



## Chapter 5. Style Points and Conventions

Please let me know if any of the style points seem awkward to you. What works in one field may not be as useful in another. A style point is something that is decided upon by the person(s) in charge or by a committee (that decides to do something because of the historical use of doing something more one way than another or after analysis of the needs of all members of the group), but by whatever means agreement is reached for style points, the main use of style points is to improve consistency.

Please note: Styling of [Appendices](#), [Boxes](#), [Figures](#), [References](#), and [Tables](#) is in [Special Content](#).

### Amino Acids

There are only four nucleotides (A, C, G, and T), but a sequence of three of the four nucleotides specifies an amino acid. The three nucleotides, sometimes called triplets, make up the codon.

Use the three-letter abbreviations for amino acids in chains, as well as with codon numbers and mutations.

#### Examples:

Arg-506

Arg506Gln

Met-Phe-Val-Asn-Gln-His

**Note:** Single-letter abbreviations should be used sparingly.

#### Example:

MFVNQH (instead of Met-Phe-Val-Asn-Gln-His)

[Amino acid abbreviations](#) are determined by [IUPAC](#).

### Capitalization

See CBE, pp. 149-165.

Note: When a sentence begins with an abbreviation or designation that has an initial lowercase letter, the abbreviation or designation retains its lowercase.

#### Example:

Our study determined the protein to be p53. p53 has been studied extensively.

## Centrifugation

Use *g* force (3000*g*), not rpm. Note that there is no “times” sign.

## Colons

According to CBE, there are four general uses of colons and nine specialized uses (pp. 44-46). Here, I will provide you with a few examples of the usual things.

### Example:

Note: The images were scanned from photographs.

In the example above, note the capital “T” after the colon.

### Example (from a Table title):

Table 2. LocusLink query terms: controlled terms.

In the example above, note the lowercase “c” after the colon.

### Example (from a Reference):

Al-Shahrour F, Diaz-Uriarte R, Dopazo J. FatiGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics*. 2004 Mar 1;20(4):578-80. Epub 2004 Jan 22.

In the example above, note the lowercase “a” after the colon.

### Example (from a Title):

GenBank: The Nucleotide Sequence Database

In the example above, note the capital “T” after the colon.

## Commas

Refer to CBE, pp. 48-52, for a complete discussion. The purpose of the comma is to make the sentence clear. It is not a symbol of conversational pause.

In general, use a “serial comma” to avoid ambiguity. A serial comma sets off a “simple series of more than two elements” (see the commas set in bold in the next example).

### Example:

MIM is organized into autosomal, X-linked, Y-linked, and mitochondrial catalogs, and MIM numbers are assigned sequentially within each catalog.

The comma not set in bold in the example above is a comma used to separate independent clauses (complete sentences). This comma makes the sentence easier to read and understand.

Another comma error is the use of a comma between the subject and the verb. The separation of the subject from the verb by an intervening comma is called “comma separation”.

### Example:

WRONG: Suggested corrections and additions are welcome for future updates of these pages, and should be sent to the author.

RIGHT: Suggested corrections and additions are welcome for future updates of these pages and should be sent to the author.

## Concentrations

Show liter as “l” with combinations of units but written out as “liter” otherwise. Show gram as “g” (and kilogram as “kg”). Abbreviate mol/l as “M”.

## Dashes

There are two types of dashes, the en dash (–) and the em dash (—). (A hyphen is shorter than an en dash.)

### en dash

used to separate ranges and words of equal weight

#### Examples:

63–65, Mann–Whitney, dose–response curve

### em dash

used to break a thought in a sentence (which is often a long pause, an aside, or a title fragment)

#### Example:

Once he made the decision—did I say that already?—nothing stopped him from reaching his goal.

Contrast these uses with hyphens (see Hyphens).

## Decimal Points

Insert 0 in numbers less than 1 (0.05).

## Degrees

Close up symbol to number (55°C).

For educational degrees, such as MD and PhD, do not use periods.

## e.g.

Use to mean “for example”. Do not italicize; use periods and follow with a comma.

See also Italics.

