

Chemical structures with ppch \TeX

Alan Braslau

Email alan.braslau@cea.fr

Address Service de Physique de l'État Condensé
CEA/Saclay
Orme des Merisiers
91191 Gif-sur-Yvette cedex FRANCE

Abstract Chemical formulas and chemical structures can be included in a \LaTeX or a Con \TeX t document easily using the ppch \TeX macros. We present here a simple introduction to their use. Additionally, a more extensive tutorial is available in the documentation of the package.

1 Introduction

In a scientific/academic environment, one may find the need to represent chemical formulas and draw chemical structures. This can be easily reached and integrated directly in a document by taking advantage of the beauty of \TeX output and its logical construction!

There are many commercial applications that are extensively used by chemists to produce graphics that can be included in a \TeX document using standard external graphics macros. **CTAN** contains a number of free macro packages aimed at writing chemical formulas, drawing reaction arrows, numbering reactions and even drawing structures such as **bpchem**, **chemarrow**, **chemcompounds**, **chemcono**, **chemsym**, **mhchem**, and others¹. One of the most sophisticated packages is XyM \TeX written by Shinsako Fujita, available from the [XyM \$\TeX\$ homepage](#).

Whereas XyM \TeX appears to be a great package, I was never able to successfully learn the complexity of it. Thus, this paper talks about a simpler yet very powerful macro package,

1. These macro packages are available in \TeX Live under many distributions in the `texlive-science` package.

ppchTeX, that was developed for ConTeXt and may be successfully used under L^ATeX as well².

The goal here is not to present a tutorial, much less to provide a complete exposition of the possibilities of ppchTeX. Nevertheless, this article aims only at giving a small taste of how you can easily draw chemical structures using TeX.

A few details: Under L^ATeX, positioning is performed by using P_ICT_EX and line drawing using pstricks; Under ConTeXt MkII, positioning is performed by using P_ICT_EX and line drawing using METAPOST; Under ConTeXt MkIV under LuaTeX, the internals of ppchTeX have been re-written to be included natively as core macros and no longer relies on P_ICT_EX, and both positioning and drawing are performed using METAPOST. This re-implementation has led to a considerable improvement in the performance and flexibility of the macro programming, so I expect a continued development of its possibilities.

Resources

- A more complete manual (currently being updated and re-written) can be found at: [ppchTeX manual](#)
- A presentation with many ppchTeX examples can be found at: [ppchTeX examples](#)
- Other resources can be found on the web, such as the ConTeXt wiki: [ConTeXt Chemistry wiki](#)
- Finally, a very nice set of examples of drawing amino-acids: [drawing organic molecules in laTeX II - amino acids](#)

2 Drawing a simple molecule

2>1 A first example using L^ATeX2_ε

Let's draw a simple chemical structure. First, you need to load the following macro definitions³:

². Indeed, it was through my learning to use ppchTeX under L^ATeX that I was first introduced to ConTeXt!
³. These macros are available in any ConTeXt distribution, such as from TeX Live.

```

\documentclass{article}
\usepackage{m-ch-en} % English interface
% m-ch-nl for a Dutch interface.
% m-ch-de for a German interface.

```

The $\text{P}\text{T}\text{E}\text{X}$ macros are loaded automatically (via $\text{\usepackage{m-pictex}}$) if not called previously.

We conclude the example above by drawing the following simple molecule below:

```

\documentclass{article}
\usepackage{m-ch-en}
\begin{document}
\startchemical
\chemical [ONE,Z0357,SB1357,MOV1,Z037,SB137,MOV1,Z01,SB1]
          [C,H,H,H,C,H,H,O,H]
\stopchemical
\end{document}

```

Yielding $\text{C}_2\text{H}_5\text{OH}$:

$$\begin{array}{ccccccc}
 & & \text{H} & & \text{H} & & \\
 & & | & & | & & \\
 \text{H} & - & \text{C} & - & \text{C} & - & \text{O} - \text{H} \\
 & & | & & | & & \\
 & & \text{H} & & \text{H} & &
 \end{array}$$

which may make your head turn!

This code looks pretty complicated at first glance. But once you begin using $\text{ppch}\text{T}\text{E}\text{X}$, you should be able to write the above line of code without even looking at the documentation.

You will notice that the chemical structure was drawn between a \startchemical - \stopchemical pair, rather than within a proper $\text{\begin{chemical}}$ - $\text{\end{chemical}}$ $\text{L}\text{A}\text{T}\text{E}\text{X}$ environment. Whereas the $\text{ppch}\text{T}\text{E}\text{X}$ package works perfectly well under $\text{L}\text{A}\text{T}\text{E}\text{X}$, its syntax more closely matches that of plain TEX and $\text{Con}\text{T}\text{E}\text{X}\text{t}$.

