XyM Markup Language (XyMML) for Electronic Communication of Chemical Documents Containing Structural Formulas and Reaction Schemes

Shinsaku Fujita

Department of Chemistry and Materials Technology, Kyoto Institute of Technology, Matsugasaki, Sakyoku, Kyoto 606-8585 Japan

Received April 5, 1999

The XyM markup language (XyMML) is proposed as a tool for the electronic communication of chemical structural formulas and reaction schemes. The XyMML system regards a structural formula as a derivative of a skeleton, where substitution and replacement are designated by using a substitution list (a ⟨subslist⟩ tag), an atom list (an ⟨atomlist⟩ tag), and a bond list (a ⟨bondlist⟩ tag). Structural formulas generated by XyMMLs can be collected to give a reaction scheme, in which cross-references to every XyMML are available by using the technique of SGML attributes.

1. INTRODUCTION

Markup languages such as SGML (standard generalized markup language)[1−3] have been significant influences in the publication processes of academic journals and books. One such significant influence is that a collection of SGML documents for articles has been regarded as a database besides being used for printing processes. For example, *Bulletin of the Chemical Society of Japan* (BCSJ) has been published on the basis of the SGML technique,[4] where it has adopted TeX/LaTeX[5,6] as a typesetting system after conversion from SGML to TeX/LaTeX. The SGML activities of the American Chemical Society have also been reported.[7] The effective preparation of electronic manuscripts based on SGML requires a well-defined document type definition (DTD). The DTD used by BCSJ, which has been developed by the cooperation of the Chemical Society of Japan, the American Chemical Society, and the Royal Society of Chemistry, is different from the ISO 12083 DTDs[8] because of careful analysis of the chemical journal data.

On the other hand, the progress of the world-wide web (WWW) system for the Internet has become remarkable in the last 5 years of this decade, where home pages and related items are based on HTML (hypertext markup language).[9−12] The next version of the Internet language will be XML (extensible markup language),[13] which takes over the strong distinction between context and presentation in SGML so that XML has a more logical nature due to the original specification of SGML than the layout-oriented HTML does. Since the HTML or XML is an application of SGML, SGML-based journals such as BCSJ have become ready to be offered for public perusal after conversion from SGML to HTML/XML.[14]

The trends described above[15] mean that a database, publication, and electronic network can be integrated into a comprehensive discipline by virtue of SGML-based techniques, where the effective electronic communication of chemical documents becomes more and more important. However, two of the most difficult problems remain unsolved for SGML-based application systems for chemistry: how we represent mathematical equations and how we represent chemical structural formulas. The former problem has been discussed so as to be answered by the proposed recommendation for MathML,[16] which defines an XML compliant markup language for describing equation content and presentation. Chemical markup language (CML)[17] has been proposed to treat the latter problem, where a chemical structural formula is represented as a kind of connection table. However, the CML does not explicitly provide for reaction schemes. The public DTDs[8] and BCSJ's DTD described above treat structual formulas and reaction schemes as binary image data, because they have no elements for describing these chemical items.

We have reported the XyMTeX system for typsetting chemical structural formulas.[18−22] A limited set of commands defined in the XyMTeX has been adopted in a WWW browsing system.[23] More recently, we have extended the XyMTeX command system into the XyM notation system in order that it serves as a common language for electronic communication and documentation.[24] Although it has been implemented as a TeX/LaTeX application, the XyM notation system has the potentiality of applying in a WWW browsing system in light of the Java-applet[25] or Active-X technique,[26] where the effective electronic communication of structural formulas will be realized. In contrast, a more direct approach is available if the XyM notation system is restructured to obey the SGML specification. By this approach, we have arrived at a new markup language named XyM markup language (XyMML), which is able to be applied to organic structures as well as to reaction schemes. The proposal of XyMML is the main subject of the present paper,[27] where synonyms representing a chemical structure will be discussed in terms of XyMML elements. The implementation of XyMML as an SGML-TeX/LaTeX application will be reported as an additional subject.

2. XYM ELEMENT

2.1. General Structure of XyM Element. The XyM markup language (XyMML) proposed here regards an

10.1021/ci990029q CCC: $18.00
Published on Web 08/24/1999

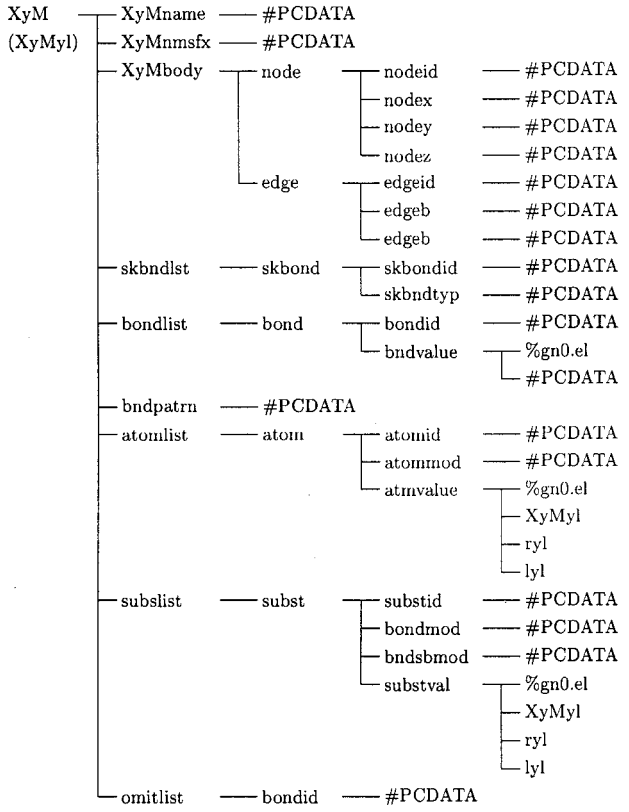


Figure 1. Structures of XyM and XyMyl.

organic structure as a derivative of a skeleton. Such an organic structure (i.e. a derivative of a skeleton) is considered to be represented by an appropriate diagram selected from a set of possible diagrams. In other words, the XyMML system conceptually discriminates an organic structure from a structural diagram. Then, the latter is assigned to a XyM element (⟨XyM⟩ and ⟨/XyM⟩), shown in Figure 1. The XyM element consists of a skeleton name (XyMname) for designating a generic, specific, or carbocyclic aromatic skeleton; a skeleton suffix (XyMnmsfx) for giving a suffix; a skeleton body (XyMbody) for giving connection data; a skeleton bond list (skbndlst) for showing the stereochemistry of skeletal bonds; a bond list (bondlist) for specifying skeletal unsaturation; a bond pattern (bndpatrn) for describing a bond pattern; an atom list (atomlist) for denoting atom replacement; a substituent list (subslist) for setting substituents; and an omitted bond list (omitlist) for specifying skeletal bonds to be omitted. The skeleton body (XyMbody), which is not always required because of implicit definition, represents the *x*,*y*-coordinates and connectivities of the skeleton represented by XyMname and XyMnmsfx.

2.2. Derivation Synonyms and Skeleton-Suffix Synonyms. An organic structure can be represented by one or more IUPAC names that are equivalent chemically. This nature of the IUPAC nomenclature is referred to as being nonunique but unambiguous.[28] It is convenient for traditional (oral and literal) communication, since a structure may be variously discussed from different points of view. For example, 3-carboxysulfanilic acid and 5-sulfoanthranilic acid are synonymous IUPAC names for designating the structure of a sole compound. These types of synonyms are tentatively called "derivation synonyms". In light of the methodology of XyMML, derivation synonyms are specified by means

```
<XyM>
<XyMname>sixhetero</XyMname><XyMnmsfx>v</XyMnmsfx>
<atomlist><atom><atomid>1</atomid>
<atmvalue>S</atmvalue></atom>
<atom><atomid>4</atomid>
<atmvalue>S</atmvalue></atom></atomlist>
<subslist></subslist>
</XyM>
```

```
<XyM>
<XyMname>sixhetero</XyMname><XyMnmsfx>v</XyMnmsfx>
<atomlist><atom><atomid>2</atomid>
<atmvalue>S</atmvalue></atom>
<atom><atomid>5</atomid>
<atmvalue>S</atmvalue></atom></atomlist>
<subslist></subslist>
</XyM>
```

```
<XyM>
<XyMname>sixhetero</XyMname><XyMnmsfx>v</XyMnmsfx>
<atomlist><atom><atomid>3</atomid>
<atmvalue>S</atmvalue></atom>
<atom><atomid>6</atomid>
<atmvalue>S</atmvalue></atom></atomlist>
<subslist></subslist>
</XyM>
```

Figure 2. Derivation synonyms due to an atom list.

