XΥMTEX for Drawing Chemical Structural Formulas

Shinsaku Fujita

Abstract

XΥMTEX,[1] a macro package of combined LATEX style files, has been developed for drawing a wide variety of chemical structural formulas. The commands of XΥMTEX have a set of systematic arguments for specifying substituents and their positions, endocyclic double bonds, and bond patterns. In some cases, they have an additional argument for specifying hetero-atoms on the vertices of heterocycles. As a result of this systematic feature, XΥMTEX works effectively as a practical tool within the "device-independent" concept of TEX.

1 Introduction

A few years ago, in order to expand the use of TEX into various fields of chemistry, I decided to write a book [1] that would use LATEX style files and BIBTEX `bst` files suitable for such scientific journals as *Journal of the American Chemical Society*, *Science*, and *Nature* (potentially there were some thirty journals in the field which might benefit). However, while preparing the book I encountered difficulties in introducing methods for drawing formulas of chemical structures. Although I had dealt with ChemTEX [2], epic [3] and PICTEX [4] as device-independent methods, as well as with PostScript [5] and tpic [6] as

[1] The Xs of XΥMTEX should be pronounced as a Greek chi or simply as 'k' in 'kyumtek'.

device-dependent methods, it was difficult to recommend any one of them as a standard method.

ChemTEX typesets structural formulas of high quality in a device-independent manner. Its commands, however, should be replaced by more systematic ones in order to cover structures with a wide range of substitution. As for epic and PICTEX, in themselves they have no facilities for drawing chemical structures. Moreover, they produce output of lesser quality than ChemTEX, especially in printing chemical bonds. Among device-dependent methods, encapsulated PostScript is now recognized as the predominant method since chemical structural formulas are usually drawn with tools whose storing and printing processes are based on PostScript. However, it is still desirable to develop a convenient method according to the device-independent concept, since this is a fundamental philosophy in TEX typesetting, and encourages electronic submission and exchange of information.

2 Features of XYMTEX

Therefore, it seemed necessary to take all of these issues into consideration and devise something new. XYMTEX was developed as a device-independent method with systematic commands (control sequences) for drawing structural formulas [7]. The features and advantages of XYMTEX are summarized below:

1. The name XYMTEX is the uppercase form of $\chi\upsilon\mu\tau\epsilon\chi$, in which $\chi\upsilon\mu$ is the Greek counterpart of the stem 'chem' of 'chemistry'. When the logo XYMTEX is unavailable, you should type `XyMTeX`.
2. XYMTEX requires the LATEX `picture` environment only, ensuring portability (since LATEX is part of most TEX distributions). Thus, wide adaptations for personal computers are available and a variety of printers can be used as output devices.
3. XYMTEX should be used within a large version of LATEX.
4. Structural formulas written with XYMTEX produce high-quality output, since they use LATEX fonts.
5. Each command name corresponds to a master template to be drawn. It can be easily remembered, since it stems from the familiar nomenclature of organic compounds.
6. The invariant part of a structure (the master template containing fixed bonds and atoms) is automatically printed with no designation.
7. The variant parts of a structure (substituents, additional bonds and atoms) are designated by up to four arguments: SUBSLIST, OPT, BONDLIST, and ATOMLIST.
8. Substituents and their positions are given by a single argument (SUBSLIST) in which they are listed consecutively, with semicolons as delimiters. It follows that an arbitrary number of substituents can be written in the SUBSLIST.
9. A command of frequent occurrence has an optional argument (OPT) of one or two characters for showing a pattern of bonds or aromatization.
10. Additional endocyclic bonds are designated by an optional argument (BONDLIST) in which one character corresponds to each of the bonds.
11. A more general command for drawing heterocycles takes an additional argument (ATOMLIST), so that a set of hetero-atoms are typeset on the vertices of the master template after truncation of edges.
12. Commands with a common stem but different suffixes ('v', 'vi', 'h', 'hi') are provided for drawing the same structure in different ways.
13. Each structure created by a XYMTEX command is regarded as a letter, or more exactly, as a TEX box. Thus, it is controlled by the inherent mechanism of TEX in breaking paragraphs (containing such structures) into lines as well as in making the lines into pages.
14. The recognition of a XYMTEX structure as a TEX box permits us to use XYMTEX commands in various LATEX environments such as `center`, `equation` and `tabular`.
15. In extreme cases, a XYMTEX command can be used in the argument of another command. For example, it may be utilized in the argument of `\section` in conjunction with the `\protect` command. However, such modes of usage should not be recommended to regular users, since they may cause unexpected errors.

