PPCHTEX: typesetting chemical formulas in TEX

J. Hagen and A.F. Otten

Pragma P.O.Box 125 8000 AC Zwolle The Netherlands

Abstract

This article is about a package for typesetting chemical formulas. The primary interface of this package is in the dutch language. Because PPCHTEX has a multilingual interface, all commands and keywords can be toggled to english. The Dutch version of this article is published in $\mathcal{N}TG$'s MAPS (94.2) and is translated to English by H. de Weert.

1 Introduction

The macro-package PPCHTEX can be used to typeset chemical formulas. The macros are based on PiCTEX, a macropackage that was created to facilitate the drawing of graphics and other line diagrams. PiCTEX was brought to the public domain by M.J. Wichura. We consider a second implementation, using PSTRICKS of T. van Zandt.

The macros can be used within different TEX-environments and only depend on Knuth's plain TEX. In addition, some general macros of our ConTEXt-library are utilized. Besides, macros are typeset in such a way that further development is quite easy.

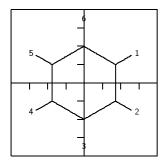
At first, macros are created to typeset chemical structure formulas. Moreover, reaction mechanisms can be reflected. Chemical structures can be typeset in different sizes and comparable formulas can be linked optically. Structures of frequent occurrence can be predefined and recalled.

During the development of the macros, processing speed is subordinated to flexibility, simplicity and quality. No use has been made of the mechanism (available in P_ICT_EX) to store parts of figures in a file. It turned out that this mechanism does not produce a gain of time.

The macros are still being developed. For example, the mechanism to place texts still has to be refined and some structures like CHAIR still have to be added.

2 Structures

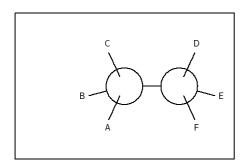
The number of commands that is used to typeset chemical structure formulas is reduced to four.¹ In the following example all of these commands are used.



Example 1

\setupchemical[axis=on,border=on]
\startchemical
 \chemical[SIX,B,R,RZ][1,2,3,4,5,6]
\stopchemical

Different features of the typesetting can be set up with \setupchemical. If something is set up in this way, the setups are valid for all the following formulas.²



Example 2

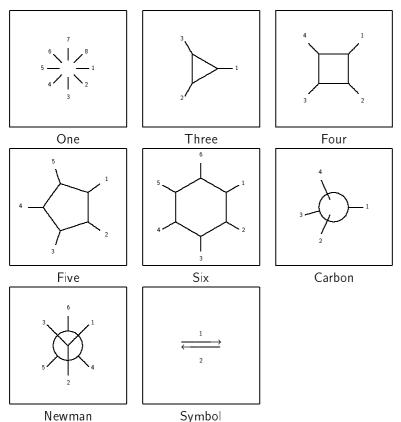
\startchemical[border=on,width=6000] \chemical[CARBON,CB1][A,B,C,D,E,F] \stopchemical

As can be seen from both examples, \chemical is the central command. This command, that can be typed many times within a \start-\stop-pair, gets one or two arguments. These arguments are given between []. The first argument refers to the bonds that are to be drawn. The second argument contains the atoms or molecules that are to be reflected. Text is typesetted in a mathematical mode, so everything that is normally allowed between \$ \$ can be given.

We work out the first example. First of all the keyword SIX is given. By using this word we can indicate that we want to draw a sixring structure. In the same way we can use the keywords ONE, THREE, FOUR, FIVE, NEWMAN and CARBON.

^{1.} The concept structure in this manual only refers to the chemical structure. It is not related to the structure of the text that is used to typeset the formula.

^{2.} Obviously the scope can be confined by using { } and the grouping macros ..group. The setups can also be given immediately after \startchemical. In that case the setups are applied to one formula.