For the rest of this article I will work with $\text{Con}\text{T}\text{E}\text{X}\text{t}$ only. However, keep in mind that the syntax is identical under $\text{L}\text{A}\text{T}\text{E}\text{X}$ once the packages are loaded as described above. The macro \chemical , either alone or within a \startchemical \stopchemical pair, produces a chemical structure object (a TEX box) that can be used with $\text{L}\text{A}\text{T}\text{E}\text{X}$ or plain TEX .

2▷2 A first example using ConT_EXt

For those unfamiliar with ConT_EXt, I present an easy introduction. ConT_EXt is currently distributed in two flavors: MkII that is compiled using pdfT_EX and MkIV that uses LuaT_EX. MkII is very stable and MkIV is under very active development. Nevertheless, it has now reached a point of stability that I find suitable for my daily production work, and that presents many advantages. The present article was typeset using ConT_EXt MkII, using the macro set in common with L^AT_EX. ConT_EXt is distributed with T_EX Live and may already be installed along with T_EX. The most up-to-date version is available through the [ConTeXt garden](#). A ConT_EXt source file is processed using the scripts `texexec` or `context`. The first is a Ruby script that, by default, uses the MkII macros; the second is pure LuaT_EX and defaults to the MkIV macros. Both handle the processing of multiple T_EX runs, the automatic processing of METAPOST code, and the running of bibT_EX, if necessary. Therefore, creating `filename.pdf` from a ConT_EXt source `filename.tex` is as simple as typing:

```
texexec filename
or
context filename
```

In order to use `ppchTEX`, you may first need to load these macro definitions:

```
\usemodule [chemic]
\setupcolors [state=start]
```

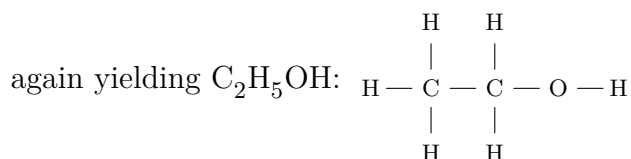
P_IC_T_EX is automatically loaded in this case (ConT_EXt MkII), as is the language variant. Under ConT_EXt MkIV `\usemodule [chemic]` can be skipped altogether as the macros are included natively; indeed, it is ignored. `\setupcolors [state=start]` is also enabled by default in MkIV and it is probably a good idea to include this if using MkII. The order of the two lines is unimportant.

The example above can be completed by drawing the same simple molecule below:

```
\usemodule[chemic]
\starttext
\startchemical
```

```
\chemical [ONE,Z0357,SB1357,MOV1,Z037,SB137,MOV1,Z01,SB1]
[C,H,H,H,C,H,H,O,H]
```

```
\stopchemical
\stoptext
```

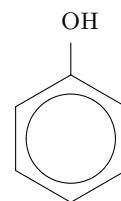


Note that the chemical formula in the text above was written using `\chemical{C_2H_5OH}` (C_2H_5OH) which differs from `C_2H_5OH` (C_2H_5OH) both in their source-code typesetting and readability.

3 Drawing another simple chemical structure

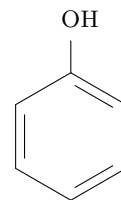
Another, perhaps, easier example is the chemical compound phenol, a six-atom ring structure that is very important in molecular biology.

```
\usemodule[chemic]
\starttext
\startchemical
\chemical [SIX,B,C,R6,RZ6] [OH]
\stopchemical
\stoptext
```



A second representation is drawn using

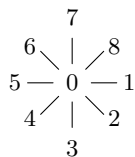
```
\usemodule[chemic]
\starttext
\startchemical
\chemical [SIX,B,EB246,R6,RZ6] [OH]
\stopchemical
\stoptext
```



4 Chemical syntax

The syntax of the command `\chemical` has two optional arguments. The first argument defines the structure and the second presents substituents to be placed in the structure, namely atoms. This allows you to easily define a general molecular form that can be reused, including different substituents to draw different molecules. The inconvenience is keeping track of the correspondence between positions and substituents for a very complicated structure.