3 Drawing Benzene and Naphthalene Derivatives

Let us first draw benzene and naphthalene derivatives as examples. XYMTEX contains nine style files for drawing various categories of chemical structural formulas (Table 1). Since the macros we will require are stored in `carom.sty`, the document file begins as follows:

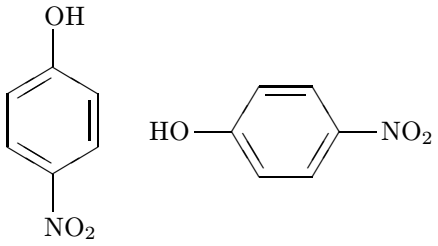
```
\documentstyle[epic,carom]{article}
\begin{document}
(body)
\end{document}
```

where `epic.sty` is also included in any order as an option for drawing dotted lines. This is the usual form of LaTeX documents.

The constitution of a command is quite simple. To draw the structure of 4-nitrophenol, you write the following simple statements in the body of your document:

```
\bzdrv{1==OH;4==NO$_2$}
\bzdrh{1==HO;4==NO$_2$}
```

where each argument is a SUBSLIST, listing substituents with their bonds. This generates the following:



A semicolon separates each mode of substitution, where a double equality symbol (==) is used as a delimiter between a substitution position and a substituent. Thus, the two arguments state that position 1 takes a hydroxyl group through a single bond ('1==OH') and position 4 takes a nitro group through a single bond ('4==NO_2'), where each single bond is automatically drawn without explicit declaration. Since statements in SUBSLIST arguments follow the nomenclature of organic compounds, as shown in these examples, most organic chemists and secretaries with appropriate training can write them down easily. The suffixes 'v' and 'h' generally indicate vertical and horizontal forms of printed formulas.

To draw three structures with different aromatic expressions for 1-bromo-4-chlorobenzene, you use `\bzdrv[OPT]{SUBSLIST}` as follows:

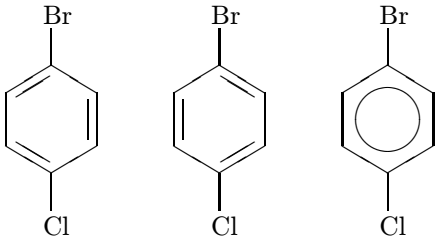
```
\bzdrv[r]{1==Br;4==Cl}
\bzdrv[l]{1==Br;4==Cl}
\bzdrv[c]{1==Br;4==Cl}
```

where the letters in brackets are optional arguments for representing patterns of double bonds. The standard mode of displaying alternant double bonds (to the right-hand side of the diagram) is the default (e.g., when no optional argument is used, as in the first example); specifying 'r' in square brackets will also yield the right-handed mode. The letter 'l' in brackets generates an alternative (left-handed) mode of alternant double bonds; 'c' in square brackets expresses an inner circle. As a result, you get

Table 1: Style Files in XyMTeX

file name	printed structures
`aliphat.sty`	aliphatic compounds
`carom.sty`	vertical and horizontal types of cyclic compounds
`ccycle.sty`	bicyclic compounds etc.
`chemstr.sty`	basic commands
`hcycle.sty`	pyranoses and furanoses
`hetarom.sty`	vertical types of heterocyclic compounds
`hetaromh.sty`	horizontal types of heterocyclic compounds
`locant.sty`	locant numbers
`lowcycle.sty`	five-or-less-membered carbocycles

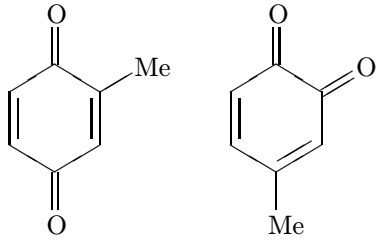
three structural formulas with different bond patterns for the same compound:



The `\bzdrv` command is also used to typeset *p*- and *o*-benzoquinone derivatives. Thus, the code:

```
\begin{center}
\bzdrv[p]{1D==O;4D==O;2==Me}
\bzdrv[o]{1D==O;2D==O;4==Me}
\end{center}
```

produces



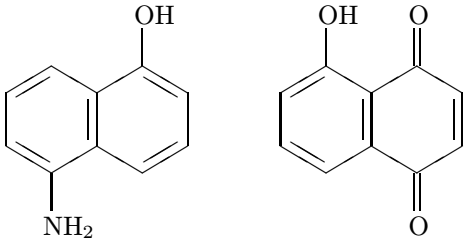
where the optional arguments 'p' and 'o' control patterns of endocyclic double bonds. The carbonyl double bonds are designated by means of a bond modifier 'D' coupled with a preceding locant number; thus, the string '1D==O' represents an oxygen atom through an exocyclic double bond.