## et al

Do not italicize. Provide all names in the “References” section. In other formats (where the display of information is an issue because of space), reference citations may use et al. with only one author.

See also Italics.

## etc

Substitution of “and others” or removal of the term is preferred, if possible.

## Emphasis

Underlining is not preferred for emphasis within the text to avoid confusion with underlining that is used to show Web links. (If hyperlinks are shown in a different color of underlining or in a different manner so that no confusion would exist, underlining for emphasis would be OK.)

Use quote marks to show that a word is used in an unusual way. If emphasis is needed to show how an acronym was derived, use boldface letters for the beginning of each word, e.g., **D**eleted in **C**olorectal **C**arcinoma (DCC). When the derivation of the acronym is obvious, do not use the boldface letters. See also [Voice: Active versus Passive](#).

## Equations

If equations are numbered, cite them as Equation 1, etc. Capitalize the word Equation and write it out in the text, figures, tables, and boxes. A sentence (or sentence fragment) introducing an equation ends with a colon. The equation itself does not receive any type of sentence punctuation, such as a period, semicolon, or comma, at its end.

Italicize variables, but use roman type for SD, SE, CV, etc.

See also [Math and Statistics](#).

## Foreign Words

See [Italics](#).

## Genes and Protein Designations

Genes are italicized, and proteins are set in roman type (not italic). Use where reasonable. Web pages with long lists of genes in italics would be distracting. In this case, use roman type.

## Hyphens

Hyphens are used for many purposes, such as with:

- prefixes and suffixes
- compound terms
- compound modifiers of several types
- spelled-out fractions
- written-out forms of numbers from 21 through 99
- representation of single bonds in chemical or molecular formulas, peptide bonds in residues, and linkage of nucleotides

See CBE pp. 61-64 for many examples of the specific uses of hyphens. Below are a few common examples to make your search quicker.

## Unit Modifiers Are Hyphenated

e.g., best-case scenario, false-positive results, high-risk behavior, model-fitted values

## Note

Two-word phrases that are easily understood as a unit are not hyphenated.

No Hyphen: amino acid residues

## Use a Hyphen with Some Prefixes

e.g., anti-oncogenes, mis-segregation, non-redundant, pre-existing, BUT nonlinear, pretreatment

Most “co” prefixes are closed up: coeluted, comigrate, copolymerize

## Exceptions

co-injected, co-worker

## Compounds with Phosphorus

No hyphen if the first word in the name is a noun. A group or adjectival form takes a hyphen, e.g., glucose 1-phosphate, cytosine 3[prime]-phosphate, glycerol-1-phosphate (see [IUPAC-IUBMB](#)).

Contrast hyphens with en dash and em dash (see Dashes). See also Amino Acids.

## -ical

English uses a couple of adjective endings that are sometimes confused: **-ic** and **-ical**. There are differences in meaning, and examples will best illustrate these differences.

Take the pair of words **classic/classical**. You may own a classic car, but you listen to classical music. “Classic” refers to something being traditional and enduring, or serving as a standard of excellence—perhaps an example of an outstanding model. “Classical”, on the other hand, refers to something that is a historically important form or genre of architecture, music, art, etc. Of the two forms, “classic” seems to apply more to something specific, and “classical” seems to apply to something broader, more general but distinguishable.

Consider **graphic/graphical**. If you watch TV, you may see a lot of graphic violence, where “graphic” more refers to the effects on the emotions, but if you choose, instead, to read [The NCBI Handbook](#), you will be learning from its many graphical images, where “graphical” more refers to the format of the material. Merriam-Webster's Dictionary does not make this particular distinction, so this is a style point. Here we are using “graphic” to mean providing details of perhaps a shocking nature, whereas “graphical” is used to mean a drawing, illustration, or art product.

## i.e.

Note punctuation. Do not italicize.

See also Italics.

## in vitro, in vivo

Use italics.

See also Italics.

## Italics

There are general and scientific uses of italics. [On the use of italic and roman fonts for symbols in scientific text](#) is available, as well CBE coverage on pages 169-171.

