Proposal for Ten Chemical Symbols L2/23-193R

Submitted by: Neil Soiffer, Murray Sargent, Asmus Freytag
Updated: 2023 July 26

1 Acknowledgements

The authors gratefully acknowledge additional input from David Carlisle from the W3C Math Working Group and Joseph Wright (School of Chemistry, University of East Anglia). Joseph Wright is part of the International Union of Pure and Applied Chemistry which is in the process of updating the IUPAC Green Book (symbols for physical chemistry). That group is waiting on this proposal to support their efforts in standardizing notation, which would benefit immensely from having assigned character codes.

2 Overview

Unicode has many arrow symbols (see arrows in the blocks U+2190..U21FF, U+27F0..U+27FF, U+2900..U+297F, U+2B00..U+2BFF, and U +1F800..U+1F8FF). Nevertheless, there are nine stretchable arrow characters plus a circle-based symbol that are used in chemistry and are not yet encoded in the Standard and we feel should be. These 10 characters are proposed here for encoding.

3 Reaction arrows

Reaction arrows connect reactants with products in a formula showing a chemical reaction. Absent a standard character encoding they are typically produced by special packages for chemical layouts using TeX or LaTeX. Here's a partial listing from the documentation for the "mchem" package (See page 10 of https://texdoc.org/serve/mhchem.pdf/0).

$\ce{A -> B}$	`\ce{A -> B}`
$\ce{A <- B}$	`\ce{A <- B}`
$\ce{A <-> B}$	`\ce{A <-> B}% not to be used according to IUPAC`
$\ce{A <--> B}$	`\ce{A <--> B}`
$\ce{A <=> B}$	`\ce{A <=> B}`
$\ce{A <=>> B}$	`\ce{A <=>> B}`
$\ce{A <<=> B}$	`\ce{A <<=> B}`

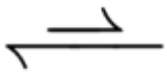
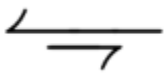
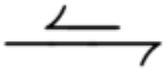
The first three of these can be unified with existing long arrows in Unicode (see "Unifications" below).

The fourth arrow is used for reactions that can be driven either way, while the last three arrow symbols are commonly used for equilibrium reactions. The first of these arrows indicates a balanced equilibrium while the other two indicate equilibria where either reactants or products are favored. For readability, the

distinguishing part of the proposed names have been emphasized. The glyphs represent actual examples found. For the representative glyphs the “standard” arrowheads or harpoon barbs should be used

.

Table 1: Six Proposed Long Arrow Pairs (for use in chemistry)

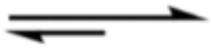
LONG RIGHTWARDS ARROW **OVER** **LONG** LEFTWARDS ARROW	
LONG RIGHTWARDS HARPOON **OVER** **LONG** LEFTWARDS HARPOON	
LONG RIGHTWARDS HARPOON ABOVE **SHORT** LEFTWARDS HARPOON	
SHORT RIGHTWARDS HARPOON ABOVE **LONG** LEFTWARDS HARPOON	
LONG LEFTWARDS HARPOON ABOVE **SHORT** RIGHTWARDS HARPOON	
SHORT LEFTWARDS HARPOON ABOVE **LONG** RIGHTWARDS HARPOON	

The last character in this list is not directly attested, but would be needed to allow correct mirroring of the unbalanced equilibrium arrows.

3.1 Alternate convention

There is an alternate convention that can be documented in use, but is considered non-preferred by practitioners in the field and possibly limited to a single source (however, that single-source conjecture cannot be proven). The two conventions differ in whether the non-favored direction (short arrow) is *centered* or not. (Additional permutations of arrangements are not observed).

Table 2: Three Glyph Variants for Long Arrows (unified with proposed)

LONG RIGHTWARDS HARPOON **OVER** **LONG** LEFTWARDS HARPOON	
LONG RIGHTWARDS HARPOON ABOVE **SHORT** LEFTWARDS HARPOON	
LONG LEFTWARDS HARPOON ABOVE **SHORT** RIGHTWARDS HARPOON	

We investigated proposing both conventions for the reason that Unicode is descriptive and not prescriptive, and therefore arrows that are in actual use are eligible, whether or not they fit some style guide. The other reason is that we don't want to see existing practice implemented as "font variant", because that would violate the encoding principles for this type of character which is based on appearance over semantics.

However, our interpretation of the naming conventions is that they do not call out horizontal alignments and we take that as an indication that the difference here is a glyph variant.

Note: the first character in Table 2 is attested but would only differ in barb style and wound thus always be considered a glyph variant.