Naphthalenes and naphthoquinones are typeset by using the `\naphdrv` command and so on. For example, the code:

```
\begin{center}
```

```
\naphdrv{1==OH;5==NH$_2$}
\naphdrv[p]{1D==O;4D==O;8==OH}
\end{center}
```

prints the following fused structures:



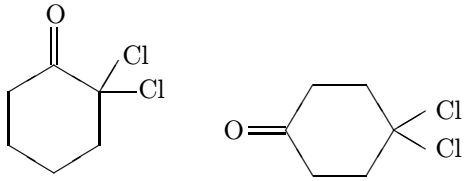
The `\naphdrv` command also takes optional arguments ('oa' to 'of') in order to print all possible structures of *o*-naphthoquinones. The command is also capable of drawing other naphthoquinones such as 2,6-naphthoquinones.

4 Drawing Cyclohexane Derivatives

The command `\cyclohexanev` and related ones are used to typeset cyclohexane derivatives. These commands are also contained in `carom.sty`. They are capable of drawing geminal substituents by using appropriate bond modifiers such as 'Sa' and 'Sb'. For example, the code:

```
\begin{center}
\cyclohexanev{1D==O;2Sa==Cl;2Sb==Cl}
\cyclohexaneh{1D==O;4Sa==Cl;4Sb==Cl}
\end{center}
```

produces the following structures:

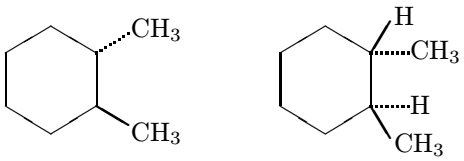


Note that the suffixes 'v' and 'h' are in accord with the general convention described above.

For specifying the stereochemistries of cyclohexanes more explicitly, we use bond modifiers of single type ('A' and 'B') as well as those of geminal type ('SA' and 'SB'). For example,

```
\begin{center}
\cyclohexanev{2A==CH$_3$;3B==CH$_3$}
\cyclohexanev{2SA==CH$_3$;2SB==H;%
 3SB==CH$_3$;3SA==H}
\end{center}
```

yields the following alternative expressions of *trans*-1,2-dimethylcyclohexane:

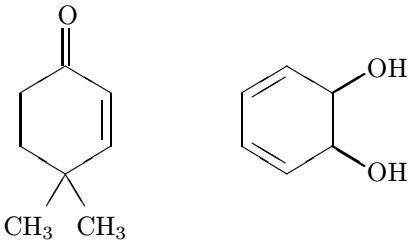


where the α-bonds ('A' and 'SA') are represented by dotted lines and the β-bonds ('B' and 'SB') are printed with boldfaced lines.

Commands such as `\cyclohexanev` take an optional argument BONDLIST that contains one or more letters ('a' to 'f') for designating endocyclic double bonds in a bond-by-bond fashion. Thus, we use `\cyclohexanev[BONDLIST]{SUBSLIST}`. For example,

```
\begin{center}
\cyclohexanev[b]{1D==O;%
 4Sa==CH$_3$;4Sb==CH$_3$}
\cyclohexanev[df]{2B==OH;3B==OH}
\end{center}
```

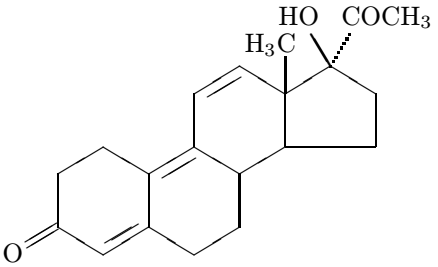
generates the following structural formulas with endocyclic double bonds:



XƎMTEX is capable of drawing more complicated structures such as steroids in a similar way. Let us write the code:

```
\begin{center}
\steroid[dim]{3D==O;%
 {{13}B}==\lmoiety{H$_3$C};{{17}SB}==HO;%
 {{17}SA}==COCH$_3$}
\end{center}
```

where the optional argument 'dim' gives three endocyclic double bonds, giving us a steroid derivative:



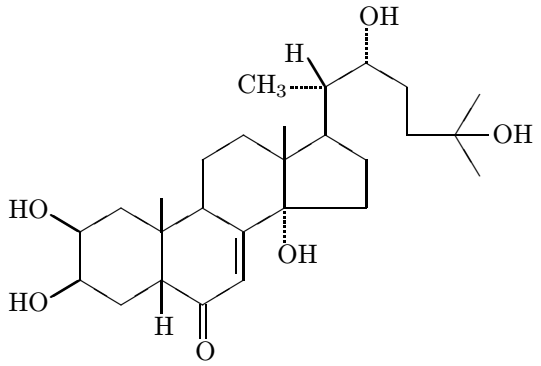
the methyl group of which attaches to the fused 13-position in such a manner that the right terminal

carbon of the methyl is linked to the corresponding bond by using the command `\lmoiety`.

The command `\steroidchain` is used to draw an insect hormone α-ecdysone just by designating a set of substituents in the SUBSLIST argument. Thus, we write the code:

```
\begin{center}
\unitlength.09pt
\steroidchain[g]{%
 2B==HO;3B==HO;{{10}B}==;5B==H;%
 6D==O;{{13}B}==;{{14}A}==OH;%
 {{20}SA}==CH$_3$;{{20}SB}==H;%
 {{22}A}==OH;{{25}}==H}
\end{center}
```

in which empty substituents are allowed to show implicit methyl groups at the 10- and 13-positions. This results in the following complex formula. In this example, I have slightly reduced the size of the structural formula by setting `\unitlength.09pt`— the default setting in XeTeX for unit length is 0.1pt.



It should be noted that further reduction of sizes is usually unsuccessful because the LaTeX `picture` environment is incapable of drawing lines of short lengths and of arbitrary slopes.

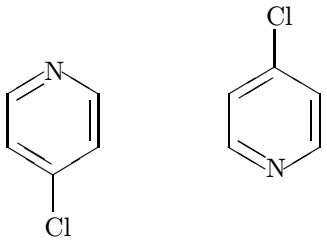
In the future, this restriction concerning line lengths and slopes should be overcome so that one could draw chemical structures more conveniently. XeTeX would then become an automatic tool linked with chemical drawing software. This linkage would mean that XeTeX codes could be created automatically in the future rather than manually as in the present situation. The future aim for XeTeX is therefore to implement a device-independent method whose codes would be created automatically with some kind of graphical user interface.

5 Drawing Heterocycles

For the purpose of drawing heterocyclic compounds, `\documentstyle` must contain `hetarom.sty` in the optional argument. Let us draw pyridine derivatives, using the following code:

```
\begin{center}
\pyridinev{4==Cl}
\pyridinevi{4==Cl}
\end{center}
```

This yields two pyridine structures of inverse types:

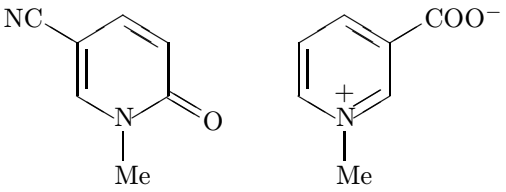


To show a ring nitrogen at the top of the pryidine derivative, the suffix 'v' is used; to mark the presence of a nitrogen atom at the bottom position of the ring, 'vi' is used. As a result, the corresponding positions in the alternative structures have common locant numbers, starting from the respective nitrogen atoms; thus, the 4-chloropyridine has been drawn in two ways without any changes to the SUBSLIST argument. This convention is also applied to commands for drawing other heterocycles.

The `\pyridinev` and related commands take an optional argument BONDLIST to designate endocyclic double bonds other than the default settings for alternant double bonds. Thus, we use `\pyridinev[BONDLIST]{SUBSLIST}`, in which one or more characters selected from 'a' to 'f' are involved in the BONDLIST. For example, the code:

```
\begin{center}
\pyridinevi[ce]{2D==O;1==Me;5==NC}
\pyridinevi[ace{1+}]{1==Me;3==COO$^{-}$}
\end{center}
```

produces the following structures:

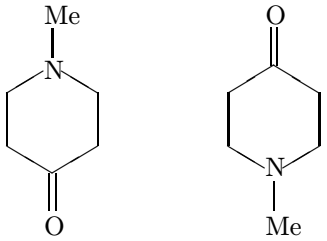


Note that the BONDLIST argument in the latter example contains a descriptor '1+' for denoting a plus charge on the ring nitrogen. Such a descriptor consists of a locant number and a character to be printed, both of which are bundled with braces according to the TeX grammar.

