

IUPAC Recommendations

Albert J. Dijkstra*, Karl-Heinz Hellwich, Richard M. Hartshorn, Jan Reedijk and Erik Szabo

End-of-line hyphenation of chemical names (IUPAC Recommendations 2020)

<https://doi.org/10.1515/pac-2019-1005>

Received October 16, 2019; accepted January 21, 2020; published online November 19, 2020

Abstract: Chemical names can be so long that, when a manuscript is printed, they have to be hyphenated/divided at the end of a line. Many names already contain hyphens, but in some cases, using these hyphens as end-of-line divisions can lead to illogical divisions in print, as can also happen when hyphens are added arbitrarily without considering the ‘chemical’ context. The present document provides guidelines for authors of chemical manuscripts, their publishers and editors, on where to divide chemical names at the end of a line, and instructions on how to avoid these names being divided at illogical places. Readability and chemical sense should prevail when authors insert hyphens. The software used to convert electronic manuscripts to print can now be programmed to avoid illogical end-of-line hyphenation and thereby save the author much time and annoyance when proofreading. The Recommendations also allow readers of the printed article to determine which end-of-line hyphens are an integral part of the name and should not be deleted when ‘undividing’ the name. These Recommendations may also prove useful in languages other than English.

Keywords: Chemical nomenclature; dividing chemical names; end-of-line hyphenation; systematic chemical names; typesetting; word processing.

CONTENTS

Hyp-1	Introduction	49
Hyp-2	Definitions	50
Hyp-2.1	Name components	50
Hyp-2.2	Prefixes	50
Hyp-2.3	Suffixes	51
Hyp-2.4	Endings	51
Hyp-2.5	Locants and other structural indicators	51
Hyp-2.6	Stereodescriptors	51
Hyp-2.7	Hyphens and other punctuation marks	51
Hyp-2.8	Special symbols and practices used in the past	52

Article note: This manuscript was prepared in the framework of IUPAC project 2014-003-2-800.

***Corresponding author: Albert J. Dijkstra**, Ajuinlei 13c, 9000 Gent, Belgium, e-mail: albert@dijkstra-tucker.be.

<https://orcid.org/0000-0001-9031-6292>

Karl-Heinz Hellwich, Beilstein-Institut zur Förderung der Chemischen Wissenschaften, Trakehner Str. 7–9, 60487 Frankfurt, Germany. <https://orcid.org/0000-0002-4811-7254>

Richard M. Hartshorn, School of Physical and Chemical Sciences, University of Canterbury, Private Bag 4800, Christchurch, New Zealand. <https://orcid.org/0000-0002-6737-6200>

Jan Reedijk, Leiden Institute of Chemistry, PO Box 9502, 2300 RA Leiden, The Netherlands.

<https://orcid.org/0000-0002-6739-8514>

Erik Szabo, Department of Physical and Theoretical Chemistry, Comenius University in Bratislava, Slovakia.

<https://orcid.org/0000-0002-5976-983X>

Hyp-3	General approach to dividing chemical names at the end of a line	52
Hyp-3.1	Introduction	52
Hyp-3.2	Division at spaces	53
Hyp-3.3	Insertion of end-of-line hyphens	53
Hyp-3.4	Division at hyphens	53
Hyp-3.5	Line breaks at the punctuation marks: solidus, en dash, and em dash	54
Hyp-3.6	'Undividing' divided systematic chemical names	54
Hyp-4	Correct typesetting practices	54
Hyp-4.1	Introduction	54
Hyp-4.2	Unicode encoding and inserting Unicode characters	55
Hyp-4.3	Special characters with typesetting functions	55
Hyp-4.3.1	Introduction	55
Hyp-4.3.2	Non-breaking space	56
Hyp-4.3.3	Non-breaking hyphen	56
Hyp-4.3.4	Optional hyphen	56
Hyp-4.3.5	No-width non break	57
Hyp-4.4	Correct typesetting in specific situations	57
Hyp-4.4.1	Minus sign vs. the hyphen-minus	57
Hyp-4.4.2	Colon and center dot	57
Hyp-4.4.3	Solidus	57
Hyp-4.4.4	Hyphens	58
Hyp-4.4.5	Dashes	58
Hyp-5	Dividing chemical names in a chemically meaningful manner	59
Hyp-5.1	Unacceptable insertion points for end-of-line hyphens	59
Hyp-5.2	Division between name components	59
Hyp-5.3	Minimum number of non-divided consecutive characters	60
Hyp-5.4	No division after locants or descriptors	60
Hyp-5.5	Division between name components and prefixes	60
Hyp-5.6	Division before suffixes	61
Hyp-5.7	Division before endings	61
Hyp-5.8	Generally accepted divisions	61
Hyp-5.9	Division in accordance with pronunciation and chemical meaning	62
Hyp-5.10	Division within a set of stereodescriptors or locants	63
Hyp-5.11	Unacceptable division at spaces	62
Hyp-6	InChI and InChIKey	63
Hyp-7	Linear chemical formulae	63
Hyp-7.1	Molecular formulae	63
Hyp-7.2	Shorthand notation	63
Hyp-7.3	Structural formulae	64
Hyp-7.4	Linear formulae	64
Hyp-7.5	Proteins (three letter abbreviation)	64
Hyp-7.6	Proteins (one letter abbreviation)	64
Hyp-7.7	Nucleotides	64
Hyp-8	Annex	64
Hyp-8.1	Role of the author	65
Hyp-8.2	Role of the publisher/typesetter	65
	Membership of sponsoring bodies during preparation of the Provisional Recommendations	67
	References	67

Hyp-1 Introduction

Chemical compounds can be described in print in several different ways.

- (1) There are the various types of chemical names:
 - a. Systematic names such as those recommended by the International Union of Pure and Applied Chemistry (IUPAC).
 - b. Accepted or retained trivial names such as the names of some amino acids or carboxylic acids, ketones such as ‘acetone’, alcohols such as ‘ethylene glycol’, and many other names.
 - c. Trivial names such as ‘soda ash’ (sodium carbonate).
 - d. Trade names such as ‘IMODIUM®’, a registered trademark for loperamide which is the International Nonproprietary Name (INN) for ‘4-[4-(4-chlorophenyl)-4-hydroxypiperidinyl]-N,N-dimethyl-2,2-diphenylbutanamide’, or ‘Teflon®’ (polytetrafluoroethene).
 - e. Other kinds of names such as the INNs for pharmaceutical substances, for example ‘trandolapril’, a compound for which the systematic name reads: ‘(2S,3aR,7aS)-1-[(2S)-2-[[[(2S)-1-ethoxy-1-oxo-4-phenylbutan-2-yl]-amino]propanoyl]octahydro-1H-indole-2-carboxylic acid’.
- (2) By using abbreviations or acronyms such as ‘DCM’ (‘dichloromethane’), but which need to be explained on their first use within a document [1].
- (3) By using InChI (IUPAC International Chemical Identifier) and InChIKey.
- (4) By using molecular or structural notations such as:
 - a. Linear formulae, like $C_{24}H_{34}N_2O_5$ or Na_2HPO_4 .
 - b. Structural formulae for linear compounds, such as ‘acrylamide’, $CH_2=CH-CONH_2$ and ‘bis(trifluoromethanesulfonic)anhydride’, $(CF_3SO_2)_2O$.
 - c. Shorthand structural formulae, such as NaOMe for ‘sodium methoxide’.
 - d. Shorthand notations for proteins showing the amino acid order by using either the three-letter symbols or the one-letter symbols [2].
 - e. Shorthand notations for nucleotides that show the sequence of nucleic acid fragments in DNA by using a string of one-letter symbols, like in d(AGAGCTAGCTCT).

Chemical names, and in particular systematic chemical names, can be so long in print that they have to be divided over more than one line of text. With respect to trivial names, dictionaries such as *Merriam-Webster's New Collegiate® Dictionary* [3], the *New Oxford Spelling Dictionary* [4], or the *Dictionary of Contemporary English* [5] indicate at which points words can be hyphenated. However, it should be noted that these dictionaries often fail to take the chemistry into account. Also, there are sometimes division differences between UK English and US English.

On the other hand, systematic names in many cases already contain hyphens, but sometimes not in a suitable position for dividing at the end of a line. This means that hyphens have to be inserted into the name between characters. Since there is a fair amount of confusion about the appropriate positions, the present document provides pertinent recommendations and guidelines. While in principle being general, it mainly focuses on the hyphenation of systematic chemical names.

When an article is about to be published, the author sends the electronic manuscript to the publisher who uses typesetting software to convert the manuscript into print. This conversion may entail dividing chemical names at the end of a line. The typesetting software used by the publisher, like the word processing software used by the authors, recognizes a hyphen as a location where a word can be divided. So, a name such as ‘1-methyl-1,2-dihydronaphthalene’ may get divided at the third hyphen in this name. However, this does not make sense chemically. In chemical terminology, the ‘1,2’ locants are closely associated with the ‘dihydro’. Just listen to yourself when you pronounce the name. You stop after ‘1-methyl’ and again after ‘1,2-dihydro’ and you pronounce ‘1,2-dihydro’ as if it were a single word. There is, therefore, a need to prevent software from using existing hyphens, or inserting new hyphens, to divide names at illogical places.