3.2 Glyph variants and representative glyphs

The differences in the shape of the barb, the height of the gap, or differences in the length shown in the examples seem to be obvious targets for unification. When these symbols are mocked up in drawing packages, they would typically assume a half triangle shape. Practitioners questioned stated that no semantic difference attaches and that "fishhook" is the way these are generically referred to.

It should be noted for completeness that none of the barb shapes matches those of other harpoon characters encoded and that Unicode has started to distinguish the use of triangular arrowheads for some arrows. Nevertheless, we think it would be acceptable to unify these and use the "standard" barb or arrowhead variant matching the other harpoon characters for the representative glyph. Ultimately, the scope of unification of arrowheads or harpoon barbs would be a question for the committee to decide, however.

3.3 Names

The proposed character names do not mention the orientation of the barbs as that matches the "unmarked" state as seen from the example of already encoded characters, such as

21CC ⇌ RIGHTWARDS HARPOON OVER LEFTWARDS HARPOON

On the other hand, the "centered" alignment of the harpoon in the alternate convention is opposite to that of existing characters pairing long and short arrows, such as

2942 ⥂ RIGHTWARDS ARROW ABOVE SHORT LEFTWARDS ARROW

which raises the issue of whether that it needs to be expressed in the character name. However, we note that other centered symbols are also named "above".

2B40 ⥀ EQUALS SIGN ABOVE LEFTWARDS ARROW

Our conclusion is that the horizontal alignment is considered a glyph variation.

Finally, while the exact length of these characters may not be fixed-length in use (stretchable), they are invariably shown with a default length longer than typical arrows and therefore should be marked as LONG to match the example given by

27F6 ⟶ LONG RIGHTWARDS ARROW
→ 2192 → rightwards arrow

Note that the existing characters exhibit an alternation in the use of ABOVE vs. OVER. We propose ABOVE except where there is a direct analog arrow of standard length using OVER. However, we will happily defer to the committee's interpretation of the precedent.

3.4 Layout

When laying out chemical formulas, text and description can be added above or below a reaction arrow. This may lead to the arrow being rendered in an elongated form, not unlike the resizing applied to some mathematical symbols in mathematical typesetting.

$a \underset{\text{below}}{\overset{\text{above}}{\rightleftharpoons}} b$

vs.

$\mathrm{A} \xrightarrow[\text{text below}]{\text{text above}} \mathrm{B}$

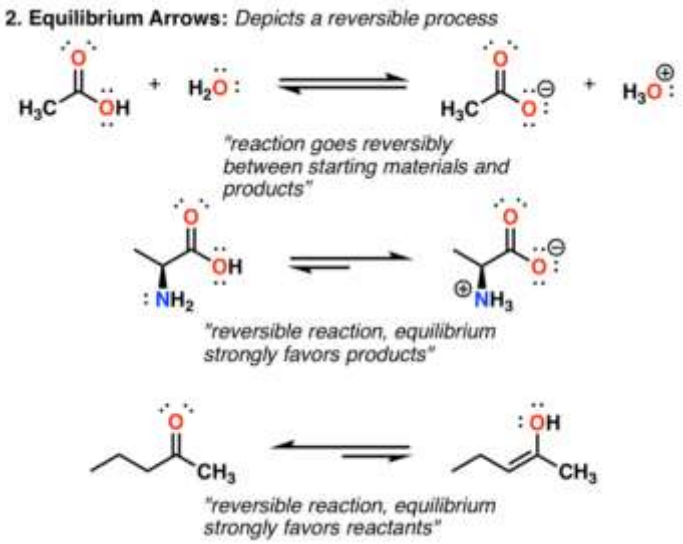
The details are outside the realm of character coding, but these arrows would be longer than "standard" arrows even outside special layout. Therefore, we are proposing them as an extension of the set of "long" arrows, but without specifying a particular length beyond the contrast to "standard" arrows.

3.5 Glyphs

Chemical publications are typically prepared using common TeX or LaTeX macro packages, which implement the reaction arrows as macros built from primitive shapes. (They also take care of additional layout issues). It would be unusual to find these as glyphs in fonts, limiting the variety of observed glyphs.

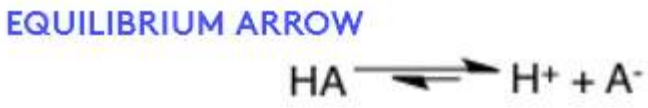
The images come from the commonly used mhchem TeX macro package (see "REACTION ARROWS") A search for "equilibrium arrows" will turn up many examples, but here are three examples:

https://www.masterorganicchemistry.com/2011/02/09/the-8-types-of-arrows-in-organic-chemistry-explained/.