The default pattern of endocyclic double bonds for drawing six-membered heterocycles is an alternant pattern to complete their aromaticity. If you intend to typeset saturated heterocycles, you should give an empty optional argument:

```
\begin{center}
\pyridinev[]{4D==O;1==Me}
\pyridinevi[]{4D==O;1==Me}
\end{center}
```

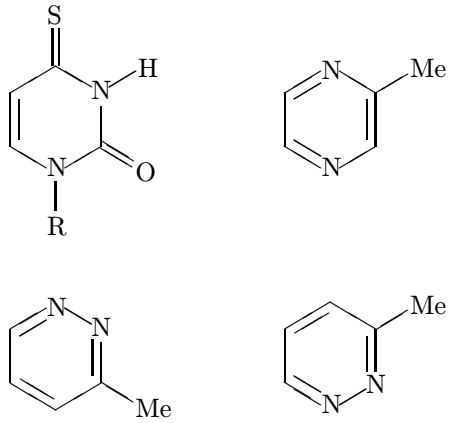
which yields the following structures:

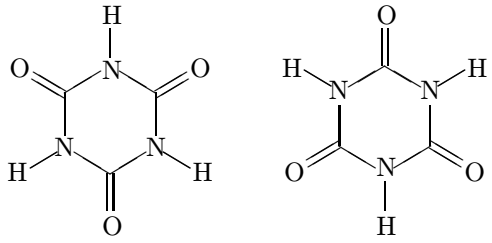


In order to show a wide variety of commands for drawing heterocyclic compounds, let us test the following statements:

```
\begin{center}
\pyrimidinevi[e]{%
 1==R;3==H;2D==O;4D==S} \qquad
\pyrazinev{2==Me} \\
\pyridazinev{3==Me} \qquad
\pyridazinevi{3==Me} \\
\triazinev[]{2D==O;4D==O;%
 6D==O;1==H;3==H;5==H} \qquad
\triazinevi[]{2D==O;4D==O;%
 6D==O;1==H;3==H;5==H}
\end{center}
```

Note that the pair of codes before each \\ constructs a text line during LaTeX processing, since XeTeX views each structure as a letter (or a TeX box). The following heterocycles result:

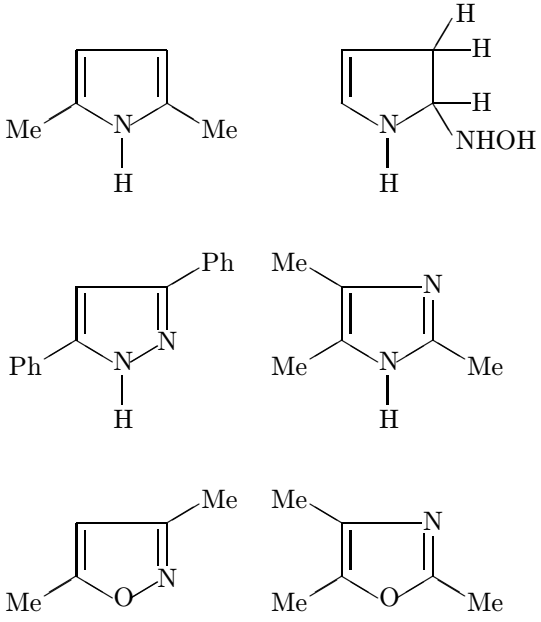




Since 5-membered heterocycles comprise a predominant family of organic compounds, XeTeX has versatile facilities for drawing them:

```
\begin{center}
\pyrrolev{1==H;2==Me;5==Me}
\pyrrolev[d]{1==H;2Sa==H;%
 2Sb==NHOH;3Sa==H;3Sb==H} \\
\pyrazolev{1==H;3==Ph;5==Ph}
\imidazolev{1==H;2==Me;4==Me;5==Me} \\
\isoxazolev{3==Me;5==Me}
\oxazolev{2==Me;4==Me;5==Me}
\end{center}
```

The command names come from those of master templates, i.e., pyrrole, pyrazole, imidazole, isoxazole, and oxazole. As before, the suffix conventions ('v', 'vi', 'h' and 'hi') are also effective in these commands. The above code yields a variety of 5-membered cyclic compounds:

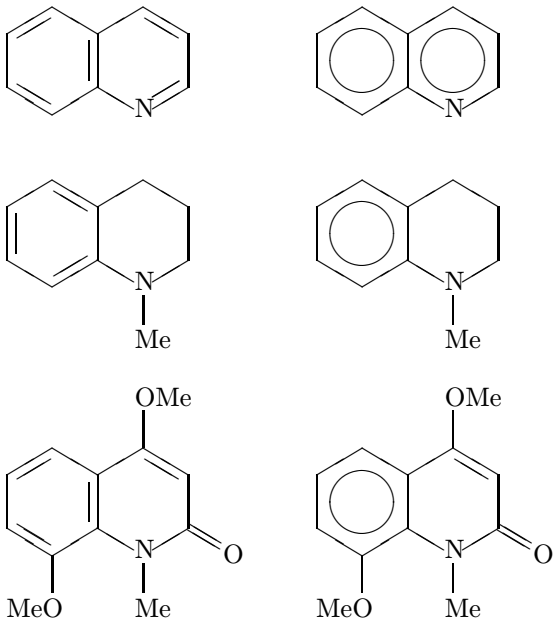


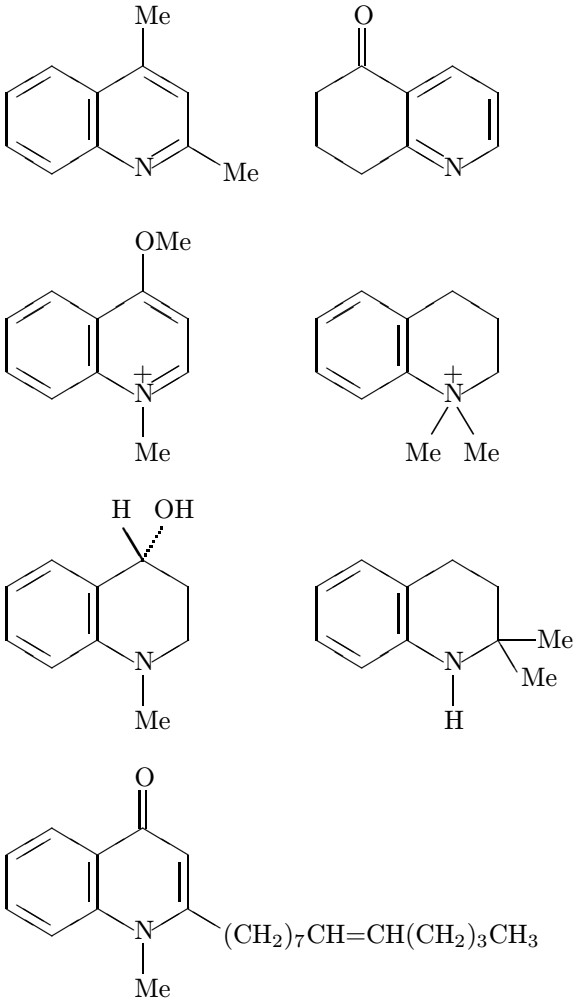
XeTeX is capable of also drawing fused heterocyclic compounds. These compounds can take a wide variety of substitutions and bond patterns.

Every command in XeLaTeX has therefore been designed to be able to deal with such diversity by using BONDLIST and SUBSLIST arguments. To illustrate the flexibility of XeLaTeX commands, let us test `\quinolinevi` in various situations:

```
\begin{center}
\quinolinevi{}
\quinolinevi[AB]{} \\[-16pt]
\quinolinevi[egi]{1==Me}
\quinolinevi[A]{1==Me} \\[16pt]
\quinolinevi[cfhk]{1==Me;2D==O;%
 4==OMe;8==\lmoiety{MeO}}
\quinolinevi[Ac]{1==Me;2D==O;%
 4==OMe;8==\lmoiety{MeO}} \\[16pt]
\quinolinevi{2==Me;4==Me}
\quinolinevi[bdj]{5D==O} \\
\quinolinevi[r{1+}]{1==Me;4==OMe}
\quinolinevi[fhk{1+}]{%
 1Sa==Me;1Sb==Me} \\[16pt]
\quinolinevi[fhk]{1==Me;4SA==OH;4SB==H}
\quinolinevi[fhk]{%
 1==H;2Sa==Me;2Sb==Me} \\[16pt]
\quinolinevi[bfhk]{1==Me;4D==O;%
 2==(CH$_2$)$_7$CH=CH(CH$_2$)$_3$CH$_3$}
\phantom{\quinolinevi{}}
\end{center}
```