The basic molecular forms are `ONE`, `THREE`, `FOUR`, `FIVE`, `SIX` and `EIGHT`.⁴



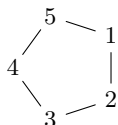
```
\chemical
[ONE,SB1..8,Z0..8]
[0,1,2,3,4,5,6,7,8]
```



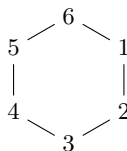
```
\chemical
[THREE,SB,Z1..3]
[1,2,3]
```



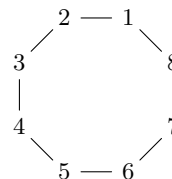
```
\chemical
[FOUR,SB,Z1..4]
[1,2,3,4]
```



```
\chemical
[FIVE,SB,Z1..5]
[1,2,3,4,5]
```



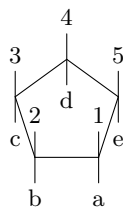
```
\chemical
[SIX,SB,Z1..6]
[1,2,3,4,5,6]
```



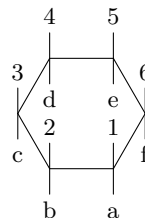
```
\chemical
[EIGHT,SB,Z1..8]
[1,2,3,4,5,6,7,8]
```

`FIVE` and `SIX` have `FRONT` variants:

⁴ Curiously, with ConT_EXt MkII (or L^AT_EX) in this last form, `EIGHT`, the numbering of atoms runs counter-clockwise rather than clockwise as for the other structures. This is, in fact, a bug that will probably *not* be fixed in order to maintain compatibility with existing documents. Under ConT_EXt MkIV, the numbering of atoms runs clockwise coherently for all structures.



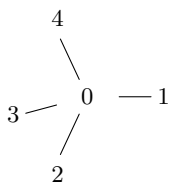
```
\chemical [FIVE,FRONT,B,R]
\chemical [+RZ1..5] [1,2,3,4,5]
\chemical [-RZ1..5] [a,b,c,d,e]
```



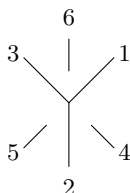
```
\chemical [SIX,FRONT,B,R]
\chemical [+RZ1..6] [1,2,3,4,5,6]
\chemical [-RZ1..6] [a,b,c,d,e,f]
```

The syntax is rather flexible and it can be broken down to multiple `\chemical` calls or else written as one long chain; both approaches are equivalent. For readability, I constructed the last molecules with three calls to `\chemical`.

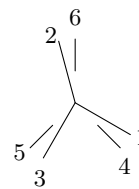
Further structures `CARBON`, `NEWMANSTAGGER` and `NEWMANECLIPSE` and `CHAIR` also are presented:



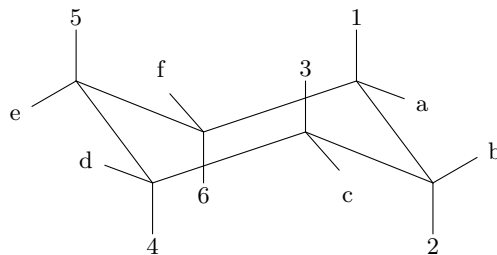
```
\chemical
[CARBON,B,Z0..4]
[0,1,2,3,4]
```



```
\chemical
[NEWMANSTAGGER,B,Z1..6]
[1,2,3,4,5,6]
```



```
\chemical
[NEWMANECLIPSE,B,Z1..6]
[1,2,3,4,5,6]
```



```
\chemical [CHAIR,B]
\chemical [+R,+RZ1,+RZ2,+RZ3,+RZ4,+RZ5,+RZ6] [1,2,3,4,5,6]
\chemical [-R,-RZ1,-RZ2,-RZ3,-RZ4,-RZ5,-RZ6] [a,b,c,d,e,f]
```

5 Bounding box

`\startchemical \stopchemical` creates a \TeX box having a certain bounding box. By default, this is a fixed and standard size: a square of dimensions corresponding to four bond lengths centered on the molecular “origin”. Often, one would like this box to bound the real extension of the molecular structure, and this can be obtained by setting the following options:

```
\setupchemical [width=fit,height=fit]
```

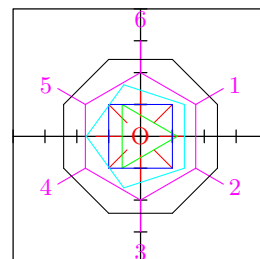
Alternatively, you can select options unique to each chemical structure drawn as:

```
\startchemical [width=fit,height=fit]
\stopchemical
```

Another option to be mentioned is `scale=small` that I have employed in this present article.

