

Chemical formula

From Wikipedia, the free encyclopedia

A **chemical formula** is a way of expressing information about the proportions of atoms that constitute a particular chemical compound, using a single line of chemical element symbols, numbers, and sometimes also other symbols, such as parentheses, dashes, brackets, commas and *plus* (+) and *minus* (−) signs. These are limited to a single typographic line of symbols, which may include subscripts and superscripts. A chemical formula is not a chemical name, and it contains no words. Although a chemical formula may imply certain simple chemical structures, it is not the same as a full chemical structural formula. Chemical formulas can fully specify the structure of only the simplest of molecules and chemical substances, and are generally more limited in power than are chemical names and structural formulas.

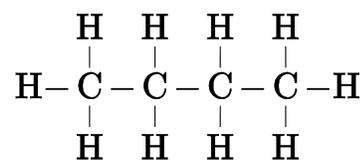
The simplest types of chemical formulas are called *empirical formulas*, which use letters and numbers indicating the numerical *proportions* of atoms of each type. *Molecular formulas* indicate the simple *numbers* of each type of atom in a molecule, with no information on structure. For example, the empirical formula for glucose is CH₂O (twice as many hydrogen atoms as carbon and oxygen), while its molecular formula is C₆H₁₂O₆ (12 hydrogen atoms, six carbon and oxygen atoms).

Sometimes a chemical formula is complicated by being written as a condensed formula (or condensed molecular formula, occasionally called a "semi-structural formula"), which conveys additional information about the particular ways in which the atoms are chemically bonded together, either in covalent bonds, ionic bonds, or various combinations of these types. This is possible if the relevant bonding is easy to show in one dimension. An example is the condensed molecular/chemical formula for ethanol, which is CH₃-CH₂-OH or CH₃CH₂OH. However, even a condensed chemical formula is necessarily limited in its ability to show complex bonding relationships between atoms, especially atoms that have bonds to four or more different substituents.

Since a chemical formula must be expressed as a single line of chemical element symbols, it often cannot be as informative as a true structural formula, which is a graphical representation of the spatial relationship between atoms in chemical compounds (see for example the figure for butane structural and chemical formulas, at right). For reasons of structural complexity, there is no condensed chemical formula (or semi-structural formula) that specifies glucose (and there exist many different molecules, for example fructose and mannose, have the same molecular formula C₆H₁₂O₆ as glucose). Linear equivalent chemical *names* exist that can and do specify any complex structural formula (see chemical nomenclature), but such names must use many terms (words), rather than the simple element symbols, numbers, and simple typographical symbols that define a chemical formula.



Aluminium sulfate has the chemical formula Al₂(SO₄)₃. The form of aluminium sulfate hexadecahydrate is Al₂(SO₄)₃•16H₂O.



Structural formula for *n*-butane. This is **not** a chemical formula. Examples of chemical formulas for *n*-butane are the *empirical formula* C₂H₅, the **molecular formula** C₄H₁₀ and the *condensed* (or *semi-structural*) *formula* CH₃CH₂CH₂CH₃.

Chemical formulas may be used in chemical equations to describe chemical reactions and other chemical transformations, such as the dissolving of ionic compounds into solution. While, as noted, chemical formulas do not have the full power of structural formulas to show chemical relationships between atoms, they are sufficient to keep track of numbers of atoms and numbers of electrical charges in chemical reactions, thus balancing chemical equations so that these equations can be used in chemical problems involving conservation of atoms, and conservation of electric charge.

Contents

- 1 Overview
- 2 Empirical formula
- 3 Molecular formula
- 4 Condensed formula
 - 4.1 In organic chemistry implying molecular geometry and structural formulas
- 5 Law of composition
 - 5.1 Chemical names in answer to limitations of chemical formulas
 - 5.2 Polymers in condensed formulas
 - 5.3 Ions in condensed formulas
- 6 Isotopes
- 7 Trapped atoms
- 8 Non-stoichiometric chemical formulas
- 9 General forms for organic compounds
- 10 Hill system
- 11 See also
- 12 References
- 13 External links

Overview

A chemical formula identifies each constituent element by its chemical symbol and indicates the proportionate number of atoms of each element. In empirical formulas, these proportions begin with a key element and then assign numbers of atoms of the other elements in the compound, as ratios to the key element. For molecular compounds, these ratio numbers can all be expressed as whole numbers. For example, the empirical formula of ethanol may be written C_2H_6O because the molecules of ethanol all contain two carbon atoms, six hydrogen atoms, and one oxygen atom. Some types of ionic compounds, however, cannot be written with entirely whole-number empirical formulas. An example is boron carbide, whose formula of CB_n is a variable non-whole number ratio with n ranging from over 4 to more than 6.5.