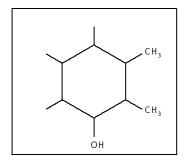
Within these structures chemical bonds between C-atoms can be indicated in a comparable way. For instance, in this example we use B and R. Bonds are numbered and can be indicated in different ways:

\chemical[SIX,B1,B2,B3,B4,B5,B6]
\chemical[SIX,B135]
\chemical[SIX,B1..5]

These commands create parts of a sixring structure. R enables us to add substituents to the sixring structure. The command R draws the beginning of a bond with a substituent from an angular point in the sixring structure (\angle 120°). The concerning angular point is indicated with a number.

\chemical[SIX,B1..6,R1..6]

The above mentioned command only places bonds to substituents. Substituents themselves are indicated with RZ. Therefore in this case, numbers are being used to mark the position. In the second optional argument substituents are given as text.



Example 3

\startchemical[border=on,width=4500] \chemical[SIX,B1..6,R1..6,RZ1..3][CH_3,CH_3,OH] \stopchemical

If the second argument is omitted, no text is placed, so the command RZ1..3 has no effect.

3 Definitions

It is possible to build a library of structures. As we wish, we can recall these structures at a later point of time and provide them with extra components. Furthermore they can serve as building blocks for more complex structures. Structures can be predefined with the TFX-primitive \def.

If a structure, for example [SIX,B,R,RZ], is often used, it is practical to predefine this structure.

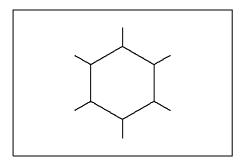
\def\sixring{\chemical[SIX,B,R,RZ]}

Instead of \def the following command can be used. In this case an already existing definition will be announced.

\definechemical[sixring]
{\chemical[SIX,B,R,RZ]}

Although both ways of defining are allowed, the second way is more robust. Protective measures are taken to avoid conflicts with existing commands.

The commands \chemical[sixring] provides a sixring structure without substituents. No second argument is given.

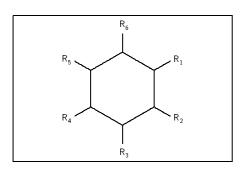


Example 4

\definechemical[sixring]
{\chemical[SIX,B,R,RZ]}

\startchemical[border=on,width=6000] \chemical[sixring] \stopchemical

If we want to add six substituents, we have to carry out the following actions:

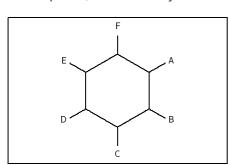


Example 5

\definechemical[sixring]
{\chemical[SIX,B,R,RZ]}

 $\label{lem:condition} $$ \operatorname{[border=on,width=6000]} $$ \operatorname{[sixring]%} $$ [R_1,R_2,R_3,R_4,R_5,R_6] $$ $$ \operatorname{stopchemical} $$$

The sixring structure can also be defined without substituents (RZ). In this case no substituents are expected if the command \chemical is given. Even now substituents can be placed, as is shown by the following example.



Example 6

\definechemical[sixring]
{\chemical[SIX,B,R]}

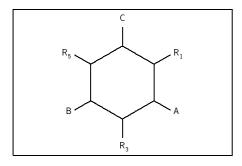
\startchemical[border=on,width=6000] \chemical[sixring,RZ][A,B,C,D,E,F] \stopchemical

Essentially the number of possibilities is unlimited. One should be aware of the fact that the atoms and molecules of the second argument are raised in the sequence of the first argument

In a definition atoms and molecules (texts) can also be placed.

```
\definechemical[sixring]
{\chemical[SIX,B,R,RZ135][R_1,R_3,R_5]}
```

So in this definition always three substituents are added. If we decide to add more substituents, we have to explicitly state that we are dealing with a sixring structure (SIX).



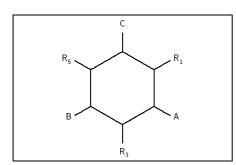
Example 7

\definechemical[sixring]
{\chemical[SIX,B,R,RZ135][R_1,R_3,R_5]}

\startchemical[border=on,width=6000] \chemical[sixring,SIX,RZ246][A,B,C] \stopchemical

So in definitions, the command \chemical has a global character and the command \chemical[][] has a local character. The idea behind this is that in the first case a series of commands is inserted and in the second case a complete independent structure is inserted.

