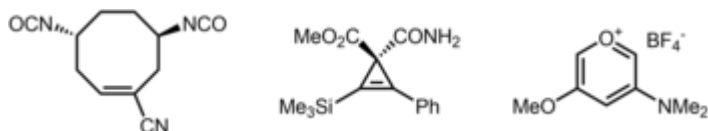
Implicit carbon and hydrogen atoms

For example, the skeletal formula of hexane (top) is shown below. The carbon atom labeled C_1 appears to have only one bond, so there must also be three hydrogens bonded to it, in order to make its total number of bonds four. The carbon atom labelled C_3 has two bonds to other carbons and is therefore bonded to two hydrogen atoms as well. A Lewis structure (middle) and ball-and-stick model (bottom) of the actual molecular structure of hexane, as determined by X-ray crystallography, are shown for comparison.
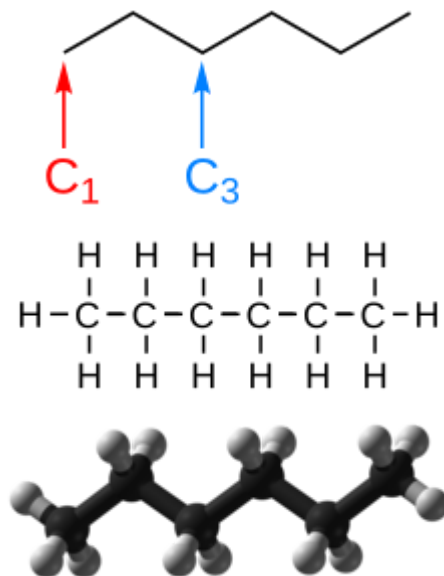
It does not matter which end of the chain one starts numbering from, as long as consistency is maintained when drawing diagrams. The condensed formula or the IUPAC name will confirm the orientation. Some molecules will become familiar regardless of the orientation.

Explicit heteroatoms and hydrogen atoms

All atoms that are not carbon or hydrogen are signified by their chemical symbol, for instance Cl for chlorine, O for oxygen, Na for sodium, and so forth. In the context of organic chemistry, these atoms are commonly known as **heteroatoms** (the prefix hetero- comes from Greek ἕτερος héteros, meaning "other").

Any hydrogen atoms bonded to heteroatoms *are* drawn explicitly. In ethanol, C_2H_5OH, for instance, the hydrogen atom bonded to oxygen is denoted by the symbol H, whereas the hydrogen atoms which are bonded to carbon atoms are not shown directly.

Lines representing heteroatom-hydrogen bonds are usually omitted for clarity and compactness, so a functional group like the hydroxyl group is most often written −OH instead of −O−H. These bonds are sometimes drawn out in full in order to accentuate their presence when they participate in reaction mechanisms.

Shown below for comparison are a skeletal formula (top), its Lewis structure (middle) and its ball-and-stick model (bottom) of the actual 3D structure of the ethanol molecule in the gas phase, as determined by microwave spectroscopy.
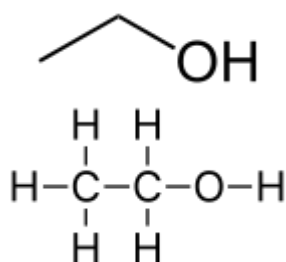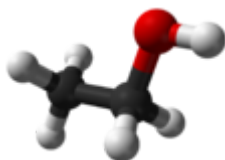
Pseudoelement symbols

There are also symbols that appear to be <u>chemical element symbols</u>, but represent certain very common substituents or indicate an unspecified member of a group of elements. These are called pseudoelement symbols or organic elements and are treated like univalent "elements" in skeletal formulae.[4] A list of common pseudoelement symbols:

General symbols

- X for any (<u>pseudo</u>)<u>halogen</u> atom (in the related <u>MLXZ notation</u>, X represents a one-electron donor ligand)
- L *or* L_n for a <u>ligand</u> or ligands (in the related MLXZ notation, L represents a two-electron donor ligand)
- M *or* Met for any <u>metal</u> atom ([M] is used to indicate a ligated metal, ML_n, when the identities of the ligands are unknown or irrelevant)
- E *or* El for any <u>electrophile</u> (in some contexts, E is also used to indicate any <u>p-block</u> element)
- Nu for any <u>nucleophile</u>
- Z for conjugating electron-withdrawing groups (in the related MLXZ notation, Z represents a zero-electron donor ligand; *in unrelated usage, Z is also an abbreviation for the carboxybenzyl group*.)
- D for <u>deuterium</u> (^{2}H)
- T for <u>tritium</u> (^{3}H)

Alkyl groups

- R for any <u>alkyl</u> group or even any <u>organyl</u> group (Alk can be used to unambiguously indicate an alkyl group)
- Me for the <u>methyl group</u>
- Et for the <u>ethyl group</u>
- Pr, *n*-Pr, *or* ^{n}Pr for the (*normal*) <u>propyl</u> group (*Pr is also the symbol for the element <u>praseodymium</u>. However, since the propyl group is monovalent, while praseodymium is nearly always trivalent, ambiguity rarely, if ever, arises in practice.*)
- *i*-Pr *or* ^{i}Pr for the <u>isopropyl</u> group
- All for the <u>allyl group</u> (uncommon)
- Bu, *n*-Bu *or* ^{n}Bu for the (*normal*) <u>butyl</u> group
- *i*-Bu *or* ^{i}Bu (*i* often italicized) for the <u>isobutyl</u> group
- *s*-Bu *or* ^{s}Bu for the *secondary* <u>butyl</u> group
- *t*-Bu *or* ^{t}Bu for the *tertiary* <u>butyl</u> group
- Pn for the <u>pentyl</u> group (*or* Am for the synonymous <u>amyl</u> group, *although Am is also the symbol for <u>americium</u>.*)

- Np *or* Neo for the neopentyl group (*Warning: Organometallic chemists often use Np for the related neophyl group, PhMe$_2$C–. Np is also the symbol for the element neptunium.*)
- Cy *or* Chx for the cyclohexyl group
- Ad for the 1-adamantyl group
- Tr *or* Trt for the trityl group

Aromatic and unsaturated substituents

- Ar for any aromatic substituent *(Ar is also the symbol for the element argon. However, argon is inert under all usual conditions encountered in organic chemistry, so the use of Ar to represent an aryl substituent never causes confusion.)*
- Het for any heteroaromatic substituent
- Bn *or* Bzl for the benzyl group (*not to be confused with Bz for benzoyl group; However, old literature may use Bz for benzyl group.*)
- Dipp for the 2,6-diisopropylphenyl group
- Mes for the mesityl group
- Ph, Φ, *or* φ for the phenyl group (*the use of phi for phenyl has been in decline*)
- Tol for the tolyl group, usually the *para* isomer
- Is *or* Tipp for the 2,4,6-triisopropylphenyl group (*the former symbol is derived from the synonym isityl*)
- An for the anisyl group, usually the *para* isomer (*An is also the symbol for a generic actinoid element. However, since the anisyl group is monovalent, while the actinides are usually divalent, trivalent, or even higher valency, ambiguity rarely, if ever, arises in practice.*)
- Cp for the cyclopentadienyl group (*Cp was the symbol for cassiopeium, a former name for lutetium*)
- Cp* for the pentamethylcyclopentadienyl group
- Vi for the vinyl group (uncommon)

Functional groups

- Ac for the acetyl group *(Ac is also the symbol for the element actinium. However, actinium is almost never encountered in organic chemistry, so the use of Ac to represent the acetyl group never causes confusion)*;
- Bz for the benzoyl group; OBz is the benzoate group
- Piv for the pivalyl (*t*-butylcarbonyl) group; OPiv is the pivalate group
- Bt for the 1-benzotriazolyl group
- Im for the 1-imidazolyl group
- NPhth for the phthalimide-1-yl group

Sulfonyl/sulfonate groups

Sulfonate esters are often leaving groups in nucleophilic substitution reactions. *See the articles on sulfonyl and sulfonate groups for further information.*