For our most common purposes, use italics for genus/species names, genes, loci, and alleles; parts of chemical names as appropriate (including *cis*, *trans*, *ortho*, *meta*, and *para*); all variables (e.g., probability (*P* or *p*)); and written-out Latin forms (such as *a priori*, *ad libitum*, *de novo*, *in situ*, *in utero*, *in vitro*, and *in vivo*).

CBE considers whether a Latin word is common and suggests that "in vitro" and "in vivo" not be italicized. This, then, becomes a style point. I think it would be more confusing to selectively choose some written-out Latin words to be italic than to go along with the old standard of italicizing all of the Latin words.

What of "e.g.," "et al.," "i.e.," and "etc."? Well, these are not written-out forms, AND they are very common; therefore, there is no need to italicize "e.g.," "et al.," "i.e.," and "etc.,". More basically, the handling of Latin and other foreign words is influenced by the "acceptance" of a term as "common" as to whether an italic form is used. As specific foreign words become more common, the trend is for the italic style to be dropped, as a simplification. For example, uncommon French words used in English would still be shown in italics, but French words that have been used in English for a long time, such as "à la carte", are readily understood by most people and would not be set in italics (and some go so far as to drop the accent mark). Because differences of opinion can arise as to whether a foreign word has become common enough in English to go without italics, style points are then described in style guides, for the sake of consistency.

When working with long lists of genus/species names, genes, loci, and alleles, particularly in tables, the italics become more of a distraction than a help for readers. In these cases, use your best judgment about the use of italics.

Per changes by IUPAC in 2003, do not italicize the first three letters of restriction enzymes (EcoRI, BamHI).

## Latin

See Italics.

## Lists

**Paragraph format:** Within paragraphs, use (a), (b), (c), etc. and then (i), (ii), (iii), etc.

**Outside of the paragraph format:**

- "Bulleted" lists are fine. Be sure that all items with a similar weight use the same symbol.

In bulleted lists, entries of phrases, words, or sentence fragments do not use a period at the end of the entry. Do not connect list items with conjunctions such as "and" or "or".

If the bulleted entries are complete sentences, use sentence capitalization and a period at the end of each entry.

- Use a "numbered" list only when the items must follow a sequence. When steps of a procedure are being referred to, use Arabic numbers (OK as a subsection within the bulleted list).

For more information, see [Itemized Lists](#) in [Writing for the Web Environment](#).

## Math

Follow CBE.

## Numbers/Numerals

Write out at the beginning of a sentence. Write out one to nine except with units of measure. Use numerals for 10 and above.

Note: If a study is of 10 or more patients, items, things, etc., use numerals in sentences referring to parts of the total group, even when less than 10.

### Example:

We studied 36 patients. In the first group of 6 patients, only 2 were anemic.

## Quick List

1. Written-out numbers *versus* as numerals:
  - a. For numbers of 10 or more that do not refer to units of measure, use numerals.
  - b. Units of measure require numerals (3 mm, 4 months).
  - c. Avoid beginning a sentence with a numeral.
2. Fractions: one-half, two-thirds, one-tenth, 1/32
3. Ordinal numbers: 1st, 2nd, 3rd
4. Series: if the series contains numbers 10 or higher, use all numerals (3 mice, 4 rats, and 11 hamsters)
5. Either side of the decimal point: use 0.1 not .1 (all cases); use 6.0 not 6 (only if significant)
6. Commas: 3,000; 14,000; 3,333,331
7. Time 0 (contrast zero time)
8. Decimals instead of fractions with measurements: 0.5 volume (not one-half volume)
9. Bases use numerals (even those fewer than 10): 5 bases

## Probability

Use an italic *P* or *p*. Use spaces around math symbols.

See also Statistics.

## Proofreaders' Marks

Some of you wonder about the shorthand-looking marks made on hard copies of your Web pages. This is a way of showing corrections and changes. There are just a few marks to learn, and they come in handy!

Here is a nice link to view the marks: [Proofreaders' Marks](#).

## Punctuation

See CBE, pp. 36-71.

See also Colons, Commas, Dashes, Hyphens, Quotation Marks, and Semicolons.

## Quotation Marks

Place quotation marks inside of punctuation.

Use double quote marks. For quote marks inside of quote marks, use single quotes on the inside of the phrase and double quote marks on the outside of the phrase.