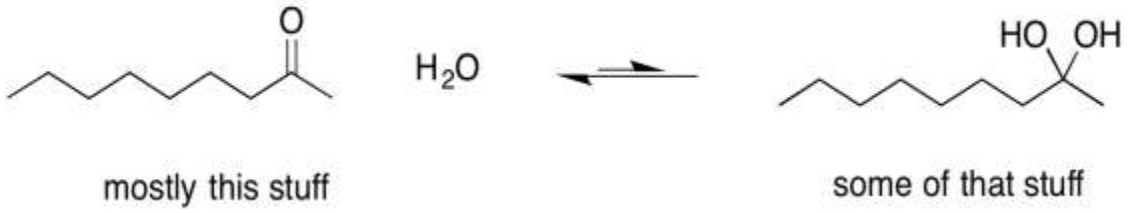
While the literature refers to these as reaction arrows and names them in terms of reactants and products, a modified set of names that uses directions seems more appropriate given the precedents for naming arrows.

The other sources show glyph variants which center the arrow for the non-favored reaction. This convention matches what is found in the “chemarr” package from which the list of arrows at the top of this section was extracted, modulo some differences in barb style.

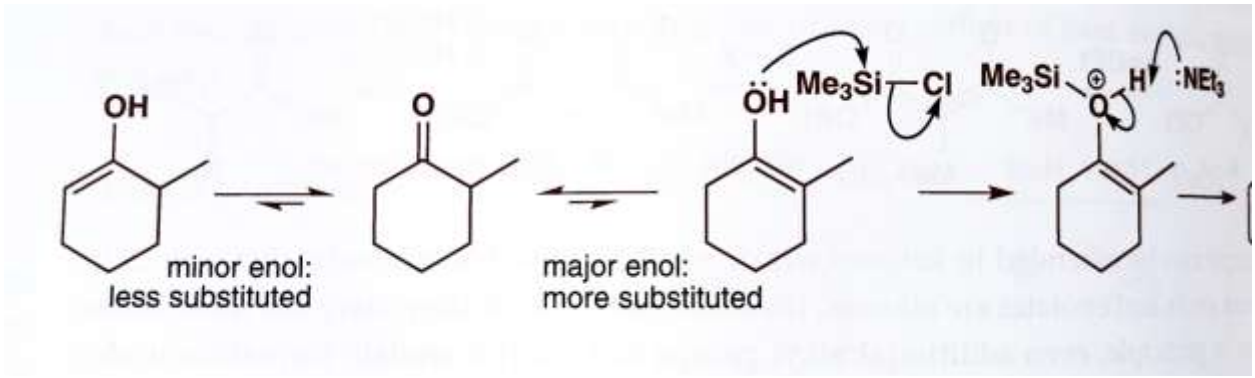


https://blog.cambridgecoaching.com/guide-to-deciphering-chemistry-arrows

These are not used contrastively. The same applies to a glyph variant with the shorter arrow on top as shown in the worksheet cited below.



https://chem.libretexts.org/Ancillary_Materials/Worksheets/Worksheets%3A_Inorganic_Chemistry/Structure_and_Reactivity_in_Organic_Biological_and_Inorganic_Chemistry_(Chem_315)/3%3A_Understanding_Mechanisms/3.3%3A_Arrow_Conventions



.

This last example (from 'Organic Chemistry', 2nd Ed, J. Clayden, N. Greeves and S. Warren, OUP, 2012.) places the LONG arrow consistently above the SHORT one, while other examples consistently have the RIGHTWARDS arrow on top. The convention used for character names for arrows fixes the relative position (ABOVE) of long and short arrows/harpoons. We feel that unification of glyphs that would be named differently is unattractive. We therefore propose a descriptive approach that recognizes a distinction in use, but treats the arrows with the LONG LEFTWARDS part on top as distinct characters. This also aids in mirroring.

In contrast, treating the differences in barb style as well as whether to center the shorter arrow as glyph differences seems more appropriate – particularly since character naming does not describe horizontal alignment for other symbols placed on arrows and long arrows or harpoons have not been encoded based on barb style.