In a definition the command \chemical can occur more than once. The last example can also be recalled with:



Example 8

\define chemical [sixring]
{\chemical [SIX,B,R,RZ135] [R_1,R_3,R_5]
\chemical [SIX,RZ246]}

\startchemical[border=on,width=6000] \chemical[sixring][A,B,C] \stopchemical

If PPCHTEX makes mention of an unknown command, one has probably forgotten to type a structure command, like SIX or FIVE.

4 Bonds

In this chapter we show the bonds that can be found in the different chemical structures. The meaning of the commands will be explained by the reviews that are stated further in this article.

In the left column the complete bonds are shown, in the right column only the shortened bonds. Due to these shortened bonds, atoms and molecules can be attached to a bond. Bonds can be shortened on both sides, left (-) as well as right (+).

В	Bond	SB	Single Bond
		-SB	Left Single Bond
		+SB Right Single Bo	

Table 1: Saturated bonds

A bond can be followed by one or more numbers or a range, for instance: B1, B135 and B1..5. If all bonds are necessary, only B can be given.

In a ring structure an extra bond can be given and furthermore double or triple bonds can be introduced between atoms and molecules.

EB	Extra Bond	DB	Double Bond
		ТВ	Triple Bond

Table 2: Unsaturated bonds

A bond, in a sixring structure for example, can be shortcut. In this case the atom is given that has to be omitted, therefore a circle can be drawn in a sixring structure.

S Shortcut C Circle

Table 3: Special bonds

Substituents can be attached to the angular points. Depending on the presence of atoms and molecules, bonds can be short or long.

R	Radical	SR	Single Radical
-R	Left Radical	-SR	Single Left Radical
+R	Right Radical	+SR	Single Right Radical

Table 4: Bonds to substituents

It is possible to bind substituents to the structure by double bonds.

Text can be linked to bonds. These texts are collected from the second set behind \chemical in the sequence that is given.

The atoms—/—molecules are numbered clockwise. In this case, combinations are allowed. With Z0 (z zero) a text can be placed in the middle of a structure.

FR	Evtra Radical	DR	Double Radical
L11	Extra Radical		Double Radical

Table 5: Double bonds to substituents

Z	Atom	RZ	Radical Atom
		-RZ	Left Radical Atom
		+RZ	Right Radical Atom

Table 6: Atoms and molecules (radicals)

While positioning the atoms and molecules in the text, their (possible) dimensions are taken into account. In this case the width of C and the height of C_m^n play a prominent role. However, this mechanism can still be refined.

5 Combinations

Structures can be combined to complex compounds. Structures can be moved to other structures by using MOV, ROT, ADJ and SUB.

MOV	Move	move the same kind of structure in the
		direction of a bond
ADJ	Adjace	move a different kind of structure in the
		direction of the x - or y -axis, linked to a
		bond
SUB	Substitute	move a structure relative to another one
		in the direction of the x- or y-axis
ROT	Rotate	rotate a structure

Table 7: Displacements and rotations.

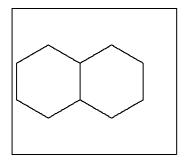
These four commands have another effect within the different structures. For example, the angle used to rotate at \chemical[FIVE,ROT1,B] differs from the angle that is used at \chemical[SIX,ROT1,B].

In addition, within CARBON it is possible to mirror a structure. This can be done with $\mathtt{MIR}.$

NALD	N.A.	
MIK	IVIIrror	mirror a structure

Table 8: Mirroring

The direction of a displacement or the amount of the rotation is indicated by a number. Since these commands are closely related to the actual structure, they must be given before bonds and texts are drawn. It makes a difference whether \chemical[FIVE,B,ROT1,R] is given or \chemical[FIVE,ROT1,B,R]. The first call delivers an unwanted result.