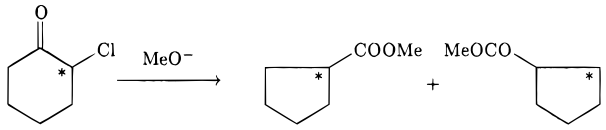


Figure 3. Two formulas in a Favorskii rearrangement.

```
<XyM>
<XyMname>sixhetero</XyMname><XyMnmsfx>v</XyMnmsfx>
<atomlist></atomlist>
<subslist></subslist>
</XyM>
```

```
<XyM>
<XyMname>sixhetero</XyMname><XyMnmsfx>h</XyMnmsfx>
<atomlist></atomlist>
<subslist></subslist>
</XyM>
```

Figure 4. Skeleton-suffix synonyms for drawing diagrams of vertical and horizontal cyclohexane.

of a substitution list (a ⟨subslist⟩ element), an atom list (an ⟨atomlist⟩ element), and/or a bond list (a ⟨bondlist⟩ element). Figure 2 illustrates derivation synonyms due to an atom list in generating the formulas of 1,4-dithiane. Note that the XyMML system discriminates a structure (e.g. 1,4-dithiane) from its structural diagrams (e.g. the three diagrams in Figure 2).

It should be noted here that two or more formulas for representing the same molecule are necessary to describe organic reactions in various contexts. For example, a Favorskii rearrangement involves the scrambling of an isotopic carbon label (∗),[29] where the products are the same but should be differentiated, as shown in Figure 3.

Since the XyMML system discriminates a structure from its structural diagrams, the structure of a skeleton is also discriminated from the corresponding structural diagrams. This provides other types of synonyms named "skeleton-suffix synonyms", because XyMML takes account of structural diagrams (not structures directly). For example, cyclohexane as a skeleton is illustrated by means of hexagonal diagrams of vertical, horizontal, and, if necessary, other intermediate directions, as found in Figure 4. These structural diagrams are regarded as coming from skeleton-suffix synonyms, since they represent a sole structure "cyclohexane". Any one of them may be selected as a representation of the structure for the purpose of electronic communication. In other words, the structural diagrams of a skeleton should be discriminated from each other, though they represent a sole structure in an abstract sense.

Table 1. Generic Skeletons

	generic skeleton for cyclic compounds	
ring size	skeleton name (XyMname)	skeleton suffix (XyMnmsfx)
3	threehetero	v, vi, h, hi
4	fourhetero	(none)
5	fivehetero	v, vi, h, hi
6	sixhetero	v, vi, h, hi
5, 6	nonahetero	v, vi, h, hi
6, 6	decahetero	v, vi, vb, vt, h, hi

	generic skeleton for polymethylenes	
length	skeleton name (XyMname)	skeleton suffix (XyMnmsfx)
2	dimethylene	(none), i
3	trimethylene	(none), i
4	tetramethylene	(none), i
5	pentamethylene	(none), i
6	hexamethylene	(none), i
7	heptamethylene	(none), i
8	octamethylene	(none), i
9	nonamethylene	(none), i
10	decamethylene	(none), i

Table 2. Specific Skeletons for Carbocyclic Compounds

ring size	skeleton name (XyMname)	skeleton suffix (XyMnmsfx)
3	cyclopropane	v, vi, h, hi
4	cyclobutane	(none)
5	cyclopentane	v, vi, h, hi
6	cyclohexane	v, vi, h, hi
5, 6	indane	v, vi, h, hi
6, 6	decaline	v, vi, vb, vt, h, hi
6, 6, 6	hanthracene	v
	hphenanthrene	v

Table 3. Specific Skeletons for N-Containing Heterocycles

nitrogens	skeleton name (XyMname)	skeleton suffix (XyMnmsfx)
1	pyridine	v, vi, h, hi
2, 4	pyrazine	v, h
1, 3	pyrimidine	v, vi, h, hi
1, 2	pyridazine	v, vi, h, hi
1, 3, 5	triazine	v, vi, h, hi

Table 4. Carbocyclic Aromatic Skeletons

ring size	skeleton name (XyMname)	skeleton suffix (XyMnmsfx)
6	bzdr	v, h
6, 6	naphdr	v, vb, vt, h
	tetraline	v, vb, vt, h
6, 6, 6	anthracene	v
	phenanthrene	v

To realize this methodology, a skeleton is represented by the combination of a skeleton name (XyMname) and a skeleton suffix (XyMnmsfx), as shown in Tables 1−4. The XyMname element refers to a structure but not to a structural diagram, while the XyMnmsfx element assures the one-to-one correspondence between a XyMML and a structural diagram as long as derivation synonyms and skeleton-suffix synonyms are taken into consideration.

2.3. Generic and Specific Skeletons. A skeleton is constructed by combining a skeleton name (a XyMname element) with a skeleton suffix (a XyMnmsfx element). Table 1 lists generic skeletons that generate three- to six-membered cyclic compounds, fused cyclic compounds, and two- to ten-carbon polymethylenes, where an arbitrary set of atoms in each skeleton can be replaced by heteroatoms. For example, the XyMname element "fivehetero" with the XyMnmsfx element "v" produces cyclopentane, tetrahydrofuran, and 1,3-dioxolane, as shown in Figure 5. Thus, these five-membered cyclic compounds are considered to belong to a derivative family derived from the generic skeleton, "fiveheterov".

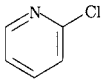
```
<XyM>
<XyMname>fivehetero</XyMname><XyMnmsfx>v</XyMnmsfx>
<atomlist></atomlist>
<subslist></subslist>
</XyM>
```

```
<XyM>
<XyMname>fivehetero</XyMname><XyMnmsfx>v</XyMnmsfx>
<atomlist><atom><atomid>1<atomid>
<atmvalue>0</atmvalue></atom></atomlist>
<subslist></subslist>
</XyM>
```

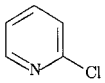
```
<XyM>
<XyMname>fivehetero</XyMname><XyMnmsfx>v</XyMnmsfx>
<atomlist><atom><atomid>2<atomid>
<atmvalue>0</atmvalue></atom>
<atom><atomid>5<atomid>
<atmvalue>0</atmvalue></atom></atomlist>
<subslist></subslist>
</XyM>
```

Figure 5. Derivative family coming from the fiveheterov skeleton.

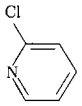
```
<XyM>
<XyMname>pyridine</XyMname><XyMnmsfx>v</XyMnmsfx>
<subslist><subst><substid>2</substid>
<substval>Cl</substval></subst></subslist>
</XyM>
```



```
<XyM>
<XyMname>pyridine</XyMname><XyMnmsfx>vi</XyMnmsfx>
<subslist><subst><substid>2</substid>
<substval>Cl</substval></subst></subslist>
</XyM>
```



```
<XyM>
<XyMname>pyridine</XyMname><XyMnmsfx>h</XyMnmsfx>
<subslist><subst><substid>2</substid>
<substval>Cl</substval></subst></subslist>
</XyM>
```



```
<XyM>
<XyMname>pyridine</XyMname><XyMnmsfx>hi</XyMnmsfx>
<subslist><subst><substid>2</substid>
<substval>Cl</substval></subst></subslist>
</XyM>
```

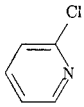


Figure 6. Skeleton-suffix synonyms of XyMMLs for drawing 2-chloropyridine diagrams.

The XyMnmsfx element is used to designate the directions of structural diagrams. For example, the upper instance in Figure 4 generates a structural formula of cyclohexane, where the skeleton "sixheterov" is generated by the combination of "sixhetero" (for XyMname) and "v" (for XyMnmsfx); and its atom list (atomlist) and substituent list (subslist) are vacant. On the other hand, the horizontal counterpart "sixheteroh" is generated by changing " v" (for XyMnmsfx) into "h", as shown at the bottom of Figure 4. As a result, problems concerning skeleton-suffix synonyms are solved by using the XyMnmsfx element.

Table 2 lists specific skeletons that generate three- to six-membered carbocyclic compounds and fused carbocyclic compounds. Since each vertex of these skeletons is fixed to have an implicit carbon atom, a XyM element corresponding to a skeleton of this type has not taken an atom list (atomlist).

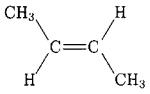
Table 3 lists specific skeletons for generating N-containing six-membered cyclic compounds. Since each vertex of these skeletons is fixed to take a given heteroatom or an implicit carbon atom, a XyM element corresponding to a skeleton of this type takes no atom list. In other words, the atom list is fixed so as to give a distinct compound.