- Bs for the brosyl (*p*-bromobenzenesulfonyl) group; OBs is the brosylate group
- Ms for the mesyl (methanesulfonyl) group; OMs is the mesylate group
- Ns for the nosyl (*p*-nitrobenzenesulfonyl) group *(Ns was the chemical symbol for*

nielsbohrium, but that was renamed bohrium, Bh); ONs is the nosylate group

- Tf for the triflyl (trifluoromethanesulfonyl) group; OTf is the triflate group
- Nf for the nonaflyl (nonafluorobutanesulfonyl) group, $CF_3(CF_2)_3SO_2$; ONf is the nonaflate group
- Ts for tosyl (*p*-toluenesulfonyl) group *(Ts is also the symbol for the element tennessine. However, tennessine is too unstable to ever be encountered in organic chemistry, so the use of Ts to represent tosyl never causes confusion)*; OTs is the tosylate group

Protecting groups

A protecting group or protective group is introduced into a molecule by chemical modification of a functional group to obtain chemoselectivity in a subsequent chemical reaction, facilitating multistep organic synthesis.

- Boc for the *t*-butoxycarbonyl group
- Cbz *or* Z for the carboxybenzyl group
- Fmoc for the fluorenylmethoxycarbonyl group
- Alloc for the allyloxycarbonyl group
- Troc for the trichloroethoxycarbonyl group
- TMS, TBDMS, TES, TBDPS, TIPS, ... for various silyl ether groups
- PMB for the 4-methoxybenzyl group
- MOM for the methoxymethyl group
- THP for the 2-tetrahydropyranyl group

Multiple bonds

Two atoms can be bonded by sharing more than one pair of electrons. The common bonds to carbon are single, double and triple bonds. Single bonds are most common and are represented by a single, solid line between two atoms in a skeletal formula. Double bonds are denoted by two parallel lines, and triple bonds are shown by three parallel lines.

In more advanced theories of bonding, non-integer values of bond order exist. In these cases, a combination of solid and dashed lines indicate the integer and non-integer parts of the bond order, respectively.
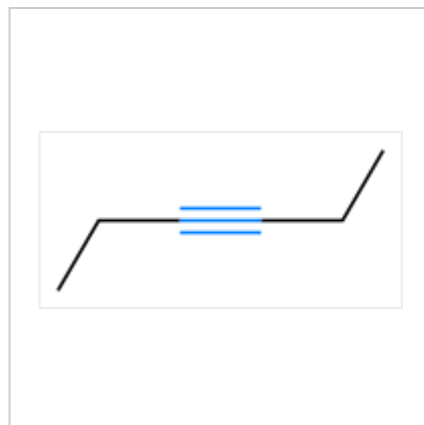
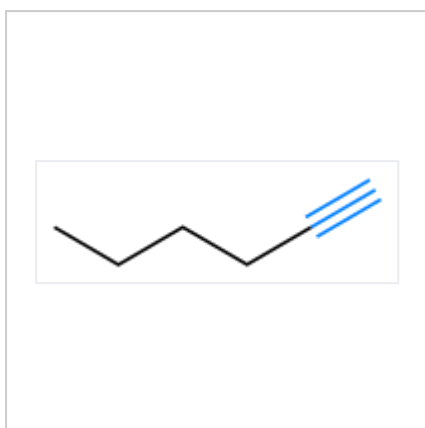**Examples of multiple bonds in skeletal formulae[e]**
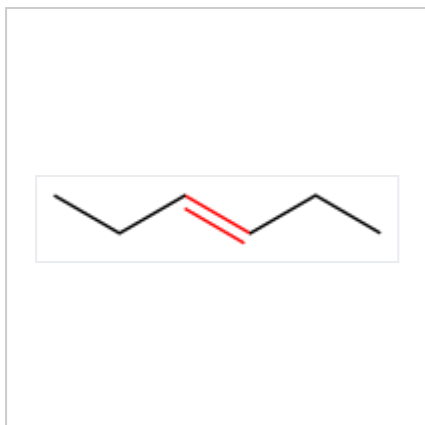
Hex-3-ene has an internal carbon–carbon double bond
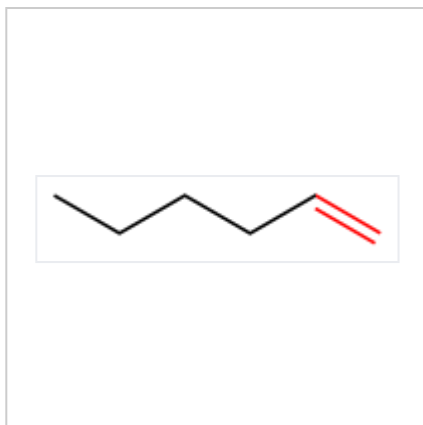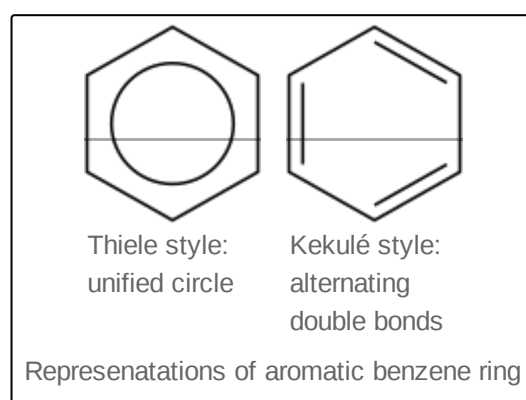