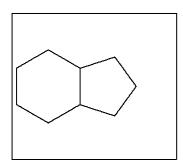


Example 9

\startchemical[border=on,width=4500,right=3500] \chemical[SIX,B,MOV1,B] \stopchemical

Successively a sixring structure is drawn: SIX,B, a displacement is realized in the direction of bond 1: MOV1, and a second sixring structure is drawn: B. A displacement with MOV concerning a sixring structure can be realized in six directions, as opposed to a displacement with ADJ, which is realized in the four axis-directions (x, -x, y, -y). In a sixring structure some of these displacements coincide. The above example also could have been achieved with: [SIX,B,ADJ1,B].

It is also possible to combine different structures. For instance, SIX can be linked to a structure FIVE. The mechanism that is responsible for this linking is for the greater part hidden from the user. In the following example a sixring structure is successively drawn: SIX,B, a displacement along the positive *x*-axis is achieved: ADJ1, and a rotated fivering structure is drawn: FIVE,ROT3,B.

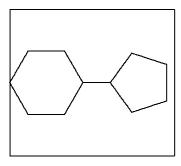


Example 10

\startchemical[border=on,width=4500,right=3500] \chemical[SIX,B,ADJ1,FIVE,ROT3,B] \stopchemical

A transition to a connected structure can be achieved with ADJ. To get a good connection, one of the two structures have to be rotated with ROT. If a structure is not directly linked, but through a bond, one uses SUB. Rotations are made in steps

of 90 degrees, clockwise. Displacements with ADJ and SUB are achieved in the four axis-directions.



Example 11

\startchemical[border=on,width=4500,right=3500] \chemical[SIX,ROT2,B,R6,SUB1,FIVE,B,R4] \stopchemical

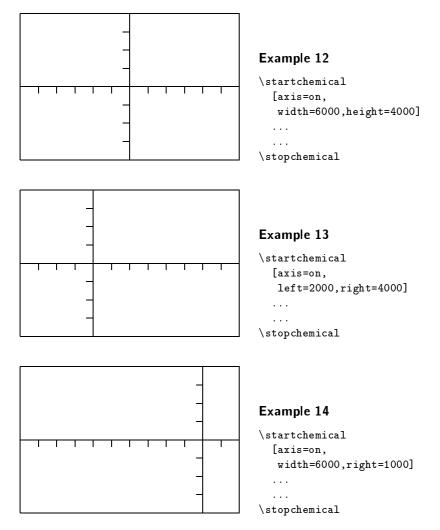
We can therefore conclude that the sequence of the given commands is very important. An obvious sequence of commands is as follows:

As a rule, the connection of structures is reduced to some translations and rotations. Although it may not seem so, a certain systematic is enclosed. In fact, the process could be simplified. The automation that was already achieved in former versions, has been undone: it turned out that 'hidden' rotations induce misunderstandings with regard to the place of bonds. Furthermore, it is easier to provide a structure that is not rotated with bonds, atoms and molecules than to provide a rotated structure. It is better to define the parts of a complex structure first, possibly with translations, and to rotate the complete complex structure later.

6 Axis

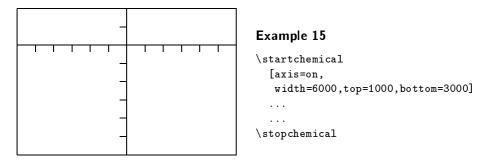
Structures are typeset in a bounded space, for convenience indicated by axis. The dimensions of the axis and the location of the origin can be defined in the setup. In addition

the axis can be made visible (for the sake of the location in the text) and a border can be drawn.



The dimensions of the structure are determined by the dimensions of the axis. However, if width=fit+ and/or height=fit is given, the dimensions of the total structure are determined by the real dimensions. Whatever is chosen is depending on the way structures are placed in the text. Side by side, on top of each other, etc. Example 12 shows the standard setups.

Within a start - stop -pair $\mathsf{P}_{\mathsf{CTE}}\mathsf{X}$ -macros can be used. Of course, some caution must be taken into account.



