Typesetting Chemical Equations using LaTeX

Contents

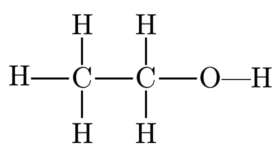
Chapter 1

$\TeX$/$\LaTeX$ code for components of organic chemical structure diagrams

1.1 Conventions for drawing the diagrams

The chemical structure of a molecule is defined by the spatial arrangement of the atoms and the bonding between them. Chemists use several standard methods for representing the structures two-dimensionally by diagrams called structural formulas; and this thesis will develop mechanisms for printing such diagrams using the $\TeX$/$\LaTeX$ system.

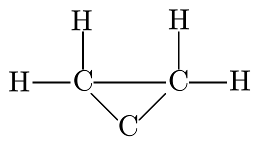
A very common structure representation, sometimes called a dash structural formula, uses the element symbols for the atoms and a dash for each covalent bond in the compound. Thus the dash represents the pair of shared electrons that constitutes the bond. Two dashes (=) represent a double bond and three dashes (≡) a triple bond. — It is usually neither necessary nor practical to represent each bond in a molecule explicitly by a dash. Some molecules and some bonds are so common that a complete dash formula would not be used except at a very introductory level of presenting chemical information. A condensed structural formula is one alternative. It does not contain dashes but uses the convention that atoms bonded to a carbon are written immediately after that carbon and otherwise atoms are written from left to right in the order in which they occur in the real structure. The following two structural formulas are a dash formula and a condensed formula, respectively, for the same compound, ethanol.

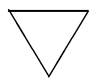


CH_3CH_2OH

Multiple bonds are usually not implied unless a very common group, such as the cyano group, is shown. It can be found as $-C{\equiv}N$ or simply as $-CN$.

Another alternative to a complete dash formula is a diagram where the symbols for carbon and for hydrogen on carbon are not shown. Each corner and each open-ended bond in these diagrams implies a carbon atom with as many hydrogen atoms bonded to it as there are free valences. This representation is the customary one for ring structures (structures with a closed chain of atoms). Thus, the following two diagrams both represent the compound cyclopropane.





The three different kinds of structure representation can be combined in one diagram, such that in part of the diagram all bonds are represented by dashes and all atoms by an element symbol, in another part a condensed structural formula fragment is used, and in still another part a cyclic fragment with implied carbon and hydrogen atoms occurs.

The rest of this chapter describes how LaTeX can be used to position and typeset the bond lines and condensed formula strings that are the components of structure diagrams.

It should be mentioned that there are no binding rules for many aspects of the two-dimensional representations of a chemical structure. Structures and fragments of structures can be oriented in different ways depending on the availability of space, the emphasis given to a certain part of a structure, or the spatial relationship of the parts to each other. Thus, a cyclopropane ring can be represented in various orientations, $\triangle$, $\bigtriangledown$, and others. Also, the angles between the bond lines can be different in different representations of one and the same compound. Since most molecules do not have all their atoms lying in one plane it would not even be possible to reproduce all bond angles in a two-dimensional representation. The structures shown in this thesis adopt the orientations and bond angles found to prevail in Solomons' textbook (Solomons 84), the organic chemistry text used for several years at the University of Tennessee.

There are some methods to indicate the real, three-dimensional structure (the stereochemistry) of a molecule in the two-dimensional representation: A dashed line and a wedge instead of a full bond line mean that the real bond extends below or above the plane, respectively.

1.2 Bond line drawing and positioning

1.2.1 Review of TeX/LaTeX facilities for line-drawing

The easiest way to produce horizontal and vertical lines representing chemical bonds is by the use of keyboard characters and simple control sequences provided by TeX. By typing one, two, or three hyphens, a normal hyphen, a medium dash designed for number ranges, and a punctuation dash are produced, - – —, respectively. When a hyphen is typed in TeX's math mode, it is interpreted as a minus sign and the spacing around it will be different from text mode. — The equal sign can represent a double bond for chemistry typesetting. It can be typed in text mode and in math mode, again resulting in different spacing around the symbol. — The control sequence `\equiv` can be used as a triple bond ($\equiv$). It has to be typed in math mode.

Vertical lines are available through the keyboard character or the control sequences `\vert` and `\mid`, all three to be entered in math mode. A double vertical bar is produced by `\|` or `\Vert`, again both in math mode.

The spacing around all these symbols can be controlled by adding extra (positive or negative) space with the horizontal spacing commands. The symbols, just as any other part of a line, can also be raised or lowered respective to the normal baseline. The length and height of the symbols however depend on the font currently in use.

Where control of length and height of the bond lines is needed, TeX's or LaTeX's command sequences for printing horizontal and vertical "rules" can be used. The systems recognize several length units, including the inch, centimeter, millimeter, and printer point (Knuth 84, p. 57). One printer point (pt), an often used unit in typesetting, measures about 0.35 mm. — LaTeX's rule-printing command has the format

```
\rule[raise-length]{width}{height}
```

Thus it can be used to produce horizontal and vertical rules. Using the `\rule` command one can also print multiple bond lines of user-controlled length, e. g. $=$, $\equiv$, with the short control sequences `\dbond` and `\tbond` defined in this thesis. The vertical spacing between the bonds depends on the current line spacing in the document and may have to be adjusted. The control sequences are set up for math mode.

When bond lines other than horizontal and vertical ones are to be printed, and when a coordinate system is needed to control placement of structure components relative to one another, LaTeX's picture environment (Lamport 86, pp. 101–111) is a necessity.

A picture environment uses length units which are dimensionless and have to be defined by the user before entering the environment. This is done by the `\setlength` command. In this study, `\setlength{\unitlength}{0.1pt}` is the definition used for most diagrams. Such a small unitlength was chosen to have fine control over the appearance of the diagram.

The picture environment starts with the statement

```
\begin{picture}(width,height)
```

where picture width and height reserve space on the page and are specified in terms of unitlengths. Optionally, one can include the coordinates of the lower left corner of the picture:

$$\texttt{\textbackslash begin\{picture\}(width,height)(}x_i\texttt{,}y_i\texttt{)}$$

The default value for these coordinates is (0,0). Objects are placed into the picture with the `\put` command with their reference point at the coordinates (x,y): `\put(x,y){picture object}`.

The picture objects of most interest to this study are straight lines. They are drawn by the `\line` command:

$$\texttt{\textbackslash line(}x_s\texttt{,}y_s\texttt{)\{length\}}$$

where the coordinate pair specifies the slope of the line, and the nonnegative value of length specifies the length of the projection of the line on the x-axis for all nonvertical lines, and the length of the line for vertical lines. The reference point of a line is one of its ends. Thus the statement

$$\texttt{\textbackslash put(x,y) \{\textbackslash line(}x_s\texttt{,}y_s\texttt{)\{len\}\}}$$

draws a line that begins at (x,y), has a slope of y_s/x_s, and extends for length len as explained above.

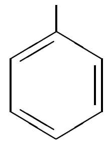
Only a limited number of slopes is available through the line fonts in LaTeX. The possible values for x_s and y_s are integers between -6 and $+6$, inclusive. These values translate into 25 different absolute angle values, which are listed in Appendix A.

a $CH \equiv C - CH = CH_2$

b $CH{\equiv}C - CH = CH_2$

c $CH \equiv C — CH = CH_2$

Figure 1.1: One-line structural formulas



$$CH_3 - CH - CH_2 - CH_2 - CH_2 - CH_3$$

Figure 1.2: One-line structure in picture environment

1.2.2 Bonds in structural formulas written on one line

The application of some of the bond-drawing mechanisms for this simplest type of structural diagrams is illustrated in Figure 1.1.

For Figure 1.1a only keyboard characters and the TEX command `\equiv` were used to produce the bonds. Figure 1.1b shows a slight improvement through raising the triple bond. Figure 1.1c was printed using the `\sbond`, `\dbond`, and `\tbond` command sequences from this thesis, choosing a length of 14pt for the bonds. It can be seen that each of the formulas in Figure 1.1 is a creditable representation of the structure. Depending on the design of the page, the reason for displaying the structure at a particular place, and the emphasis put on features of the structure in the text, one would choose shorter or longer bonds and take more or less trouble to produce the structure.

The picture environment is not needed for one-line structural formulas, unless one of these formulas has to be attached to another structural fragment, as in Figure 1.2 Then the coordinate system of the picture environment makes it possible to fit the two fragments together.

1.2.3 Bonds in acyclic structures with vertical branches

Structure diagrams with vertical, single- or double-bonded, branches, going up or down, are frequently seen. Several experiments with TEX and LATEX were made to see how this type of structure can be handled. One method is to align the vertical bonds by using the mechanisms for tabbing or for printing tables and matrices. Here a structure such as the one shown in Figure 1.3 is treated as a set of columns as indicated by the vertical dividing lines drawn into the second version of this structure in Figure 1.3.

The structure diagram in Figure 1.3 uses `\vert` for the vertical bonds and LATEX's tabbing environment for the alignment. One can also use "rules" as the vertical bonds in order to give the horizontal and vertical bonds the same lengths. Furthermore, vertical bonds can also be double

$$\begin{array}{c}
\quad Br \qquad\qquad CH_3 \\
\quad | \qquad\qquad\quad | \\
CH_3CH_2CHCHCH_2CHCH_2CH_3 \\
| \\
CH_2CH_3
\end{array}
\qquad\qquad
\begin{array}{l}
\qquad\qquad\qquad |Br\ | \qquad\qquad |CH_3 \\
\qquad\qquad\qquad ||\ \ | \qquad\qquad\ \ |\ | \\
CH_3CH_2|CH|CHCH_2|CHCH_2CH_3 \\
\qquad\qquad\qquad | \qquad ||\qquad\qquad | \\
\qquad\qquad\qquad | \qquad |CH_2CH_3
\end{array}$$

Figure 1.3: Vertical branches

bonds. The following examples illustrate these features.