Figure 6 illustrates four diagrams of 2-chloropyridine, which are generated from the XyMname element "pyridine" with the respective XyMnmsfx elements, "v", "vi", "h", and "hi". These XyMMLs exemplify skeleton-suffix synonyms in a specific skeleton for representing N-containing heterocycles. Note that the four XyMMLs are different only in

Table 5. Aliphatic Skeletons

skeleton name (XyMname)	skeleton suffix (XyMnmsfx)	function
tetrahedral	(none), d	tetravalent carbon
trigonal	r, R, l, L, u, U, d, D	trigonal carbon
ethylene	v, h, V, H	ethylene

```
<XyM>
<XyMname>ethylene</XyMname><XyMnmsfx>h</XyMnmsfx>
<atomlist><atom><atomid>1</atomid>
<atmvalue>C</atmvalue></atom>
<atom><atomid>2</atomid>
<atmvalue>C</atmvalue></atom></atomlist>
<subslist><subst><substid>1</substid>
<substval>CH<inf>3</inf></substval></subst>
<subst><substid>2</substid>
<substval>H</substval></subst>
<subst><substid>3</substid>
<substval>CH<inf>3</inf></substval></subst>
<subst><substid>4</substid>
<substval>H</substval></subst></subslist>
</XyM>
```




```
<XyM>
<XyMname>ethylene</XyMname><XyMnmsfx>h</XyMnmsfx>
<atomlist><atom><atomid>1</atomid>
<atmvalue>C</atmvalue></atom>
<atom><atomid>2</atomid>
<atmvalue>N</atmvalue></atom></atomlist>
<subslist><subst><substid>1</substid>
<substval>CH<inf>3</inf></substval></subst>
<subst><substid>2</substid>
<substval>H</substval></subst>
<subst><substid>3</substid>
<substval>OH</substval></subst></subslist>
</XyM>
```

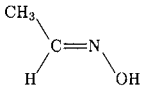


Figure 7. XyMMLs for drawing *cis*-2-butene and *syn*-acetaldehyde oxime.

their XyMnmsfx elements. Thus, the orders of locant numbering in the pyridinev (or pyridineh) and in the pyridinevi (or pyridinehi) are decided to be inverse so that the resulting XyMMLs take the same subslist element.

Table 4 lists specific skeletons for generating carbocyclic aromatic compounds, since they frequently appear in the chemical literature. A XyM element corresponding to a skeleton of this type takes a bond pattern (bndpatrn) in place of a bond list (bondlist).

Table 5 lists specific skeletons for drawing aliphatic skeletons. The "tetrahedral" skeleton is used to draw a tetrahedral carbon with horizontal and vertical valencies, while the "tetrahedrald" skeleton is used to draw the one with diagonal valencies. The XyMname elements "trigonal" with the XyMnmsfx elements "r", "l", "u", and "d" are used to draw trigonal carbons of right-hand, left-hand, upward, and downward types with a narrow bond angle. The corresponding uppercase XyMnmsfx elements indicate trigonal carbons with a broad bond angle. These skeletons take only a subslist element.

The XyMname elements "ethylene" with the XyMnmsfx elements "h" and "v" (Table 5) are used to draw ethylene with a narrow bond angle. The corresponding uppercase XyMnmsfx elements indicate ethylene with a broad bond angle. This skeleton takes a bondlist, an atomlist, and a subslist element. Figure 7 shows *trans*-2-butene and *syn*-acetaldehyde oxime drawn by the ethyleneh skeleton.

2.4. Substitution Lists. A ⟨subslist⟩ tag represents the substitution of a skeleton, where each substituent is placed by using a ⟨subst⟩ tag, as illustrated in Figure 8. For example, the first substituent (NO_2) is designated by the code,

```
<subst>
<substid>1</substid>
<bondmod>S</bondmod>
<bndsbmod>A</bndsbmod>
<substval>NO<inf>2</inf></substval>
</subst>
```

```
<XyM>
<XyMname>cyclohexane</XyMname><XyMnmsfx>v</XyMnmsfx>
<subslist><subst><substid>1</substid>
<bondmod>S</bondmod><bndsbmod>A</bndsbmod>
<substval>NO<inf>2</inf></substval></subst>
<subst><substid>1</substid><bondmod>S</bondmod>
<bndsbmod>B</bndsbmod>
<substval>CH<inf>3</inf></substval></subst>
<subst><substid>4</substid>
<substval>CH<inf>3</inf></substval></subst>
</subslist>
</XyM>
```

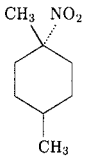


Figure 8. Substitution list for 1-nitro-1,3-dimethylcyclohexane.

```
<XyM>
<XyMname>decahetero</XyMname><XyMnmsfx>v</XyMnmsfx>
<atomlist>
<atom><atomid>6</atomid>
<atmvalue>O</atmvalue></atom>
<atom><atomid>2</atomid>
<atmvalue>NH</atmvalue></atom>
<atom><atomid>8</atomid>
<atmvalue>S</atmvalue></atom>
</atomlist>
<subslist></subslist>
</XyM>
```

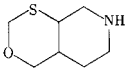


Figure 9. Atom list for a fused heterocycle.

```
<XyM>
<XyMname>cyclohexane</XyMname><XyMnmsfx>v</XyMnmsfx>
<bondlist>
<bond><bondid>a</bondid></bond>
<bond><bondid>c</bondid></bond>
</bondlist>
<subslist></subslist>
</XyM>
```

Figure 10. Bond list for representing unsaturation.

The letter l in the tag ⟨substid⟩ represents a locant number. The letter S in the tag ⟨bondmod⟩ and the letter A in the tag ⟨bndsbmod⟩ are combined to give a bond modifier "SA", which indicates that the bond to be drawn is an α single bond. Such a bond modifier is selected from the list for the XyM notation reported previously.[24] The substituent NO_2 is placed in the ⟨substval⟩ tag.

2.5. Atom Lists. A generic skeleton is capable of generating heterocyclic compounds by using its ⟨atomlist⟩ tag, as illustrated by tetrahydrofuran shown in Figure 5, where an oxygen atom (O) along with its locant number is placed by using an ⟨atom⟩ tag, i.e.,

```
<atomlist><atom><atomid>1<atomid>
<atmvalue>O</atmvalue></atom></atomlist>
```

1,3-Dioxolane in Figure 5 illustrates the use of two ⟨atom⟩ tags in an atom list. A fused heterocyclic case is shown in Figure 9, in which three heteroatoms along with their locant numbers are placed by using ⟨atom⟩ tags.

2.6. Bond Lists. A bond list treats information on skeletal bonds, e.g. bond unsaturation and ring fusion at a bond. The mode of each skeletal bond is described by using a ⟨bond⟩ tag, which contains a ⟨bondid⟩ tag and sometimes a ⟨bndvalue⟩ tag. A ⟨bond⟩ tag without a ⟨bndvalue⟩ tag indicates bond unsaturation. For example, a structural formula of cyclohexane-1,3-diene is generated by the XyMML shown in Figure 10, where the bond list

```
<bondlist>
<bond><bondid>a</bondid></bond>
<bond><bondid>c</bondid></bond>
</bondlist>
```

specifies bonds "a" and "c" to be unsaturated bonds.

2.7. Stereochemistries of Skeletal Bonds. The stereochemistry of a skeletal bond is designated by using an "skbndlst" element, in which an "skbondid" element represents a skeletal bond identifier (alphabet) and an "skbndtyp" element (A or B) indicates whether the skeletal bond at issue is an α- or β-bond. Figure 11 shows the use of the skbndlst

```
<XyM>
<XyMname>fivehetero</XyMname><XyMnmsfx>v</XyMnmsfx>
<skbndlst><skbond><skbondid>a</skbondid>
<skbndtyp>A</skbndtyp></skbond>
<skbond><skbondid>e</skbondid>
<skbndtyp>B</skbndtyp></skbond></skbndlst>
<atomlist><atom><atomid>4</atomid>
<atmvalue>O</atmvalue></atom></atomlist>
<subslist>
<subst><substid>5</substid><bondmod>D</bondmod>
<substval>O</substval></subst>
<subst><substid>1</substid><bondmod>S</bondmod>
<bndsbmod>a</bndsbmod>
<substval>CH<inf>3</inf></substval></subst>
<subst><substid>1</substid><bondmod>S</bondmod>
<bndsbmod>b</bndsbmod>
<substval>CH<inf>3</inf>CH<inf>2</inf></substval>
</subst></subslist>
</XyM>
```

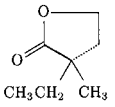


Figure 11. Skeletal bond list for representing skeletal stereochemistry.

```
<XyM>
<XyMname>decahetero</XyMname><XyMnmsfx>v</XyMnmsfx>
<atomlist></atomlist><subslist>
<subst><substid>1</substid><bondmod>D</bondmod>
<substval>O</substval></subst></subslist>
<omitlist><bondid>k</bondid></omitlist>
</XyM>
```

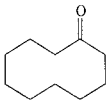
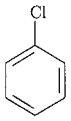


Figure 12. Omit list for representing bond deletion.

```
<XyM id="PhCl" x="400" y="750"
xshift="220" yshift="200">
<XyMname>bzdr</XyMname><XyMnmsfx>v</XyMnmsfx>
<subslist><subst><substid>1</substid>
<substval>Cl</substval></subst>
</subslist>
</XyM>
```



5

Figure 13. Attributes for the XyM element.

element, where the code

```
<skbndlst><skbond><skbondid>a</skbondid>
<skbndtyp>A</skbndtyp></skbond>
<skbond><skbondid>e</skbondid>
<skbndtyp>B</skbndtyp></skbond></skbndlst>
```

specifies bond "a" to be an α-bond (a dotted bond) and bond "e" to be a *β*-bond (a boldfaced bond). The designation by means of the "skbondlst" does not involve a canonical-naming process so that it aims at unambiguity but not at uniqueness. The pursuit of unambiguity and uniqueness is one of the future subjects in the application of XyMML.