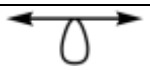
The output of various TeX and LaTeX packages can be viewed using https://texlive.net. Here's a link showing a live view of some of the characters discussed. (https://texlive.net/run?%5Cdocumentclass%7Barticle%7D%0A%5Cusepackage%5BT1%5D%7Bfontenc%7D%0A%5Cusepackage%7Bmhchem%7D%0A%5Cusepackage%7Bchemarr%7D%0A%5Cbegin%7Bdocument%7D%0A%5C%5Ba%5Cxrightleftharpoons%5B%5Ctext%7Bbelow%7D%5D%7B%5Ctext%7Babove%7D%7Db%5C%5D%0A%5C%5Ba%5Crightleftharpoons%20b%5C%5D%0A%5C%5B%5Cce%7BA%20%3C=%3E%3E%20B%7D%5C%5D%0A%5C%5B%5Cce%7BA%20%3C%3C=%3E%20B%7D%5C%5D%0A%5Cend%7Bdocument%7D)

4 Three additional arrows

Two arrow symbols covers the symbols for unsuccessful reactions, while another one is used in case where the two sides have the same arrangement of electron lobes ("isolobal"):

Table 3: Three proposed long arrows (for use in chemistry)

LONG RIGHTWARDS ARROW WITH THROUGH X	⟞
LONG RIGHTWARDS ARROW WITH DOUBLE SLASH	⟋

LONG LEFT RIGHT ARROW WITH DEPENDENT LOBE	

These are derived from the standard LONG LEFT RIGHT ARROW at U+27F7 and RIGHTWARDS ARROW found at U+27F6 (notwithstanding the small difference between the arrowhead designs compared to the representative glyphs in the standard, which we judge to be unifiable).

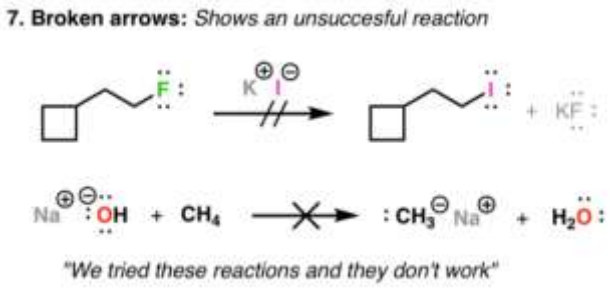
4.1 Glyphs

The glyphs shown here match existing use. For the representative glyphs we propose matching both the length and the arrowhead designs to that of the “long” arrows already encoded.

The first two symbols in Table 3 can be produced in the popular (in chemistry) WYSWYG ChemDraw package, while the third one is rare enough that it is typically supported with an ad-hoc solution, often by construction in a graphical editor.

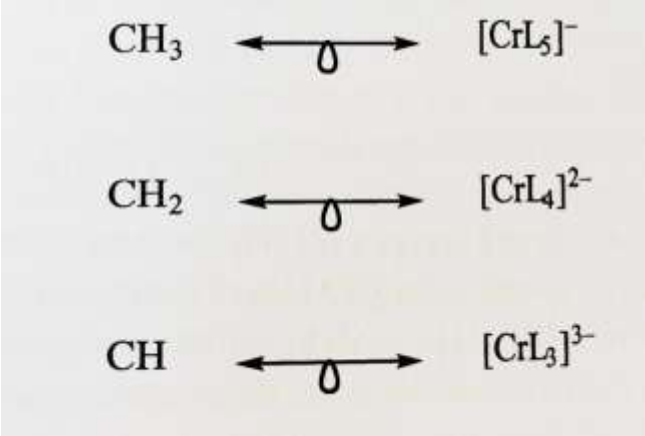
Here are a few sites that show their use:

https://www.masterorganicchemistry.com/2011/02/09/the-8-types-of-arrows-in-organic-chemistry-explained/. For example (lower arrow)



https://www.thoughtco.com/chemical-reaction-arrows-overview-609203

https://blog.cambridgecoaching.com/guide-to-deciphering-chemistry-arrows



This arrow is attested in

1. 'Molecular orbitals of Transition Metal Complexes', Y. Jean, OUP, 2005 (research level book)
2. 'Inorganic Chemistry', 7th Ed., A. Weller, T. Overton, J. Rourke and F. Armstrong, OUP, 2014 (core undergrad teaching)

5 Standard state symbol ("Plimsoll" mark)

The last proposed symbol is the standard state symbol:

Table 4: One proposed symbol (for use in chemistry)

MEDIUM **SMALL** WHITE CIRCLE WITH HORIZONTAL BAR	⊖

In TeX, this symbol is typically produced by overlaying characters. There is a slightly similar existing character: U+29B5, which has been identified by nameslist annotation as the character to use for standard state. However, as can be seen, this identification appears to be an example of arms' length unification. The glyph for U+29B5 is much too large, and the standard state symbol should also have a significantly longer horizontal line (as mentioned in the Wikipedia link).