Hex-1-ene has a terminal double bond

Hex-3-yne has an internal carbon–carbon triple bond

Hex-1-yne has a terminal triple bond

Benzene rings

In recent years, benzene is generally depicted as a hexagon with alternating single and double bonds, much like the structure Kekulé originally proposed in 1872. As mentioned above, the alternating single and double bonds of "1,3,5-cyclohexatriene" are understood to be a drawing of one of the two equivalent canonical forms of benzene (the one explicitly shown and the one with the opposite pattern of formal single and double bonds), in which all carbon–carbon bonds are of equivalent length and have a bond order of exactly 1.5. For aryl rings in general, the two analogous canonical forms are almost always the primary contributors to the structure, but they are nonequivalent, so one structure may make a slightly greater contribution than the other, and bond orders may differ somewhat from 1.5.

Thiele style: unified circle

Kekulé style: alternating double bonds

Represenatations of aromatic benzene ring

An alternate representation that emphasizes this delocalization uses a circle, drawn inside the hexagon of single bonds, to represent the delocalized pi orbital. This style, based on one proposed by Johannes Thiele, used to be very common in introductory organic chemistry textbooks and is still frequently used in informal

settings. However, because this depiction does not keep track of electron pairs and is unable to show the precise movement of electrons, it has largely been superseded by the Kekuléan depiction in pedagogical and formal academic contexts.[f]
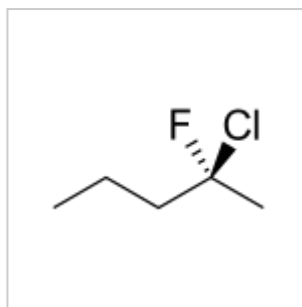
Stereochemistry

Stereochemistry is conveniently denoted in skeletal formulae:[5]
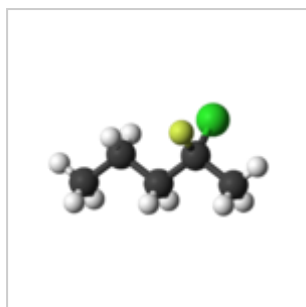
Ball-and-stick model of
(R)-2-chloro-2-fluoropentane
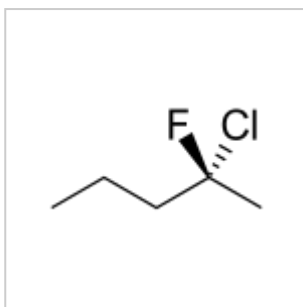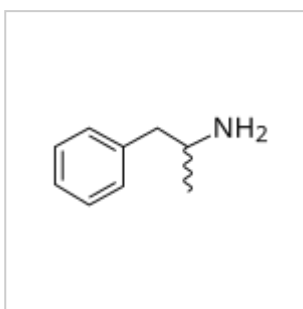