### Example:

The NCBI Handbook addresses PubMed's handling of Stopwords, specifically that "PubMed ignores Stopwords, such as 'about', 'of', or 'what?..'", when processing simple searches.

Use double quotation marks around phrases that some of the words begin with lowercase.

**Example:**

You may also want to display your citations using "Send to Text" to eliminate the sidebar menu and toolbars before printing your results.

In the example above, the "Send to Text" phrase would not need quote marks if it were a hotlink because of the way links show up on the Web pages. The blue font would be enough to make it distinctive.

Also use double quotation marks to avoid confusion, as well as to show that you are implying a meaning somewhat different than that usually assumed (or that you doubt the validity of the usual meaning).

## Restriction Enzymes

Do not use italics for the first three letters and close up the entire name, e.g., AccI, HaeII. Removal of italics is a change made by IUPAC in 2003.

### Helpful Tip

Visit [The Restriction Enzyme Database \(REBASE\)](#). Select "Enzyme Navigation Tables" and "All enzyme names..." for a complete list of restriction enzymes.

## Semicolons

See CBE, pp. 46-48, for the four general uses and five specialized uses of semicolons. Here are examples of common uses:

**Example:**

Certain treatments were more effective than others in eliminating the contaminating DNA; however, to achieve this there was a decrease in sensitivity.

**Example:**

Links are provided to related Web sites including: chromosome databases (e.g., the Mitelman database); other NCI (e.g., CGAP and CCAP) and NCBI (e.g., the Map Viewer (Chapter 20) and LocusLink (Chapter 19) resources and PubMed (Chapter 2)) sites; The Jackson Laboratory; and several other CGH sites.

## Spelling

American spellings are preferred over British spellings.

Differences in spellings, from British to American, usually involve:

- swap of the letters "r" and "e", such as "centre" (British) and "center" (American)
- addition of the letter "l", such as "cancelled" (British) and "canceled" (American)
- use of letter "s" instead of letter "z", such as "characterise"/"analyse" (British) and "characterize"/"analyze" (American)
- addition of letter "a" or "o" with letter "e", such as "haematology"/"foetus" (British) and "hematology"/"fetus" (American)
- addition of letter "u", such as "tumour"/"mould" (British) and "tumor"/"mold" (American)



- use of “ph” for “f”, such as “sulphate” (British) and “sulfate” (American)

Table 1 gives a few examples of British and American spellings.

**Table 1. British and American spellings.**

British	American
analyse	analyze
artefacts	artifacts
cancelled	canceled
centre	center
characterise	characterize
colour	color
criticise	criticize
defence	defense
emphasise	emphasize
encyclopaedia	encyclopedia
favour	favor
foetus	fetus
fulfill	fulfil
grey	gray
haematology	hematology
immortalise	immortalize
judgement	judgment
licence	license
mould	mold
neighbour	neighbor
organisation	organization
oxidise	oxidize
prioritise	prioritize
realise	realize
recognise	recognize
specialise	specialize
stylised	stylized
sulphate	sulfate
summarise	summarize
symbolise	symbolize
tumour	tumor

## Interesting Fact

PubMed “translates” users' queries with British spellings into American spellings behind the scenes. For more information, read the [NLM Technical Bulletin](#).

## Statistics

In general, variables are italicized, such as *P* for probability test, *t* test, *U* test; *R* for regression coefficient, *r* for correlation coefficient; *n* for total number.

Distributions could be:

- exponential (duration)
- geometric (time to first success)
- Poisson (counting, rare events)
- normal (measurement error, asymptotic, approximations)
- uniform
- Bernoulli (coin toss)
- binomial (sum of coin tosses)

Some common variables used in bioinformatics are: *CR* (critical region); *D* (divergence); *E* (*E*-value), *err* (average training error); *Err* (true test error); *F* (distribution function); *g*; *H* (entropy); *I* (mutual information); *K*, *k*; *P*, *p* (probability); *S* (sample space); *T*, *t*; *V* (variance); *X*, *x*; *Y*, *y*; *z*.