7 Setups

The behavior of the macro's and the layout of the formulas can be adapted to personal needs. After both \startchemical and \setupchemical setups can be given.

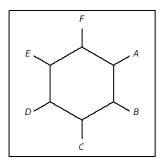
variable	values	default
width	number	4000
height	number	4000
left	number	
right	number	
top	number	
bottom	number	
resolution	number	see manual
corps	8pt 9pt 10pt etc.	see manual
style	\rm \bf etc.	see manual
scale	number small medium big	medium
size	small medium big	medium
status	start stop	start
option	test	
axis	on off	off
border	on off	off
alternative	1 2	1
offset	HIGH LOW	LOW

Table 9: Setups for structures

The axis reaches from -2000 till +2000, in height as well as in width. The point Z0 is situated at (0,0). Other setups can be set up with left, right, above and/or under in combination with width and height.

The dimensions of the characters can be set up with size. In doing so, the TEX-primitives \textsize, \scriptsize and \scriptscriptsize are used. The dimensions of the structure itself can be set up with scale (1..1000). The scale is also determined by corps. The keywords small, medium and big are attuned.

In the mathematical mode commands like \mbox{rm} , \mbox{bf} and \mbox{sl} can be utilized in TEX as well as in ConTEXt. PPCHTEX by default uses \mbox{rm} . An alternative command can be set up with style. For instance in example 16 the substituents are typeset slanted.



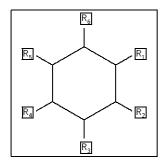
Example 16

\startchemical[border=on,style=\s1] \chemical[SIX,B,R,RZ][A,B,C,D,E,F] \stopchemical

For the time being style is valid for chemical formulas in the text as well as in a figure. The sub- and superscripts also change, as can be seen from: CH_4 , CH_4 and CH_4 . Successively the setups are: \rm , \bf and \sl . When formulas are typeset in italic (\tl), baseline-distances can be greater than normal. Within ConTeXt bold-slanted (\bf) and bold-italic (\tl) are also available. All these commands automatically adjust to the actual style (\sl s, \rm , \tl t): CH_4 , CH_4 etc.

The time-consuming calculations can be short-circuited with status. The variables border and axis speak for themselves. A border round a text can be drawn with option=test. By doing so, one can see how things are aligned. The quality of the lines is set up with alternative. By default, PICTEX uses a 5 point period to draw the lines. As one chooses alternative 2 smaller points are used. Therefore thinner lines are drawn.

The offset refers to the position of the sub- and superscripts. By using HIGH the subscripts are positioned high (H_2O) . Self-evident, with LOW the subscripts are positioned a little lower (H_2O) .

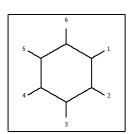


Example 17

 $\label{lem:cal_border} $$ \operatorname{LSIX_B_R_RZ}[R_1,R_2,R_3,R_4,R_5,R_6] $$ \operatorname{LSIX_B_R_RZ}[R_1,R_2,R_3,R_4,R_5,R_6] $$$

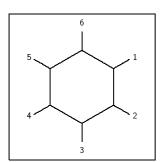
8 Dimensions

A structure can be shown in different sizes. Dimensions can be set up with size and scale.



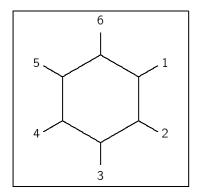
Example 18

\startchemical[scale=small,size=small] \chemical[SIX,B,R,RZ][1,2,3,4,5,6] \stopchemical



Example 19

\startchemical[scale=medium, size=medium]
\chemical[SIX,B,R,RZ][1,2,3,4,5,6]
\stopchemical



Example 20

\startchemical[scale=big,size=big]
\chemical[SIX,B,R,RZ][1,2,3,4,5,6]
\stopchemical

Scaling can take place between 1 and 1000. The values belonging to the keywords small, medium and big are attuned.