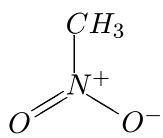
$$\begin{array}{c} O \\ \| \\ H_2N{-}C - NH_2 \end{array}
\qquad
\begin{array}{c} CH_3 - CH_2{-}C - CH_3 \\ \| \\ NH \end{array}
\qquad
\begin{array}{c} H{-}C = C - H \\ |\quad\ \ | \\ Br\ \ \ Br \end{array}$$

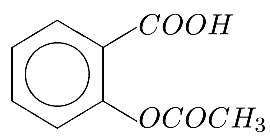
Similar structures were also generated with TEX's `\halign` mechanism which forms templates for the columns rather than setting tab stops. For the purpose of printing the structure diagrams, no clearcut advantage was seen in one or the other method of alignment. In each case the vertical spacing depends on the line spacing in the document.

The alternative method of producing these structures is the use of the picture environment. It provides better control over horizontal and vertical spacing and over bond lengths. Also, as illustrated in Section 1.2.2, using a picture environment makes it possible to attach one structural fragment to another at a specific place. Thus, although the picture environment is not necessary for drawing structures with vertical branches, it has several advantages, and writing LATEX code for this implementation is not more difficult than writing the code for the tabbing method of alignment.

1.2.4 Bonds in Structures Containing Slanted Bond Lines

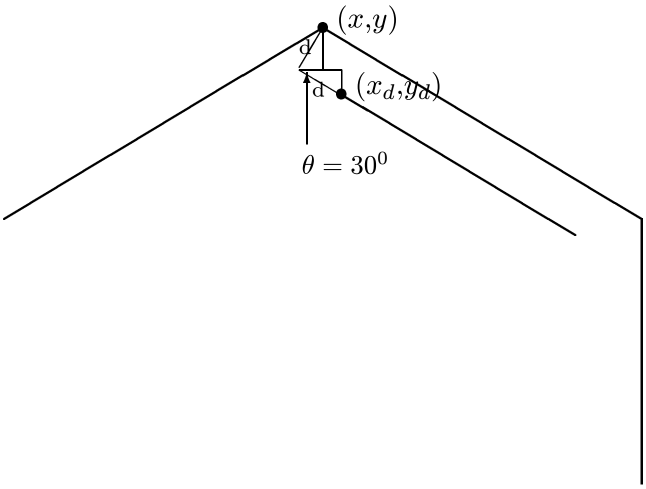
Structure diagrams with slanted bond lines are frequently used for acyclic compounds and have to be used to depict almost all cyclic structures. Two examples are shown here:





In developing diagrams for such structures in this thesis the conventions described in Section 1.1 are followed. Thus the symbol for carbon is not printed for the carbons that are ring members, but it is usually printed in acyclic structures, unless the acyclic structure fragment is a long chain, or space for the diagram is limited.

The picture environment is always needed for slanted lines. It was explained in Section 1.2.1 that LATEX can draw lines only with a finite number of slopes. This is not a severe limitation for creating the structure diagrams, since the conventions for structure representation allow variations in the angles. The representation does not have to reflect the true atomic coordinates. In fact many

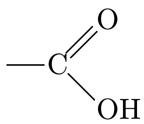


$$x_d = x - d\sin\theta + d\cos\theta$$

$$y_d = y - d\sin\theta - d\cos\theta$$

Figure 1.4: Calculating position and length of double bond.

chemistry publications contain structure diagrams with angles significantly deviating from the real bond angles, even where those could have been used easily. Thus, Solomons' text (Solomons 84) shows the carboxylic acid group often in this form



with an angle of about 90○ between the OH and doublebonded O, whereas the true angle is close to 120○. — The angles used in this thesis for the regular hexagon of the sixring deviate by $\pm 1^\circ$ from 120° because of LaTeX's limited number of slopes. This difference is not big enough to be detected as a flaw.

To write the LaTeX statement for a slanted bond line, one chooses the origin and the slope and then uses trigonometric functions to calculate the LaTeX "length" of the line for the desired real length. Once the LaTeX length is determined, the coordinates of the end point of the line can be calculated in case the end point is needed as the origin of a connecting line. — The origin and length of slanted double bonds were also calculated with standard methods from trigonometry. As an example, Figure 1.4 shows how coordinates of the origin were calculated for the inside part of a ring double bond that is at a distance d from the outside bond.

The LaTeX command `\multiput` is similar to `\put` and provides a shortcut for the coding of structures where several bond lines of the same slope and length occur at regular intervals. Multiput has the format

```
multiput(x,y)(xΔ,Δy){n} {object}
```

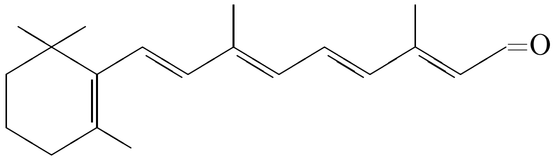


Figure 1.5: Diagram using \multiput

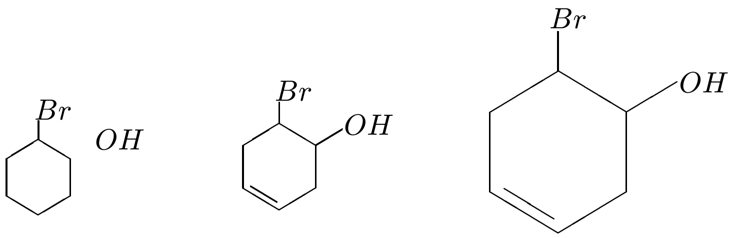


Figure 1.6: Scaling (unitlength=0.07pt, 0.08pt, 0.15pt)

where n is the number of objects, *e.g.,* lines. A structure diagram for which several `\multiput` statements are appropriate is the structure of vitamin A shown in Figure 1.5

The size of objects in a picture environment can be scaled in a simple way by changing the unitlength. Figure 1.6 illustrates scaling and two problems associated with it. Changing the unitlength changes the length of the lines only, not the width of the lines or the size of text characters. Thus "it does not provide true magnification and reduction" (Lamport 86, p. 102). However, the size of the text characters can be varied separately, as will be discussed in the next section of this chapter.

The smallest diagram in Figure 1.6 illustrates a limitation that is unfortunate for the printing of structure diagrams. The shortest slanted line that can be printed by LaTeX's line fonts is one with an x-axis projection of about 3.6mm. If a shorter slanted line is requested, LaTeX just prints nothing. A chemist would occasionally want to draw shorter lines, especially for the purpose of generating dashed lines indicating stereochemical features.

1.3 Atomic symbols and condensed structural fragments

Special considerations for the printing of condensed structural fragments are required since many of them contain subscripts. TeX considers the printing of subscripts a part of mathematics typesetting which has to be done in the special math mode. As it is known, typesetting of mathematics documents is one of the strong points of TeX; the fonts of type for the math mode are designed to agree with all conventions of high quality mathematics publishing. Each typestyle in math mode consists of a family of three fonts (Knuth 84, p. 153), a textfont for normal symbols, a scriptfont for first-level sub- and superscripts, and a scriptscriptfont for higher-level sub- and superscripts. When structural fragments such as C_2H_5 are typeset, the textfont is used for the C and the H.

As TEX enters math mode it selects `\textfont1` as the textfont unless otherwise instructed. `\textfont1` is defined by the TEX macros as math italic, a typestyle that prints letters (not numbers) similar to the italic style, but with certain features adapted for mathematics typesetting. The italic style letters, lower and upper case, are the ones commonly seen in typeset mathematical formulas. Chemical formulas on the other hand are not usually printed with slanted letters. In this thesis, two methods were employed to produce chemistry-style letters in TEX's math mode which has to be used because of the presence of subscripts.

For a document that contains many chemical formulas it is convenient to redefine `\textfont1` at the beginning of the TEX input file. The statement `\textfont1=\tenrm` was used at the beginning of the input file that produced this document and causes TEX to select the roman font as the textfont in math mode. The roman typestyle is the one normally used by TEX outside of math mode and it is the style in which this thesis is printed. The ten point size, which is slightly smaller than the eleven point size of the text in this document, was chosen because it appears to look better for the chemical formulas which consist largely of capital letters. When different typesizes are used in this way, all the atomic symbols and formulas in any one structure, even those without subscripts, have to be printed in math mode so that they all have the same size. It could be a problem with this method of selecting the roman font for math mode that the lowercase Greek letters (and some other symbols used in mathematics) are not available in this font. To print these one can temporarily redefine textfont1 to math italic with the statement `\textfont1=\tenmi`. One can also switch to a math font different from the default `\textfont1`. Using one of LATEX's font definitions, `\small`, a statement `{\small$\theta$}` will print the Greek letter.

Another method for avoiding the math italic style for letters in chemical formulas is to select the roman style in each individual instance where a formula has to be printed in math mode. A statement such as `${\rm C_2H_5}$` produces $\mathrm{C_2H_5}$ at the size of type currently used in the document. When the typestyle is thus selected within math mode, enclosed by dollar signs, TEX changes the style of the letters of the alphabet only; the lowercase Greek letters and math symbols remain available.

