WIKIPEDIA

Substituent

A **substituent** is one or a group of atoms that replaces (one or more) hydrogen atoms on the parent chain of a hydrocarbon, thereby becoming a moiety in the resultant (new) molecule. (In organic chemistry and biochemistry, the terms *substituent* and *functional group*, as well as *side chain* and *pendant group*, are used almost interchangeably to describe those branches from the parent structure,[1] though certain distinctions are made in polymer chemistry.[2] In polymers, side chains extend from the backbone structure. In proteins, side chains are attached to the alpha carbon atoms of the amino acid backbone.)

The suffix *-yl* is used when naming organic compounds that contain a single bond replacing one hydrogen; *-ylidene* and *-ylidyne* are used with double bonds and triple bonds, respectively. In addition, when naming hydrocarbons that contain a substituent, positional numbers are used to indicate which carbon atom the substituent attaches to when such information is needed to distinguish between isomers. The polar effect exerted by a substituent is a combination of the inductive effect and the mesomeric effect. Additional steric effects result from the volume occupied by a substituent.

The phrases *most-substituted* and *least-substituted* are frequently used to describe or compare molecules that are products of a chemical reaction. In this terminology, methane is used as a reference of comparison. Using methane as a reference, for each hydrogen atom that is replaced or "substituted" by something else, the molecule can be said to be more highly substituted. For example:

- Markovnikov's rule predicts that the hydrogen atom is added to the carbon of the alkene functional group, which has the greater number of hydrogen atoms (fewer alkyl substituents).
- Zaitsev's rule predicts that the major reaction product is the alkene with the more highly substituted (more stable) double bond.

Contents

Nomenclature

The suffix *-yl* is used in organic chemistry to form names of radicals, either separate species (called *free radicals*) or chemically bonded parts of molecules (called *moieties*). It can be traced back to the old name of methanol, "methylene" (from Ancient Greek: μέθυ *méthu*, 'wine' and ὕλη *húlē*,[3] 'wood',

'forest'), which became shortened to "methyl" in compound names, from which *-yl* was extracted. Several reforms of chemical nomenclature eventually generalized the use of the suffix to other organic substituents.

The use of the suffix is determined by the number of hydrogen atoms that the substituent replaces on a parent compound (and also, usually, on the substituent). According to the 1993 IUPAC recommendations:[4]

- *-yl* means that one hydrogen is replaced.
- *-ylidene* means that two hydrogens are replaced by a double bond between parent and substituent.
- *-ylidyne* means that three hydrogens are replaced by a triple bond between parent and substituent.

The suffix *-ylid**ine*** is encountered sporadically, and appears to be a variant spelling of "-ylidene";[5] it is not mentioned in the IUPAC guidelines.

For multiple bonds of the same type, which link the substituent to the parent group, the prefixes *di-*, *tri-*, *tetra-*, etc., are used: *-diyl* (two single bonds), *-triyl* (three single bonds), *-tetrayl* (four single bonds), *-diylidene* (two double bonds).

For multiple bonds of different types, multiple suffixes are added: ***-**ylylidene* (one single and one double), *-ylylidyne* (one single and one triple), *-diylylidene* (two single and one double).

The parent compound name can be altered in two ways:

- For many common compounds the substituent is linked at one end (the 1 position) and historically not numbered in the name. The IUPAC 2013 Rules[6] however *do* require an explicit locant for most substituents in a preferred IUPAC name. The substituent name is modified by stripping *-ane* (see alkane) and adding the appropriate suffix. This is "recommended only for saturated acyclic and monocyclic hydrocarbon substituent groups and for the mononuclear parent hydrides of silicon, germanium, tin, lead, and boron". Thus, if there is a carboxylic acid called "*X*-ic acid", an alcohol ending "*X*-anol" (or "*X*-yl alcohol"), or an alkane called "*X*-ane", then "*X*-yl" typically denotes the same carbon chain lacking these groups but modified by attachment to some other parent molecule.
- The more general method omits only the terminal "e" of the substituent name, but requires explicit numbering of each *yl* prefix, even at position 1 (except for -ylidyne, which as a triple bond must terminate the substituent carbon chain). *Pentan-1-yl* is an example of a name by this method, and is synonymous with *pentyl* from the previous guideline.

Note that some popular terms such as "vinyl" (when used to mean "polyvinyl") represent only a portion of the full chemical name.

Methane substituents

According to the above rules, a carbon atom in a molecule, considered as a substituent, has the following names depending on the number of hydrogens bound to it, and the type of bonds formed with the remainder of the molecule:

CH_4	methane	no bonds
$-CH_3$	methyl group or methanyl	one single bond to a non-hydrogen atom
$=CH_2$	methylene group or methanylidene or methylidene	one double bond
$-CH_2-$	methylene bridge or methandiyl or methdiyl	two single bonds
≡CH	methanylidyne group or methylidyne	one triple bond
=CH−	methine group or methanylylidene or methylylidene	one single bond and one double bond
>CH−	methantriyl group or methtriyl	three single bonds
≡C−	methanylylidyne group or methylylidyne	one triple bond and one single bond
=C=	methandiylidene group or methdiylidene	two double bonds
>C=	methandiylylidene group or methdiylylidene	two single bonds and one double bond
>C<	methantetrayl group or methtetrayl	four single bonds

Structures

In a chemical structural formula, an organic substituent such as methyl, ethyl, or aryl can be written as *R* (or R^1, R^2, etc.) It is a generic placeholder, the *R* derived from *radical* or *rest*, which may replace any portion of the formula as the author finds convenient. The first to use this symbol was Charles Frédéric Gerhardt in 1844.[7]