When the chemical compound of the formula consists of simple molecules, chemical formulas often employ ways to suggest the structure of the molecule. These types of formulas are variously known as **molecular formulas** and **condensed formulas**. A molecular formula enumerates the number of atoms

to reflect those in the molecule, so that the molecular formula for glucose is $C_6H_{12}O_6$ rather than the glucose empirical formula, which is CH_2O . However, except for very simple substances, molecular chemical formulas lack needed structural information, and are ambiguous.

For simple molecules, a condensed (or semi-structural) formula is a type of chemical formula that may fully imply a correct structural formula. For example, ethanol may be represented by the condensed chemical formula CH_3CH_2OH , and dimethyl ether by the condensed formula CH_3OCH_3 . These two molecules have the same empirical and molecular formulas (C_2H_6O), but may be differentiated by the condensed formulas shown, which are sufficient to represent the full structure of these simple organic compounds.

Condensed chemical formulas may also be used to represent ionic compounds that do not exist as discrete molecules, but nonetheless do contain covalently bound clusters within them. These polyatomic ions are groups of atoms that are covalently bound together and have an overall ionic charge, such as the sulfate $[SO_4]^{2-}$ ion. Each polyatomic ion in a compound is written individually in order to illustrate the separate groupings. For example, the compound dichlorine hexoxide has an empirical formula ClO_3 , and molecular formula Cl_2O_6 , but in liquid or solid forms, this compound is more correctly shown by an ionic condensed formula $[ClO_2]^+[ClO_4]^-$, which illustrates that this compound consists of $[ClO_2]^+$ ions and $[ClO_4]^-$ ions. In such cases, the condensed formula only need be complex enough to show at least one of each ionic species.

Chemical formulas as described here are distinct from the far more complex chemical systematic names that are used in various systems of chemical nomenclature. For example, one systematic name for glucose is *(2R,3S,4R,5R)-2,3,4,5,6-pentahydroxyhexanal*. This name, interpreted by the rules behind it, fully specifies glucose's structural formula, but the name is not a chemical formula as usually understood, and uses terms and words not used in chemical formulas. Such names, unlike basic formulas, may be able to represent full structural formulas without graphs.

Empirical formula

In chemistry, the empirical formula of a chemical is a simple expression of the relative number of each type of atom or ratio of the elements in the compound. Empirical formulas are the standard for ionic compounds, such as $CaCl_2$, and for macromolecules, such as SiO_2 . An empirical formula makes no reference to isomerism, structure, or absolute number of atoms. The term **empirical** refers to the process of elemental analysis, a technique of analytical chemistry used to determine the relative percent composition of a pure chemical substance by element.

For example, hexane has a molecular formula of C_6H_{14} , or structurally $CH_3CH_2CH_2CH_2CH_2CH_3$, implying that it has a chain structure of 6 carbon atoms, and 14 hydrogen atoms. However, the empirical formula for hexane is C_3H_7 . Likewise the empirical formula for hydrogen peroxide, H_2O_2 , is simply HO expressing the 1:1 ratio of component elements. Formaldehyde and acetic acid have the same empirical formula, CH_2O . This is the actual chemical formula for formaldehyde, but acetic acid has double the number of atoms.

Molecular formula

Molecular formulas indicate the simple numbers of each type of atom in a molecule of a molecular substance. They are the same as empirical formulas for molecules that only have one atom of a particular type, but otherwise have larger numbers. An example of the difference is the empirical formula for glucose, which is CH_2O (*ratio* 1:2:1), while its molecular formula is $\text{C}_6\text{H}_{12}\text{O}_6$ (*number of atoms* 6:12:6). For water, both formulas are H_2O . A molecular formula provides more information about a molecule than its empirical formula, but is more difficult to establish.