The size of the letters in chemical formulas can be changed with the ten size declarations provided by LATEX (Lamport 86, p. 200) or with TEX's declarations. (Some of TEX's declarations are not defined in LATEX (Lamport 86, p. 205)). The size declaration has to be written outside of math mode. One place in chemistry typesetting where a smaller typesize is desirable is the writing on reaction arrows. The size in the following example is scriptsize:

$$HC \equiv CH + H_2O \xrightarrow[\text{18\% H}_2\text{SO}_4,\ 90^\circ]{\text{Hg}^{++}} CH_3 - CHO$$

Finally, condensed structural formulas sometimes have to be right-justified to be attached to the main structural diagram. Figure 1.7 illustrates this for the positioning of the substituent in the 4-position of the pyrazole ring. LATEX makes this positioning convenient with the `\makebox` command, especially in the picture environment where the command has the format

```
\makebox(width,height)[alignment]{content}
```

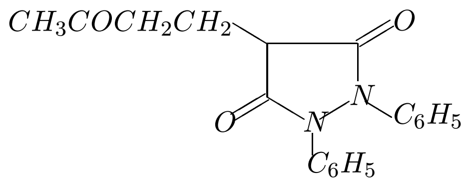


Figure 1.7: Right-justification of substituent formula

(Lamport 86, p. 104). The one-line piece of text that constitutes the content of the (imaginary) box can be aligned with the top, bottom, left side, or right side of the box.

Chapter 2

Macros for typesetting chemical structure fragments

2.1 General organization of a structure macro

The macro facility in TEX/LATEX was used to define mnemonics for typesetting frequently occurring structure fragments such as common ring structures and branching patterns. Chapter 3 describes the complete system of macros designed for this thesis. All of the macros are defined with the LATEX declaration `\newcommand` which has the format

```
\newcommand{\commandname}[n]{replacement text}
```

(TEX calls these definitions macros, whereas LATEX just uses the more general word “command.”) In the definition, n is an integer from 1 to 9 and gives the number of arguments if any are used. The arguments are represented in the replacement text by parameters of the form #1, #2 etc.

A macroname, like the name of any control sequence in TEX, can contain letters only, not numerals. Where it was considered important in this thesis to indicate a numbering scheme in a macroname, either the full word for the number or Roman numerals in lower case letters were used. Thus `\hetisix` is a mnemonic for a hetero sixring with one hetero atom, and `\fuseiv` indicates a fusing fragment with four atoms. — In some macronames the mnemonic as such is preceded by “c” or “chem,” for “chemistry.” This was mainly done where the mnemonic was already used for a control sequence in TEX or LATEX. Thus, chemical structure fragments that point “right” or “left” are drawn by the macros `\cright` and `\cleft` since TEX employs control sequences `\right` and `\left` with different meanings.

All the structure-drawing macro definitions in this thesis follow the same pattern in their organization. The actual structure-drawing LATEX code is preceded by a “box” declaration. The diagrams are produced in a box because it is sometimes desirable to move the diagram as a whole. In all but one of the macros, the box is the LATEX “picture”; in the macro `\tbranch`, which uses the tabbing environment, the box is the LATEX minipage. Box dimensions in the macros are made flexible through the use of global variables. — The structure-drawing code itself consists of unconditional and of conditional statements. The macro arguments are used to vary parts of the structure diagram, such as substituents and multiple bonds. Those features of a structure macro that have not been discussed before are described in more detail in the following sections.

2.1.1 Box constructions with global variables

Integer variables such as the ones used here for box dimensions have to be stored in one of TEX's 256 numerical registers and can be given symbolic names with TEX's `\newcount` declaration (Knuth 84, pp. 118–121). The variables used in the macros of this thesis are defined and initialized in the macro `\initial` which should be part of the preamble of an input file containing chemistry typesetting using this system (see Figure 3.1 for a summary of the preamble). A user can then change the variables by simple assignment, e. g. `\xi=400`. When several variables have been changed, it is convenient to reset all of them, including the unitlength, to their initial values with the macro `\reinit`.

The role of the LATEX picture for line-drawing was discussed in Chapter 1, but the picture is also a box. As such it is processed in horizontal mode, as part of a line. Within a horizontal box, line breaks can never occur. The macros use variables for all numerical parameters in the picture declaration. Thus, the picture declaration in the macros has the form

```
\begin{picture}(\pw,\pht)(\-xi,\-yi)
```

The picture width and height, `\pw` and `\pht`, specify the nominal size used by TEX to determine how much room to leave for the box. The diagram in the box can extend beyond these dimensions, but an adjoining box is typeset next to the preceding one according to the specified width. The user needs control over the picture width in cases where several such boxes are put on one line, especially for the horizontal connection of structure fragments. Control over the picture height is important because some chemical structures take up more vertical space than others. The picture width and height are initialized to 400 and 900 respectively, which is about 1.4×3.2cm with the unitlength of 0.1points.

Variables are used for the coordinates of the lower left corner, `\xi` and `\yi`, so that the user can change the placement of the diagram within the picture window. This is not often necessary for individual structures since they can be conveniently centered by the display mechanisms discussed later in this chapter; and also the whole picture can be shifted horizontally by adding horizontal space in front of it with the `\hspace` command. It was considered to be most convenient to put the minus signs in front of `\xi` and `\yi` in the declaration, since one thinks of the lower left corner of a coordinate system as having negative coordinates. With this declaration, an increase in the absolute `\xi` and `\yi` values shifts the diagram to the right and up. The coordinates `\xi` and `\yi` are initialized to 0 and 300 respectively, which places the coordinate origin about 1cm above the bottom of the picture window with the unitlength of 0.1points.

The minipage, the box used in the macro `\tbranch`, is a paragraph box, which allows line breaks. Only the width is specified for a paragraph box since the height is controlled by the number of lines that will be produced by a given amount of text. The variable used in this system for the width of paragraph boxes is `\xbox`. The number value assigned to `\xbox` is interpreted as printer points.

2.1.2 Use of TEX's conditional facility

TEX's conditional facility is very similar to those of other high-level languages; it has the form:

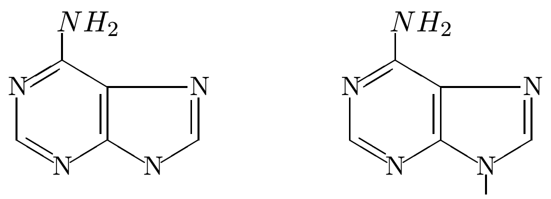
`\if`⟨condition⟩⟨true text⟩ `\else`⟨false text⟩`\fi`

(Knuth 84, p. 207ff.). Nesting is possible. The TEX `\if` primitive has over ten different forms for testing numbers, processing modes, or tokens. The form used in the chemical structure macros is `\ifx`⟨token$_1$⟩⟨token$_2$⟩, which tests for the equality of the (character code, category code) pair of two tokens.

In the structure macros, the `\ifx` tests the arguments. When the arguments are single characters, such as "S", "D", or "C" for single bond, double bond, and circle, respectively, the application is straightforward. The character "Q" is used as argument where "no action" — no substituent, no additional bond — is a desired option at a particular place in a structure diagram. Thus, the coding for ring positions where substituents are an option is:

```
\ifx#nQ
\else   draw a bond line
        print the substituent #n \fi
```

The parameter n represents the substituent formula. — The explicit no-action symbol makes it possible to distinguish three different cases at a particular ring position: no action, just a bond line extending from the ring, and a bond line with a substituent at the position. As an example, the purine macro was used with an argument of Q for the 9-position in the left-hand diagram and with an empty set argument in the right-hand diagram:



One would use the bond-line-only option in cases where another structure fragment in a picture is to be attached to the bond. — The character Q was chosen because it is not part of any element symbol and is not commonly used as a structural symbol otherwise.

When the parameter after the `\ifx` is substituted by a text string representing a multi-character substituent, TEX actually compares the first character of the string with its second character, since these are the first two tokens encountered. Thus, a substituent that begins with two identical characters always makes the condition true. In such cases, one has to ensure that the string as a whole is compared by enclosing it in a box, *e.g.,* `\mbox{$\rm NNHC_{6}H_{5}$}`.

TEX's conditional facility can also be used to impart some chemical intelligence to a macro by causing screen messages to be generated when the user supplies a combination of arguments that is chemically not possible. Such a combination would be a ring double bond (argument 7="D") and a circle denoting aromaticity (argument 9="C") for the carbon sixring. A section of code

```
\ifx#7D \ifx#9C
                \message{Error: ...}
\fi \fi
```

will produce the message on the screen while the input file is processed by TEX to give the DVI file. The user can then correct the mistake and reprocess the input file before sending the DVI file to the output device. Error messages were not placed into all macros, just into the `\sixring` macro to demonstrate this feature.

In addition to the simple `\if` statements TEX has an `\ifcase` construction of the form

`\ifcase`(number)(text for case 0) `\or` (text for case 1) `\or` ...

`\or` (text for case n) `\else` (text for all other cases) `\fi`

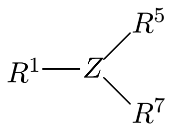
(Knuth 84, p. 210). When the `\ifcase` statement is used in a macro and the case number is passed as an argument, many different actions can be requested through one argument. For a larger number of cases, the TEX code with `\ifcase` is somewhat more elegant than a series of individual `\if` statements. A suitable application for the chemical structure macros is the placement of double bonds in various positions of a structure, where the position number, or a numeric code for a combination of positions, is passed as the case number. One version of the sixring macro, `\sixringb`, contains an `\ifcase` construct.