In general, word processing and typesetting software makes use of pre-loaded dictionaries that will show how to divide words at the end of a line. Accordingly, it will contain words like: ‘an-a-lyse’, ‘anal-y-sis’,

‘an-a-lyst’ or ‘ana-lyst’,¹ and so forth. It will contain some chemical names like ‘ac-et-al-de-hyde’ and ‘acet-amide’, but ‘1-methyl-1,2-di-hy-dro-naph-tha-lene’ is less likely to have been included. Consequently, names like that will be divided manually or by some kind of computer assumption in accordance with the typographical needs at the time, and the result of that division will then be included in the database. It may therefore contain ‘1-methyl-1,2-dihydro-naphthalene’ and/or ‘1-methyl-1,2-di-hydronaphthalene’ and/or ‘1-methyl-1,2-dihydronaphthalene’, *etc.* Consequently, there is clearly a need to provide databases with names that are divided at logical places and to prevent names that have been divided at illogical places from being included in these databases.

Authors may themselves have divided chemical names in their manuscripts for the simple reason that they were too long and didn’t fit on the line. When this division is no longer necessary in print, the typesetter will have to know whether or not to maintain this end-of-line hyphen. Is it part of the name, or is it merely an end-of-line division? There is, therefore, a need to distinguish between these two kinds of hyphen in a manuscript.

In a later stage of publishing, the publisher sends the proofs to the author who is likely to encounter such illogical end-of-line divisions. If the author corrects these illogical divisions, these corrections may lead to subsequent illogical divisions further down the paragraph (or even on other pages), necessitating a second proofreading. Authors would therefore be well served if proofs did not contain such illogical end-of-line divisions. These can be avoided by asking authors to indicate in their manuscript where a systematic chemical name may be divided in accordance with the present Recommendations, and which hyphens form part of the systematic name and should not be used for end-of-line divisions. Fortunately, word processing software in use provides means to indicate this difference, but this requires that such indications are correctly interpreted by the typesetting software used by the publisher. More detailed instructions for authors, publishers and typesetters are given in the Annex (Hyp-8) to these Recommendations.

Hyp-2 Definitions

Hyp-2.1 Name components

Systematic chemical names describe the structure of a compound by listing, for example, its parent name, its substituents and/or added terms, or the names of the ligands and the central atom of a complex. In the present Recommendations, these name parts will be referred to as name ‘components’. Name components refer to atoms or groups of atoms that have a name. So, the name ‘chloroethane’ consists of two name components: the group of atoms forming the constituent ‘ethane’, a parent hydride name, and the chlorine atom as the substituent ‘chloro’, a substituent prefix. When the ‘chloroethane’ itself becomes a substituent and is then referred to as ‘2-chloroethyl’, this substituent also consists of two name components: ‘chloro’ and ‘ethyl’. When more than one chlorine substituent is present as in ‘1,2-dichloroethane’, the multiplicative prefix (di) and the locants (1 and 2) indicating the positions of the chlorine atoms form an integral part of that name component (‘1,2-dichloro’).

Hyp-2.2 Prefixes

Besides substituent prefixes, chemical names can also contain various other prefixes. The multiplicative prefix indicates the number of times a name component occurs, *i.e.*, ‘mono’, ‘di’, ‘tri’, ‘tetra’, ‘penta’, ‘hexa’, *etc.*, for non-substituted name components, and ‘bis’, ‘tris’, ‘tetrakis’, *etc.*, for identically substituted name components. These prefixes can distinguish between isomers, *e.g.*, ‘iso’, ‘tert’; they can indicate a ring structure or

¹ These examples have been chosen to illustrate the unpredictability of where to divide normal words [3, 5]. Other Dictionaries may provide even other hyphenations, *e.g.* [4].

modification, e.g., ‘cyclo’, ‘spiro’, ‘homo’, ‘nor’, ‘seco’, *etc.* Names of polymers, in most cases, have the prefix ‘poly’.

Prefixes can also specify the configuration, in which case they are called stereodescriptors (see Hyp-2.6). When used as locants in, e.g., stereodescriptors (see Hyp-2.5), they indicate a position.

Prefixes or part of them, depending on the kind of prefix, can be joined to the name component they are often followed by a hyphen, as in ‘*tert*-butyl’, but not necessarily so, as in ‘isopropyl’.

Hyp-2.3 Suffixes

Suffixes in chemical names always indicate a functional group, usually the principal characteristic group, or the free valence of a substituent group such as ‘yl’ or ‘ylidene’. Suffixes can be short as the suffixes ‘ol’ indicating an alcohol group, ‘ide’ or ‘ate’ for an anionic position and ‘one’ indicating a ketone, but also longer suffixes can occur, like ‘amide’ or those consisting of more than one name component, e.g. ‘oic acid’.

Hyp-2.4 Endings

Endings are part of the parent name and are therefore regarded as part of the name component. The most common examples of endings indicate the number of multiple bonds such as ‘ane’ in hexane, ‘ene’ in propene and ‘yne’ in ethyne. Other examples are ‘ine’ in pyridine, or in the names of the halogens, or ‘olidine’ in pyrrolidine.

Hyp-2.5 Locants and other structural indicators

A locant is a numeral or a letter, or a combination of both that identifies position(s) in a structure (modified from [6, 7]). Typical examples are ‘1’, ‘4a’, ‘N’ or ‘ α ’, ‘ β ’, *etc.* For organic-chemical compounds, more details on the different kinds and the use of locants can be found in P-14.3 [8]. Other structural indicators include the bridging indicator, ‘ μ ’, which may be followed by a numerical subscript denoting the number of bridged metals, and ‘ κ ’ and ‘ η ’, which are used to indicate connectivity and bonding in inorganic compounds as explained in IR-2.14 and in IR-9.2.5.2, IR-9.2.4, and IR-10.2.3 [9], and for polymers for example in [6, 7, 10]. In the vast majority of cases, locants are separated from each other by commas and from surrounding text by hyphens. A locant can also appear in a general expression, such as ‘2-substituted’ or ‘ α -hydrogen’.

Hyp-2.6 Stereodescriptors

A stereodescriptor is a prefix that specifies the configuration (absolute or relative) or conformation [11] of a compound or a stereogenic unit of the compound. Typical examples are ‘R’, ‘S’; ‘r’, ‘s’; ‘P’, ‘M’; ‘A’, ‘C’; ‘R_a’, ‘S_a’; ‘E’, ‘Z’; ‘cis’, ‘trans’; ‘fac’, ‘mer’; ‘D’, ‘L’; ‘ap’, ‘sp’. Most stereodescriptors are directly preceded by their corresponding locant and enclosed by parentheses together with this locant. For more details on the use of stereodescriptors, see P-9 [8] and IR-9.3 [9].

Hyp-2.7 Hyphens and other punctuation marks

In normal text, the hyphen (-) is used to link words which belong closely together (e.g., ‘end-of-line’). However, it is also used as an end-of-line character to link two parts of a word at a line break. In that case, the hyphen is somewhat conditional, *i.e.*, if the text is rewritten, the word may have to be “undivided” by removing the hyphen. However, it can also be an integral part of the word that may not be removed.

It is this double function of the hyphen that is the crux of problems with end-of-line division in chemistry. Unfortunately, unlike some other languages, English is not very precise in terms of what to do when the two different functions of the hyphen coincide. Indeed, plain English seldom encounters cases where it is not

self-evident whether or not an end-of-line hyphen was a part of the word before its division (*e.g.*, consider the end-of-line division of homographs like ‘re-treat’ and ‘retreat’).

In chemistry, the context does not reveal the origin of a hyphen in the same way as in plain language. In chemical nomenclature, the hyphen is used in several contexts, most importantly to separate locants, stereodescriptors, or italic prefixes from other parts of the name. In general, it is much more abundant than in plain language. Full details on the use of hyphens can be found in sections IR-2.3.1 of [9] and P-16.2.4 of [8]. It is obvious that a chemist needs clear rules to unambiguously determine the origin of each hyphen.

It is the very purpose of these Recommendations to make a clear distinction between the “virtual” end-of-line hyphens that denote division of words and that have to be deleted on “undividing” and the true “chemical” hyphens that are integral parts of chemical names.

In addition to the hyphen, two more symbols represented by a horizontal line are in use: the en dash and the em dash. They differ in length: hyphen (–), en dash (—), and em dash (—). For details on the use of the em dash see IR-2.3.3 of [9] and P-16.2.5 of [8]. The en dash has a very distinct meaning in typography. It means a junction between one item/value and another. Examples for the use of the en dash can be found in IR-10.2.5.1 of [9], and in [10] (The use of the en dash as originally prescribed in polymer nomenclature has not been continued in the new edition of [7].) A related symbol is the minus sign. The minus sign is very similar, but not identical to the en dash (see for example IR-2.3.2 in [9]). It is used to indicate the charge of an ion, or in a name, like in ‘tetracarbonylcobaltate(1–)’ and therefore never functions as a hyphen.