Here are four typical attempts at rendering this symbol. In order, the sample shows two fallbacks (the degree sign and a manual overlay of a degree with a minus sign), the desired shape, and finally a superscripted U+29B5.

$$\Delta_r H^\circ \; \Delta_r H^{\circ\!\!-} \; \Delta_r H^{\ominus} \; \Delta_r H^{\circ\!\!\!\!\ominus}$$

In chemistry, the **standard state** of a material (pure substance, mixture or solution) is shown with a superscript symbol. We propose encoding a regular sized symbol which can then be superscripted as needed as part of layout. The ratio of diameters to U+29B5 (1:2) in the samples is larger than the ratio between the medium white circle (at U+26AA) and regular circle symbols for Unicode, and instead closer to the ratio obtained for U+26AC MEDIUM SMALL WHITE CIRCLE. We have indicated that preference by adding "(SMALL)" in the proposed character name.

Alternatively, the committee may prefer to encode this semantically as its own symbol under the name "plimsoll mark" or "standard state symbol", which then sidesteps the question whether the aspect ratio of bar and smaller circle is generic and part of the series of sizes for circled characters with overlays. (See also the discussion in Section 2.11 of UTR#25 on sizes of geometric shapes).

More discussion from the related Wikipedia article:

> A superscript circle ° (degree symbol) or a Plimsoll (⦵) character is used to designate a thermodynamic quantity in the standard state, such as change in enthalpy (Δ*H*°), change in entropy (Δ*S*°), or change in Gibbs free energy (Δ*G*°).[1][2] The degree symbol has become widespread, although the Plimsoll is recommended in standards, see discussion about typesetting below.

See also https://iupac.org/wp-content/uploads/2019/05/IUPAC-GB3-2012-2ndPrinting-PDFsearchable.pdf (search for "standard state") and https://www.coursehero.com/study-guides/introchem/standard-states-and-standard-enthalpy-changes/.

While the IUPAC currently condones the practice of using a superscript small circle (or degree sign) as a fallback due to lack of ready alternatives, the major drawback of that is the confusion with superscript zero. A new edition of the IUPAC Green Book, which makes recommendation for symbols in chemistry, is in preparation, and with a move to digital-first there is a strong desire to be able to use the dedicated symbol here in a way that can be recognised both electronically and visually.

One small macro package that is dedicated to producing just this symbol is plimsoll.pdf. (see https://ctan.org/pkg/plimsoll.) This can be used to access an actual outline.

6 General Notes

6.1 Unifications

The following existing characters can be unified with reaction symbols. Note that U+27F7 is not recommended for equilibrium reactions, but can be found in use. This also underscores that it is not the role of the Unicode Standard to standardize chemical notation, but to provide the code points to match existing practice, and leave it up to chemistry publishers to enforce style guides.

6.2 Description in chapter 22

If this proposal is accepted, the text in chapter 22 would need to be revised by adding additional information.

Existing text:

> **Long Arrows**
>
> The long arrows encoded in the range U+27F5..U+27FF map to standard SGML entity sets supported by MathML. Long arrows represent distinct semantics from their short counterparts, rather than mere stylistic glyph differences. For example, the shorter forms of arrows are often used in connection with limits, whereas the longer ones are associated with mappings. The use of the long arrows is so common that they were assigned entity names in the ISOAMSA entity set, one of the suite of mathematical symbol entity sets covered by the Unicode Standard.

Additional text:

> In chemistry, various long arrows are used to indicate reactions. The set includes U+27F5..U+27F7 as well as extensions in other block(s) (TBD). When typesetting chemical formulas, such arrows may be further decorated by annotations placed above or below, which can result in their length being adjusted, not unlike the resizing applied to certain mathematical symbols in mathematical typesetting. The Unicode Standard encodes a single, default length that contrasts with the normal “short” arrows, and represents an acceptable default length for use in plain text.

6.3 Namelist annotations and cross references

The erroneous annotation should be removed from U+29B5 and moved to the newly encoded character for the plimsoll mark.

The usage of both the proposed and the unified long arrows for chemical reactions should be noted in namelist annotations. For long versions of existing composite arrows, cross references to the standard length ones would be appropriate. As would a set of notices tying together the existing range of long arrows with the ranges for these proposed extensions.

6.4 Properties for arrows and additional Symbol

The character properties would conform in all respects to those of analogous arrow characters, in particular any long arrow characters. In the case of the circle symbol (plimsoll mark), properties match those of U+29B5. Once names and tentative code points have been settled, we can prepare the usual extract of UnicodeData.txt, if that is considered helpful.

6.5 Code positions

No code point locations have been proposed. There is an existing set of three holes in the 2B00 block, One of them may be used for the additional circle-based symbol (plimsoll mark), but it may be more useful to keep the arrows together as they are all “long” arrows.