2.2 Use of structure macros

2.2.1 Invoking the structure macros

In order to typeset a chemical structure through invoking one of the macros in this system, the user must know how the structure will be oriented on the page, besides knowing, of course, the function of each argument. The structures in this system can not be rotated; they are oriented according to common practices in chemistry, but occasionally the user will have to adapt a model structure to the given orientation. Chapter 3 shows for each macro a typical structure produced by it, to illustrate the orientation and the position numbers. In this section it will be demonstrated with two representative macros how this information is to be used.

The macro `\cright` typesets structures or structure fragments of the general form



.

The arguments 1, 3, 5, and 7 are the substituents or groups R^1, Z, R^5, and R^7. Arguments 2, 4, and 6 trigger the drawing of bonds between R^1 and Z, between Z and R^5, and between Z and R^7, respectively. The bonds can be single or double bonds, from arguments "S" and "D" respectively, and the bond between R^1 and Z does not have to be present. A typical structure is shown in Figure 2.1; it was drawn using `\cright{$CH_3$}{Q}{$CH$}{S}{$COOCH_3$}S{$COOCH_3$}`. Since the second argument for Figure+2.1 is "Q," no bond is drawn between R^1 and Z; and R^1, which is in a `\makebox` (see Section 1.3), is moved next to Z. Figure 2.2 shows two additional structures drawn with the `\cright` macro.

The macro `\sixring` typesets the very common carbon sixring. For this structure and all the ring structures, the user has to know how the system of macros assigns position numbers to the ring atoms. The assignment follows chemical nomenclature rules where applicable. However in the case of single-ring structures where all ring atoms are carbons, the assignment of position

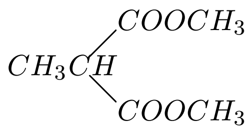
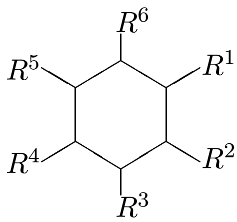


Figure 2.1: Figure drawn using `\cright`

Figure 2.2: Structures drawn with the `\cright` macro

number 1 is arbitrary. In this system of macros, the sixring is numbered as follows:



.

The first six arguments are the formulas for the optional substituents in the respective positions. The user has to refer to the position assignment to supply the text strings in the correct form, e. g. a sulfonic acid group for positions 4 and 5 would be typed in as HO_3S, whereas it would be SO_3H for all other positions. The remaining three arguments can produce alternating ring double bonds, but each of them has a second function. Argument 7 can produce a second substituent at position 1, argument 8 an outside double bond with substituent in position 3, and argument 9 a circle inside the ring denoting aromaticity. Where arguments have two functions in this way, the different structural features generated are of course mutually exclusive chemically. Figure 2.3 shows a sixring structure typeset with the command `\sixring{$OH$}{Q}{Q}{Q}{$NC$}{Q}{$CH_{3}$}{$NH$}{D}`. The substituent in position 3 is passed as argument 8 since it is not the regular single-bonded substituent represented by argument 3.

2.2.2 Displaying macro-generated diagrams within a document

Since the macro-generated diagrams constitute TEX boxes, the code for a diagram can be included anywhere in the input file and TEX will try to find a place for the diagram as a whole in the line and on the page. The diagrams, however, take up more space horizontally and vertically than a

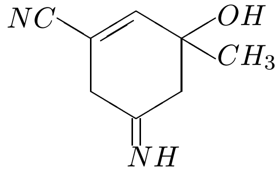


Figure 2.3: Structure typeset using \sixring

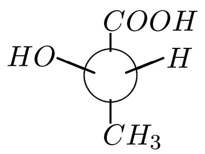


Figure 2.4: (S)−lactic acid

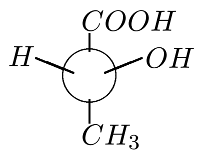


Figure 2.5: (R)−lactic acid

box of text; therefore TEX's line- and page breaking mechanisms would be strained, sometimes to such a degree that text squeezing or spreading would be apparent on the printed page.

A convenient way to display one or several structures on one line by themselves and centered, is the math display environment, enclosed by LATEX with brackets in the form `\[...\]`. A modification of this environment puts consecutive equation numbers at the right edge of the line. When using math display one has to remember that the only spacing in math mode is around math operators, space inserted by the user is ignored. Also, a new paragraph can not be started in math display. — The diagrams in math display are printed at the place in the document where they are coded in the input file. Thus the display can still cause problems with pagebreaking, especially since the diagrams usually take up more vertical space than a math equation for which the environment is designed.

The LATEX figure environment was specially designed for larger displays, such as the structure diagrams (Lamport 86, pp. 59, 60, 176, 177). A figure will usually not appear at the place in the document where the user has coded it; instead LATEX finds space for it on the current page or the next one in such a way that an overfull page is never produced. The user has some control over the placement of figures with optional parameters such as top or bottom of a page, but LATEX still makes the final decision, often producing surprising results. Since a figure is moved to a convenient place by LATEX , the figure is called a "float." In addition to the automatic space-finding, the figure environment has the advantage that it makes captions possible. Furthermore, several displayed objects, each with a caption, can be included in one figure environment; Figures 2.2.2 and 2.2.2 present an example, produced by the LATEX code in Figure 2.6 The code shows that each picture with its caption is enclosed in a `\parbox`.

It should also be mentioned that all the display mechanisms discussed above can be used in a two-column document style which is the format of most scientific chemistry journals.

If the same structure is to be printed many times in a document, processing time can be

```
\begin{figure}
  \parbox{.4\textwidth}{\centering
    \ccirc{$COOH$}{$HO$}{$H$}{$CH_{3}$}
    %ccirc is a macro
  \caption{${\rm (S)-lactic acid}$}}
  \hfill
  \parbox{.4\textwidth}{\centering
    \ccirc{$COOH$}{$H$}{$OH$}{$CH_{3}$}
  \caption{${\rm (R)-lactic acid}$}}
\end{figure}
```

Figure 2.6: LaTeX code for two captions in one figure

saved by storing it in a LaTeX `\savebox`. The structure-drawing code then has to be processed once only. Applied to the chemistry macros, a statement

`\savebox1{\macroname{arg`$_1$`}{arg`$_2$`}`

will process the code and save the typeset structure. A `\usebox1` statement, enclosed in a math display or figure environment, is then placed into the input file whereever the structure is to be printed. A `\savebox` can also be given a symbolic name (Lamport 86, p. 101).

Chapter 3

The complete system of macros—its design and its use

3.1 General design criteria

LaTeX code can be written to typeset a structure diagram for any chemical compound in such a way that the diagram conforms to accepted practices in chemistry publications. The purpose of the macros in this thesis is to reduce the amount of low-level, bond-by-bond coding necessary to typeset a particular structure. The problem of designing a generally useful system of macros for this purpose has to be seen in the context of the large number of possible structures: More than 7 million chemical compounds are registered with the Chemical Abstracts Service, including 60,000 different ring systems, and innumerable additional structures are possible. The fragments to be typeset by the macros in this thesis were selected such that they could be helpful in producing the more common types of structures. The arguments of each macro in turn were selected with the goal of making the respective fragment as flexible as possible, such that the more common known structures of this type can be typeset by a particular macro. The selection criteria were informal, using the "expert knowledge" of the writer. Where a related approach to displaying chemical structure diagrams was taken in previous work, a similar selection of fragments was made (Zimmerman 84)(Bendall 80,85). Both authors are chemists and present a collection of fragments without any attempt to justify their choices.

For more flexibility, the user of the system of macros presented in this thesis can effect many structural variations by defining an outer picture and placing supplemental lines and atomic symbols into it in addition to the fragment(s) produced by the macros. It is estimated that the macros can provide shortcuts to the drawing of more than 50% of the structures shown in the widely used textbook by Solomons (Solomons 84).

The next section lists the individual macros in this system, each with a typical generic structure and directions for the use of all arguments. Specific selection criteria for the fragment as such and the arguments are mentioned in many cases. The fragments are listed in the traditional categories of organic chemistry: acyclic structures, alicyclic structures (rings where all ring members are carbon atoms), and heterocyclic structures. The number of arguments is given in brackets behind the macro name.

3.2 Macros of the system

3.2.1 Macros for acyclic fragments

Macro `\cbranch`[9]

This macro typesets structural fragments with vertical branches:

$$\begin{array}{ccccc} & & R^1 & & \\ & & | & & \\ R^3 & — & Z & — & R^7 \\ & & | & & \\ & & R^9 & & \end{array}$$

Arguments 1, 3, 7, 9: The substituent formulas for R^1, R^3, R^7, and R^9.

Argument 2: The bond between R^1 and Z, "S" for a single bond and "D" for a double bond. No action is taken for any other value of the argument.

Argument 4: The bond between R^3 and Z, "S" for a single bond and "D" for a double bond. When the argument is "Q", no bond is drawn and R^3 is moved next to Z.

Argument 5: The center atom(s), Z. When the argument is a string of more than one character, argument 6 should not be "S" or "D," and argument 7 should be an empty set.

Argument 6: The bond between Z and R^7, "S" for a single bond and "D" for a double bond. No action is taken for any other value of the argument.

Argument 8: The bond between Z and R^9, "S" for a single bond and "D" for a double bond. No action is taken for any other value of the argument.

Macro `\tbranch`[7]

This macro typesets structural fragments with vertical branches similar to `\cbranch`. The main reason for including `\tbranch` is to show the use of the LaTeX tabbing mechanism for printing structural fragments. Macro `\cbranch` is the preferred macro for structures of this type. In contrast to the other macros, `\tbranch` provides the math mode for the substituent formulas in the macro code. Therefore substituent formula arguments do not have to be enclosed by `$` symbols.

$$\begin{array}{cccc} & & R^1 & \\ & & | & \\ R^3 & - & Z & - \\ & & | & \\ & & R^6 & \end{array}$$

Arguments 1 and 6: The substituent formulas for R^1 and R^6.