Chemical formulae may also contain a solidus (oblique stroke, slash, /) in polymer descriptions like ‘poly-[(chloromethylene)/methylene]’, or to separate the Arabic numerals which indicate the proportion of individual constituents in a compound like ‘boron trifluoride—water (1/2)’. The center dot (also known as middle dot) is also used in compound formulae, as shown in $3\text{Cd}(\text{SO}_4)\cdot 8\text{H}_2\text{O}$ and in $\text{Cu}(\text{SO}_4)\cdot 5\text{H}_2\text{O}$. Finally, the colon can be used in compound names, such as: ‘di- μ -hydroxido- μ -nitrito- $\kappa\text{N}:\kappa\text{O}$ -bis(triamminecobalt)(3+)’.

Hyp-2.8 Special symbols and practices used in the past

In some documents, specific symbols have been used instead of, or in addition to, a hyphen to indicate a division of a name at a position where a hyphen is not an integral part of the name. In other documents, the equals sign (=) was used for that purpose, for example in [12] and in the German translation of the 1990 edition of the Red Book [13].

The 2005 edition of the Red Book [9], *e.g.*, on page 265, used the symbol \supset which is similar to the symbol used in printed CAS Index Guides [14].

In the *Principles of Chemical Nomenclature* [15], the very conspicuous symbol \blacktriangleright was added when a chemical name was broken at the end of a line, even after a hyphen. Similarly, the symbol \neg is used in *British Approved Names*, BAN [16]. The same rationale is followed by other authors who always add a hyphen, even after an existing hyphen, when a name was broken at the end of a line, like in amino acid sequences (see below in Hyp-7.5). This last approach has also been adopted in the present Recommendations, as will be explained in the following sections.

Hyp-3 General approach to dividing chemical names at the end of a line

Hyp-3.1 Introduction

In this section, the overall rules are established for end-of-line division of chemical names as applicable not only to printed text but also to handwriting, for instance. Rules concerning how the desired results should be achieved in print by correct type-setting practices, as well as rules concerning whether a division is chemically allowed or not at specific positions, will be discussed in the subsequent sections.

A chemist needs to make a clear and unambiguous distinction between those hyphens that are added to denote an end-of-line division (and may be omitted after rewriting in undivided form), and those hyphens that are an integral part of a chemical name. The general approach to ensure this distinction is to make sure that the two different functions of the hyphen never overlap or interfere.

A hyphen appearing at the end of a line should always be considered as added, only to denote division, and should be omitted upon rewriting the lines in undivided form. This role should be the only function and interpretation of any hyphen appearing at the end of a line, and any end-of-line hyphen should never be considered to have any other function, *e.g.*, as a part of a chemical name. In contrast, if a line ends with any symbol other than a hyphen, the lines are connected on rewriting using a space.

Similarly, a hyphen that was present in a chemical name before its division always remains in that original, grammatical function and never serves as an end-of-line hyphen. An end-of-line hyphen may be added or removed, but hyphens that are a part of chemical name must never appear in the end-of-line position.

Hyp-3.2 Division at spaces

A chemical name consisting of two or more separate words, *e.g.*, ‘benzoic acid’ or ‘ethyl acetate’, can generally be divided at an existing space when necessary. In that case no additional end-of-line hyphen is used to indicate this division. For exceptions, see Hyp-5.11.

For example, when divided by this method, the name ‘benzoic acid’ will appear as:

(text text text text text text text) ... benzoic
acid ... (text text text text text text text text text)

Hyp-3.3 Insertion of end-of-line hyphens

Although not the preferred way, a chemical name not containing a hyphen at a suitable position for dividing can be divided at a chemically meaningful position by inserting an end-of-line hyphen and putting the rest of the name onto the next line. Dividing a chemical name at the end of a line without adding an end-of-line hyphen that indicates continuation of the name on the next line is, like in any other word, unacceptable – except at a space in the name (see Hyp-3.2). In particular, an end-of-line hyphen must be inserted if a division is made after enclosing marks, or other symbols in a name.

For example, ‘2-(chloromethyl)benzoic acid’ can be divided after ‘benzoic’ without an additional end-of-line symbol (see Hyp-3.2), but if it had to be divided before ‘benzoic’, an end-of-line hyphen would have to be inserted after the ‘2-(chloromethyl)’, *i.e.*,

(text text text text text text text) ... 2-(chloromethyl)-
benzoic acid ... (text text text text text text text text text)

The fact that the name component of the divided name on the lower line does not start with a hyphen shows that there was no hyphen in the chemical name in that position before the name was divided. The hyphen at the end of the upper line serves one function only, *i.e.*, to indicate that a word has been divided; it is not part of the chemical name. This is in fact an example of the general rule, as explained in Hyp-3.1, that a hyphen at the end of a line should always be deleted on ‘undividing’.

Hyp-3.4 Division at hyphens

A chemical name containing hyphens should preferably be divided at a hyphen that is part of the name. However, not every hyphen in a name is suitable for dividing that name (see Hyp-5 for such situations).

If a hyphen is considered to be suitable for dividing the name, the break is indicated, as described in Hyp-3.3, by adding a new end-of-line hyphen at the end of the upper line, while the original hyphen from the name is repeated at the beginning of the lower line. This is in accordance with the way proteins are divided at the end of a line when written as a series of amino acids (*cf.* Hyp-7.5) and it makes sense to follow this procedure. This additional hyphen also greatly facilitates the ‘undividing’ (*cf.*, Hyp-3.6) of divided names: just delete one hyphen when there are two.

For example, the name ‘1,2-dihydro-1- λ^5 -phosphine’ may be conveniently divided at the position of the second hyphen, after the ‘dihydro’ prefix. Divided this way, the name will appear as:

(text text text text text text text) ... 1,2-dihydro-
-1- λ^5 -phosphine ... (text text text text text text text)

Hyp-3.5 Line breaks at the punctuation marks: solidus, en dash, and em dash

When a chemical name contains a solidus, an en dash, or an em dash, and these characters are used in a function similar to a hyphen, as for example in names of addition compounds, like ‘boron trifluoride—water (1/2)’, they may be used to divide a chemical name at that location.

In these cases, the approach outlined above is extrapolated. Division is denoted by adding a new end-of-line hyphen and the name is continued on the new line. To avoid confusion, the special character that was identified as a suitable position to divide the word, should not be left on the upper line, but should be moved to the beginning of the new line by inserting an optional hyphen in front of the special character.

For example, the name ‘boron trifluoride—water (1/2)’ may be conveniently divided at the position of the em dash. Divided this way, the name will appear as:

(text text text text text text text) ... boron trifluoride-
—water (1/2) ... (text text text text text text text)

As an example of dividing a name in the position of a solidus, ‘poly[(chloromethylene)/methylene]’ can be divided to appear as:

(text text text text text) ... poly(chloromethylene)-
/methylene ... (text text text text text text text)

Hyp-3.6 ‘Undividing’ divided systematic chemical names

‘Undividing’ systematic chemical names can be summarized by the simple rule that all hyphens at the end of a line have been inserted for division purposes only, and must therefore be deleted on ‘undividing’.

Hyp-4 Correct typesetting practices

Hyp-4.1 Introduction

The general approach described in Hyp-3 pertains to the physical appearance of end-of-line divisions as required for their unambiguous interpretation, and so it applies even to handwriting. However, professional texts are, nowadays, always prepared by using digital word-processing software. Every author thus needs some basic knowledge about how this software works, including how the text is formatted by correct typesetting.

While publishers use professional software and very advanced functions that authors need not be aware of, the typesetter has no way of knowing the intended behavior of a chemical name if authors neglect specifying this information at their end. Therefore, authors need to know how the text they prepared in their word-processing software encodes (or misencodes, if neglected) this information.

Hyp-4.2 Unicode encoding and inserting Unicode characters

Digital text files may have their characters encoded according to various standards of encoding. By far the most modern and unified standard of encoding is Unicode. Accordingly, characters described in the subsequent sections will be referred to by their Unicode names and hexadecimal codes, *e.g.*, the ‘Greek small letter kappa’ is the Unicode character with code 03BA. The shorthand notation for this is ‘U+03BA’. Please note, that this is not a keyboard shortcut, but a shorthand denoting the unique identity of a Unicode character.

Modern operating systems and modern word-processing software usually include a general method of inserting any Unicode character based on its code. This is, indeed, true for Microsoft Windows (Windows) and Microsoft Word (MS Word), that provide users with the following two general procedures for inserting Unicode characters:

- In MS Word (as of the 2019 edition), any Unicode character (except for so-called ‘control characters’, see Hyp-4.3.1) can be inserted by typing its hexadecimal four-digit code (*e.g.*, for the small letter kappa either ‘03ba’ or ‘03BA’), followed by pressing Alt + X. This replaces the four-digit code with the corresponding Unicode character.
- Alternatively, Windows provides another procedure of inserting Unicode characters. While the Alt key is being held, typing the code of the Unicode character in its decimal representation leads to the insertion of the intended character. Hexadecimal 03BA corresponds to decimal 954 and this code can be used to insert the small letter kappa by the Windows procedure. A leading zero is sometimes required in the decimal, and in that case, it must not be omitted.