2.8. Skeletal Bond Deletion. The "omitlist" element is used to designate bonds to be deleted from a generic skeleton. Figure 12 shows the use of the omitlist element, where the bond "k" (the fused bond) of a decaline ring is deleted by writing ⟨omitlist⟩⟨bondid⟩k⟨/bondid⟩⟨/omitlist⟩.

2.9. Compound Numbers and Their Cross-References. The XyM element takes several attributes if necessary: a compound identifier "id" for giving its compound number, "x" and "y" for giving its size, and "xshift" and "yshift" for adjusting *x*,*y*-coordinates. Figure 13 shows the use of these attributes. When such a compound identifier is written, a compound number is automatically generated under a structural formula. The compound number is referred to by the XyMref element. In this case, the code ⟨XyMref refid="PhCl"⟩ gives such a reference number as 5.

3. XYMYL ELEMENT

3.1. Substitution Derivation for Large Substituents. A structure generated by a XyM element is used as a substituent after an entity "&ylfunc;" is written as a hypothetical substituent (yl function) in its "subslist". For such a case, a XyMyl element is used in place of a XyM element in order to discriminate the substituent from the structure. Figure 14 illustrates a cyclohexyl substituent generated by the ⟨XyMyl⟩ tag, which is placed in the ⟨substval⟩ tag of the outer skeleton generated by a XyM element.

```
<XyM>
<XyMname>cyclohexane</XyMname><XyMnmsfx>v</XyMnmsfx>
<subslist><subst><substid>3</substid>
<substval>
<XyMyl>
<XyMname>cyclohexane</XyMname><XyMnmsfx>v</XyMnmsfx>
<subslist><subst>
<substid>6</substid><substval>&ylfunc;</substval>
</subst></subslist>
</XyMyl>
</substval>
</subst></subslist>
</XyM>
```

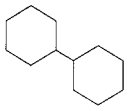


Figure 14. XyMyl element for the cyclohexyl group of bicyclohexyl.

```
<XyM>
<XyMname>cyclohexane</XyMname><XyMnmsfx>v</XyMnmsfx>
<subslist><subst><substid>3</substid>
<substval>
<XyMyl>
<XyMname>sixhetero</XyMname><XyMnmsfx>v</XyMnmsfx>
<atomlist>
<atom><atomid>6</atomid><atmvalue>N</atmvalue></atom>
</atomlist>
<subslist><subst>
<substid>6</substid>
<substval>&ylfunc;</substval></subst>
</subslist>
</XyMyl>
</substval>
</subst>
</subslist>
</XyM>
```

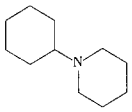


Figure 15. XyMyl element for a heterocyclic substituent.

```
<XyM>
<XyMname>sixhetero</XyMname><XyMnmsfx>v</XyMnmsfx>
<atomlist>
<atom><atomid>3</atomid><atommod>s</atommod>
<atmvalue><XyMyl>
<XyMname>cyclohexane</XyMname><XyMnmsfx>v</XyMnmsfx>
<subslist>
<subst>
<substid>6</substid><substval>&ylfunc;</substval>
</subst>
</subslist>
</XyMyl></atmvalue></atom>
</atomlist>
<subslist></subslist>
</XyM>

<XyM>
<XyMname>sixhetero</XyMname><XyMnmsfx>v</XyMnmsfx>
<atomlist>
<atom><atomid>3</atomid><atommod>h</atommod>
<atmvalue>
<XyMyl>
<XyMname>sixhetero</XyMname><XyMnmsfx>v</XyMnmsfx>
<atomlist>
<atom><atomid>6</atomid>
<atmvalue>N</atmvalue></atom>
</atomlist>
<subslist><subst>
<substid>6</substid>
<substval>&ylfunc;</substval></subst>
</subslist>
</XyMyl>
</atmvalue></atom>
</atomlist>
<subslist></subslist>
</XyM>
```

Figure 16. XyMyl element in an atom list for generating spiro compounds.

A heterocyclic structure generated by a XyM element with a filled atom list is used as a substituent by using a yl function (an entity &ylfunc;). The resulting XyMyl element is placed in the ⟨substval⟩ tag of the outer XyM element. Figure 15 illustrates a morphorino substituent generated by the ⟨XyMyl⟩ tag.

3.2. Atom Derivation for Spiro Ring Replacement. The XyMyl element can be written in the ⟨atomlist⟩ tag of an outer XyM element, as illustrated in Figure 16. The cyclohexyl substituent used in Figure 14 (written in a ⟨subslist⟩ tag) is used alternatively as a cyclohexane spiro group in the first example of Figure 16 (written in an ⟨atomlist⟩ tag), where the code ⟨atommod⟩s⟨/atommod⟩ indicates that the spiro position accommodates an implicit carbon atom. On the other hand, the second example of Figure 16 shows that a XyMyl element in an atom list corresponds to a morphorine spiro group, while it is used as a substituent in a ⟨subslist⟩ tag in Figure 15. The code ⟨atommod⟩h⟨/atommod⟩ indicates that the spiro position accommodates a heteroatom.

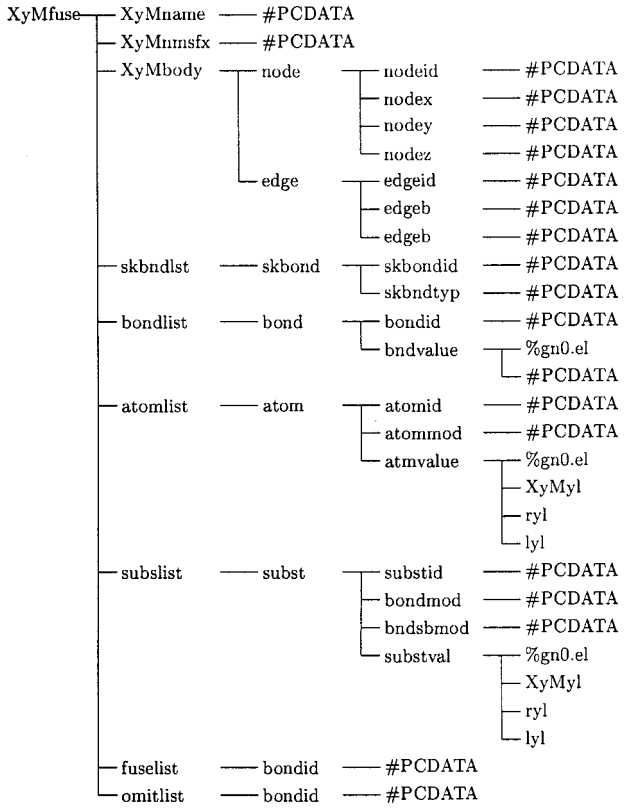


Figure 17. Structures of XyMfuse.

4. XYMFUSE ELEMENT

4.1. Structure of XyMfuse Element. In the XyMML system, a fused ring is regarded as a derivative generated by a bond-derivation process in which a fusing unit is attached to an appropriate bond of a skeleton. Such a fusing unit is generated by a XyMfuse element (⟨XyMfuse⟩ and ⟨/XyMfuse⟩) shown in Figure 17. The XyMfuse element consists of a skeleton name (XyMname), a skeleton suffix (XyMnmsfx), a skeleton body (XyMbody), a skeleton bond list (skbndlst), a bond list (bondlist), an atom list (atomlist), a substituent list (subslist), a fuse-bond list (fuselist), and an omitted bond list (omitlist). The XyMfuse element has no bond pattern (bndpatrn) but takes a fuse-bond list (fuselist) as compared with the XyM element shown in Figure 1. The fuse-bond list is used to designate a bond to be fused. The skeleton name (XyMname) of a XyMfuse element is selected from the fusing skeletons listed in Table 6.

4.2. Bond Lists for Ring Fusion. To treat information on ring fusion, a fusing unit generated by a XyMfuse element is placed in a ⟨bndvalue⟩ tag. For example, a structural formula of a tricyclic compound is generated by the XyMML shown in Figure 18. The six-membered fusing unit generated by the following code

```
<XyMfuse>
<XyMname>sixfuse</XyMname><XyMnmsfx>vi</XyMnmsfx>
<atomlist></atomlist>
<subslist></subslist>
<fuselist><bondid>f</bondid></fuselist>
</XyMfuse>
```

has a reference site "f" designated in the ⟨fuselist⟩ tag, which carries out a ring fusion at the younger-numbered terminal (vertex 6) of bond "f" (the bond between vertices 6 and 1). Note that the numbering of locants starts from the bottom vertex of the fusing unit (a vertical hexagon) in an anti-clockwise manner (1-2-3-4-5-6-1) and that the term "younger" refers to the appearance in the order of the locants. On the other hand, the five-membered fusing unit has a reference site "E" designated in the ⟨fuselist⟩ tag: ⟨fuselist⟩⟨bondid⟩E-⟨/bondid⟩⟨/fuselist⟩, which carries out a ring fusion at the older-numbered terminal (vertex 1) of bond "e" (the bond between vertices 5 and 1). The numbering of locants starts from the top vertex of the fusing unit (a vertical pentagon) in a clockwise manner.