A molecular formula shows the number of elements in a molecule, and determines whether it is a binary compound, ternary compound, quaternary compound or has even more elements.

Condensed formula

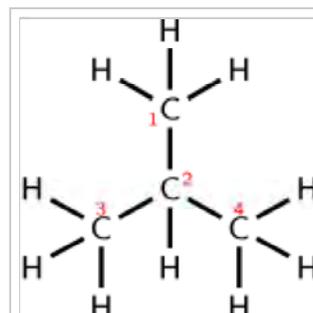
In organic chemistry implying molecular geometry and structural formulas

The connectivity of a molecule often has a strong influence on its physical and chemical properties and behavior. Two molecules composed of the same numbers of the same types of atoms (i.e. a pair of isomers) might have completely different chemical and/or physical properties if the atoms are connected differently or in different positions. In such cases, a structural formula is useful, as it illustrates which atoms are bonded to which other ones. From the connectivity, it is often possible to deduce the approximate shape of the molecule.

A condensed chemical formula may represent the types and spatial arrangement of bonds in a simple chemical substance, though it does not necessarily specify isomers or complex structures. For example, ethane consists of two carbon atoms single-bonded to each other, with each carbon atom having three hydrogen atoms bonded to it. Its chemical formula can be rendered as CH_3CH_3 . In ethylene there is a double bond between the carbon atoms (and thus each carbon only has two hydrogens), therefore the chemical formula may be written: CH_2CH_2 , and the fact that there is a double bond between the carbons is implicit because carbon has a valence of four. However, a more explicit method is to write $\text{H}_2\text{C}=\text{CH}_2$ or less commonly $\text{H}_2\text{C}::\text{CH}_2$. The two lines (or two pairs of dots) indicate that a double bond connects the atoms on either side of them.

A triple bond may be expressed with three lines or pairs of dots, and if there may be ambiguity, a single line or pair of dots may be used to indicate a single bond.

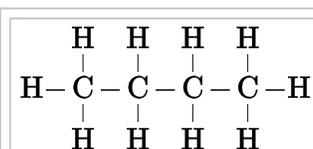
Molecules with multiple functional groups that are the same may be expressed by enclosing the repeated group in round brackets. For example, isobutane may be written $(\text{CH}_3)_3\text{CH}$. This condensed structural formula



Isobutane structural formula

Molecular formula:
 C_4H_{10}

Condensed or semi-structural chemical formula: $(\text{CH}_3)_3\text{CH}$



n-Butane structural formula

Molecular formula: C_4H_{10}

Condensed or semi-structural formula:
 $\text{CH}_3\text{CH}_2\text{CH}_2\text{CH}_3$

implies a different connectivity from other molecules that can be formed using the same atoms in the same proportions (isomers). The formula $(\text{CH}_3)_3\text{CH}$ implies a central carbon atom connected to one hydrogen atom and three CH_3 groups. The same number of atoms of each element (10 hydrogens and 4 carbons, or C_4H_{10}) may be used to make a straight chain molecule, *n*-butane: $\text{CH}_3\text{CH}_2\text{CH}_2\text{CH}_3$.

Law of composition

In any given chemical compound, the elements always combine in the same proportion with each other. This is the law of constant composition.

The law of constant composition says that, in any particular chemical compound, all samples of that compound will be made up of the same elements in the same proportion or ratio. For example, any water molecule is always made up of two hydrogen atoms and one oxygen atom in a 2:1 ratio. If we look at the relative masses of oxygen and hydrogen in a water molecule, we see that 94% of the mass of a water molecule is accounted for by oxygen and the remaining 6% is the mass of hydrogen. This mass proportion will be the same for any water molecule.^[1]

Chemical names in answer to limitations of chemical formulas

The alkene called **but-2-ene** has two isomers, which the chemical formula $\text{CH}_3\text{CH}=\text{CHCH}_3$ does not identify. The relative position of the two methyl groups must be indicated by additional notation denoting whether the methyl groups are on the same side of the double bond (*cis* or *Z*) or on the opposite sides from each other (*trans* or *E*). Such extra symbols violate the rules for chemical formulas, and begin to enter the territory of more complex naming systems.

