

# database application

## A Database for PPCH<sub>T</sub>E<sub>X</sub>

Richard Müller  
(r.a.mueller@cityweb.de)

### abstract

A database with PPCH<sub>T</sub>E<sub>X</sub> code would ease the work for those who have to draw chemical formulas only occasionally and have no time to become acquainted with it. But experts could also make use of it. The establishing of such a database is discussed and proposed.

### keywords

PPCH<sub>T</sub>E<sub>X</sub>, chemistry, structural formulas, database

Some time ago I came around to draw some chemical formulas. Since I am using T<sub>E</sub>X for the work which should look nicer than with ordinary textsystems I looked around to find something to set these formulas with. I found out that PPCH<sub>T</sub>E<sub>X</sub> would do the job quite well. Then I started to try to understand the language you use for PPCH<sub>T</sub>E<sub>X</sub>. This was not as simple as it first seemed. Why does, for example, that sidegroup not appear on the screen? What was wrong with the code? In my profession (I am a biology teacher) I don't use heaps of chemical formulas, but the occasional one or other. So you don't get very acquainted with the PPCH<sub>T</sub>E<sub>X</sub>-language. In OTTEN (1998) and CONT<sub>E</sub>XT (1998) I found help. But both publications have disadvantages: The manual (OTTEN 1998) is essential to understand and write proper PPCH<sub>T</sub>E<sub>X</sub>-code. But complete examples are rare and can be used unchanged only in very seldom cases. And of course this is not the intention of the article; it's a manual and not a dictionary of formulas. CONT<sub>E</sub>XT up-to-date 1998/2 is a fine PDF-document with estimated 170 examples of nice-looking formulas. But unfortunately no names are given, so you can find the structures you are searching for only by clicking from one page to the next through the whole document. In case names were given you could use Acrobat's search-function. And as far as I know it is not intended that further examples become added to this collection, be it by the authors or be it by the users. One day as I was figuring out a new structure I thought "Are you the only one who uses this wicked language?" The idea came on to me, that all users of PPCH<sub>T</sub>E<sub>X</sub> should join their codes and establish one pool which is used and also being contributed to by everyone who uses PPCH<sub>T</sub>E<sub>X</sub>. It is obvious that the right place to

publish such a code collection is the internet. To facilitate search and navigation you would need a hypertext system as PDF or HTML. The choice fell on HTML because it is integrated in every browser and office software. I simply used StarOffice to set up my example pages quite quickly. (Perhaps I could have used PDF<sub>T</sub>E<sub>X</sub>, but I have no experience with it). What should the pages of the database contain? The answer is quite simple: name, formula and structure! To cover synonyms, every known and used name of a chemical substance covered in the database should appear in the alphabetical index, be it perhaps the systematical IUPAC name, be it its common name, or be it a name used in former times. These different names for the same substance should be linked with the same page: *propanetriol* and *glycerol* are linked with the page `glycerine.html`. On the other hand, in every case you can write the same substance in different manners: vertically or horizontally orientated, with carbon atoms in the corners or without, in ring or in chair form and so on. All these different examples of the same substance should be collected on the same page, with the typography of the formula shown together with its specific code. In a discussion with Hans HAGEN the idea was born to publish only the bodies of the structures. The user should then complete the code of the desired structure himself. In my opinion this is possible with some substances which differ e.g. only in sidegroups: Most types of *dioxines* have nearly the same structure. The sidegroups are represented by letters in alphabetical order which can be found in the published code. So it is simple to replace the letters in the code with the desired groups or atoms. But how to print the exact typography of the formula in the database? I found no other way than to perform a L<sub>A</sub>T<sub>E</sub>X run on the code and to print the structure on the screen. I used Windvi and Ghostscript with Ghostview. As far as I could distinguish, the formula printed by Ghostscript looked a little bit nicer on the screen. With screen capturing I made a GIF of the formula which became a part of my HTML page. To save space (and loading time) the GIF should be as small as possible. The GIFs I present have filesizes between 2 and 5 kB.

This database cannot live solely with my own contributions. In the beginning it will be quite empty, but with a little help of PPCH<sub>T</sub>E<sub>X</sub>' friends it should become filled in a certain amount of time. I want to find out, if there is need for such a project and if other PPCH<sub>T</sub>E<sub>X</sub> users are willing to share their codes with the rest of the community. In case

this is true, the following topics should become discussed: Can the proposed form of the database suit the needs of the user? Or is it perhaps of more advantage to create a kind of mailinglist restricted for publishing PPCH<sub>T</sub>E<sub>X</sub> code, where everyone simply can contribute by mailing the name, code and GIF? In case we vote for establishing an internet-based database as pointed out above, should it be open for direct writing into it by everybody? I am sure you find a lot more points worth discussion. You can navigate to my example pages of the proposed database at:

<http://www.caw.de/walram/ppchtex/00index.html>.

I appreciate your comments.

### Literature

CON<sub>T</sub>E<sub>X</sub>T 1998: CON<sub>T</sub>E<sub>X</sub>T up-to-date 1998/2. Pragma ADE, Hasselt 1998. (<http://www.ntg.nl/context/uptodate.htm>)

OTTEN, T. 1998: PPCH<sub>T</sub>E<sub>X</sub> manual. MAPS 20, 150, 1998