9 Symbols

To typeset reaction equations some symbols are available. In the following figure an equation is shown. This equation is defined as follows:

```
\setupchemical
  [width=fit,
  height=5500,
   under=1500]
\hbox
  {\startchemical
     \chemical[SIX,B,ER6,RZ6][0]
   \stopchemical
   \startchemical
     \chemical[SPACE, PLUS, SPACE]
   \stopchemical
   \startchemical
      \chemical [FIVE, ROT4, B125, +SB3, -SB4, Z4, SR4, RZ4] [N, H]
   \stopchemical
   \startchemical
      \chemical[SPACE,GIVES,SPACE][?]
   \stopchemical
   \startchemical
      \chemical [SIX,B,EB6,R6,SUB4,FIVE,ROT4,B125,+SB3,-SB4,Z4] [N]
   \stopchemical}
```

The command \hbox is necessary to position the structures behind each other. The symbols GIVES and PLUS speak for themselves. Extra space can be achieved with SPACE.

An equilibrium can be shown with EQUILIBRIUM. A text can be set above GIVES and EQUILIBRIUM. In the example this text is a question mark.

10 Special features

By using ONE, ZO can consist of more than one atom. In this case, the reserved space is insufficient. If more space is needed for the command ZO, the bonds 1, 2 and 8 can be moved with the command OFF, which means 'offset'. Below, an example of the use of this command is given.

```
\startchemical[width=fit]
\chemical
[SIX,B,C,ADJ1,
FIVE,ROT3,SB34,+SB2,-SB5,Z345,DR35,SR4,CRZ35,SUB1,
ONE,OFF1,SB258,Z0,Z28]
[C,N,C,0,0,
CH,COOC_2H_5,COOC_2H_5]
\stopchemical
```

In this case the offset is 1, which means that we use one extra character. Similar, at first sight quite complicated, definitions can best be constructed by defining the separate parts first. The rotation can best be made last.

$$\bigcap_{C} \bigvee_{N} - CH \bigvee_{COOC_{2}H_{5}}$$

A new command is seen in the above example: CRZ. This command can be used to position an atom or a molecule in line with a bond, which is desirable in this case. This positioning could have been achieved with the command RZ, since one can influence the spacing with the second set: {\,0} instead of {0}.

	OFF	Offset	CRZ	Centered Radical Atom
--	-----	--------	-----	-----------------------

Table 10: Special commands

11 Textformulas

Along with the typesetting of structure formulas, the typesetting of reaction equations is supported too. Therefore, the former described command \chemical has another two versions:

```
\chemical{formula}
\chemical{formula}{text}
```

This command will fit to the position in the text. This means that a difference is made between:

- a text-mode
- a mathematical text-mode
- a mathematical display-mode

If the command is given in the current text, automatically \$ \$ are set round the command. In this way the command $\hline \hline \hlin$

The same result is achieved by placing the command between \$ \$. Therefore in both cases the second argument is omitted. If we type the command between \$\$ \$\$, the second argument is obliged. The second argument may be empty:

```
\ \chemical{2H_2}{} \chemical{PLUS}{} \chemical{0_2}{} \chemical{GIVES}{} \chemical{2H_20}{} $$
```

gives:

$$2H_2 + O_2 \longrightarrow 2H_2O$$

This formula can also be defined in a shorter way:

or even:

$$\$$
 \chemical{2H_2,+,0_2,->,2H_20}{}\$\$

The more experienced TEX-user will notice that the plus-sign as well as the arrow are located at the base-line. If you just compare + and +. Independent of the size of reflection, the + and the \longrightarrow are aligned.

In addition to PLUS and GIVES, EQUILIBRIUM can be given, which delivers double arrows (<->).

This formula can also be typeset in the current text, where a smaller layout is chosen: $2H_2 + O_2 \longrightarrow 2H_2O$. It is also possible to show bonds. For example, the rather long sequence \chemical{H,SINGLE,CH,DOUBLE,HC,SINGLE,H} gives H - CH = HC - H. This can also be achieved in a shorter way: \chemical{H,-,CH,--,HC,-,H}. A triple bond can be called with TRIPLE or --: HC = CH.