Argument 2: The bond between R^1 and Z, "S" for a single bond and "D" for a double bond. No action is taken for any other value of the argument.

Argument 3: Atom symbols and bonds to the left of Z. Single bonds have to be typed in as hyphens, double bonds as equal signs.

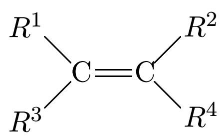
Argument 4: The center atom Z and any bonds and atom symbols to its right. Single bonds have to be typed in as hyphens, double bonds as equal signs.

Argument 5: The bond between Z and R^6, "S" for a single bond and "D" for a double bond. No action is taken for any other value of the argument.

Argument 7: An integer number which is interpreted as printer points of negative space between lines. The correct number for a document with double spacing is 13.

Macro `\ethene`**[4]**

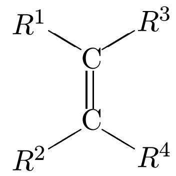
This macro typesets an ethene fragment with four variable substituents:



Arguments 1–4 are the substituent formulas represented by R^1, R^2, R^3, and R^4.

Macro `\upethene`**[4]**

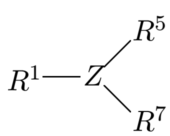
This macro is similar to `\ethene`, but it draws the ethene double bond vertically:



The arguments have the same meaning as they do for `\ethene`.

Macro `\cright`**[7]**

This macro typesets the following fragment which is often used for carboxylic acids and their derivatives:



Arguments 1,5,7: The substituent formulas R^1, R^5, and R^7.

Argument 2: The bond between R^1 and Z, "S" for a single bond and "D" for a double bond. For an argument of "Q", no bond is drawn and R^1 is moved next to Z.

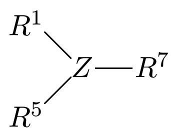
Argument 3: The center atom(s) Z.

Argument 4: The bond between Z and R^5, "S" for a single bond and "D" for a double bond. No action is taken for any other value of the argument.

Argument 6: The bond between Z and R^7, "S" for a single bond and "D" for a double bond. No action is taken for any other value of the argument.

Macro `\cleft`**[7]**

This macro typesets a fragment similar to the one produced by `\cright`, but opening to the left:



Arguments 1, 5, 7: The substituent formulas R^1, R^5, and R^7.

Argument 2: The bond between R^1 and Z, "S" for a single bond and "D" for a double bond. No action is taken for any other value of the argument.

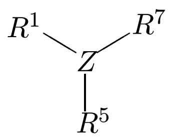
Argument 3: The center atom(s) Z. When the argument is a string of more than one character, argument 6 should not be "S" or "D", and argument 7 should be an empty set.

Argument 4: The bond between R^5 and Z, "S" for a single bond and "D" for a double bond. No action is taken for any other value of the argument.

Argument 6: The bond between Z and R^7, "S"for a single bond and "D" for a double bond. No action is taken for any other value of the argument.

Macro `\chemup`**[7]**

This macro typesets the following fragment which can be used for small molecules with trigonal geometry:



Arguments 1, 5, 7: The substituent formulas R^1, R^5, and R^7.

Argument 2: The bond between R^1 and Z, "S" for a single bond and "D" for a double bond. No action is taken for any other value of the argument.

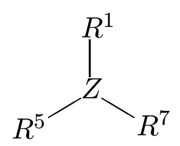
Argument 3: The center atom Z.

Argument 4: The bond between Z and R^5, "S" for a single bond and "D" for a double bond. No action is taken for any other value of the argument.

Argument 6: The bond between Z and R^7, "S" for a single bond and "D" for a double bond. No action is taken for any other value of the argument.

Macro`\cdown`[7]

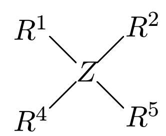
This macro typesets a fragment similar to the one produced by `\chemup`, but opening downwards:



The arguments have the same meaning as they do for `\chemup`.

Macro `\csquare`[5]

This macro typesets a fragment that is sometimes used when all four substituents on a center atom have to be shown explicitly:

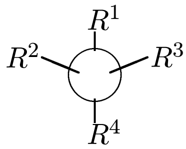


Arguments 1–4: The substituent formulas R^1, R^2, R^4, and R^5.

Argument 3: The center atom Z.

Macro `\ccirc`[4]

. This macro typesets a fragment used to show the actual configuration at a tetrahedral atom. The tetrahedral atom itself is not shown and is assumed to be in the middle of the sphere represented by the circle. The bonds typeset as heavier lines and intersecting the circle are directed out of the plane of the paper, towards the viewer.

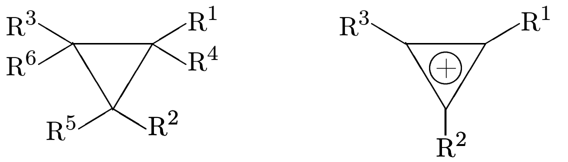


The arguments 1–4 are the substituent formulas R^1, R^2, R^3, and R^4.

3.2.2 Macros for alicyclic ring systems

Macro `\threering`[9]

This macro typesets the cyclopropane ring. The aromatic cyclopropenyl cation is drawn with a circle enclosing a plus sign inside the ring. The ring positions to which R^1, R^2, and R^3 are attached are designated as position 1, 2, and 3, respectively:



Arguments 1–6: An argument of "Q" causes no action. All other argument values are used as the respective substituent formulas R^1, R^2, R^3, R^4, R^5, and R^6. The bond line to R^2 is straight if the circle is in the ring or if there is no second substituent at position 2, and slanted otherwise.

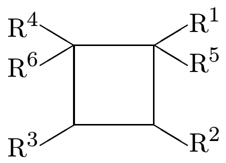
Argument 7: An argument of "D" typesets a second bond between ring positions 1 and 3. No action is taken for any other value of the argument.

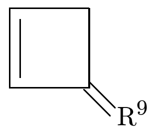
Argument 8: An argument of "Q" causes no action. All other argument values cause an outside double bond to be drawn from ring position 2, and the argument itself to be put at the end of the double bond as the substituent formula.

Argument 9: An argument of "C" typesets a circle enclosing a plus sign inside the ring. All other argument values cause no action.

Macro \fourring[9]

This macro typesets the cyclobutane ring. The ring positions to which R^1, R^2, R^3, and R^4 are attached are designated position 1, 2, 3, and 4, respectively.





Arguments 1–6: An argument of “Q” causes no action. All other argument values are used as the respective substituent formulas R^1, R^2, R^3, R^4, R^5, and R^6.

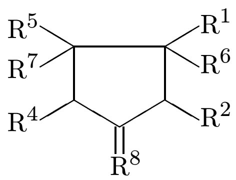
Argument 7: An argument of “D” typesets a second bond between ring positions 1 and 2. No action is taken for any other value of the argument.

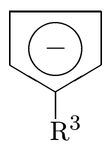
Argument 8: An argument of “D” typesets a second bond between ring positions 3 and 4. No action is taken for any other value of the argument.

Argument 9: An argument of “Q” causes no action. All other argument values cause an outside double bond to be drawn from ring position 2, and the argument itself to be put at the end of the double bond as substituent formula R^9.

Macro \fivering[9]

This macro typesets the cyclopentane ring. The aromatic cyclopentadienyl anion is drawn with a circle enclosing a minus sign inside the ring. The ring positions to which R^1, R^2, R^3, R^4, and R^5 are attached are designated as position 1, 2, 3, 4, and 5, respectively:





Arguments 1–5: An argument of “Q” causes no action. All other argument values are used as the respective substituent formulas R^1, R^2, R^3, R^4, and R^5.

Argument 6: An argument of “D” typesets a second bond between ring positions 1 and 2. An argument of “S” causes no action. All other argument values are used as the substituent formula R^6.

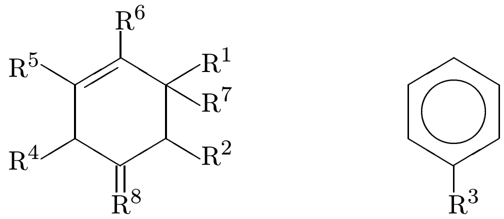
Argument 7: An argument of “D” typesets a second bond between ring positions 4 and 5. An argument of “S” causes no action. All other argument values are used as the substituent formula R^7.

Argument 8: An argument of “Q” causes no action. All other argument values cause an outside double bond to be drawn from ring position 3, and the argument itself to be put at the end of the double bond as substituent formula R^8.

Argument 9: An argument of “C” typesets a circle enclosing a minus sign inside the ring. No action is taken for any other value of the argument.

Macro `\sixring`**[9]**

This macro typesets a carbon sixring as a regular hexagon. A benzene ring can be drawn with alternating double bonds or with a circle inside the ring. The ring positions to which R^1, R^2, R^3, R^4, R^5, and R^6 are attached are designated as position 1, 2, 3, 4, 5, and 6, respectively:



Arguments 1–6: An argument of “Q” causes no action. All other argument values are used as the respective substituent formulas R^1, R^2, R^3, R^4, R^5, and R^6.

Argument 7: An argument of “D” typesets a second bond between ring positions 1 and 2. An argument of “S” causes no action. All other argument values are used as the substituent formula R^7.

Argument 8: An argument of “D” typesets a second bond between ring positions 3 and 4. An argument of “S” causes no action. All other argument values cause an outside double bond to be drawn from ring position 3 and the argument itself to be put at the end of the double bond as substituent formula R^8.