The symbol *X* is often used to denote electronegative substituents such as the halides.[8][9]

Statistical distribution

One cheminformatics study identified 849,574 unique substituents up to 12 non-hydrogen atoms large and containing only carbon, hydrogen, nitrogen, oxygen, sulfur, phosphorus, selenium, and the halogens in a set of 3,043,941 molecules. Fifty substituents can be considered common as they are found in more than 1% of this set, and 438 are found in more than 0.1%. 64% of the substituents are found in only one molecule. The top 5 most common are the methyl, phenyl, chlorine, methoxy, and hydroxyl substituents. The total number of organic substituents in organic chemistry is estimated at 3.1 million, creating a total of 6.7×10^{23} molecules.[10] An infinite number of substituents can be obtained simply by increasing carbon chain length. For instance, the substituents methyl ($-CH_3$) and pentyl ($-C_5H_{11}$).

See also

- Functional groups are a subset of substituents

References

1. D.R. Bloch (2006). *Organic Chemistry Demystified* (https://books.google.com/books?id=yVPcSIn5xjAC&pg=PT88). ISBN 978-0-07-145920-4.
2. Jenkins, A. D.; Kratochvíl, P.; Stepto, R. F. T.; Suter, U. W. (1996). "PAC, 1996, 68, 2287. Glossary

of basic terms in polymer science (IUPAC Recommendations 1996)" (http://goldbook.iupac.org/src_PAC1996682287.html). *Pure and Applied Chemistry*. **68** (12): 2287–2311. doi:10.1351/pac199668122287 (https://doi.org/10.1351%2Fpac199668122287). This distinguishes a *pendant group* as neither oligomeric nor polymeric, whereas a *pendant chain* must be oligomeric or polymeric.

3. This name came through a Greek language error: ὕλη (*hȳlē*) means "wood" ("forest"), ξυλο- (*xylo-*) means "wood" (the substance)
4. IUPAC (1997) [1993]. "R-2.5 Substituent Prefix Names Derived from Parent Hydrides" (http://www.acdlabs.com/iupac/nomenclature/93/r93_271.htm). *A Guide to IUPAC Nomenclature of Organic Compounds (Recommendations 1993)*. Blackwell Scientific Publications; Advanced Chemistry Development, Inc.
5. The PubChem database lists 740,110 results for *-ylidene*, of which 14 have synonyms where the suffix is replaced by *-ylidine*. Another 4 results contain *-ylidine* without listing *-ylidene* as a synonym.
6. *Nomenclature of Organic Chemistry. IUPAC Recommendations and Preferred Names 2013*. Favre, Henri A.,, Powell, Warren H., 1934–, International Union of Pure and Applied Chemistry. Cambridge, UK: Royal Society of Chemistry. 2013. ISBN 9781849733069. OCLC 865143943 (https://www.worldcat.org/oclc/865143943).
7. See:
   - Charles Gerhardt, *Précis de chimie organique* (Summary of organic chemistry), vol. 1 (Paris, France: Fortin et Masson, 1844), page 29. From page 29: (https://books.google.com/books?id=sjlBAAAAYAAJ&pg=PA29) "*En désignant, par conséquent, les éléments combustibles par R, sans tenir comptes des proportions atomiques de carbone et d'hydrogène, on peut exprimer d'une manière générale: Par R. — Les hydrogènes carbonés.*" (Consequently, by designating combustible components by *R*, without considering the atomic proportions of carbon and hydrogen, one can express in a general way: By *R* — hydrocarbons.)
   - William B. Jensen (2010) "Ask the Historian: Why is R Used for Hydrocarbon Substituents?," *Journal of Chemical Education*, **87**: 360–361. Available at: University of Cincinnati. (http://www.che.uc.edu/jensen/W.%20B.%20Jensen/Reprints/163.%20Origin%20of%20R.pdf)
8. Jensen, W. B. (2010). "Why Is "R" Used To Symbolize Hydrocarbon Substituents?". *Journal of Chemical Education*. **87** (4): 360–361. Bibcode:2010JChEd..87..360J (https://ui.adsabs.harvard.edu/abs/2010JChEd..87..360J). doi:10.1021/ed800139p (https://doi.org/10.1021%2Fed800139p).
9. The first use of the letter *X* to denote univalent electronegative groups appeared in:
   - Stanislao Cannizzaro (1858) "Sunto di un corso di filosofia chimica, fatto nella R. Universita di Genova" (https://books.google.com/books?id=VQOrAAAAIAAJ&pg=PA321) (Sketch of a course of chemical philosophy, offered at the Royal University of Genoa), *Il Nouvo Cimento* (The New Experiment), **7** : 321–366. From page 355: (https://books.google.com/books?id=VQOrAAAAIAAJ&pg=PA355) " … *X indica tutto ciò che vi è nella molecola, oltre l'idrogeno metallico …* " (… *X stands for all that is in the molecule, apart from metallic hydrogen …*).
10. Ertl, P. (2003). "Cheminformatics Analysis of Organic Substituents: Identification of the Most Common Substituents, Calculation of Substituent Properties, and Automatic Identification of Drug-like Bioisosteric Groups". *Journal of Chemical Information and Modeling*. **43** (2): 374–380. doi:10.1021/ci0255782 (https://doi.org/10.1021%2Fci0255782). PMID 12653499 (https://pubmed.ncbi.nlm.nih.gov/12653499).
