CHEMICAL FORMULA CONSTRUCTION

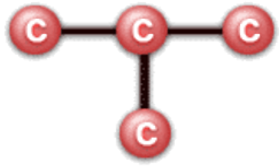
A chemical formula (also called molecular formula) is a concise way of expressing information about the atoms that constitute a particular chemical compound. It identifies each type of chemical element by its element symbol and identifies the number of atoms of such element to be found in each discrete molecule of that compound. The number of atoms (if greater than one) is indicated as a subscript (although 19th-century books often used superscripts). For non-molecular substances the subscripts indicate the ratio of elements in the empirical formula. 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. The Hill system is a common convention for writing and sorting formulas.

Molecular and structural formulas

For example methane, a simple molecule consisting of one carbon atom bonded to four hydrogen atoms has the chemical formula CH_4 and glucose with six carbon atoms, twelve hydrogen atoms and six oxygen atoms has the chemical formula $C_6H_{12}O_6$.

A chemical formula may also supply information about the types and spatial arrangement of bonds in the chemical, though it does not necessarily specify the exact isomer. For example ethane consists of two carbon atoms single-bonded to each other, each having three hydrogen atoms bonded to it. Its chemical formula can be rendered as CH_3CH_3. If there were a double bond between the carbon atoms (and thus each carbon only had two hydrogens), the chemical formula may be written: CH_2CH_2, and the fact that there is a double bond between the carbons is assumed. However, a more explicit and correct method is to write $H_2C{:}CH_2$ or $H_2C{=}CH_2$. The two dots or lines indicate that a double bond connects the atoms on either side of them.

A triple bond may be expressed with three dots or lines, and if there may be ambiguity, a single dot or line may be used to indicate a single bond. Molecules with multiple functional groups that are the same may be expressed in the following way: $(CH_3)_3CH$. However, this implies a different structure from other molecules that can be formed using the same atoms (isomers). The formula $(CH_3)_3CH$ implies a chain of three carbon atoms, with the middle carbon atom bonded to another carbon and the remaining bonds on the carbons all leading to hydrogen atoms.



However, the same number of atoms (10 hydrogens and 4 carbons, or C_4H_{10}) may be used to make a straight chain:

$CH_3CH_2CH2CH_3$.

The alkene but-2-ene has two isomers which the chemical formula $CH_3CH{=}CHCH_3$ 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*).

Polymers

For polymers, parentheses are placed around the repeating unit. For example, a hydrocarbon molecule that is described as: $CH_3(CH_2)_{50}CH_3$, is a molecule with 50 repeating units. If the number of repeating units is unknown or variable, the letter n may be used to indicate this: $CH_3(CH_2)nCH_3$.

Ions

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: hydronium, H_3O+ or sulfate, SO_4^{2-}.

Isotopes

Although isotopes are more relevant to nuclear chemistry or stable isotope chemistry than to conventional chemistry, different isotopes may be indicated with a left-hand superscript in a chemical formula. For example, the phosphate ion containing radioactive phosphorus-32 is $^{32}PO_4^{3-}$. Also a study involving stable isotope ratios might include 18O:16O.

A left-hand subscript is sometimes used to indicate redundantly, for convenience, the atomic number.

Empirical formulas

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 it. 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 could have a chemical formula of $CH_3CH_2CH_2CH_2CH_2CH_3$, implying that it has a straight chain structure, 6 carbon atoms, and 14 hydrogen atoms. However the empirical formula for the same molecule would be C_3H_7.

Non-stoichiometric formulas

Chemical formulas most often use natural numbers for each of the elements. However, there is a whole class of compounds, called non-stoichiometric compounds that cannot be represented by well-defined natural numbers. Such a formula might be written using real numbers, as in $Fe_{0.95}O$, or it might include a variable part represented by a letter, as in $Fe_{1-x}O$.

The Hill system is a system of writing chemical formulas 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. 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 numbers being treated as more significant than differences in any later number — like sorting text strings into lexicographic order — it is possible to collate chemical formulas into the Hill system order.

The Hydroxyl Group (OH):

The hydroxyl group (a hydrogen single bonded to an oxygen) gives polarity to the molecules by allowing it to form hydrogen bonds with other molecules around it. Mostly all proteins carbohydrates, and nucleic acids contain at least one hydroxyl group. Any molecule that does contain a hydroxyl group can be classified as an alcohol.