Argument 9: An argument of “D” typesets a second bond between ring positions 5 and 6. An argument of “C” typesets a circle inside the ring. No action is taken for any other value of the argument.

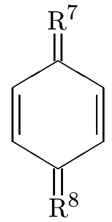
Macro `\sixringa`**[9]**

This macro differs from `\sixring` only in the positions of the double bonds. A value of “D” for arguments 7, 8, and 9 puts a double bond between ring positions 1 and 6, ring positions 2 and 3, and ring positions 4 and 5, respectively. Since the carbon sixring is so common, more options are needed for it than for the other rings.

Macro `\sixringb`**[9]**

This macro is also very similar to `\sixring`, but it allows all 17 chemically possible combinations of double bonds, including the three quinoid structures that can not be typeset with `\sixring` or

`\sixringa`.



Arguments 1–6: These arguments have the same meaning as in \sixring+.

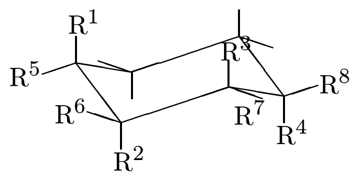
Argument 7: An argument of “Q” causes no action. All other argument values cause an outside double bond to be drawn from ring position 6 and the argument itself to be put at the end of the double bond as substituent formula R^7.

Argument 8: An argument of “Q” causes no action. All other argument values cause an outside double bond to be drawn from ring position 3 and the argument itself to be put at the end of the double bond as substituent formula R^8.

Argument 9: An integer number. The number zero causes the circle to be drawn inside the ring. All other integers are interpreted as a combination of ring double bonds according to the bit pattern corresponding to the decimal integer: A bit pattern of 000001 is interpreted as a double bond beginning at ring position 1, a bit pattern of 100000 (integer 32) as a double bond beginning at ring position 6. Thus, argument 9 for the diagram shown above is 9 (001001). No action occurs for argument values that correspond to a combination of double bonds which is chemically not possible, namely any combination with two adjoining double bonds.

Macro `\chair`[8]

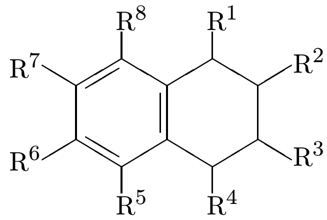
This macro typesets the saturated carbon sixring in its most favorable conformation. The axial and equatorial bond lines to the substituents are always drawn by this macro, even when there is no substituent in a particular position. This is the usual practice in drawing the chair form.



The eight arguments represent the respective substituent formulas R^1–R^8.

Macro `\naphth`9]

This macro typesets the aromatic naphthalene ring system, the fully saturated decalin ring system, and the 1,2,3,4-tetrahydronaphthalene shown in the diagram. The position numbers 1–8 are specified by the nomenclature rules of chemistry.

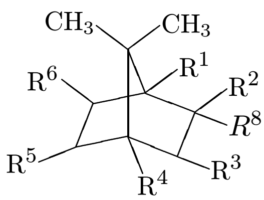


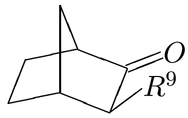
Arguments 1–8: An argument of Q causes no action. All other argument values are used as the respective substituent formulas R¹–R⁸.

Argument 9: A value of “S” typesets the ring system with no double bonds (decalin). A value of “D” typesets the aromatic system naphthalene with alternating double bonds. All other argument values draw the partially saturated system shown above.

Macro `\terpene`**[9]**

This macro typesets the bicyclo(2.2.1)heptane ring system found in such terpenes as borneol, camphor, and fenchol. The position numbers 1–7 are specified by the nomenclature rules of chemistry.





Arguments 1–6: An argument of “Q” causes no action. All other argument values are used as the respective substituent formulas R¹–R⁶.

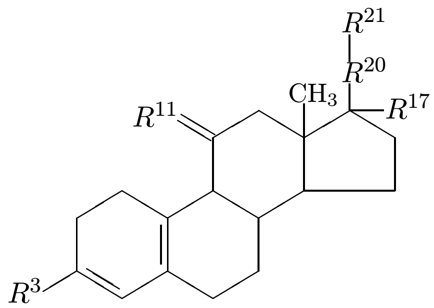
Argument 7: An argument of “M” prints two methyl groups on bonds extending from carbon #7. All other arguments cause no action.

Argument 8: An argument of “Q” causes no action. An argument of “O” prints an oxo group at carbon #2. All other argument values are used as a second substituent on carbon #2, shown as R⁸.

Argument 9: An argument of “Q” causes no action. An argument of “D” prints a second bond between positions 2 and 3. All other argument values are used as a second substituent on carbon #3, shown as R⁹.

Macro `\steroid`**[9]**

This macro typesets the steroid skeleton. The position numbers are specified by the nomenclature rules. The arguments are selected such that common types of steroids can be printed. Cholesterol, estradiol, progesterone, and cortisone are some of the steroids that can be produced.



Argument 1: An argument of “D” prints a second bond between positions 1 and 2. An argument of “Q” causes no action. All other argument values cause an outside double bond to be drawn from position 11 and the argument itself to be put at the end of the double bond as substituent formula R^{11}.

Argument 2: An argument of “D” prints a second bond between positions 3 and 4 (this double bond is shown in the diagram). An argument of “Q” causes no action. All other argument values cause an outside double bond to be drawn from position 3 and the argument itself to be put at the end of the double bond.

Argument 3: An argument of “Q” causes no action. All other argument values cause a single bond to be drawn from position 3 and the argument itself to be put at the end of the bond as substituent formula R^3.

Argument 4: An argument of “D” prints a second bond between positions 4 and 5. All other argument values cause no action.

Argument 5: An argument of “D” prints a second bond between positions 5 and 6. An argument of “Q” causes no action. All other argument values cause an outside double bond to be drawn from position 17 and the argument itself to be put at the end of the double bond.

Argument 6: An argument of “D” prints a second bond between positions 5 and 10 (shown in the diagram). An argument of “M” prints the methyl group containing carbon #19 and the bond to it. No action is taken for any other value of the argument.

Argument 7: An argument of “Q” causes no action. All other argument values print the substituent formula beginning with carbon #20, represented by R^{20} in the diagram, and the bond to it.

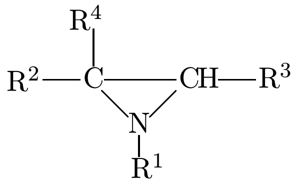
Argument 8: An argument of “Q” causes no action. All other argument values print the substituent formula beginning with carbon #21, represented by R^{21} in the diagram, and the bond to it.

Argument 9: An argument of “Q” causes no action. All other argument values print the second substituent on carbon #17 and the bond to it. This substituent is shown in the diagram as R^{17}.

3.2.3 Macros for heterocyclic ring systems

Macro `\hetthree`[8]

This macro typesets a 3-membered ring with one hetero atom. The common ring structures of this type are epoxides (oxirane) and ethylene imine (aziridine). Ring positions 1, 2, and 3 are the positions to which R^1, R^2, and R^3 are attached.



Arguments 1–5: An argument of "Q" causes no action. All other argument values are used as the respective substituent formulas R^1–R^5.

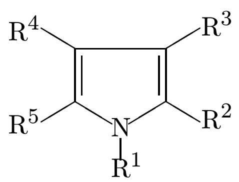
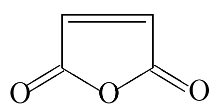
Argument 6: An argument of "S" typesets a bond to the left of ring atom #2. An argument of "H" puts —H to the left of ring atom #2. For all other argument values, no bond is drawn and R^2 is moved next to ring atom #2.

Argument 7: An argument of "S" typesets a bond to the right of ring atom #3. An argument of "H" puts H— to the right of ring atom #3. For all other argument values, no bond is drawn and R^3 is moved next to ring atom #3.

Argument 8: The atom symbol for the hetero atom.

Macro `\hetifive`[9]

This macro typesets 5-membered rings with one hetero atom. Thus it can be used to print the pyrrole, furan, and thiophene ring systems, and their hydrogenated versions. The arguments are selected such that common compounds like proline, pyrrolidone, maleic anhydride, and vitamin C can be printed. Ring positions 1, 2, 3, 4, and 5 are the positions to which R^1–R^5 are attached.

Arguments 1,3,4: An argument of "Q" causes no action. All other argument values are used as the respective substituent formulas R^1, R^3, and R^4.

Argument 2: An argument value of "Q" causes no action. An argument value of "O" puts an outside double bond with an O in ring position 2. All other argument values are used as the substituent formula R^2 with a single bond.

Argument 5: An argument value of “Q” causes no action. An argument value of “O” puts an outside double bond with an O in ring position 5. All other argument values are used as the substituent formula R^5 with a single bond.

Argument 6: An argument of “D” draws a second bond between ring positions 2 and 3. All other argument values cause no action.

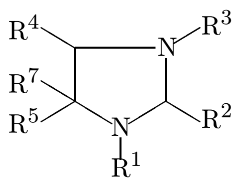
Argument 7: An argument of “D” draws a second bond between ring positions 3 and 4. All other argument values cause no action.

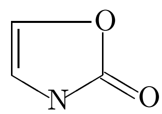
Argument 8: An argument of “D” draws a second bond between ring positions 4 and 5. All other argument values cause no action.

Argument 9: The atomic symbol of the hetero atom.

Macro `\heticifive`[9]

This macro typesets a 5-membered ring with 2 hetero atoms separated by a carbon atom. Thus it can be used to print ring systems such as imidazole, thiazole, and oxazole. The arguments were selected by considering actually occurring compounds containing these ring systems. Ring positions 1, 2, 3, 4, and 5 are the positions to which R^1–R^5 are attached.