This gives us the following test structures:





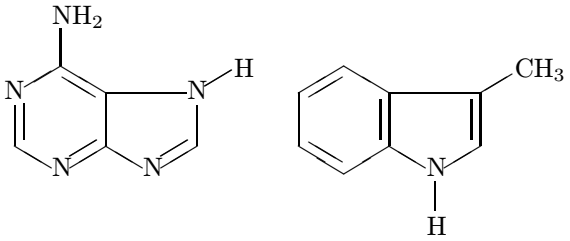
This last example contains a long-chain substituent with a double bond represented by a single equality symbol (=). For this reason, we use double equality (==) as a delimiter in the SUBSLIST argument.

XeLaTeX includes other commands for drawing fused heterocycles with two 6-membered rings: isoquinolines, quinoxalines, quinazolines, cinnolines, and pteridines. These commands can be used in the same way as described for the `\quinolinevi` command, where 'v', 'vi', 'h', and 'hi' are also effective.

XeLaTeX has various commands for drawing heterocycles with fused 5- and 6-membered rings: purines, indoles, indolizines, isoindoles, benzofuranes, isobenzofuranes, and benzoxazoles. For example, adenine and 3-methylindole are drawn by using the following code:

```
\begin{center}
\purinev[adfh]{3==H;4==NH$_2$}
\indolev{1==H;3==CH$_3$}
\end{center}
```

which gives the following results:

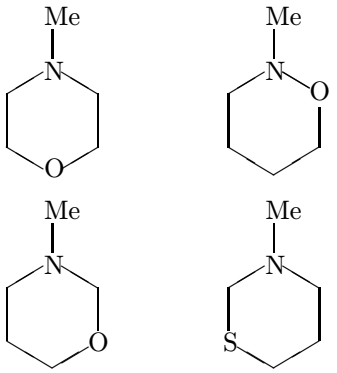


In the preceding paragraphs of this section, each mother skeleton is associated with a specific command. This approach is combinatorially explosive in nature since there are further categories of heterocyclic master templates. More general commands have therefore been designed to have the additional function of specifying inner atoms on rings, as in `\sixheterovi[BONDLIST]{ATOMLIST}{SUBSLIST}`.

Each of the following code examples contains two arguments in addition to an empty optional BONDLIST argument. Among them, the second argument is an ATOMLIST for specifying the positions and species of hetero-atoms.

```
\begin{center}
\sixheterovi[]{1==O;4==N}{4==Me}
\sixheterovi[]{3==O;4==N}{4==Me} \\
\sixheterovi[]{2==O;4==N}{4==Me}
\sixheterovi[]{6==S;4==N}{4==Me}
\end{center}
```

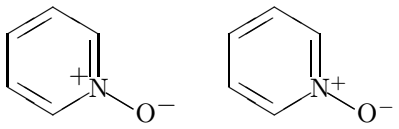
These codes generate the following structures:



The next examples illustrate two ways of drawing pyridine-N-oxide:

```
\sixheterovi[r{2+}]%
 {2==N}{2==O$^{\displaystyle -}$}
\sixheterovi{2==N$^{+}$}{2==O$^{-}$}
```

which yield the following results:

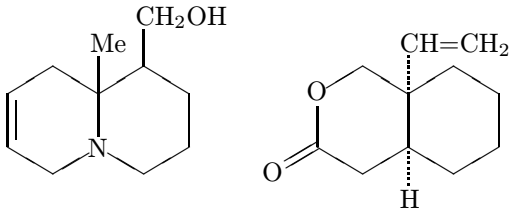


Note that pyridine nuclei with a ring nitrogen at a position other than the top or the bottom cannot be drawn by using such a specific command as `\pyridinev`.