It should be noted, however, that while Unicode is the most developed way of specifying and distinguishing various characters, it is only universally reliable for those characters that are defined by their form, not function. Implementation of special characters used for formatting and controlling the text (see the next section) is still not universally based on Unicode characters, even if Unicode characters with the expected functions might already exist. In the Annex, two Tables are given with relevant Unicode codes and related characters (Tables 1 and 2).

Hyp-4.3 Special characters with typesetting functions

Hyp-4.3.1 Introduction

Besides characters that function solely as printable representations of letters, symbols *etc.*, there are also characters, that provide specific functions used to format the text during typesetting. Many of these characters are normally hidden from the user and appear only in a special mode of viewing the document.

The list of the most common functions that special characters fulfill in documents is quite small compared to the list of printable characters. Below, just four functions of special characters will be described that are needed for the correct typesetting of end-of-line hyphenation. These four special characters are common to the vast majority of word-processing software, but they are not always implemented through the same Unicode characters, and their specific names might vary between various software packages.

Publishers should always specify the word-processing software that their authors are expected to use, and include instructions on which particular characters and practices are expected to secure the four basic special-character functions listed below. On the other hand, an author should be expected to be familiar with viewing hidden special characters in the word-processing software and using at least the four basic special-character functions identified below.

Since the most common word-processor is MS Word, the following descriptions use the names of special characters as they are used in this software. In MS Word, many special characters can be inserted using keyboard shortcuts, but the most general way is through the dialog window accessed from the main menu item Insert > Symbol > More Symbols... and its ‘Special Characters’ tab. If other word processors than MS Word are

allowed by the publisher, it should be up to the publisher to determine the practices that correspond to those discussed below and to provide clear instructions to the author in this respect.

Hyp-4.3.2 Non-breaking space

As will be illustrated in Hyp-5.11, there are occasions where two name components that are separated by a space should not be separated from each other at a line break. So the name ‘boron trifluoride—water (1/2)’ should not be divided between ‘water’ and ‘(1/2)’.

In MS Word, the non-breaking space is represented by a Unicode character called ‘no-break space’ (U+00A0, decimal code 0160). It can be inserted either through the ‘Special Characters’ dialog, or by the general procedures for entering Unicode characters (see Hyp-4.2), but also with a keyboard shortcut **Ctrl** + **Shift** + **Space**.

In this text, the “°” symbol will be used to indicate non-breaking space, *i.e.*, the text of the above example would be shown as ‘boron trifluoride—water°(1/2)’.

Hyp-4.3.3 Non-breaking hyphen

The non-breaking hyphen special character, also known as the hard hyphen, is a hyphen at which no line break is allowed. It is not treated as a hyphen by word-processing software, rather as a letter. Consequently, a line never ends on a non-breaking hyphen.

In MS Word, the non-breaking hyphen is represented by a software-specific control character (U+001E, decimal code 030), but *cannot* be inserted by the general procedure for entering Unicode characters of MS Word (using the hexadecimal code), only by the procedure for Windows (using the decimal code, see Hyp-4.2). A Unicode character called the ‘non-breaking hyphen’ (U+2011) also exists, and even though it would work if inserted into MS Word manually, this is not the character that MS Word uses as its non-breaking hyphen by default. The native MS Word non-breaking hyphen (encoded as U+001E) should be preferred to inserting U+2011. It is normally inserted either through the ‘Special Characters’ dialog, or by using a keyboard shortcut **Ctrl** + **Shift** + **=** (but this only works with the hyphen key, not with the numeric minus key **Num-**).

In this text, the equals sign ‘=’ will represent the non-breaking hyphen, *i.e.*, the typesetting of ‘1,2-dichloroethane’ would be shown as ‘1,2=dichloroethane’.

Hyp-4.3.4 Optional hyphen

The optional or soft hyphen is used to indicate at what point a word can be divided if this were to be convenient. In MS Word, it shows on screen as ¬, when hidden characters are on display. If the word is divided at the point indicated by the optional hyphen, a normal hyphen appears in print at the end of the line. Otherwise, printers omit the optional hyphen in print as they omit any other hidden character.

In MS Word, the optional hyphen is represented by a software-specific control character (U+001F, decimal code 031). It cannot be inserted by using the hexadecimal code (see Hyp-4.2), but only by the procedure using the decimal code (see Hyp-4.2). A Unicode character called ‘soft hyphen’ (U+00AD) does exist, but if inserted in MS Word, it does not work as an optional hyphen, but is always displayed (and letter-like, *i.e.*, non-breaking). The MS Word optional hyphen is normally inserted either through the ‘Special Characters’ dialog or with a keyboard shortcut **Ctrl** + **-** (but this only works with the hyphen key, not with the numeric minus key **Num-**).

In this text, the ‘~’ (tilde) symbol will be used to indicate an optional hyphen, *i.e.*, the typesetting of ‘cyclohexane’ would be shown as ‘cyclo~hexane’.

Hyp-4.3.5 No-width non break

A no-width non break, is a non-printing special character that prevents text from breaking at the position where this character has been inserted, even if it falls at the end of a line. In chemical nomenclature, this is especially important when using the en dash or the em dash, since word processors treat them as normal hyphens (*i.e.*, as dividing characters) and there are instances, like ‘bis(pentacarbonylmanganese)(Mn–Mn)’, in which name components joined by a dash should not be divided.



In MS Word, the no-width non break is represented by a Unicode character called ‘zero-width joiner’ (U+200D, decimal code 8205). It can be inserted either through the ‘Special Characters’ dialog or by either of the general procedures for entering Unicode characters (see Hyp-4.2).

A Unicode character called ‘zero-width no-break space’ (U+FEFF, decimal code 65279) also exists. It can be inserted in MS Word by both general methods of entering Unicode characters (see Hyp-4.2), and the character currently works as intended. However, its use is deprecated since Unicode 3.2 (released 2002) and was supposed to be replaced by the character called ‘word joiner’ (U+2060, decimal code 8288). This character really does work as word joiner in MS Word, but is always zero-width, even when special characters are displayed, thus practically invisible (still, one can tell that the character is present by the movements of the cursor). Out of the three options, it is recommended to use the default MS Word no-width non break (encoded as U+200D).

In the present text, the ‘*’ (asterisk) symbol will be used to indicate a no-width non break, *i.e.*, the typesetting of ‘bis(pentacarbonylmanganese)(Mn–Mn)’ would be shown as ‘bis(pentacarbonylmanganese)(Mn*—*Mn)’.



Hyp-4.4 Correct typesetting in specific situations

Hyp-4.4.1 Minus sign vs. the hyphen-minus


Properly inserting a minus sign character is often underestimated. Keyboards typically have a separate hyphen key  and a separate numeric minus sign key , but they usually yield the same character, called a hyphen-minus. Devised as a compromise to save space in the earliest character encodings, this hyphen-minus was supposed to represent both, the hyphen and the minus. But, in fact, it is not properly designed to represent either. Using this unspecific hyphen-minus character should be avoided.

The proper Unicode ‘minus sign’ character – (U+2212, decimal code 8722) should be used to make it distinct from other characters that have the form of a horizontal line (see Hyp-4.4.4 and Hyp-4.4.5). In MS Word, the proper minus sign character is treated like a letter and does not allow words to be divided at its position. This is the only desired behavior of a minus sign in chemical names, and MS Word does not require any additional considerations to secure it. See also Hyp-4.3.3.

Hyp-4.4.2 Colon and center dot

Unicode character ‘colon’ (U+003A, decimal code 58) is usually available on keyboards on the same key as the semicolon, and thus normally inserted by pressing  + . The center dot is inserted as the Unicode character U+00B7 (see Hyp-4.2). Like the minus sign, the colon and the center dot are also treated like letters in MS Word, and do not cause words to be divided at their positions. Consequently, no specific action is required to prevent end-of-line division.

Hyp-4.4.3 Solidus

The Unicode ‘solidus’ character (U+002F, decimal code 47) is usually available on keyboards as the  key. Like the colon and the center dot, the solidus character is treated in MS Word like a letter and does not cause words to be divided at its position. However, in the case of solidus, this might not be its only desired behavior, as a

solidus in a chemical name can also represent a suitable point for dividing it. Depending on whether this is true or not, the solidus needs to be typeset in either of the following two ways:

- (1) If a solidus *is not* a suitable point for dividing the name, it is typeset simply as the solidus character, relying on its default non-breaking nature to prevent division of the chemical name at its position.
- (2) If a solidus in a chemical name *is* a suitable point for dividing the name, it should be typeset as a sequence of the optional hyphen followed by the solidus, in this text shown as '~/'. This assures that in case the word does get divided, the optional hyphen becomes visible at the end of the line, while the solidus starts the new line, as intended.