4.3. Nested XyMyl and XyMfuse Elements. XyMyl and XyMfuse elements can be nested in a XyM element. For example, the structure shown in Figure 19 is generated by writing the following code:

```
<XyM>
<XyMname>sixhetero</XyMname><XyMnmsfx>v</XyMnmsfx>
<bondlist>
<bond><bondid>e</bondid>
<bndvalue>
<XyMfuse>
<XyMname>sixfuse</XyMname><XyMnmsfx>v</XyMnmsfx>
<bondlist><bond><bondid>a</bondid></bond>
<bond><bondid>c</bondid></bond>
<bond><bondid>e</bondid></bond>
<bond><bondid>d</bondid>
<bndvalue><XyMfuse>
<XyMname>sixfuse</XyMname><XyMnmsfx>v</XyMnmsfx>
<bondlist><bond><bondid>c</bondid></bond>
<bond><bondid>e</bondid></bond></bondlist>
<atomlist></atomlist><subslist></subslist>
<fuselist><bondid>A</bondid></fuselist>
</XyMfuse></bndvalue></bond></bondlist>
<atomlist></atomlist><subslist></subslist>
<fuselist><bondid>B</bondid></fuselist></XyMfuse>
</bndvalue></bond></bondlist>
<atomlist>
<atom><atomid>3</atomid><atommod>h</atommod>
<atmvalue><XyMyl>
<XyMname>sixhetero</XyMname><XyMnmsfx>v</XyMnmsfx>
<atomlist>
<atom><atomid>6</atomid><atmvalue>N</atmvalue></atom>
<atom><atomid>2</atomid><atommod>s</atommod>
<atmvalue><XyMyl>
<XyMname>sixhetero</XyMname><XyMnmsfx>v</XyMnmsfx>
<bondlist><bond><bondid>b</bondid>
<bndvalue><XyMfuse>
<XyMname>sixfuse</XyMname><XyMnmsfx>v</XyMnmsfx>
<bondlist><bond><bondid>b</bondid></bond>
<bond><bondid>d</bondid></bond>
<bond><bondid>f</bondid></bond></bondlist>
<atomlist></atomlist><subslist></subslist>
<fuselist><bondid>E</bondid></fuselist>
</XyMfuse></bndvalue></bond></bondlist>
<atomlist></atomlist>
<subslist><subst>
<substid>5</substid><substval>&ylfunc;</substval>
</subst></subslist></XyMyl></atmvalue>
</atom></atomlist>
<subslist><subst>
<substid>6</substid><substval>&ylfunc;</substval>
</subst></subslist></XyMyl>
</atmvalue></atom></atomlist>
<subslist><subst>
<substid>1</substid><bondmod>S</bondmod>
<bndsbmod>A</bndsbmod>
<substval>NO<inf>2</inf></substval></subst>
<subst><substid>1</substid><bondmod>S</bondmod>
<bndsbmod>B</bndsbmod>
<substval>CH<inf>3</inf></substval></subst>
<subst><substid>4</substid>
<substval>CH<inf>3</inf></substval></subst>
</subslist>
</XyM>
```

5. RYL AND LYL ELEMENTS

5.1. Structures of Ryl and Lyl Elements. The XyMyl element is a versatile tool for generating a substituent to be directly linked to a substitution site of a mother skeleton. On the other hand, the ryl and lyl elements are designed to treat cases in which a substituent is linked to a substitution site through an intervening unit such as $NHSO_2$ (Figure 20). The first tag ⟨lunit⟩ indicates a linking unit (⟨lunitval⟩ tag) with a locant (⟨locant⟩ tag), showing the slope (0−8) of a left or right incidental bond. The designation of the slope is based on the method described previously for \ryl and \lyl

Table 6. Fusing Skeletons

ring size	skeleton name (XyMname)	skeleton suffix (XyMnmsfx)
3	threefuse	v, vi, h, hi
4	fourfuse	(none)
5	fivefuse	v, vi, h, hi
6	sixfuse	v, vi, h, hi

```
<XyM>
<XyMname>cyclohexane</XyMname><XyMnmsfx>v</XyMnmsfx>
<bondlist>
<bond><bondid>a</bondid>
<bndvalue>
<XyMfuse>
<XyMname>sixfuse</XyMname><XyMnmsfx>vi</XyMnmsfx>
<atomlist></atomlist>
<subslist></subslist>
<fuselist><bondid>f</bondid></fuselist>
</XyMfuse>
</bndvalue>
</bond>
<bond><bondid>c</bondid>
<bndvalue>
<XyMfuse>
<XyMname>fivefuse</XyMname><XyMnmsfx>vi</XyMnmsfx>
<atomlist></atomlist>
<subslist></subslist>
<fuselist><bondid>E</bondid></fuselist>
</XyMfuse>
</bndvalue>
</bond>
</bondlist>
<subslist></subslist>
</XyM>
```

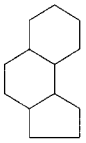


Figure 18. XyMfuse elements in a bond list for representing ring fusion.

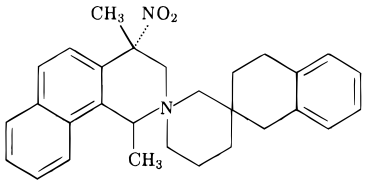


Figure 19. Nested XyMyl and XyMfuse elements.

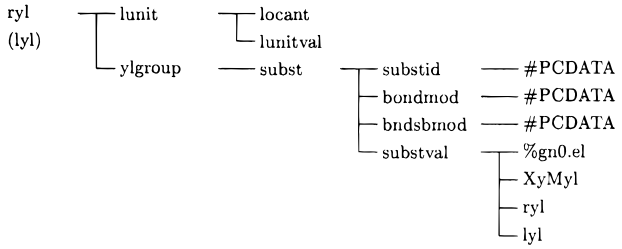


Figure 20. Structures of ryl and lyl elements.

```
<XyM>
<XyMname>naphdr</XyMname><XyMnmsfx>h</XyMnmsfx>
<subslist><subst>
<substid>4</substid>
<substval>
<ryl>
<lunit><locant>4</locant>
<lunitval>NHSO<inf>2</inf></lunitval></lunit>
<ylgroup>
<subst><substid>4</substid>
<substval>
<XyMyl>
<XyMname>bzdr</XyMname><XyMnmsfx>h</XyMnmsfx>
<subslist><subst><substid>1</substid>
<substval>&ylfunc;</substval></subst>
</subslist>
</XyMyl>
</substval></subst>
</ylgroup>
</ryl>
</substval></subst></subslist>
</XyM>
```

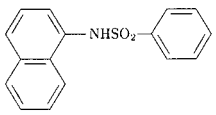


Figure 21. Ryl element in a substitution list.

notations of the XyM notation system.[24] The second tag ⟨ylgroup⟩ accommodates a substituent produced by a XyMyl element.

5.2. Ryl and Lyl Elements in Substitution Lists. A substituent generated by an "ryl" or "lyl" element is placed into a ⟨subslist⟩ tag in order to draw a derivative having an intervening unit. Figure 21 illustrates the usage of an ⟨ryl⟩ tag for drawing the $NHSO_2$ unit of a benzenesulfonamido group.

```
<XyM>
<XyMname>decahetero</XyMname><XyMnmsfx>hi</XyMnmsfx>
<atomlist><atom><atomid>1</atomid>
<atommod>h</atommod>
<atmvalue>
<ryl>
<lunit><locant>4</locant>
<lunitval>NSO<inf>2</inf></lunitval></lunit>
<ylgroup>
<subst><substid>4</substid>
<substval>
<XyMyl>
<XyMname>bzdr</XyMname><XyMnmsfx>h</XyMnmsfx>
<subslist><subst><substid>1</substid>
<substval>&ylfunc;</substval></subst>
</subslist>
</XyMyl>
</substval></subst>
</ylgroup>
</ryl>
</atmvalue></atom></atomlist>
<subslist></subslist>
</XyM>
```

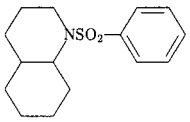


Figure 22. Ryl element in an atom list.

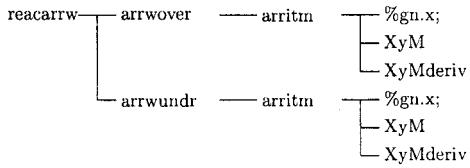


Figure 23. Structure of the reacarrw element.

5.3. Ryl and Lyl Elements in Atom Lists. A substituent generated by an "ryl" or "lyl" element can be placed into an ⟨atomlist⟩ tag in order to draw another type of derivative having an intervening unit. Figure 22 illustrates this technique, where an ⟨ryl⟩ tag for drawing the NSO_2Ph unit is placed at a vertex of a "decahetero" ring.