*Common terms are:*

- aggregating, model averaging
- bagging, bootstrap aggregating
- Bayes theorem
- best-case scenario
- BIC, Bayesian information criterion
- bootstrap, replicate learning sets
- Central limit theorem
- Chebyshev's inequality
- CI, confidence interval
- COV, covariance
- CV, coefficient of variance
- cross-validation
- eigen value
- FDA, flexible discriminant analysis
- Gaussian
- Gini index
- GLM, generalized linear model
- Hessian
- hyperplane

- K-fold cross-validation
- kernel functions
- kNN,  $k$  nearest neighbor
- IRLS, iterated reweighted least squares
- Kolmogorov–Smirnov
- Lagrangian multipliers
- Laplace approximation
- LDA, linear discriminant analysis
- least squares
- LLR, likelihood ratio
- logodds
- Mahalanobis distance
- MAR, missing at random
- MCAR, missing completely at random
- Markov chains
- MDL, minimum description length
- MLE, maximum likelihood error
- model-fitted values
- MoM, method of moments
- MSE, mean square error
- Newton–Rhapson
- Neyman–Pearson
- nonlinear
- QQ-plots
- PDA, penalized discriminant analysis
- SD, standard deviation
- SE, standard error of the mean
- Softmax, dummy regression
- SRS, simple random sample
- WLLN, weak law of large numbers
- z-score

## Symbols

An abbreviated list of quantities, units, and symbols used in physical chemistry and related disciplines is available from IUPAC.

See pp. 815-818 in the CBE Index for a list of symbols.

CBE provides the correct symbol representations for many diverse fields of science.

In Chapter 15 (**Subatomic Particles, Chemical Elements, and Related Notations**, pp. 251-259), you learn that IUPAC recommends roman letters for particle symbols, but the Particle Data Group uses italics for the elementary particles. By IUPAC standards, “e” stands for “electron”, and “*e*” stands for “elementary charge”. The Particle Data Group uses *e* to mean “electron”.

Chapter 16 (**Chemical Names and Formulas**, pp. 260-304), p. 292, describes “Protein Designations Based on Gene Names”.

### Example:

“Names for proteins representing a mutant that is characterized by replacement of a single amino acid are sometimes derived from the corresponding gene symbol. For example, an amino acid replacement with valine at position 12 of the Ras protein may be indicated by Ras<sup>Val12</sup>.” (from CBE, p. 292)

More specific information on genes in particular can be found in Chapter 20 (**Cells, Chromosomes, and Genes**) on pp. 334-380. In general, genes are shown in italics and proteins are shown as roman, but when listing many, many genes, it is not helpful to use this style. You must use your judgment when deciding on when to show genes in an italic style.

**Viruses** are covered in Chapter 21 (pp. 381-394), **Bacteria** in Chapter 22 (pp. 395-410), **Plants, Fungi, Lichens, and Algae** in Chapter 23 (pp. 411-445), and **Human and Animal Life** in Chapter 24 (pp. 446-485). Within Chapter 24 is a detailed description of taxonomy.

There are other interesting chapters in the “Special Scientific Conventions” section of CBE as well.

## Time

Convert military time (0600–1800) to AM/PM time (6 a.m.–6 p.m.; 6:30 a.m.–6:30 p.m.). Include time zone and daylight/standard time designations, if necessary.

Units of time: year, month, week, day, hour, min, sec. Only “min” and “sec” are abbreviated.

## Trademarks, Trade/Supplier Names

Use superscript TM or ® when known, at first occurrence only.

At first mention of a company name, use its full form (e.g., Sigma Chemical Co.).

It is OK to abbreviate Co., Inc., Ltd., etc. Thereafter, refer to the company with a shortened version (e.g., Sigma). No need to supply city/state information. If items have been provided as gifts, provide city/state information (no postal code).

## Units of Measure

See the CBE Index of “symbols(s)”, beginning on page 815.

Use “ml” for milliliter, “mol” for mole, and “M” for mol/l. “Liter” is written out when used alone.

## **versus**

Use italics.

Write out “versus” in text, but it is OK to use “vs.” in tables.

## **X-ray**

Use a capital X wherever the term appears.