Hyp-4.4.4 Hyphens

Just as the hyphen-minus (see Hyp-4.4.1) character is not suitable to represent the minus sign, it is also not suitable to represent the hyphen. This use of the hyphen-minus is discouraged. In addition, even though there exists a proper hyphen character, its use in correct typesetting of chemical names is discouraged as being similarly unspecific. There are only two roles a hyphen can fulfill in a chemical name, and neither is typeset by a regular hyphen. These two roles are:

- (1) If a hyphen in a chemical name *is not* a suitable point for dividing the name, it must be typeset as the non-breaking hyphen, in this text shown as '≡'.
- (2) If a hyphen in a chemical name *is* a suitable point for dividing the name, it should be typeset as a sequence of a pair of hyphens, *i.e.*, an optional hyphen followed by a non-breaking hyphen, in this text shown as '≡'. This assures that in case the word does get divided, the optional hyphen becomes visible at the end of the line, while the non-breaking hyphen starts the new line, as intended.

N.B. As will be explained in Hyp-8.1, if authors wish to avoid ambiguity, they should refrain from using regular hyphens when typing chemical names. Instead, they should use non-breaking hyphens where the name shows a hyphen and insert optional hyphens and no-width non breaks at a later stage.

Hyp-4.4.5 Dashes

Dashes can also be inserted in different ways. By using the Unicode 'en dash' character (U+2013, decimal code 8211), by using the keyboard shortcut **Ctrl** + **Num-** in MS Word, and by typing 0150 while holding the **Alt** key in Windows. For the em dash the Unicode 'em dash' character (U+2014, decimal code 8212) can be used, in MS Word the shortcut **Ctrl** + **Alt** + **Num-** leads to insertion of the em dash as does holding the **Alt** key while typing 0151 in Windows.

It must be noted that word processors usually treat the dashes like hyphens, *i.e.*, if they appear near the end of the line, the processor may automatically divide the word by leaving the dash as the end-of-the-line character. This must be avoided. Therefore, a dash should never be typeset on its own, and special characters are always needed to provide either of the two possible desired behaviors of a dash:

- (1) If a dash in a chemical name *is* a suitable point for dividing the name, it should be typeset as a sequence of an optional hyphen followed by a dash, in this text shown as '~—'. This assures that if the word does get divided, the optional hyphen becomes visible at the end of the line, while the dash starts the new line, as is the intention.
- (2) If the dash in a chemical name *is not* a suitable point for dividing the name, it should be typeset as a sequence of a no-width non break, followed by the dash, followed by a second no-width non break, in this text shown as '*—*'. This assures that the chemical name does not get divided before or after the dash.

Hyp-5 Dividing chemical names in a chemically meaningful manner

Section Hyp-3 showed how divisions are made if they are allowed, while section Hyp-4 showed how to implement allowed division by correct typesetting of chemical names if a digital word processor is used for writing. In the present section, Hyp-5, we will discuss whether divisions are allowed as chemically meaningful.

The rules for deciding which divisions of chemical names are chemically meaningful and which are not, consist of prohibitions and prescriptions. The purpose of the rules is to make it easier to read a printed text and to identify hyphens that have only been added to divide the name; the additional guidelines are mainly advisory.

Hyp-5.1 Unacceptable insertion points for end-of-line hyphens

There are positions in a chemical name or a chemical formula at which a division or line break is unacceptable. These comprise:

- Within a consecutive series of letters unless an optional hyphen has been inserted at a point where a break is permitted. For example: The systematic name of loperamide should not be broken as for example done below after the p in: '4-[4-(4-chlorop-henyl)-4-hydroxypiperidin-yl]- N,N-dimethyl-2,2-diphenylbutanamide'.
- within a locant (e.g., '13' must not be divided as '1-3').
- within a stereodescriptor ('endo' or 'RS' should not be divided, as 'en-do' or 'R-S', respectively).
- between a stereodescriptor and its locant ['(13R)' should not be divided as '(13-R)'].
- within descriptors like '2λ⁵', '3δ²', or 'κN¹:κN²'. So a name like '1,2-dihydro-1λ⁵-phosphine' should not be divided before or after the λ, e.g., not ...1-λ⁵... or ... 1 λ⁵ (see also item i).
- directly before a closing enclosing mark; e.g., '...methyl)' should not be divided as '...methyl-)'.
- directly after an opening enclosing mark; e.g., '(chloro...' should not be divided as '(-chloro...'.
- in between opening or between closing marks. Accordingly, the compound with the formula '[(Mo₆Cl₈) Cl₃{(C₆H₅)₂PCH₂CH₂P(C₆H₅)₂py}Cl]' should not be divided as: [(Mo₆Cl₈)Cl₃{-(C₆H₅)₂PCH₂CH₂P(C₆H₅)₂py}Cl].
- directly before or after a punctuation mark, such as a comma, semicolon, or colon ('1,1,1-trichloro' should not be divided after a comma as '1,-1,1-trichloro',. and 'κN: κO' should not be divided after the colon as 'κN:- κO'). For some exceptions see Hyp-5.10.
- directly before or after an en dash that represents a bond such as in 'bis(pentacarbonylmanganese)(Mn—Mn)', or in the bridge descriptors, e.g., '[B1—B1]', in names of polycyclic macromolecules [10].
- directly before a valence or charge indication (e.g., not 'Cu-(II)' or 'Fe-(3+)'.
- directly before a right subscript or superscript (e.g., not 'SO₄'), directly after a left subscript or superscript (e.g., not '³¹P') and within subscripts or superscripts (e.g., not '²³⁵U').
- within a set of primes (e.g., so that 'N''' should always be on the same line; when more than 3 primes are needed, a non-breaking space should be inserted to assist reading.
- within an italicized prefix (e.g., *arachno* should not be divided).
- within an element symbol (e.g., not N-a) or within a group of atoms (e.g., not N-O₃).
- within a formula having a center dot (e.g., not Cu(SO₄)·5H₂O or Cu(SO₄)·H₂O).
- after a solidus within a constituent ratio (e.g., not 'boron trifluoride—water (3/-8)').
- within a common abbreviation or acronym (e.g., not in FeMoco, EXAFS, NADPH).

Hyp-5.2 Division between name components

A systematic chemical name written as one word and comprising two or more name components, can be divided between these components. Accordingly, 'chloroethane' can be divided as 'chloro~ethane' where the ~ sign indicates the optional hyphen as defined above in Hyp-4.3.4. Similarly, 'diamminedichloridoplatinum' can be divided as 'diammine~dichlorido~platinum', or 'poly(oxyethane-1,2-diyl)' as 'poly(oxy~ethane-1,2-diyl)'. If a name component contains enclosing marks as in '4-(chloromethyl)pyridine' or 'dichloridobis(urea)copper(II)',

the hyphen has to be inserted after the closing enclosing mark: ‘4-(chloromethyl)~pyridine’ and ‘dichloridobis(urea)~copper(II)’, respectively. Of course, a hyphen can always be inserted between two groups of parentheses, as in ‘tricarbonyl(triethylphosphane)~(trimethylsilyl)cobalt’.

Hyp-5.3 Minimum number of non-divided consecutive characters

The ACS Style Guide of the American Chemical Society [17] recommends leaving at least three characters on each line whereby a hyphen is counted as a character. Accordingly, it allows the name component ‘cyclohexyl’ in a name like ‘cyclohexylbenzene’ to be divided at the indicated positions: ‘cy~clo~hex~yl’. In order to improve the recognizability of divided words, these Recommendations opt for a larger number of characters and suggest a number of at least six or exceptionally five consecutive characters at the end of a line (where the dividing hyphen should not be counted as a character). This count should be made from the beginning of the word. The number of consecutive characters on the next line can be smaller, but not less than two characters, for instance ‘butan~one’.

Applying this rule to the above example will prevent its name component from being divided as ‘cyclohexyl’ but will allow it to be divided as ‘cyclo-hexyl’. On the other hand, despite the fact that dividing as ‘cyclohex-ylbenzene’ would provide sufficient characters at the end of the line, it is not recommended because the suffix ‘-yl’ is part of the name component and should not be divided from it (see Hyp-5.2 and Hyp-5.6).

Hyp-5.4 No division after locants or descriptors

In Hyp-5.1 it has already been stated that a division between a stereodescriptor and its locant is unacceptable. In addition, it is recommended that a division is generally avoided at a hyphen following a locant. In the Introduction (Hyp-1), the pronunciation of ‘1-methyl-1,2-dihydronaphthalene’ was discussed and it was concluded that the ‘1,2’ is closely associated with the name component ‘dihydro’. In general, locants are to be considered part of the name component that follows. Accordingly, the hyphen between the locant or locant set and the name component must be considered as a non-breaking hyphen (see Hyp-4.3.3). This is particularly important for contracted forms of substituent groups such as ‘2-pyridyl’. Using the notation defined in Hyp-4.3 therefore leads to ‘1=methyl~1,2=dihydro~naphthalene’ and ‘2=pyridyl’, respectively. The combination ‘~=’ after ‘methyl’ ensures that just one hyphen is printed when the name is not divided at that point and that two hyphens are printed on division.

This procedure also extends to the situation where a name component is preceded by both a stereodescriptor and a locant. Accordingly, non-breaking hyphens should be used as indicated by the symbol ‘=’ in the example ‘(2*R*)=2=chloro~butane’.