6. XYMDFORM ELEMENT FOR DRAWING REACTION EQUATIONS

6.1. Reaction Arrows. Reaction arrows of various types are involved in reaction equations, schemes, and so on. They are generated by using the "reacarrw" element, the structure of which is shown in Figure 23. The ⟨arrwover⟩ tag is to place items over a reaction arrow, while the ⟨arrwundr⟩ tag is to place items under a reaction arrow. The "%gn.x" entity appearing in Figure 23 is defined by the code

```
<!ENTITY % gn.x "#PCDATA|ul|bf|it|verb|inf|sup|reacarrw|chemrel">
```

which contains the following elements: "ul" (underlined), "bf" (boldfaced), "it" (italic), "verb" (verbatim), "inf" (subscript), "sup" (superscript), "reacarrw" (reaction arrow), and "chemrel" (relation symbol).

The "reacarrw" element takes four attributes: "arrwtype" for designating an arrow type (default type = r), "length" for designating an arrow length (default value = 400), "xshift" for adjusting an *x*-coordinate (default value = 0), and "yshift" for adjusting a *y*-coordinate (default value = 0). The attributes, length, xshift, and yshift, are positive integer parameters, which are multiplied by a unit length (default value, 0.1 pt) to give values adjusting the places to be drawn. For example, the XyMML represented by

```
<reacarrw arrwtyp="r" length="600" xshift="10" yshift="60">
<arrwover><arritm>HCl</arritm></arrwover>
<arrwundr><arritm>H<inf>2</inf>O</arritm></arrwundr>
</reacarrw>
```

generates a rightward arrow of length 600 units that is shifted rightward by 10 units and upward by 60 units, where HCl is placed over the arrow and H_2O is placed under the arrow. These attributes can be omitted so that default values are selected. We have the following arrow types: r = right

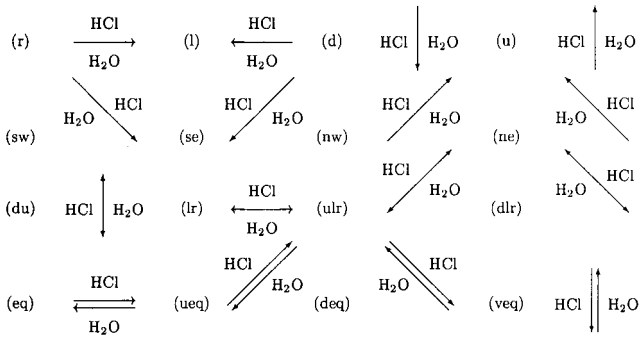


Figure 24. Reaction arrows of various types.

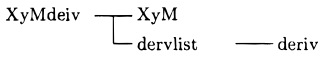
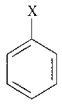


Figure 25. Structure of the XyMderiv element.

```
<XyMderiv>
<XyM id="PhX" x="400" y="750"
xshift="220" yshift="200">
<XyMname>bzdr</XyMname><XyMnmsfx>v</XyMnmsfx>
<subslist><subst><substid>1</substid>
<substval>X</substval></subst></subslist>
</XyM>
<dervlist>
<deriv>X = F</deriv>
<deriv>X = Cl</deriv>
<deriv>X = OCH<inf>3</inf></deriv>
</dervlist>
</XyMderiv>
```



6
a: X = F
b: X = Cl
c: X = OCH_3

Figure 26. Derivatives by the XyMderiv element.

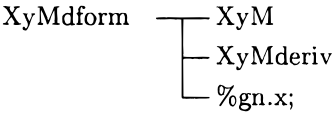


Figure 27. Structure of the XyMdform element.

arrow, l = left arrow, lr = left−right arrow, d = down arrow, u = up arrow, du = down−up arrow, eq = equilibrium arrow, veq = vertical equilibrium arrow, deq = down equilibrium arrow, leq = up equilibrium arrow, dlr = down left−right arrow, ulr = up left−right arrow, sw = southwest arrow, se = southeast arrow, nw = northwest arrow, and ne = northeast arrow. Figure 24 lists all of these reaction arrows with the attributes of default values. Although the drawing of reaction arrows more or less requires layout data, the ⟨reacarrw⟩ tag maintains the logical nature of mapping reactants onto products to a large extent by treating the layout data as attributes.

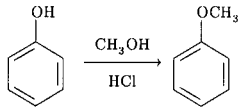
6.2. Derivatives. The "XyMderiv" element is used to specify derivatives of a structure produced by a XyM element. The structure of the XyMderiv element is shown in Figure 25, where the "dervlist" element accommodates substituents for such derivatives by means of inner "deriv" elements. Figure 26 shows an example of the use of a XyMderiv element. Compound numbers for fluorobenzene (**6a**), chlorobenzene (**6b**), and methoxybenzene (**6c**) are called by writing such codes as ⟨XyMref refid="PhX" s="a"⟩.

6.3. Display Formulas. A display formula containing structural formulas (generated by the XyM element) and reaction arrows (generated by the reacarrw element) is generated by the XyMdform element, the structure of which is shown in Figure 27.

Figure 28 shows an example use of the XyMdform element, generating a display formula (eq 1). The resulting formula is referred to by the ref element with a reference key "eq1" (⟨ref refid="eq1"⟩). Phenol (**7**) and methoxybenzene (**8**) in the display formula can be distinctly referred to by using the XyMref elements (⟨XyMref refid="PhOH"⟩ and ⟨XyMref refid="PhOMe"⟩). A display formula generated by the XyMdform element can contain simple characters, as shown in eq 2 of Figure 28. The ⟨chemrel⟩ tag is used to give spaces before and after the tag.

```
Equation <ref refid="eq1"> shows that
phenol (<XyMref refid="PhOH">)
can be easily converted into methoxybenzene
(<XyMref refid="PhOMe">).
<XyMdform id="eq1">
<XyM id="PhOH" x="400" y="750"
xshift="220" yshift="200">
<XyMname>bzdr</XyMname><XyMnmsfx>v</XyMnmsfx>
<subslist><subst><substid>1</substid>
<substval>OH</substval></subst></subslist>
</XyM><reacarrw arrwtyp="r" length="500"
xshift="10" yshift="60">
<arrwover>
<arritm>CH<inf>3</inf>OH</arritm></arrwover>
<arrwundr><arritm>HCl</arritm></arrwundr>
</reacarrw><XyM id="PhOMe" x="400" y="750"
xshift="220" yshift="200">
<XyMname>bzdr</XyMname><XyMnmsfx>v</XyMnmsfx>
<subslist><subst><substid>1</substid>
<substval>OCH<inf>3</inf></substval>
</subst></subslist>
</XyM>
```

Equation 1 shows that phenol (**7**) can be easily converted into methoxybenzene (**8**).



7 **8** (1)

```
<XyMdform id="eq2">
HCl + H<inf>2</inf>O
<chemrel><reacarrw arrwtyp="eq" length="200">
<arrwover></arrwover>
<arrwundr></arrwundr></reacarrw></chemrel>
H<inf>3</inf>O<sup>+</sup> + Cl<sup>-</sup>
</XyMdform>
```

$$HCl + H_2O \rightleftarrows H_3O^+ + Cl^- \quad (2)$$

Figure 28. Display formulas by the XyMdform element.

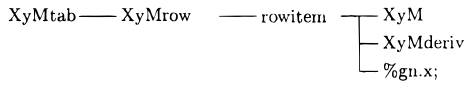


Figure 29. Structure of the XyMtab element.

7. TABULAR SCHEMES CONTAINING STRUCTURAL FORMULAS

7.1. Structure of XyMtab Element. A reaction scheme containing structural formulas is generated by using the "XyMtab" element, the structure of which is shown in Figure 29. The XyMtab element has an attribute "xymcols" for designating a column alignment, which involves the column number of letters selected from c (centering), l (flush left), and r (flush right). For example, the code ⟨XyMtab xymcols="ccccc"⟩ indicates five columns of centering mode. The XyMtab element involves several XyMrow elements, each of which contains the column number of rowitem elements. Each rowitem element may contain structural formulas generated by XyM elements.