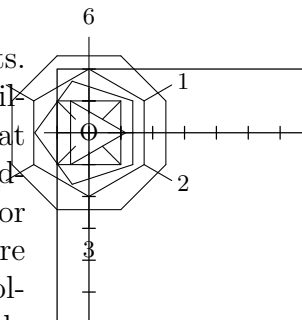
Let me better illustrate the size of the different chemical geometries⁵:

```
\usemodule [chemic]
\starttext
\startchemical [frame=on,axis=on]
  \chemical [ONE,SB,ZO] [0]
  \chemical [THREE,B]
  \chemical [FOUR,B]
  \chemical [FIVE,B]
  \chemical [SIX,B,R,RZ] [1,2,3,4,5,6]
  \chemical [EIGHT,B]
\stopchemical
\stoptext
```



⁵ In this illustration, in fact, I draw each structure using a different color by using the following options `rulecolor=` and `color=`. Each `\startchemical [rulecolor=...] \stopchemical` is then overlaid using a `collector`. I do not display the code here in order to avoid confusion, but you can find details by looking in the source code of this paper that is available for download.

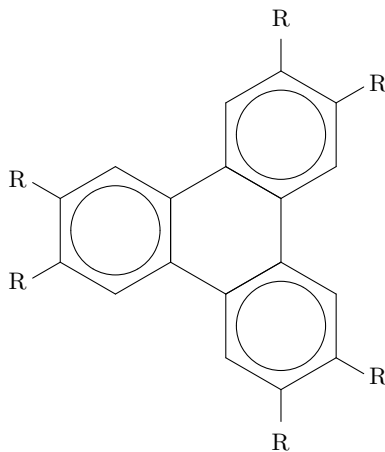
ppchTeX is programmed to use bond lengths of 1000 chemical units. Thus, the default options are `width=4000` and `height=4000`. Also available are the options `left=2000` and `top=2000` (values by default) that allows us to place the molecular origin with respect to a fixed bounding box (automatically they are adjusted when specifying `width=fit` or `height=fit`). To illustrate the bounding box, the same previous Figure is redrawn on the right exactly as shown in the code above (without colors), adding the options `left=500` and `top=1000`. Indeed, the molecule is here located partially outside of the bounding box.



6 Combinations

You also can combine basic forms and build complicated chemical structures, such as this disk-like molecule that forms columnar liquid-crystal mesophases.

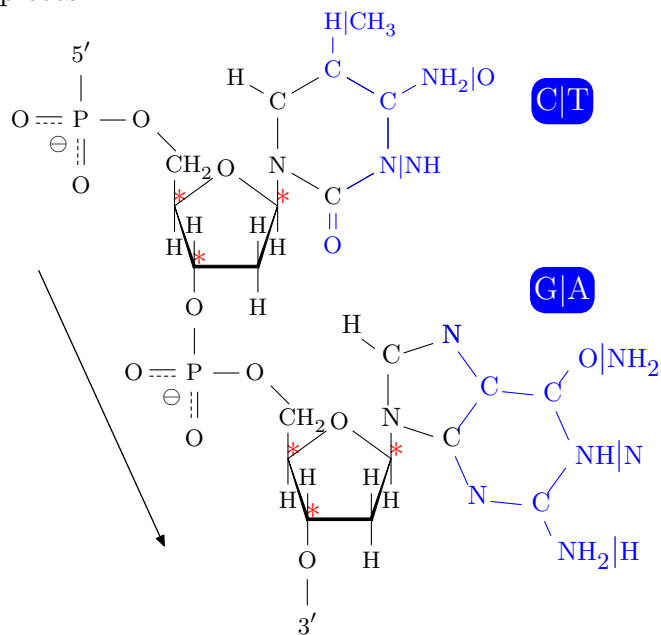
```
\startchemical
  \chemical [SIX,B,MOV2]
  \chemical [B,C,R23,RZ23,MOV5,MOV4] [R,R]
  \chemical [B,C,R45,RZ45,MOV1,MOV6] [R,R]
  \chemical [B,C,R61,RZ61] [R,R]
\stopchemical
```



Building combined structures can result in complicated forms, and I invite you to refer to the [ppchTeX manual](#) for further explanation.

7 Exercise for readers

Some structures appear to be extremely complicated to construct. So, I conclude this article with one very important, and not so easy, example as an exercise. See if you can figure out how it was done. Hint: this molecule was assembled like a jigsaw puzzle from individual pieces.



The primary structure of DNA

8 Acknowledgements

I would like to thank the creators of ppch \TeX , Hans Hagen and Ton Otten, for providing this amazing tool and encourage them and others to continue its development.