Now we just return to the display-mode. The second argument can be used to give an explanation of the formula:

\$\$

 $\label{lem:cal} $$ \operatorname{2H_2}{ \operatorname{oxygen} \subset {PLUS}_{ \subset {GIVES}_{violent} \subset {2H_20}_{ }} $$ $$ \operatorname{CHE}_{0,2}(oxygen) $$ $$ $$ $$ $$ oxygen} $$ $$ $$ oxygen} $$$

\$\$

gives:

$$2H_2 + O_2 \xrightarrow{\text{violent}} 2H_2O$$

The size of the formulas that are placed in the text can be set up with the setupcommand:

variable	values	default
textsize	small medium big	big

Table 11: Setups for textformulas

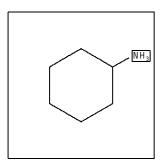
```
When big, medium or small is used, \chemical{H,SINGLE,CH,DOUBLE,HC,SINGLE,H} gives the successive formulas:

H—CH=HC—H H—CH=HC—H H—CH=HC—H
```

12 Subscripts

Sub- and superscripts are placed a little lower, as is recommended by Knuth in the TeXbook. Like this, the example on page 179 in the TeXbook, that is chemically a bit strange, can be typeset by typing $\chemical{Fe_2^{+2}Cr_20_4}$. This command gives $Fe_2^{+2}Cr_2O_4$. Without a correction the command would have given the result: $Fe_2^{+2}Cr_2O_4$.

The location of the subscript is by default determined by the offset: HIGH or LOW. The offset can be overruled by the commands of the same name.



Example 21

\startchemical[border=on,option=test,alternative=2] \chemical[SIX,B,R1,HIGH,RZ1][NH_3] \stopchemical

Although the location of the subscript can be set up by substituent, it is advisable to do this with \setupchemical. By doing so, all the formulas will be set up in the same way.

The keywords LOW and HIGH can also be used in text formulas, although this may lead to inconsistencies in layout. We can see that $\chemical{HIGH,H_20}{}$ results in H_2O and $\chemical{LOW,H_20}{}$ in H_2O .

13 Installation

The package PPCHTEX is developed as an extension to ConTEXt but can be used with other packages. The macros can be found in the file m-chemie.tex, with the m standing for module. Therefore the macros are not standard available.

Furthermore, the package can be used in combination with LATEX. In this case the file m-chemie.sty is used too. PPCHTEX can be activated by means of \documentstyle:

\documentstyle[m-chemic]{}

^{3.} The dutch word chemie stands for chemics in english.

In addition to the english interface that was described earlier, a dutch interface is available. The dutch version is loaded with:

\documentstyle[m-chemie]{}

14 Some dutch

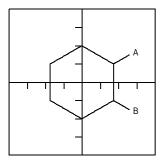
The original interface of PPCHT_FX is dutch. One can switch back to dutch with:

\resetinterface

Switching back to english can be done with:

\setinterface[english]

Although more interfaces are possible, only the dutch and english are implemented. The file m-chemie.sty shows how an interface is defined.



Example 22

\startchemie [assenstelsel=aan,kader=aan] \chemie [SIX,B,R12,RZ12] [A,B] \stopchemie

15 Extensions

Of course, all TeX-users are allowed to use the PPCHTeX-macros in another, non-commercial, way. However, some caution has to be taken into account, since the macros are still being developed, optimized and made more robust. Some macros may seem quite complicated, but appearances are deceptive. For example, commands in the shape of \setup... utilize the macros that support nesting and different ASCII-layouts. Compare for instance:

\setupchemical[size=small]

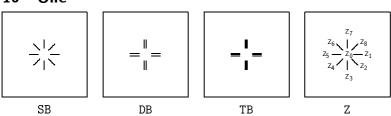
with:

```
\setupchemical
  [size=small,
  scale=500,
  textsize=big]
```

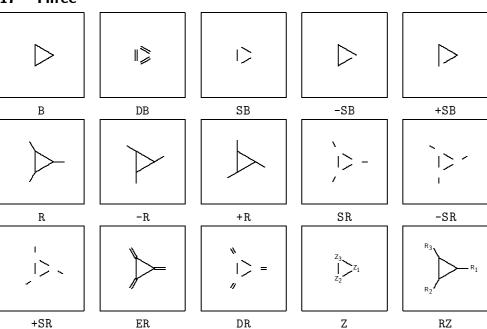
Setups can be given in an arbitrary sequence. When possible, spaces and newline-characters are suppressed and errors are announced.