7.2. Sample of XyMtab Element. Figure 30 shows a reaction scheme generated by a XyMtab element, which consists of two rows due to a XyMrow element. Each row involves five cells (due to a rowitem element), each of which contains a structural formula due to a XyM element or a reaction arrow due to a reacarrw element. The reference numbers of the compounds in the tabular scheme, i.e. **9** for chlorobenzene, **10** for phenol, **11** for methoxybenzene, and

```
<XyMtab xymcols="ccccc">
<!-- ==========the first row==================== -->
<XyMrow>
<rowitem>
<XyM id="aPhCl" x="400" y="750" xshift="220" yshift="200">
<XyMname>bzdr</XyMname><XyMnmsfx>v</XyMnmsfx>
<subslist><subst><substid>1</substid><substval>Cl</substval>
</subst></subslist></XyM>
</rowitem>
<rowitem>
<reacarrw arrwtyp="r" length="600" xshift="10" yshift="60">
<arrwover><arritm>H<inf>2</inf>O</arritm></arrwover>
<arrwundr><arritm>High press.</arritm></arrwundr></reacarrw>
</rowitem>
                    (abbreviated)
</XyMrow>
<!-- ==========the second row==================== -->
<XyMrow>
<rowitem></rowitem><rowitem></rowitem><rowitem></rowitem>
<rowitem>
<reacarrw arrwtyp="sw" length="400" xshift="0" yshift="300">
<arrwover><arritm>HNO<inf>3</inf></arritm></arrwover>
<arrwundr></arrwundr></reacarrw>
</rowitem>
<rowitem>
<XyM id="PhOHNO2" x="400" y="850" xshift="220" yshift="0">
<XyMname>bzdr</XyMname><XyMnmsfx>v</XyMnmsfx>
<subslist><subst><substid>1</substid><substval>OH</substval>
</subst><subst><substid>4</substid>
<substval>NO<inf>2</inf></substval></subst></subslist>
</XyM>
</rowitem>
</XyMrow>
</XyMtab>
```

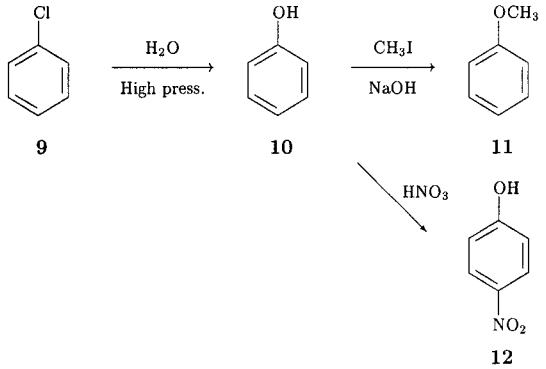


Figure 30. Tabular scheme by the XyMtab element.

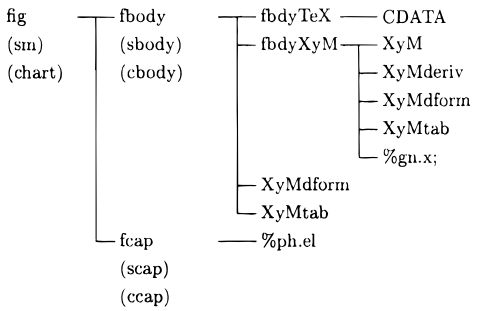


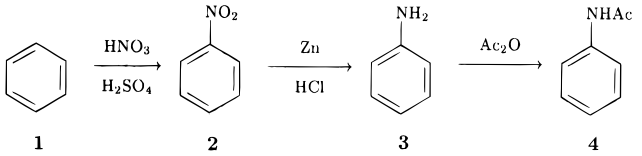
Figure 31. Structures of floating elements.

12 for nitrophenol, can be called by using the XyMref elements, e.g. ⟨XyMref refid="aPhCl"⟩ for **9**.

8. FIGURES, SCHEMES, AND CHARTS

8.1. Structures of Floating Elements. The XyMML system gives figures, schemes, and charts as floating elements. They are generated by the "fig", "sm", and "chart" elements, the structures of which are shown in Figure 31. The body of a floating element is involved in the "fbody" element of the "fig" element, the "sbody" element of the "sm" element, or the "cbody" element of the "chart" element. The caption of a floating element is written by using an fcap, scap, or ccap element. Tabular schemes generated by XyMtab elements and display formulas generated by XyMdform elements are incorporated in a floating body ("fig", "sm", or "chart") directly or through the fbdyXyM element.

Scheme 1. Sample Scheme



8.2. Samples of Floating Elements. A sample scheme is shown in Scheme 1, which is generated by the "sm" element:

```
<sm>
<scap id="scheme01">Sample Scheme</scap>
<sbody>
<fbdyXyM>
<XyM id="PhH" x="400" y="750" xshift="220" yshift="200">
<XyMname>bzdr</XyMname><XyMnmsfx>v</XyMnmsfx>
<subslist></subslist>
</XyM><reacarrw arrwtyp="r" length="400" xshift="10" yshift="60">
<arrwover><arritm>HNO<inf>3</inf></arritm></arrwover>
<arrwundr><arritm>H<inf>2</inf>SO<inf>4</inf></arritm></arrwundr>
</reacarrw><XyM id="PhNO2" x="400" y="750" xshift="220" yshift="200">
<XyMname>bzdr</XyMname><XyMnmsfx>v</XyMnmsfx>
<subslist>
<subst>
<substid>1</substid>
<substval>NO<inf>2</inf></substval>
</subst>
</subslist>
</XyM>
                    (abbreviated)
</fbdyXyM>
</sbody>
</sm>
```

The scheme is referred to by writing Scheme ⟨ref refid="scheme01"⟩, where the term "scheme01" is a reference key. Moreover, each compound contained in Scheme 1 can be referred to by using the XyMref element. For example, nitrobenzene (**2**) is referred to by writing ⟨XyMref refid="$PhNO_2$"⟩.

9. DISCUSSIONS

9.1. Characteristics of the XyMML System. Since the XyMML system takes the same methodology as the XyM notation reported previously,[24] a XyMML also has a nonunique nature. In other words, synonymous XyMMLs are available to generate a sole structure. If we focus our attention on structural diagrams, however, we find that a XyMML generates a diagram in a one-to-one fashion, as shown in the preceding sections. Thus, a XyMML is nonunique and unambiguous to a structure, but unique and unambiguous to a structural diagram when considering derivation and skeleton-suffix synonyms. These characteristics of the XyMML system are rather convenient to electronic communication of chemical documents, since various standpoints and illustrations may exist, depending upon the aspects of discussion on a chemical structure. In addition, the XyMML system satisfies the following items described in the previous paper:[24]

1. Each XyMML is written as an ASCII text, which can be generated by computer as well as by human.

2. Each XyMML is readable and understandable by computer as well as by human so that it is capable of reproducing a structural diagram.

3. Each XyMML contains a procedure of linking structural units so as to represent substitution, spiro ring replacement, and ring fusion. The procedure does not depend on layout data.

4. Each XyMML contains stereochemical information not in the form of layout data (such as *x*,*y*-coordinates) but in the form of logical designation (such designation α, β as in

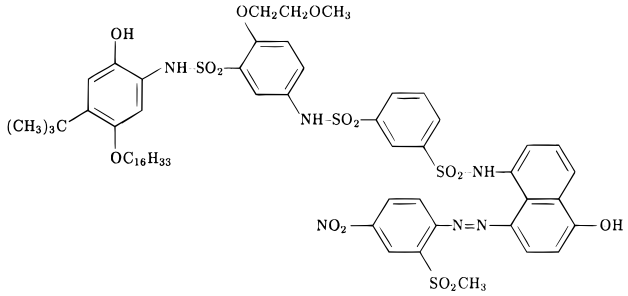


Figure 32. Cyan dye developer drawn by XyMML.

the IUPAC nomenclature). The information is reproducible by computer and by human to depict the corresponding structural diagram easily.

9.2. XyMML and XyM Notation. The XyM notation system proposed previously has been regarded as a linear notation.[24] On the other hand, XyMML proposed here is a markup language based on the SGML specification. Although their appearances are different, they stem from the common viewpoint that a structure is a derivative of a skeleton. Hence, they can be translated to each other. In fact, the components involved in the XyM element (Figure 1) correspond to the skeleton and arguments of the following XyM notations:

```
\GenSkel(SKBONDLIST)[BONDLIST]{ATOMLIST}{SUBSLIST}[OMITLIST]
\GenSkel[BONDLIST]{ATOMLIST}{SUBSLIST}
\SpecSkel[BONDLIST]{SUBSLIST}
\CaromSkel[BONDPATTERN]{SUBSLIST}
```

The first notation \GenSkel represents a generic skeleton containing a full set of arguments, while the second \GenSkel contains a set of selected arguments. The notation \SpecSkel represents a specific skeleton with a BONDLIST and a SUBSLIST. The notation \CaromSkel represents a carbocyclic aromatic skeleton with a BONDPATTERN and a SUBSLIST. Similarly, the components of a XyMfuse element (Figure 17) correspond to the skeleton and arguments of the following XyM notation:

```
\FuseSkel(SKBONDLIST)[BONDLIST]{ATOMLIST}{SUBSLIST}[OMITLIST]
```

A skeleton name "sixhetero" and a suffix "v" (Table 6), for example, are combined to give a skeleton "sixheterov", which corresponds to the \sixheterov skeleton of the XyM notation system. The correspondences between the components of the XyM element and the arguments of the XyM notation are found as follows: skbndlst vs SKBONDLIST, bondlist vs BONDLIST, atomlist vs ATOMLIST, subslist vs SUBSLIST, and omitlist vs OMITLIST.

We compare XyMML with the XyM notation by using a complicated structural formula, which represents a cyan dye releaser developed for instant color photography.[30] This formula can be drawn by the XyMML shown in Chart 1, which consists of a XyM element containing four nested XyMyl elements (numbered as =3=, =5=, =7=, and =10=).

The resulting diagram is shown in Figure 32.[31] As found by careful inspection, the mother skeleton of the cyan dye releaser (the *o*-sulfonamidophenol ring substituted by a hexadecyloxy group etc. at the upper left position) corresponds to the outer XyM element; the second ring (a benzene substituted by a 2-methoxyethoxy group etc.) is generated by the XyMyl element numbered as =3=; the third ring (a benzene substituted by two sulfamoyl groups) is drawn by the XyMyl element numbered as =5=; the fourth ring (a naphthol) is depicted by the XyMyl element numbered as =7=; and the fifth ring (a benzene substituted by a methanesulfonyl group etc.) is drawn by the XyMyl element numbered as =10=. Thus, the logical structure of the XyMML remains transparent, although its appearance becomes complicated in accord with the chemical structure which is complicated.