Arguments 1, 3, 5: An argument of “Q” causes no action. All other argument values are used as the respective substituent formulas R^1, R^3, and R^5.

Argument 2: An argument of “Q”causes no action. An argument of “O” puts an outside double bond with an O in ring position 2. All other argument values are used as the substituent formula R^2 with a single bond.

Argument 4: An argument of “Q”causes no action. An argument of “O” puts an outside double bond with an O in ring position 4. All other argument values are used as the substituent formula R^4 with a single bond.

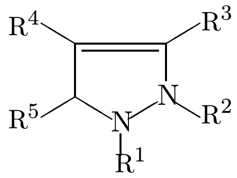
Argument 6: An argument of “D” draws a second bond between ring positions 2 and 3. All other argument values cause no action.

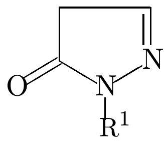
Argument 7: An argument of “Q” causes no action. An argument of “D” draws a second bond between ring positions 4 and 5. All other argument values are used as the substituent formula R^7, the second substituent at ring position 5.

Arguments 8 and 9: The atomic symbols of the hetero atoms in position 1 and 3, respectively.

Macro `\pyrazole`**[8]**

The pyrazole ring is found in a number of drugs, such as aminopyrine. Ring positions 1, 2, 3, 4, and 5 are the positions to which R^1–R^5 are attached.





Arguments 1, 2, 4: An argument of "Q" causes no action. All other argument values are used as the respective substituent formulas R^1, R^2, and R^4.

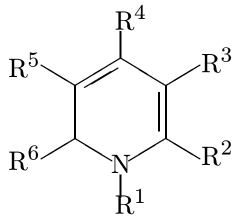
Argument 3: An argument of "Q" causes no action. An argument of "O" puts an outside double bond with an O in ring position 3. All other argument values are used as the substituent formula R^3 with a single bond.

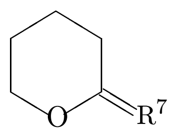
Argument 5: An argument of "Q" causes no action. An argument of "O" puts an outside double bond with an O in ring position 5. All other argument values are used as the substituent formula R^5 with a single bond.

Arguments 6, 7, 8: An argument of "D" draws a second bond between ring positions 2 and 3, ring positions 3 and 4, and ring positions 4 and 5, respectively. All other argument values cause no action.

Macro `\hetisix`**[9]**

This macro typesets a six-membered ring with one hetero atom. Thus it can be used to print ring systems such as pyridine and pyran. The arguments were selected by considering actually occurring compounds such as the B vitamins niacin and pyridoxine and the coumarin ring system. Ring positions 1–6 are the positions to which R^1–R^6 are attached.





Argument 1: An argument of "Q" causes no action. An argument of "D" prints a second bond between positions 1 and 6. All other arguments values are used as the substituent formula R^1.

Arguments 2–6: An argument of "Q" causes no action. All other argument values are used as the respective substituent formulas R^2–R^6.

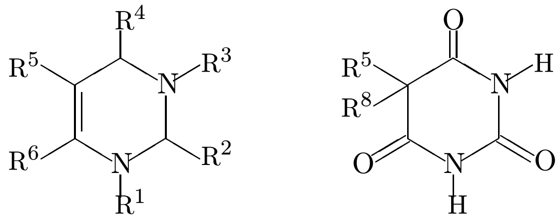
Argument 7: An argument of “Q” causes no action. An argument of “D” prints a second bond between positions 2 and 3. All other argument values cause an outside double bond to be drawn from position 2 and the argument itself to be put at the end of the double bond as R^7.

Argument 8: An argument of “D” prints a second bond between positions 4 and 5. All other argument values cause no action.

Argument 9: The atomic symbol of the hetero atom.

Macro `\pyrimidine`**[9]**

The pyrimidine ring occurs in such important compounds as cytosine, thymine, uracil, and the barbiturates. The arguments of the macro were selected such that these compounds can be typeset. Ring positions 1–6 are the positions to which R^1–R^6 are attached.



Arguments 1, 3, 5: An argument of “Q” causes no action. All other argument values are used as the respective substituent formulas R^1, R^3, and R^5.

Argument 2: An argument of “Q” causes no action. An argument of “O” causes an outside double bond with an O to be drawn at position 2. All other argument values are used as the substituent formula R^2 with a single bond.

Argument 4: An argument of “Q” causes no action. An argument of “O” causes an outside double bond with an O to be drawn at position 4. All other argument values are used as the substituent formula R^4 with a single bond.

Argument 6: An argument of “Q” causes no action. An argument of “O” causes an outside double bond with an O to be drawn at position 6. All other argument values are used as the substituent formula R^6 with a single bond.

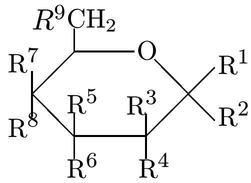
Argument 7: An argument of “D” prints a second bond between positions 1 and 2. All other argument values cause no action.

Argument 8: An argument of “Q” causes no action. An argument of “D” prints a second bond between positions 3 and 4. All other argument values are used as the second substituent in position 5, R^8.

Argument 9: An argument of “D” prints a second bond between positions 5 and 6. All other argument values cause no action.

Macro `\pyranose`[9]

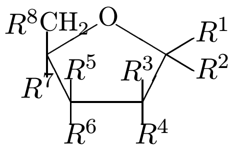
This macro was developed specifically for monosaccharide structures. Carbon #1 is at the position to which R^1 is attached. Structures from this macro look best with substituent formulas in 10 point size (shown) or even smaller.



Arguments 1–9 are used as the respective substituent formulas R^1–R^9. An argument of "Q" causes no action.

Macro `\furanose`[8]

This macro was also developed specifically for monosaccharide structures. Carbon #1 is at the position to which R^1 is attached. Structures look best with substituent formulas in 10 point size (shown) or even smaller.

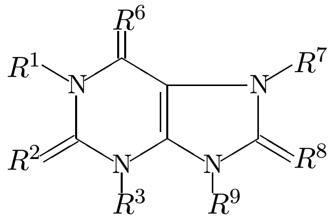


Argument 1: An argument of "Q" causes no action. An argument of "N" prints a long vertical bond at position 1, used for attachment to purine and pyrimidine bases to form nucleosides. All other argument values are used as the substituent formula R^1.

Arguments 2–8 An argument of "Q" causes no action. All other argument values are used as the respective substituent formulas R^2–R^8.

Macro `\purine`[9]

The purine ring system occurs in such important compounds as adenine, guanine, caffeine, and uric acid. The arguments of the macro were selected such that these compounds can be typeset. Positions 1, 2, 3, 6, 7, 8, and 9 are indicated in the following diagram by the respective substituent formulas.



Arguments 1, 3, 6, 7, 9: An argument of “Q” causes no action. All other argument values are used as the respective substituent formulas $R^1 \ldots R^9$.

Argument 2: An argument of “D” prints a second bond between positions 2 and 3. All other argument values cause an outside double bond to be printed at position 2 and the argument itself to be put at the end of the double bond as the substituent formula R^2.

Argument 4: An argument of “D” prints a second bond between positions 1 and 6. All other argument values cause no action.

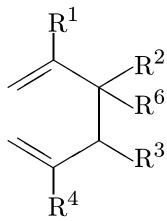
Argument 5: An argument of “Q” causes no action. All other argument values cause an outside double bond to be printed at position 6 and the argument itself to be put at the end of the double bond as the substituent formula R^6.

Argument 8: An argument of “D” prints a second bond between positions 7 and 8. All other argument values cause an outside double bond to be printed at position 8 and the argument itself to be put at the end of the double bond as the substituent formula R^8.

3.2.4 General Utility Macros

Macro `\fuseiv`[9]

This macro typesets a fragment that is designed to be connected at two places to another ring system with the effect of fusing an additional sixring to that system. The fragment can be fused to positions 1 and 2 of the carbon fivering and the carbon sixring, and to positions 2 and 3 of the `\hetifive` and `\hetisix` rings without changing the unitlength and the `\yi` coordinate.



Arguments 1–4: An argument of “Q” causes no action. All other argument values are used as the substituent formulas R^1–R^4.

Argument 5: An argument of “D” prints a second bond between the upper point of attachment and position 1 (this double bond is shown in the diagram). All other argument values cause no action

Argument 6: An argument of “Q” causes no action. An argument of “D” prints a second bond between positions 1 and 2. All other argument values are used as the substituent formula R^6.

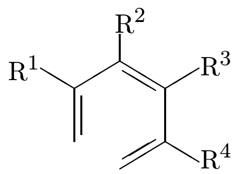
Argument 7: An argument of “D” prints a second bond between positions 2 and 3. All other argument values cause no action.

Argument 8: An argument of “Q” causes no action. An argument of “D” prints a second bond between positions 3 and 4. All other argument values are used as a second substituent in position 3 (not shown in the diagram).

Argument 9: An argument of “D” prints a second bond from the lower point of attachment to position 4 (this double bond is shown in the diagram). All other argument values cause no action.

Macro `\fuseup`[9]

This macro typesets a fragment that is designed to be connected at two places to another ring system with the effect of fusing an additional sixring to that system at an angle. The fragment can be fused to positions 1 and 6 of the carbon sixring and positions 3 and 4 of the `\hetisix` rings without changing the unitlength and the `\yi` coordinate.



Arguments 1–4: An argument of “Q” causes no action. All other argument values are used as the respective substituent formulas R^1–R^4.