Skeletal formula of
(R)-2-chloro-2-fluoropentane


Skeletal formula of
(S)-2-chloro-2-fluoropentane

Different depictions of chemical bonds in skeletal formulas

Skeletal formula of amphetamine, indicating a mixture of two stereoisomers:
(R)- and (S)-

The relevant chemical bonds can be depicted in several ways:

- Solid lines represent bonds in the plane of the paper or screen.
- Solid wedges represent bonds that point out of the plane of the paper or screen, towards the observer.
- Hashed wedges or dashed lines (thick or thin) represent bonds that point into the plane of the paper or screen, away from the observer.[g]
- Wavy lines represent either unknown stereochemistry or a mixture of the two possible stereoisomers at that point.
- An obsolescent[h] depiction of hydrogen stereochemistry that used to be common in steroid chemistry is the use of a filled circle centered on a vertex (sometimes called H-dot/H-dash/H-circle, respectively) for an upward pointing hydrogen atom and two hash marks next to vertex or a hollow circle for a downward pointing hydrogen atom.
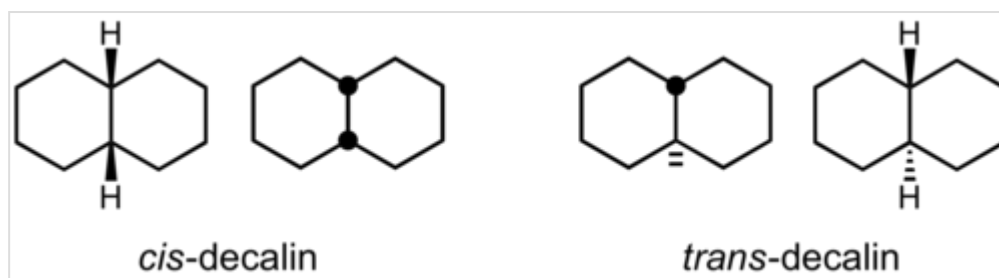
A small filled circle represented an upward pointing hydrogen, while two hash marks represented a downward pointing one.

An early use of this notation can be traced back to Richard Kuhn who in 1932 used solid thick lines and dotted lines in a publication. The modern solid and hashed wedges were introduced in the 1940s by Giulio Natta to represent the structure of high polymers, and extensively popularised in the 1959 textbook *Organic Chemistry* by Donald J. Cram and George S. Hammond.[6]

Skeletal formulae can depict *cis* and *trans* isomers of alkenes. Wavy single bonds are the standard way to represent unknown or unspecified stereochemistry or a mixture of isomers (as with tetrahedral stereocenters). A crossed double-bond has been used sometimes; it is no longer considered an acceptable style for general use but may still be required by computer software.[5]

Alkene stereochemistry
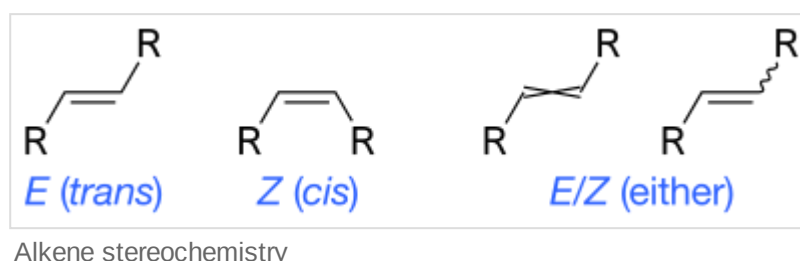
Hydrogen bonds

Hydrogen bonds are generally denoted by dotted or dashed lines. In other contexts, dashed lines may also represent partially formed or broken bonds in a transition state.
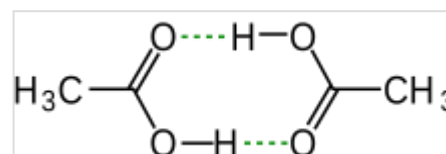
Dashed lines (green) to show hydrogen bonding in acetic acid.