Systematic names can also contain letters that are followed by numerals or *vice versa*, such as ‘indicated hydrogen’ or ‘added hydrogen’ (see P-14.7 [8]), *e.g.*, ‘2*H*’, ‘2(3*H*)’, lambda and delta descriptors (P-21.2.4 and P-25.7.2 [8]), *e.g.*, ‘2λ⁵’, ‘3δ²’, and other descriptors that do not contain hyphens, *e.g.*, ‘(¹³C, ²H)’ (Chapter P-8 [8]), ‘κN¹: κN²’ (IR-9.2.4.2 [9]). These descriptors should not be separated from their name components at an end-of-line break. See Hyp-5.1 item d.

Hyp-5.5 Division between name components and prefixes

In general, an end-of-line division between a name component and a prefix is acceptable if they are written as one word, *i.e.*, without space or hyphen. Therefore, ‘cyclohexane’ can be divided between the prefix and the compound name according to ‘cyclo~hexane’. The name ‘1,2-dichloroethane’ can also be divided after the prefix, *i.e.*, ‘1,2=dichloro~ethane’. In fact, it can also be divided after the prefix ‘di’ according to: ‘1,2=di~chloroethane’, because the number of symbols in front of the optional hyphen equals six, but a division according to ‘1,2=dichloro~ethane’ is far easier to read.

In this respect, authors should use their discretion in applying dividing rules. A systematic name starting with ‘tetrakis(....’ can be divided before the parenthesis according to ‘tetrakis~(....’ since ‘tetrakis-’ has eight consecutive characters before the hyphen, and a chemist knows that the substituted component will be described after the parenthesis. Consequently, a name starting with ‘bis(....’ or ‘tris(....’ is also allowed to be divided in a similar way, despite the fact that ‘bis-’ and ‘tris-’ have fewer than six consecutive characters. Accordingly, ‘tris~(ethane~1,2=di~amine)~cobalt(III)’, ‘tetra~ammine~aqua~cop~per(II)’ and ‘cis=di~ammine~di~chlorido~pla~ti~num(II)’ are perfectly acceptable. The same rationale holds for polymers in which a division is generally allowed after the prefix ‘poly’, e.g., ‘polystyrene’ can be divided as ‘poly~styrene’. Readability and chemical sense should prevail.

Hyp-5.6 Division before suffixes

When suffixes are preceded by a locant, the name can be divided at the hyphen before this locant (or locant set), e.g., ‘pentan~3=one’. If there is no locant, division before the suffix is similarly allowed, but a hyphen has to be inserted, e.g., ‘cyclo~hexan~one’, benzene~thiol, or ‘heptan~oic acid’.

However, in contracted forms such as ‘phenol’, ‘ethyl’, or ‘4=pyridyl’ the suffixes are regarded as part of the name component, and a division before the suffix is unacceptable (just as the locant for the suffix is not inserted in such contracted forms). Analogously, a division of ‘tetrachloridoplatinate(II)’ is possible at the following positions ‘tetra~chlor~ido~pla~tin~ate(II)’, and ‘acetone’ may be divided according to ‘acet~one’ if it is substituted as in ‘1,3=di~chloro~acet~one’, but a division according to ‘1,3=di~chloro~acetone’ is far easier to read.

Hyp-5.7 Division before endings

Endings are regarded as part of the name component and should therefore not be divided from the parent structure name. Contrary to the ACS Style Guide [17], the present Recommendations keep name components with an ending like ‘ethyne’ together by shifting them to the next line. However, division before the locant or locants is permitted when the ending is preceded by a locant, e.g., ‘hepta-1,3=diene’. The systematic name for ‘oleic acid’ can therefore be divided according to ‘(9Z)=octa~dec~9=en~oic acid’ and the systematic name for another fatty acid, ‘agonandoic acid’, can be divided according to ‘(11E)=octa~dec~11=en~9=yn~oic acid’, where the ‘-ynoic’ and the ‘-enoic’, respectively, are regarded as endings with part of a suffix (see Hyp-5.6).

According to this exception ‘(2Z)=pent~2=ene’ can be divided as shown. Consequently, it is also acceptable to divide ‘pentane’ in the same way (‘pent~ane’) provided there are enough characters in front of ‘pent~’ to justify this division, as in ‘2=methyl~pent~ane’. This preferred way of division may therefore not be in accordance with the way some people pronounce it (‘pen~tane’) and should therefore preferably be avoided (*cf.* Hyp-5.9). Another example with similar reasoning is ‘catena~tribor~ane’.

However, dividing ‘but-2-ene’ at the first hyphen is not recommended since the number of letters in front of this hyphen is too small, which leads to ‘but=2=ene’, *i.e.*, no division at all. When the number of characters is increased as in ‘1=chloro~but-2=ene’, division at the indicated hyphen would be possible, but a division according to ‘1=chloro~but=2=ene’ is far easier to read.

Hyp-5.8 Generally accepted divisions

The last two examples in Hyp-5.7 also illustrate that the readability of words does not suffer when they are divided in a way that is generally accepted and readers will recognize. So ‘octadec-’ can be divided after ‘octa~’. Similarly, ‘phosphorus’ can be divided according to: ‘phos~phorus’. The prefix ‘hydro’ can be divided according to ‘hy~dro’ provided there are at least four characters in front as in ‘1,2=di~dronaphthalene’.

Hyp-5.9 Division in accordance with pronunciation and chemical meaning

Division should not hamper pronunciation. Take the word ‘anthracene’. This consists of two chemical terms: the root ‘anthrac’ and the ending ‘ene’, but nobody says ‘anthrac-ene’, people say ‘anthra-cene’. If the line had ended on ‘anthrac-’, readers might have started to pronounce this as ‘anthrak’. So ‘anthra~cene’ is the generally accepted way of dividing this word. This way of dividing is preferred to the division according to ‘anthr~acene’. This latter division is in accordance with the way ‘poly~acenes’ can be divided, but then endings like ‘acenes’ are part of a name component or name. So ‘anthra~cene’ is preferred because that is how the name is pronounced. It is noted, however, that this means that the division of names may be both language and dialect dependent. Similarly, ‘naphthalene’ should not be divided according to ‘naphthal~ene’, but in accordance with its pronunciation as ‘naphtha~lene’, which is again preferred to ‘naphth~alene’.

On the other hand, there are instances where a division may be acceptable according to the above described Recommendations but would mislead the reader by implying a different chemical meaning. For example ‘acetaldehyde’ may be divided as ‘acet~aldehyde’, but not as ‘acetal-dehyde’ which is allowed by some dictionaries [3, 4] and also the ACS Style Guide [17], ‘formaldehyde’ may be divided as ‘form~aldehyde’, but not as ‘formal-dehyde’.

Hyp-5.10 Division within a set of stereodescriptors or locants

Special attention should be given to names that include stereochemical information in the form of a string of descriptors. The systematic name of ‘DHA’ (docosahexaenoic acid) is: ‘(4Z,7Z,10Z,13Z,16Z,19Z)-docosa-4,7,10,13,16,19-hexaenoic acid’. If this name were to be divided after the ‘do-’, the preceding string (including the hyphen) would be 27 characters long. That can lead to unsightly print. Accordingly, the stereodescriptors should be divisible, according to:

‘(4Z,7Z,~10Z,~13Z,~16Z,~19Z)=’

In this example, no optional hyphen has been inserted between the first two stereodescriptors to ensure there are at least six characters at the end of a line. Accordingly, a name starting with two stereodescriptors such as the systematic name for ‘linoleic acid’, i.e., ‘(9Z,12Z)=octa deca~9,12=dienoic acid’ should keep these stereodescriptors together and keep them attached to the prefix (‘octa’), despite the fact that this leads to a string of fourteen (including the hyphen) consecutive characters. The locant set for the double bonds in ‘DHA’ can also be divided according to:

‘do~cosa-4,7,~10,~13,~16,~19=hexa~’

The same rules apply to the stereodescriptors ‘R’ and ‘S’. Take, for instance, ‘eudesmane’. The systematic name of this compound is: ‘(1R,4aR,7R,8aS)-1,4a-dimethyl-7-(propan-2-yl)decahydronaphthalene’. This name could be presented to the typesetter in a way that illustrates the rules outlined above: ‘(1R,4aR,~7R,~8aS)~1,4a=di~methyl~7=(propan~2=yl)~deca~hy~dro~naphtha~lene’.

Hyp-5.11 Unacceptable division at spaces

As stated in Hyp-3.2, a chemical name can, in general, be divided at an existing space within the name. Exceptions to this rule are the following: A division at the space before a descriptor following a name, such as the numerical descriptor for the composition of an addition compound, or of a polymer should be avoided, in order not to split such a descriptor from the name at a line break. Similarly, spaces within such descriptors should be typeset as non-breaking spaces. For example, ‘boron trifluoride—water°(1/2)’ should not be divided before ‘(1/2)’, as indicated by the non-breaking space, and ‘(polyethene) mod~chloro°(w_{Cl}°=°0.32)’ should not be divided after ‘chloro’, and a line break within the parenthetical

descriptor '(w_{Cl} = 0.32)' is not acceptable either. Spaces before and after an equal sign, =, are also considered as non-breaking spaces.