For drawing fused heterocycles with two 6-membered rings, the `\decaheterov` command and related ones are designed to be capable of specifying any hetero-atoms in the nuclei:

```
\decaheterov[g]{9==N}{%
 1B==CH$_2$OH;{{10}B}==Me}
\decaheterov[]{7==O}{6D==O;9A==H;%
 {{10}A}==CH=CH$_2$}
```

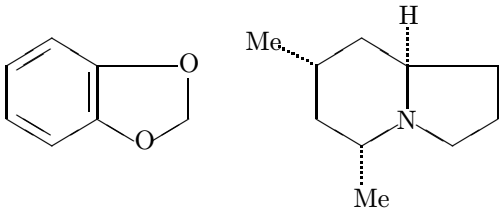
yields



The `\nonaheterov` command and related ones are also able to specify any hetero-atoms in the nuclei of fused heterocycles. For example:

```
\nonaheterov[egj]{1==O;3==O}{}
\nonaheterov[]{9==N}{5A==Me;7A==Me;8A==H}
```

produces



6 Further Techniques

More complicated structural formulas can be constructed by combining two or more structures created by XyMTEX commands. These structures are combined within an outer `picture` environment, since XyMTEX is based on the LaTeX `picture` environment and two or more `picture` environments can be nested. The technique is discussed in Chapter 14 of the on-line manual for XyMTEX [8].

Another technique for constructing complicated formulas is to use a XyMTEX command inside the

Table 2: Original XƎMTEX Files in NIFTY-Serve

no.	size (bytes)	data name
204	76093	`xymtexi.lzh`—XƎMTEX. An introduction (in Japanese)
202	77281	`xymtexj.lzh`—XƎMTEX by Example (in Japanese)
201	299053	`xymtex.lzh`—XƎMTEX for drawing chem. structures

argument of another XƎMTEX command. This technique is discussed in Chapter 15 of the same on-line manual cited above.

The book which was the original cause of all this work on TEX for chemistry [1] contains several commands for the use of chemical fields, e.g., counters for compounds and derivatives, various reaction arrows, `parbox`-like boxes for structural formulas, and chemical equation environments. These commands combined with the XƎMTEX ones are useful for drawing reaction schemes of multistep syntheses. Many illustrative examples are described in the Japanese edition of the XƎMTEX on-line manual. [9]

7 Program Availability

The original location supported by the author is the NIFTY-Serve archives (FPRINT library No. 7), from which you can take the compressed packages shown in Table 2. The XƎMTEX files as listed in Table 1 (including the reference manual of about 120 pages [8]) are also available on CTAN:[2]

```
tex-archive/macros/
latex209/contrib/xymtex
```

The present article has been typeset by using XƎMTEX within LATEX, where the top declaration of the document file is as follows:

```
\documentstyle[epic,carom,hetarom]%
{ltugboat}
```

References

[1] S. Fujita, *LATEX for Chemists and Biochemists. A Guide for Preparing Papers with Personal Computers* [in Japanese], Tokyo Kagaku Dozin, Tokyo (1993).

[2] R.T. Haas and K.C. O'Kane, "Typesetting chemical structure formulas with the text formatter TEX/LATEX", *Computers and Chemistry*, Vol. 11, No. 4, pp. 251–271 (1987).

[3] S. Podar, "Enhancements to the picture environment of LATEX", On-line manual for Version 1.2 dated July 14, 1986. Manual for Version 1.2 dated July 14, 1986.

[4] M.J. Wichura, *The PICTEX Manual*, TEX Users Group, Providence (1992).

[5] A.C. Norris and A.L. Oakley, "Electronic publishing and chemical text processing", in M. Clark, ed., *TEX. Applications, uses, methods*, Ellis Horwood, Chichester (1990), pp. 207–225; see also M. Ramek, "Chemical structure formulae and x/y diagrams with TEX", pp. 227–258.

[6] C. Kwok, "EEPIC. Extensions to epic and LATEX picture environment", On-line manual dated August 14, 1988.

[7] S. Fujita, "Typesetting structural formulae with the text formatter TEX/LATEX", *Computers and Chemistry*, Vol. 18, No. 2, pp. 109–116 (1994).

[8] S. Fujita, "XƎMTEX. A macro package for typesetting chemical structural formulas", On-line manual for XƎMTEX for Version 1.00 (1993). [Available from CTAN, see article.]

[9] S. Fujita, "XƎMTEX by example" [in Japanese], On-line manual for XƎMTEX (1994). [Available from NIFTY-Serve, see Table 2.]

◇ Shinsaku Fujita
Ashigara Research Laboratories
Fuji Photo Film Co., Ltd.
Minami-Ashigara, Kanagawa-ken, 250-01 Japan
Email: `hbh00445@niftyserve.or.jp`

[2] Thanks to the kind volunteer efforts of Ms. M. Burbank and Ms. H. Ase.