As noted above, in order to represent the full structural formulas of many complex organic and inorganic compounds, chemical nomenclature may be needed which goes well beyond the available resources used above in simple condensed formulas. See IUPAC nomenclature of organic chemistry and IUPAC nomenclature of inorganic chemistry 2005 for examples. In addition, linear naming systems such as International Chemical Identifier (InChI) allow a computer to construct a structural formula, and simplified molecular-input line-entry system (SMILES) allows a more human-readable ASCII input. However, all these nomenclature systems go beyond the standards of chemical formulas, and technically are chemical naming systems, not formula systems.

Polymers in condensed formulas

For polymers in condensed chemical formulas, parentheses are placed around the repeating unit. For example, a hydrocarbon molecule that is described as $\text{CH}_3(\text{CH}_2)_{50}\text{CH}_3$, is a molecule with fifty repeating units. If the number of repeating units is unknown or variable, the letter *n* may be used to indicate this formula: $\text{CH}_3(\text{CH}_2)_n\text{CH}_3$.

Ions in condensed formulas

For ions, the charge on a particular atom may be denoted with a right-hand superscript. For example, Na^+ , or Cu^{2+} . The total charge on a charged molecule or a polyatomic ion may also be shown in this way. For example: H_3O^+ or SO_4^{2-} .

For more complex ions, brackets [] are often used to enclose the ionic formula, as in $[\text{B}_{12}\text{H}_{12}]^{2-}$, which is found in compounds such as $\text{Cs}_2[\text{B}_{12}\text{H}_{12}]$. Parentheses () can be nested inside brackets to indicate a repeating unit, as in $[\text{Co}(\text{NH}_3)_6]^{3+}$. Here $(\text{NH}_3)_6$ indicates that the ion contains six [ammine| NH_3 groups], and [] encloses the entire formula of the ion with charge +3.

Isotopes

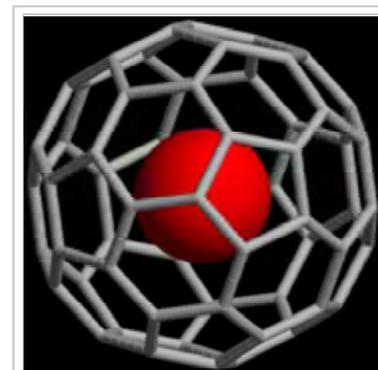
Although isotopes are more relevant to nuclear chemistry or stable isotope chemistry than to conventional chemistry, different isotopes may be indicated with a prefixed superscript in a chemical formula. For example, the phosphate ion containing radioactive phosphorus-32 is $^{32}\text{PO}_4^{3-}$. Also a study involving stable isotope ratios might include the molecule $^{18}\text{O}^{16}\text{O}$.

A left-hand subscript is sometimes used redundantly to indicate the atomic number. For example, ${}_8\text{O}_2$ for dioxygen, and $^{16}_8\text{O}_2$ for the most abundant isotopic species of dioxygen. This is convenient when writing equations for nuclear reactions, in order to show the balance of charge more clearly.

Trapped atoms

The @ symbol (at sign) indicates an atom or molecule trapped inside a cage but not chemically bound to it. For example, a buckminsterfullerene (C_{60}) with an atom (M) would simply be represented as MC_{60} regardless of whether M was inside the fullerene without chemical bonding or outside, bound to one of the carbon atoms. Using the @ symbol, this would be denoted $\text{M}@\text{C}_{60}$ if M was inside the carbon network. A non-fullerene example is $[\text{As}@\text{Ni}_{12}\text{As}_{20}]^{3-}$, an ion in which one As atom is trapped in a cage formed by the other 32 atoms.

This notation was proposed in 1991^[2] with the discovery of fullerene cages (endohedral fullerenes), which can trap atoms such as La to form, for example, $\text{La}@\text{C}_{60}$ or $\text{La}@\text{C}_{82}$. The choice of the symbol has been explained by the authors as being concise, readily printed and transmitted electronically (the at sign is included in ASCII, which most modern character encoding schemes are based on), and the visual aspects suggesting the structure of an endohedral fullerene.