Hyp-6 InChI and InChIKey

Closely related to chemical names are the InChI (IUPAC International Chemical Identifier) and the related InChIKey for a chemical structure. An InChI is produced by computer (see <https://www.inchi-trust.org/downloads/>) from structures drawn on-screen (or from derived 'mol' files) to represent a compound in a completely unequivocal manner. An InChIKey is a contracted form of defined length and format that is produced from the InChI by another algorithm. Whereas the length of an InChI depends on the compound, all InChIKeys have the same 27 character length, including two hyphens in fixed positions (positions 15 and 26).

Examples (InChI and InChIKey for chalcone):

- I. : 1S/C15H12O/c16-15(14-9-5-2-6-10-14)12-11-13-7-3-1-4-8-13/h1-12H/b12-11+
II. : DQFBYFPFKXHELB-VAWYXSNFSA-N

Both InChIs and InChIKeys are electronic codes, *i.e.*, machine-readable strings of symbols primarily intended to be interpreted by computers. Therefore, any addition of symbols, including spaces, is unacceptable for InChIs and InChIKeys. When they appear in print and do not fit into a line, they may only be divided at a suitable existing symbol, *i.e.*, after an hyphen that is already present, or a solidus. The first symbol on the next line should not be a comma. The InChIKey can only be divided at the first hyphen.

This method of dividing therefore does not involve the use of the bis-hyphen sequence. Accordingly, the hyphen at the end of part of the InChI or InChIKey must not be deleted on undividing, as would be the case for a systematic chemical name.

Hyp-7 Linear chemical formulae

The treatment of end-of line hyphenation would be incomplete without also considering formulae, even though formulae are not chemical names. Whenever possible, a line break within a line formula should be avoided. However, linear formulae can become quite long. In order to avoid unsightly appearance, formulae may also be divided at the end of a line. In that case, the rules described in Hyp-5.1 apply.

Hyp-7.1 Molecular formulae

The first type of formula listed in the introduction is the molecular formula that just lists how many carbon, hydrogen, oxygen, and other atoms a molecule contains. Given the limited number of different atoms present in most molecules for which the molecular formulae are used, it is unlikely that molecular formulae will require end-of-line division. If it turns out to be desirable, or inevitable, molecular formulae can be divided by inserting an optional hyphen in front of an atom symbol and the presence of this hyphen will indicate that the molecular formula has been divided and that the hyphen must be removed on undividing.

Hyp-7.2 Shorthand notation

Another type of linear formula is the shorthand notation, such as NaOMe for 'sodium methoxide' or EtOAc for 'ethyl acetate'. Such shorthand formulae will usually be quite short so that no end-of-line division will be necessary. In the unlikely event that this division is deemed to be desirable, the guidelines of Hyp-5.1 provide an answer where not to insert the optional hyphen, which has to be removed on undividing.

Hyp-7.3 Structural formulae

Another slightly different type of formula is the structural linear formula of which NaOCH_3 is an example. Again, it is unlikely that such formulae will have to be divided at the end of a line, but when this happens to be the case the same principle as above should be used.

Hyp-7.4 Linear formulae

Linear formulae can also show single, double and triple bonds like for instance in $\text{CH}_2=\text{CH}_2-\text{CONH}_2$ ('prop-2-enamide') or $\text{CH}_3-[\text{CH}_2]_8=[\text{CH}_2]_8-\text{COOH}$ ('oleic acid'). When such formulae have to be divided at the end of a line, they can be divided at any indicated single bond. To this end the original bond is shifted to the next line by replacing it by a non-breaking hyphen, inserting an optional hyphen in front of this non-breaking hyphen, thereby forming the bis-hyphen sequence (see Hyp-4.4.4). On undividing, the inserted hyphen is deleted.

Hyp-7.5 Proteins (three letter abbreviation)

A common type of linear formula deals with biomacromolecules such as proteins. Although two linear proteins can be linked by cystine bridges as in insulin, the linear chains themselves can be represented in print as a series of amino acid residues, like the 'Gly-Pro-Pro' repeat unit in collagen. This example shows that hyphens are dispersed at regular intervals in these amino-acid sequences and so are well suited for an end-of-line division. The IUPAC-IUB document on amino acids and peptides [2] recommends that in such cases, the hyphen is repeated at the beginning of the next line, which recommendation has also been adopted in the present Recommendations as the basis of the hyphenation rules. Example:

(text text text text text text text)... Ala-Ser-Tyr-Phe-Ser-
-Gly-Pro-Gly-Tyr-Arg ... (text text text text text text text)

Just as in systematic chemical names, the end-of-line hyphen has to be removed on undividing.

Hyp-7.6 Proteins (one letter abbreviation)

If the one-letter system for describing proteins or peptides is used, the letters representing the amino acids are presented in groups of 10 that are separated by spaces (*e.g.*, ...GTPQDRRLRL ECHETRPLRG RCGCGERRVP ...). There is no need to divide these groups at the end of a line. This has the additional advantage that in a text column, the groups are vertically aligned if the strings are very long.

Hyp-7.7 Nucleotides

The last linear formula to be discussed is used for sequences of nucleotides, denoted by single capitals notations, *e.g.*, by using a string of one-letter symbols, like d(AGAGCTAGCTCT). Dividing strings of this kind should be avoided wherever possible (*i.e.*, they should appear on one line), but if absolutely necessary, the rules on numbers of characters on either line should be obeyed. An optional hyphen should be inserted and since it is a hyphen at the end of a line, it should be deleted on undividing.

Hyp-8 Annex

The typesetting software used by the various publishing houses varies. Some software packages may differentiate between differently encoded hyphens, whereas other software may automatically remove such encodings during

import of a submitted manuscript file into the typesetting software. At worst, optional hyphens may be transformed into obligatory (regular) hyphens; otherwise, information provided by the author may simply be lost.

Both authors and publishers can contribute to an improvement of this situation. The following sections will describe the roles of both parties.

Hyp-8.1 Role of the author

Any author should realize that the end-of-line hyphenation in the manuscript will lead to problems in typesetting because in a typeset manuscript the line breaks are almost sure to be at different positions; this then may result in excess hyphens in the middle of a line. Since manually inserted end-of-line hyphens cannot be easily recognized as such and can therefore not be removed automatically, authors are advised to refrain from inserting a normal hyphen at the end of a line, but rather to insert optional hyphens if a division is felt to be desirable.

For ease of reference, below two tables are given, listing the various punctuation symbols used in this document and summarizing their names, character codes, and shortcut keys.

If an author wishes to retain some control of name and formulae division during the typesetting of a manuscript, then the following steps are advised. These steps will be illustrated by the compound name ‘tri-μ-carbonyl-bis(tricarbonyliron)(Fe–Fe)’.

- (1) During writing, the author is advised to use non-breaking hyphens for any hyphen that may occur in the chemical name: ‘tri=μ=carbonyl=bis(tricarbonyliron)(Fe–Fe)’;
- (2) When preparing the manuscript for typesetting, the author could/should insert optional hyphens in front of non-breaking hyphens if a division is allowed at that point: ‘tri=μ=carbonyl~bis(tricarbonyliron)(Fe–Fe)’;
- (3) The author could/should surround any en dash or em dash that should not act as end-of-line divisions by no-width non breaks, one before and one after the dash: ‘tri=μ=carbonyl=bis(tricarbonyliron)(Fe*–*Fe)’;
- (4) In a later stage of manuscript preparation, the author may decide to look for systematic names with long stretches of non-interrupted characters, whereby the non-breaking hyphens that have been introduced before are counted as characters. The author may then introduce optional hyphens by using the division rules listed above, but **when in doubt should refrain from inserting an optional hyphen**. ‘tri=μ=carbonyl=bis~(tri~carbonyl~iron)~(Fe*–*Fe)’.

When inserting optional hyphens, the author should use his/her discretion in applying the recommended division rules. **Readability and chemical sense should prevail.** Therefore, when consulting a desk dictionary, *e.g.*, [3–5], it should be kept in mind that in these dictionaries, chemical sense is often disregarded.

Because both typesetting and word processing software may store words in their databases and may add new words when they arise, the need to specify what type of hyphen should be used will eventually decrease. Until that time, applying the above rules will expedite proofreading and reduce the need of proofreading for a second time.

Hyp-8.2 Role of the publisher/typesetter

The publisher should, first of all, incorporate the Recommendations in the Author Guidelines or Instructions for Authors and develop appropriate dictionaries of relevant examples. Also embedding the rules in the journal templates would be very helpful, although in some cases this may be challenging to implement.

Secondly, publishers should ensure that their typesetting software and/or the software used by their sub-contractors incorporates these Recommendations and recognises the various types of hyphens used by the authors. Authors could then, in an ideal case, perhaps be advised to neither insert any manual line breaks, nor any encodings for specific hyphens. An even better solution would result if the software could also recognise and act on author intentions, as revealed by their deliberate insertion of non-breaking hyphens and any other specifications regarding hyphenation. Ideally the software should also flag situations where the author may have specified an incorrect or inappropriate hyphenation.

Tables with special characters:

Table 1: Regular characters.

