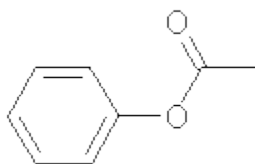


CHEMICAL NAMES, TRANSLATIONS, AND INCHIS (AND SOME KLINGON)

Roger Sayle of OpenEye describes a software tool for translating chemical names from one language into another, for example a name in English to the corresponding name in German or Japanese or Chinese.¹ One might imagine the many difficulties in this process – the strong dependency on very minor changes in the name can mean substantially different chemicals (think of propane *vs.* propene *vs.* propyne *vs.* propanol *vs.* propanal *vs.* propenal, etc.) and all this needs to be carefully recognized and translated.

My interest here is that chemical names including IUPAC names are really not the *lingua franca* of chemistry. Yes, chemists do use names but we really rely most on chemical structure drawings – that's the surest way of transmitting our meaning to one another. But images are not a very good way for computers to communicate meaning. And that's where the InChI label steps in. InChIs provide a standard means for two chemists (or two computers) to exchange chemical information without introducing errors and without the need for an intervening third-party to guarantee the meaning.

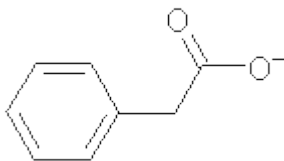
Here is a simple example of the benefit of the InChI using the first example from Sayle's article. Figure 1 in his paper has the two similarly named compounds phenylacetate **1** and phenyl acetate**2**. Note the critical importance of the blank space (actually this blank space is omitted in Figure 1 of the paper indicating just how easy it is for errors to sneak in!). Sayle points out that many languages beside English do not make use of whitespace in the way that English does, so that this translation must be done with special care.



phenyl acetate **1**

InChI=1/C8H8O2/c1-7(9)10-8-5-3-2-4-6-8/h2-6H,1H3

InChIKey=IPBVNPXQWQGGJP-UHFFFAOYAF



phenylacetate **2**

InChI=1/C8H8O2/c9-8(10)6-7-4-2-1-3-5-7/h1-5H,6H2,(H,9,10)/p-1/fC8H7O2/q-1

InChIKey=WLJVXDMOQOGPHL-CAGLWTAICB

Now the InChIs of **1** and **2** are unambiguous and will be the same in all (human) languages, eliminating the need of translation concerns. And if you're unhappy with the length of the InChI, the InChIKey provides a fixed length string that captures almost all of the information of the InChI itself.

All compounds mentioned in my blog are listed at the end with their InChIs to promote the exchange of information and to encourage search-and-retrieval through the web. The leader in this sort of technology has been the Chemspider site, and I urge you to explore it and the use of InChIs.

As an aside, for all you *Star Trek* nerds, the paper includes this quote

This relationship is even preserved in some modern synthetic languages, such as Klingon where water is “[klington glyphs omitted here but do appear in the paper!]” (bIQ) and hydrogen is “[klington glyphs omitted here]” (bIQ-SIp).

Source: <http://comporgchem.com/blog/?p=122>