Notes

a. This term is ambiguous, because "Kekulé structure" also refers to Kekulé's famous proposal of a hexagon of alternating single and double bonds for the structure of benzene.
b. To prevent a 'kink' from emerging and causing a structure to take up too much vertical space on a page, the IUPAC (Brecher, 2008, p. 352) makes an exception for long chain *cis*-olefins (such as oleic acid), allowing the *cis* double bond within them to be depicted with 150° angles, so that the zigzags on either side of the double bond can propagate horizontally.
c. Smaller rings may also be drawn as concave to show stereochemistry (such as the conformations of cyclohexane) or polycyclic molecules that cannot be drawn 'flat' without significant distortion (such as tropane and adamantane).

d. In cases where the atom has bonds coming from both the left and right (such as a secondary [amine](#) NH in the middle of a chain), some authors allow the group's formula to be stacked vertically whereas others draw an explicit vertical bond within the group.

e. In this gallery, double bonds have been shown in red and triple bonds in blue. This was added for clarity – multiple bonds are not normally coloured in skeletal formulae.

f. For instance, the acclaimed 1959 textbook by Morrison and Boyd (6th edition, 1992) uses the Thiele notation as its standard depiction of the aryl ring, while the 2001 textbook by Clayden, Greeves, Warren, and Wothers (2nd edition, 2012) uses the Kekulé notation throughout and warns students to avoid using the Thiele notation when writing mechanisms (p. 144, 2nd ed.).

g. American and European chemists use slightly different conventions for a hashed bond. Whereas most American chemists draw hashed bonds with short hash marks close to the stereocenter and long hash marks further away (in analogy to wedged bonds), most European chemists start with long hash marks close to the stereocenter that gradually become shorter moving away (in analogy to perspective drawing). In the past, the [IUPAC](#) has suggested the use of a hashed bond with hash marks of equal length throughout as a compromise but now prefers the American-style hashed bonds (Brecher, 2006, p. 1905). Some chemists use a thick bond and dotted bond (or hashed bond with equal length hashes) to depict *relative stereochemistry* and a wedged bond and hashed bond with unequal hashes to depict *absolute stereochemistry*; most others do not make this distinction.

h. The IUPAC now strongly deprecates this notation.

References

1. Stoker, H. Stephen (2012). *General, Organic, and Biological Chemistry* (6th ed.). Cengage. ISBN 978-1133103943.

2. IUPAC Recommendations 1999, Revised Section F: *Replacement of Skeletal Atoms* (http://www.chem.qmul.ac.uk/iupac/sectionF/RF5.html)

3. Brecher, Jonathan (2008). "Graphical representation standards for chemical structure diagrams (IUPAC Recommendations 2008)" (https://doi.org/10.1351%2Fpac200880020277). *Pure and Applied Chemistry*. **80** (2): 277–410. doi:10.1351/pac200880020277 (https://doi.org/10.1351%2Fpac200880020277). hdl:10092/2052 (https://hdl.handle.net/10092%2F2052). ISSN 1365-3075 (https://www.worldcat.org/issn/1365-3075).

4. Clayden, Jonathan; Greeves, Nick; Warren, Stuart; Wothers, Peter (2001). *Organic Chemistry* (1st ed.). Oxford University Press. p. 27. ISBN 978-0-19-850346-0.

5. Brecher, Jonathan (2006). "Graphical representation of stereochemical configuration (IUPAC Recommendations 2006)" (http://www.iupac.org/publications/pac/pdf/2006/pdf/7810x1897.pdf) (PDF). *Pure and Applied Chemistry*. **78** (10): 1897–1970. doi:10.1351/pac200678101897 (https://doi.org/10.1351%2Fpac200678101897). S2CID 97528124 (https://api.semanticscholar.org/CorpusID:97528124).

6. Jensen, William B. (2013). "The Historical Origins of Stereochemical Line and Wedge Symbolism". *Journal of Chemical Education*. **90** (5): 676–677. Bibcode:2013JChEd..90..676J (https://ui.adsabs.harvard.edu/abs/2013JChEd..90..676J). doi:10.1021/ed200177u (https://doi.org/10.1021%2Fed200177u).

External links

- Drawing organic molecules (http://www.chemguide.co.uk/basicorg/conventions/draw.html) from *chemguide.co.uk*

Retrieved from "https://en.wikipedia.org/w/index.php?title=Skeletal_formula&oldid=1232868349"

-