The following overview shows the available structures. We use a somewhat smaller corps to save space. The manual of PPCHTEX is a bit more extensive in its overviews.

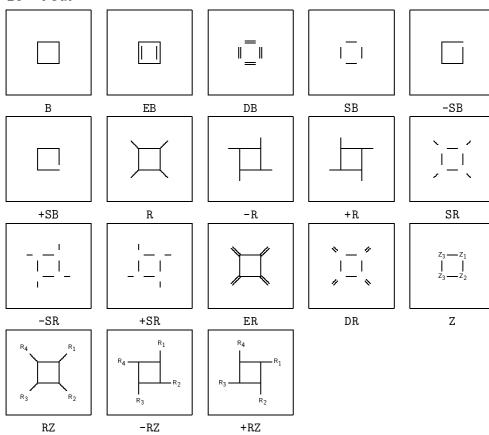
16 One



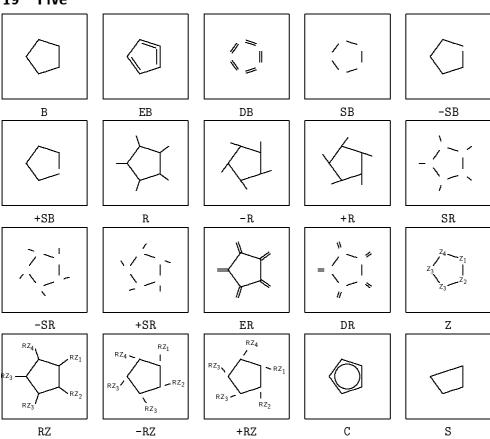
17 Three



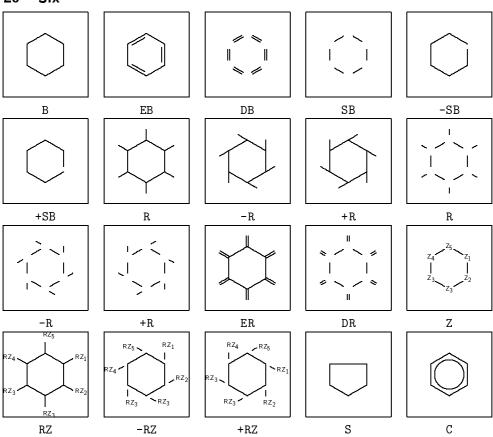
18 Four



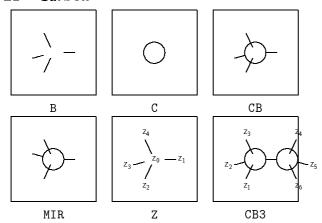
19 Five



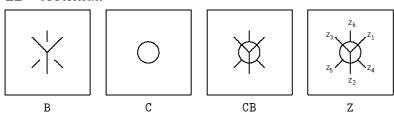
20 Six



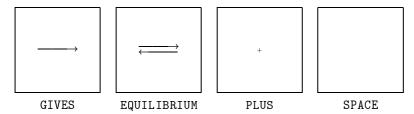
21 Carbon



22 Newman



23 Symbol



24 Other features

The following features of $\mbox{PPCHT}_{\mbox{E}}\mbox{X}$ are not described here, but one can find more on them in the manual.

- coloring parts of formulas
- active formulas in interactive texts

- automatic adapting to the corps size
- adapting to the resolution of printers
- the multi-lingual interface
- manipulating super- en subscripts

Also, the manual is a bit more extensive in its overview of the rotations and displacements.