Traditional formula: MC_{60}

The "@" notation: $\text{M}@\text{C}_{60}$

Non-stoichiometric chemical formulas

Chemical formulas most often use integers for each element. However, there is a class of compounds, called non-stoichiometric compounds, that cannot be represented by small integers. Such a formula might be written using decimal fractions, as in $\text{Fe}_{0.95}\text{O}$, or it might include a variable part represented by a letter, as in Fe_{1-x}O , where x is normally much less than 1.

General forms for organic compounds

A chemical formula used for a series of compounds that differ from each other by a constant unit is called **general formula**. Such a series is called the homologous series, while its members are called homologs.

For example, alcohols may be represented by: $\text{C}_n\text{H}_{(2n+1)}\text{OH}$ ($n \geq 1$)

Hill system

The Hill system (or Hill notation) is a system of writing empirical chemical formulas, molecular chemical formulas and components of a condensed formula such that the number of carbon atoms in a molecule is indicated first, the number of hydrogen atoms next, and then the number of all other chemical elements subsequently, in alphabetical order of the chemical symbols. When the formula contains no carbon, all the elements, including hydrogen, are listed alphabetically.

By sorting formulas according to the number of atoms of each element present in the formula according to these rules, with differences in earlier elements or numbers being treated as more significant than differences in any later element or number—like sorting text strings into lexicographical order—it is possible to collate chemical formulas into what is known as Hill system order.

The Hill system was first published by Edwin A. Hill of the United States Patent and Trademark Office in 1900.^[3] It is the most commonly used system in chemical databases and printed indexes to sort lists of compounds.^[4]

Examples: The following formulas are written using the Hill system, and listed in Hill order:

- BrI
- CCl_4
- CH_3I
- $\text{C}_2\text{H}_5\text{Br}$
- $\text{H}_2\text{O}_4\text{S}$

A list of formulas in Hill system order is arranged alphabetically, as above, with single-letter elements coming before two-letter symbols when the symbols begin with the same letter (so "B" comes before "Be", which comes before "Br").^[4]

See also

- Dictionary of chemical formulas

- Element symbol
- Nuclear notation
- Periodic table
- IUPAC nomenclature of inorganic chemistry
- Formula unit

References

1. "Law of Constant Composition". *Everything Math and Science*. SIYAVULA. Retrieved 31 March 2016.  This material is available under a Creative Commons Attribution-Share Alike 3.0 license.
 2. Chai, Yan; Guo, Ting; Jin, Changming; Haufler, Robert E.; Chibante, L. P. Felipe; Fure, Jan; Wang, Lihong; Alford, J. Michael; Smalley, Richard E. (1991). "Fullerenes with Metals Inside". *Journal of Physical Chemistry*. **95** (20): 7564–7568. doi:10.1021/j100173a002.
 3. Edwin A. Hill (1900). "On a system of indexing chemical literature; Adopted by the Classification Division of the U.S. Patent Office". *J. Am. Chem. Soc.* **22** (8): 478–494. doi:10.1021/ja02046a005.
 4. Wiggins, Gary. (1991). *Chemical Information Sources*. New York: McGraw Hill. p. 120.
- Ralph S. Petrucci; William S. Harwood; F. Geoffrey Herring (2002). "3". *General Chemistry: Principles and Modern Applications* (8th ed.). Prentice-Hall. ISBN 0-13-198825-5. OCLC 46872308.



Wikidata has the property:
chemical formula (*P274*)
(see uses)

External links

- Hill notation example (http://library.uml.edu/personal/Marion_Muskiewicz/hillorder.htm), from the University of Massachusetts Lowell libraries, including how to sort into Hill system order
- Molecular formula calculation applying Hill notation (<http://www.chemcalc.org>). The library calculating Hill notation is available on npm (<https://www.npmjs.com/package/chemcalc>).

Retrieved from "https://en.wikipedia.org/w/index.php?title=Chemical_formula&oldid=756491982"

Categories: Chemical formulas | Chemical nomenclature | Notation

- This page was last modified on 24 December 2016, at 17:00.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.