Argument 5: An argument of “D” prints a second bond from the upper point of attachment to position 1 (the resulting double bond is shown in the diagram). All other argument values cause no action

Argument 6: An argument of “D” prints a second bond between positions 1 and 2. All other argument values cause no action

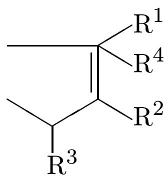
Argument 7: An argument of “D” prints a second bond between positions 2 and 3 (the resulting double bond is shown in the diagram). All other argument values cause no action

Argument 8: An argument of “D” prints a second bond between positions 3 and 4. All other argument values cause no action

Argument 9: An argument of “D” prints a second bond between position 4 and the lower point of attachment (the resulting double bond is shown in the diagram). All other argument values cause no action

Macro `\fuseiii`[6]

This macro typesets a fragment that is designed to be connected at two places to another ring system with the effect of fusing an additional fivering to that system. The fragment can be fused to positions 1 and 2 of the carbon fivering and sixring, and to positions 2 and 3 of the `\hetifive` and `\hetisix` rings without changing the unitlength and the `\yi` coordinate.



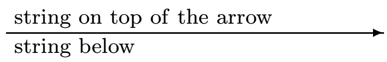
Arguments 1–4: An argument of “Q” causes no action. All other arguments are used as the respective substituent formulas R^1–R^4.

Argument 5: An argument of “Q” causes no action. All other argument values are used as a second substituent in position 2 (not shown in the diagram).

Argument 6: An argument of “D” prints a second bond between positions 1 and 2. All other argument values cause no action

Macro `\cto`[3]

This macro draws a reaction arrow and puts the requested character strings representing reagents and reaction conditions on top and below the arrow, respectively. The arrow is made long enough to accommodate the longer of the strings. The vertical position of the arrow can be changed by changing the `\yi` value.



Arguments 1 and 2: The character strings above and below the arrow, respectively.

Argument 3: An integer, the number of characters—including subscripts—in the longer string.

Macro `\sbond`[1]

This macro draws a horizontal single bond of a specified length, vertically centered on a line. It should be used for structural formulas that do not use the picture environment and are written on one line.

The argument is an integer, expressing the length of the bond in printer points (1pt = .35mm).

Macro `\dbond`[2]

This macro draws a horizontal double bond of a specified length. It should be used for structural formulas that do not use the picture environment and are written on one line.

$$=$$

Argument 1: An integer, expressing the length of the bond in printer points.

Argument 2: An integer, expressing the amount of vertical space by which the bonds have to be pushed together to give the desired vertical distance. In a document with double spacing, the value 19 produces appropriate spacing. The value 10 works for single spacing.

Macro `\tbond`[2]

This macro is similar to `\dbond`, except that it draws a triple bond:

$$\equiv$$

The meaning of the arguments is the same as in `\dbond`. The value 20 can be used as argument 2 for double spacing; 11 works for single spacing.

3.3 Common requirements for the use of the system

So far in this thesis it has been explained how to write LaTeX code to produce a chemical structure diagram at a particular place in a document. This section will discuss the mandatory and the optional statements at the beginning of an input file that make the system of macros accessible and its use more practical and convenient. Figure 3.1 contains these statements together with the two required declarations at the beginning of a LaTeX file, lines (1) and (7). (The line numbers are for reference only, they are not used in the input file.)

The part of the input file preceding the `\begin{document}` statement is called the "preamble" in the LaTeX Manual. In addition to the statements shown here, the preamble usually contains declarations pertaining to text formatting details such as margin width, text height on a page, and space between lines.

The document style option `chemtex` on line (1) of Figure 3.1 is necessary if the structure-drawing macros of this thesis are to be used for the preparation of a document. This statement reads the file `chemtex.sty` into TeX's memory, a file that contains the macros, `\initial` and `\reinit` as well as those described above. Macro `\initial` defines the command sequences `\xi`, `\yi`, `\pw`, `\pht`, `\xbox`, and `\len` as integer variables and assigns a count register to each of them. The use of the first four variables in the picture declaration and the use of `\xbox` in a minipage or parbox environment was explained in chapter II. The counter `\len` is a general purpose integer variable for the user. All the variables except `\len` are also given initial values. Furthermore, the unitlength for the picture environments is set to 0.1 printer points in `\initial`. This is the recommended unitlength for the chemical structure diagrams, but it can be changed anywhere in the document. Line (9) from Figure 3.1 calls `\initial`.

The macro `\reinit` simply resets all the parameters to their initial values from `\initial`. It is a convenience, especially for cases where more than one variable needs to be reset.

```
(1)  \documentstyle[chemtex,...]{report}
(2)  \setcounter{totalnumber}{4}
(3)  \setcounter{topnumber}{2}
(4)  \setcounter{bottomnumber}{2}
(5)  \renewcommand{\topfraction}{.5}
(6)  \renewcommand{\bottomfraction}{.5}
(7)  \begin{document}
(8) \textfont1=\tenrm
(9) \initial
```

Figure 3.1: Statements at the beginning of a LaTeX file

Note that other document style options (represented by the `...`) can also be indicated on the `\documentstyle` command. See the LaTeX manual for details.

Lines (2)–(6) in Figure 3.1 affect the placement of "floats" on the page. The only floats discussed in this thesis are the diagrams produced in the figure environment (see Chapter 2). In defining the style of a document–the report style is designated by line (1)—the LaTeX program sets default values for the maximum total number of floats on a page (three), the maximum number of floats at the top of the page (two), and at the bottom of the page (one). These values can be changed for documents with an unusually large number of figures. Thus, lines (2)–(4) increase the maximum number of floats to 4, evenly distributed on the page. It is then necessary to change the counters `\topfraction` and \bottomfraction+ to reflect the distribution of figures on the page.

Finally, line (9) is the optional redefinition of the math textfont, discussed in Chapter 1. This definition can be changed anywhere in the document.

Appendix A

Slopes of lines drawn with LaTeX

Table tb:slopes lists the first quadrant slopes of lines that can be typeset with LaTeX, together with the corresponding degrees of angle. The integers $\mathrm{x_s}$ and $\mathrm{y_s}$ represent the slope in LaTeX's line-drawing statement

```
\put(x,y){\line(xs,ys){length}}
```

Corresponding angles in the other quadrants can be generated by preceding $\mathrm{x_s}$ and/or $\mathrm{y_s}$ with a minus sign.

x_s, y_s	$\tan\theta(y_s/x_s)$	θ(degrees)
1,0	0.00	0.0
6,1	0.17	9.5
5,1	0.20	11.3
4,1	0.25	14.0
3,1	0.33	18.5
5,2	0.40	21.8
2,1	0.50	26.5
5,3	0.60	31.0
3,2	0.67	33.7
4,3	0.75	36.8
5,4	0.80	38.7
6,5	0.83	39.8
1,1	1.00	45.0
5,6	1.20	50.2
4,5	1.25	51.3
3,4	1.33	53.2
2,3	1.50	56.3
3,5	1.67	59.0
1,2	2.00	63.5
2,5	2.50	68.2
1,3	3.00	71.5
1,4	4.00	76.0
1,5	5.00	78.7
1,6	6.00	80.5
0,1	∞	90.0

Table A.1: Slopes of lines possible with LaTeX

Appendix B

Coordinates of points of attachment

The tables in this appendix list coordinates of points of attachment that will probably be used most frequently with the techniques described in chapter V. Table B.1 lists the coordinates of the six corners of the carbon sixring and the coordinates at the end of the six bonds extending from the sixring.

The ring structures typeset by the macros `\fivering`, `\naphth`, `\steroid`, `\hetifive`, `\heticifive`, `\pyrazole`, `\hetisix`, `\pyrimidine`, and `\purine` have identical coordinates at positions that are equivalent to the sixring positions in regard to the printed diagram. (The position numbers are not necessarily the same as those of the sixring.)

Table B.2 lists the coordinates of other points of attachment in structures typeset by various macros.

Sixring Position	Ring Corner	End of Bond
1	(342,200)	(470,277)
2	(342,0)	(470,-77)
3	(171,-103)	(171,-203)
4	(0,0)	(-128,-77)
5	(0,200)	(-128,277)
6	(171,303)	(171,403)

Table B.1: Points of attachment in the sixring

Macro	Position Description	Coordinates
`\cbranch`	begin of left bond	(-150,33)
`\cbranch`	end of right bond	(230,33)
`\cright`	begin of left bond	(-150,33)
`\cleft`	end of right bond	(230,33)
`\chemup`	end of vertical bond	(33,-150)
`\cdown`	end of vertical bond	(33,220)
`\threering`	ring position 1	(300,0)
`\threering`	end of bond on 1	(428,77)
`\naphth`	ring position 1	(513,303)
`\naphth`	end of bond on 1	(513,403)
`\naphth`	ring position 2	(684,200)
`\naphth`	end of bond on 2	(812,277)
`\steroid`	end of vert. bond on 17	(1026,706)
`\hetthree`	end of bond on 3	(580,30)
`\pyranose`	end of alpha bond	(688,100)
`\pyranose`	end of beta bond	(688,-100)
`\furanose`	end of alpha bond	(553,-63)
`\furanose`	end of beta bond	(553,63)
`\furanose`	end of vert. long bond	(448,380)
`\purine`	point below N(9)	(513,-130)
`\fuseiv`	upper point of attachment	(0,200)
`\fuseiv`	lower point of attachment	(0,0)
`\fuseup`	upper point of attachment	(-171,303)
`\fuseup`	lower point of attachment	(0,200)
`\fuseiii`	upper point of attachment	(0,200)
`\fuseiii`	lower point of attachment	(0,0)

Table B.2: Points of attachment in various structures