Character	Glyph	Regular connecting characters			
		Code point		Shortcuts for inserting	
		Hex	Dec	Alt Code(s) in Windows 10, while holding Alt , type the number, then release Alt ^a	MS Word 2019 Shortcut, press the following keys together ^b
Minus sign	–	2212	8722	Alt + 8722	
En-dash	–	2013	8211	Alt + 8211; Alt + 0150	Ctrl + Num– ^c
Em-dash	—	2014	8212	Alt + 8212; Alt + 0151	Ctrl + Alt + Num– ^c
Solidus	/	002F	47	Alt + 47; Alt + 047	Or use keyboard /
Center dot	·	00B7	183	Alt + 0183	
Colon	:	003A	58	Alt + 58; Alt + 058	Or use keyboard :

^aNote that in Windows 10, all printable Unicode characters can be inserted by holding the **Alt** key while pressing the numerical keys of the character's code in its decimal representation (as shown in the table). In some cases, a leading zero must be added to yield the correct character. Additional alternative Alt codes may also be available (as shown in the table). However, results of this general Windows procedure may vary depending on the currently active application.

^bNote that in addition to the shortcuts listed in the table, in MS Word 2019, all printable Unicode characters can also be inserted by typing out the four-digit hexadecimal representation of the character's code point followed by pressing **Alt** + **X**.

^cNote that in MS Word 2019, using the **–** keyboard key and the **Num–** keyboard key (numeric minus key), especially in keyboard shortcuts, may yield different results.

Table 2: Special characters.

Special char. name in MS Word and this work	In this work shown as	Unicode character representing the special character in MS Word 2019			
		Code point		Shortcuts for inserting	
		Hex	Dec	Alt code in Windows 10, while holding Alt , type the number, then release Alt ^a	MS Word 2019 Shortcut, press the following keys together ^b
Non-breaking space	°	00A0	160	Alt + 0160	Ctrl + Shift + Space
Non-breaking hyphen	=	001E	30	Alt + 030	Ctrl + Shift + – ^c
Optional hyphen	~	001F	31	Alt + 031	Ctrl + – ^c
No-width Non break	*	200D	8205	Alt + 8205	

^a Note that in all four selected special characters of MS Word can be inserted by the general Windows procedure for inserting Unicode characters, by holding the **Alt** key while pressing the numerical keys of the character's code point in its decimal representation (as shown in the table).

^b Note that in addition to the shortcuts listed in the table, in MS Word 2019, the non-breaking space and the no-width non break can also be inserted by typing out the four-digit hexadecimal representation of the character's code point followed by pressing **Alt** + **X**. This method, however, does not work for the non-breaking hyphen and the optional hyphen characters.

^cPlease note that in MS Word 2019, using the **–** keyboard key and the **Num–** keyboard key (numeric minus key), especially in keyboard shortcuts, may yield different results.

Membership of sponsoring bodies during preparation of the Provisional Recommendations

Membership of the Division Committee of the IUPAC Division of Chemical Nomenclature and Structure Representation for the period 2018–2019 is as follows:

President: A. T. Hutton (South Africa); **Past President:** K.-H. Hellwich (Germany); **Secretary:** R. S. Laitinen (Finland); **Titular Members:** M. A. Beckett (UK); E. C. Constable (Switzerland); T. Damhus (Denmark); R. T. Macaluso (USA); E. Nordlander (Sweden); A. P. Rauter (Portugal); M. M. Rogers (USA); **Associate Members:** E. Mansfield (USA); J. Nagy (Hungary); M. A. Strausbaugh (USA); K. T. Taylor (USA); C. A. Tovee (UK); J. Vohlídal (Czech Republic); **National Representatives:** F. Aricò (Italy); N. Burford (Canada); A. M. da Costa Ferreira (Brazil); S. Erdem (Turkey); S. Koo (Korea); R. Kruszyński (Poland); L. Meesuk (Thailand); M. A. Petrova (Bulgaria); E. Szabó (Slovakia); A. Yerin (Russia); **Ex Officio:** R. M. Hartshorn (New Zealand); G. P. Moss (UK).

Membership of the Division Committee of the IUPAC Division of Inorganic Chemistry for the period 2018–2019 is as follows:

President: L. Öhrström (Sweden); **Vice-President:** J. García-Martínez; **Past President:** J. Reedijk (the Netherlands); **Secretary:** M. Leskelä (Finland); **Titular Members:** L. Armelao (Italy), X. Zhu (China); P. Karen (Norway), R. Macaluso (USA); M. Hasegawa (Japan), M. Drabik (Slovakia); **Associate Members:** A. Powell (Germany), L. Meesuk, (Thailand), F. Abdul Aziz, (Malaysia) L. Colón, (Puerto Rico); T. Walczyk (Switzerland), D. Rabinovich, (USA) book-chapter; **National Representatives:** J. Galamba Correia (Portugal), S. Kalmykov (Russian Federation), P. Knaut, (France), K. Yoon (South Korea), M. Diop (Senegal), G.J. Leigh (UK), K. Sakai (Japan), N. Trendafilova (Bulgaria), V. Stilinović (Croatia), O. Metin (Turkey), M.H. Lim (South Korea), P. Kulesza (Poland).

Research funding: This work was prepared under project 2014-003-2-800 of IUPAC (Funder ID: 10.13039/100006987).

References

- [1] J. He, J. Chen, K.-H. Hellwich, M. Hess, K. Horie, R. G. Jones, J. Kahovec, T. Kitayama, P. Kratochvíl, S. V. Meille, I. Mita, C. dos Santos, M. Vert, J. Vohlídal. *Pure Appl. Chem.* **86**, 1003–1015 (2014).
- [2] IUPAC, IUB. *Pure Appl. Chem.* **56**, 595–624 (1984).
- [3] Merriam-Webster's New Collegiate® Dictionary. Merriam-Webster Inc., Springfield MA, US (2014).
- [4] New Oxford Spelling Dictionary. M. Waite (Ed.), pp. xii + 596, Oxford University Press, Oxford, UK (2014).
- [5] Dictionary of Contemporary English. P. Procter (Ed.), Longman Group Ltd., UK (1978).
- [6] R. G. Jones, T. Kitayama, K.-H. Hellwich, M. Hess, A. D. Jenkins, J. Kahovec, P. Kratochvíl, I. Mita, W. Mormann, C. K. Ober, S. Penczek, R. F. T. Stepto, K. Thurlow, J. Vohlídal, E. S. Wilks. *Pure Appl. Chem.* **88**, 1073–1100 (2016).
- [7] IUPAC. *Compendium of Polymer Terminology and Nomenclature (IUPAC Recommendations 2008) (the 'Purple Book')*, R. G. Jones, J. Kahovec, R. Stepto, E. S. Wilks, M. Hess, T. Kitayama, W. V. Metanovski (Eds.), pp. xix + 443, RSC Publishing, Cambridge, UK (2008); PDF on-line version: <https://iupac.org/what-we-do/books/purplebook/>.
- [8] IUPAC. *Nomenclature of Organic Chemistry, IUPAC Recommendations and Preferred Names 2013 (the 'Blue Book')*, H. A. Favre, W. H. Powell (Eds.), pp. xliii + 1568, RSC Publishing, Cambridge, UK (2014); A list of errata can be found on the web at the address: <https://iupac.org/what-we-do/books/bluebook/>.
- [9] IUPAC. *Nomenclature of Inorganic Chemistry (IUPAC Recommendations 2005) (the 'Red Book')*, N. G. Connelly, T. Damhus, R. M. Hartshorn, A. T. Hutton (Eds.), pp. xi + 366, RSC Publishing, Cambridge, UK (2005). <https://iupac.org/what-we-do/books/redbook/>.
- [10] M. Barón, R. B. Fox, A. Fradet, K.-H. Hellwich, P. Hodge, K. Horie, J. Kahovec, P. Kubisa, E. Maréchal, W. Mormann, R. F. T. Stepto, J. L. Schultz, E. S. Wilks. *Pure Appl. Chem.* **80**, 201–232 (2008).
- [11] G. P. Moss. *Pure Appl. Chem.* **68**, 2193–2222 (1996).
- [12] W. Mormann, K.-H. Hellwich, J. Chen, E. S. Wilks. *Pure Appl. Chem.* **89**, 1695–1736 (2017).

- [13] *Nomenklatur der Anorganischen Chemie, Deutsche Fassung der Empfehlungen 1990*, W. Liebscher (Ed.), VCH, Weinheim (1994).
- [14] *Naming and Indexing of Chemical Substances for Chemical Abstracts™*, Chemical Abstracts Services, ACS (2007).
- [15] IUPAC. *Principles of Chemical Nomenclature, A guide to IUPAC Recommendations*, 2011 Edition, G. J. Leigh (Ed.), pp. xii + 257, RSC Publishing, Cambridge, UK (2011).
- [16] *British Approved Names; BAN Book*, ISBN: 9780113230334, British Pharmacopoeia (2017).
- [17] *The ACS Style Guide*, A. M. Coghill, L. R. Garson (Eds.), American Chemical Society (2006), ISBN13: 9780841239999.