The Carbonyl Group (C=O)

The carbonyl group (carbon double bonded to an oxygen) tends to be slightly polar functional groups. Depending on where they are found bonded to the molecule, they can be classified as ketones or aldehydes. Ketones are carbonyl groups that are located in the middle of the molecule, while aldehydes are located at the end of the molecule. In simple sugars, the carbonyl group may react with a hydroxyl group to form a ring structure. Carbonyls are present in simple sugars, some proteins and some nucleotides.

The Carboxyl Group (COOH)

The carboxyl group (carbon bonded to two oxygen atoms and a hydrogen) is an extremely acidic group. This allows the functional group to ionize and releases a H+ atom decreasing the pH levels around it. Molecules that contain a carboxyl group are usually called carboxylic acids for this reason. It is also a highly polar and reactive functional group. The carboxyl group is present in amino acids and proteins.

The Amino Group (NH2)

The amino group (a nitrogen bonded to two hydrogen atoms) is the only functional group that can act as a base by accepting protons. This functional group also increases the polarity and reactivity of a molecule. The amino group allows hydrogen bonding to take place between the hydrogen and other atoms floating around near it. Between this functional group and the carboxyl group, they can bond together in amino acids to make peptide bonds, and important step in making proteins. The group is found in all amino acids, proteins and some nucleotides.

The Sulfhydryl Group (SH)

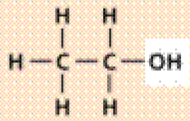
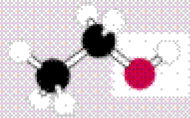
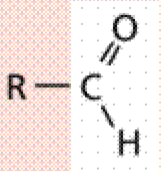
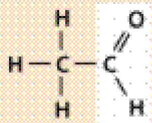
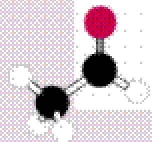
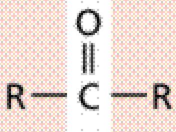
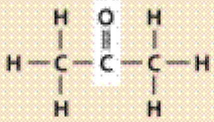
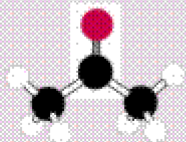
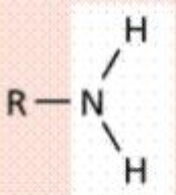
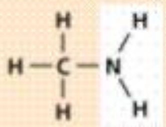
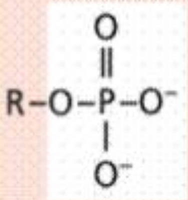
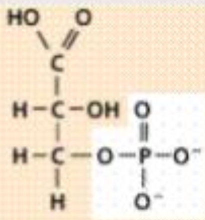
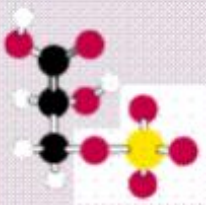
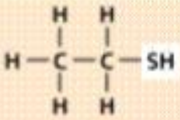
The sulfhydryl group (sulphur bonded with a hydrogen) is a reactive group that is present in the amino acid cysteine. If two of these groups come together, they are able to form a disulfide bridge strengthening and stabilizing protein structure. This group can also be known as thiols.

The Phosphate Group (OPO_3^{2-})

The phosphate group (four oxygen bound to a phosphorus atom) is highly polar, reactive, and acidic. It is important in high energy compounds such as ATP and GTP. The phosphate group is present in all nucleotides and phospholipids.

The Methyl Group (CH_3)

The methyl group (one carbon bound to three hydrogen atoms) is the only truly nonpolar functional group. The methyl group is found in alcohols, fatty acids, along with some amino acids and nucleotides. If a methyl group is added to an organic molecule, it is called methylation, and can potentially be very bad for the molecule, possibly making the molecule non-functional.

Functional group	Class of compounds	Structural formula	Example	Ball-and-stick model
Hydroxyl $-OH$	Alcohols	R — OH	 Ethanol	
Carbonyl $-CHO$	Aldehydes		 Acetaldehyde	
Carbonyl $\rangle CO$	Ketones		 Acetone	
Amino $-NH_2$	Amines		 Methylamine	
Phosphate $-OPO_3^{2-}$	Organic phosphates		 3-Phospho-glyceric acid	
Sulfhydryl $-SH$	Thiols	R — SH	 Mercapto-ethanol