The same structural formula (Figure 32) can be generated by the following XyM notation,

Chart 1

```
<XyM x="3500" y="2000" xshift="0" yshift="-850">
<XyMname>bzdr</XyMname><XyMnmsfx>v</XyMnmsfx>
<subslist>
<subst><substid>1</substid><substval>OH</substval></subst>
<subst><substid>5</substid>
<substval>(CH<inf>3</inf>)<inf>3</inf>C</substval></subst>
<subst><substid>4</substid>
<substval>OC<inf>16</inf>H<inf>33</inf></substval></subst>
<subst><!-- =====1===== -->
<substid>2</substid>
<substval>
<ryl><!-- =====2===== -->
<lunit><locant>4</locant>
<lunitval>NH--SO<inf>2</inf></lunitval></lunit>
<ylgroup>
<subst><substid>4</substid>
<substval>
<XyMyl><!-- =====3===== -->
<XyMname>bzdr</XyMname><XyMnmsfx>h</XyMnmsfx>
<subslist>
<subst><substid>1</substid><substval>&ylfunc;</substval></subst>
<subst><substid>2</substid>
<substval>OCH<inf>2</inf>CH<inf>2</inf>OCH<inf>3</inf></substval>
</subst>
<subst><substid>5</substid>
<substval>
<ryl><!-- =====4===== -->
<lunit><locant>2</locant>
<lunitval>NH--SO<inf>2</inf></lunitval></lunit>
<ylgroup>
<subst><substid>4</substid>
<substval>
<XyMyl><!-- =====5===== -->
<XyMname>bzdr</XyMname><XyMnmsfx>h</XyMnmsfx>
<subslist>
<subst><substid>1</substid><substval>&ylfunc;</substval></subst>
<subst><substid>5</substid>
<substval>
<ryl><!-- =====6===== -->
<lunit><locant>2</locant>
<lunitval>SO<inf>2</inf>--NH</lunitval></lunit>
<ylgroup>
<subst><substid>4</substid>
<substval>
<XyMyl><!-- =====7===== -->
<XyMname>naphdr</XyMname><XyMnmsfx>h</XyMnmsfx>
<subslist>
<subst><substid>1</substid><substval>&ylfunc;</substval></subst>
<subst><substid>5</substid><substval>OH</substval></subst>
<subst><!-- =====8===== -->
<substid>8</substid>
<substval>
<lyl><!-- =====9===== -->
<lunit><locant>4</locant><lunitval>N=N</lunitval></lunit>
<ylgroup><subst><substid>4</substid>
<substval>
<XyMyl><!-- =====10===== -->
<XyMname>bzdr</XyMname><XyMnmsfx>h</XyMnmsfx>
<subslist>
<subst><substid>4</substid><substval>&ylfunc;</substval></subst>
<subst><substid>1</substid>
<substval>NO<inf>2</inf></substval></subst>
<subst><substid>5</substid>
<substval>SO<inf>2</inf>CH<inf>3</inf></substval></subst>
</subslist>
</XyMyl><!-- =====10===== -->
</substval>
</subst>
</ylgroup>
</lyl><!-- =====9===== -->
</substval>
</subst><!-- =====8===== -->
</subslist>
</XyMyl><!-- =====7===== -->
</substval></subst></ylgroup>
</ryl><!-- =====6===== -->
</substval></subst></subslist>
</XyMyl><!-- =====5===== -->
</substval></subst></ylgroup>
</ryl><!-- =====4===== -->
</substval></subst></subslist>
</XyMyl><!-- =====3===== -->
</substval></subst></ylgroup>
</ryl><!-- =====2===== -->
</substval>
</subst><!-- =====1===== -->
</subslist>
</XyM>
```

```
\bzdrv{1==OH;5==CH$_{3}$;4==OC$_{16}$H$_{33}$;%
2==\ryl(4==NH--SO$_{2}$){4==\bzdrh{1==(yl);%
2==OCH$_{2}$CH$_{2}$OCH$_{3}$;%
5==\ryl(2==NH--SO$_{2}$){4==\bzdrh{1==(yl);%
5==\ryl(2==SO$_{2}$--NH){4==\naphdrh{1==(yl);5==OH;%
8==\lyl(4==N=N){4==\bzdrh{4==(yl);1==NO$_{2}$;%
5==SO$_{2}$CH$_{3}$}}}}}}}}}
```

which corresponds to the XyMML described above. In fact, the XyMML is translated into a LaTeX code that contains the latter XyM notation as a main descriptor. The translation process will be described in the Implementation section.

The XyMML and the XyM notations correspond to the following IUPAC substitution name if the phenol group is selected as a principal group (i.e. a mother skeleton in terms of the present terminology): 3-*tert*-butyl-4-hexadecyloxy-6-{5-{3-{*N*-[1-hydroxy-4-(3-methanesulfonyl-4-nitrobenzeneazo)naphth-5-yl]sulfamoyl}-2-(2-methoxyethoxy)-benzenesulfonylamino}phenol. It should be noted that the nesting structure of the IUPAC name is the same as those of the XyMML and the XyM notations.

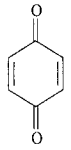
It is worthwhile to mention another type of synonym which is tentatively called a "skeleton-selection synonym". For the structure depicted in Figure 32, a synonymous IUPAC name has been reported: *N*-{3-[*N*-(4-*tert*-butyl-5-hexadecyloxy-2-hydroxyphenyl)sulfamoyl]-4-(2-methoxyethoxy)phenyl}-*N*′-[5-hydroxy-8-(2-methanesulfonyl-4-nitrophenylazo)-1-naphthyl]-1,3-benzenesulfonamide, where the benzene ring of 1,3-benzenesulfonamide is selected as a principal group.[32] According to the nesting structure of this IUPAC name, we can write another XyMML to generate the same formula as that shown in Figure 32, i.e.,

```
<XyM>
<XyMname>bzdr</XyMname><XyMnmsfx>h</XyMnmsfx>
<subslist>
<subst><substid>1</substid><substval>A</substval></subst>
<subst><substid>5</substid><substval>B</substval></subst>
</subslist>
</XyM>
```

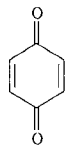
where A and B represent substituents generated by the XyMyl, ryl, or lyl elements. Thus, the formula of the cyan dye releaser can be depicted by both of the XyMMLs, which select the phenol moiety and the 1,3-benzenesulfonamide moiety as mother skeletons. It follows that, under the consideration of skeleton-selection synonyms, a XyMML is nonunique but unambiguous with respect to a structural diagram as well as a structure.

9.3. Skeleton-Name Synonyms. Derivation synonyms or skeleton-suffix ones generate different structural diagrams that are synonymous in representing a chemical structure. On the other hand, skeleton-selection synonyms depict a same structural diagram, where they are concerned with XyMname and the XyMsuffix elements. According to the methodology of the XyMML system, there are further types of synonyms that generate a same structural diagram. These types of synonyms are tentatively called "skeleton-name synonyms", since they are concerned with the XyMname element but not with the XyMnmsfx element. Figure 33 illustrates synonymous XyMMLs of this type for drawing 1,4-benzoquinone, where three types of skeletons, i.e., a sixheterov skeleton (as a generic skeleton), a cyclohexanev skeleton (as a specific skeleton), and a bzdrv skeleton (as a carbocyclic aromatic skeleton), are used, respectively. Although they contain the same set of substitution lists, they are different in other lists. Thus, the sixheterov skeleton takes

```
<XyM>
<XyMname>sixhetero</XyMname><XyMnmsfx>v</XyMnmsfx>
<bondlist><bond> <bondid>b</bondid> </bond>
<bond><bondid>e</bondid></bond></bondlist>
<atomlist></atomlist>
<subslist><subst><substid>1</substid>
<bondmod>D</bondmod>
<substval>O</substval></subst><subst>
<substid>4</substid><bondmod>D</bondmod>
<substval>O</substval></subst></subslist>
</XyM>
```



```
<XyM>
<XyMname>cyclohexane</XyMname><XyMnmsfx>v</XyMnmsfx>
<bondlist><bond> <bondid>b</bondid> </bond>
<bond><bondid>e</bondid></bond></bondlist>
<subslist><subst><substid>1</substid>
<bondmod>D</bondmod>
<substval>O</substval></subst><subst>
<substid>4</substid><bondmod>D</bondmod>
<substval>O</substval></subst></subslist>
</XyM>
```



```
<XyM>
<XyMname>bzdr</XyMname><XyMnmsfx>v</XyMnmsfx>
<bndpatrn>pa</bndpatrn>
<subslist><subst><substid>1</substid>
<bondmod>D</bondmod>
<substval>O</substval></subst><subst>
<substid>4</substid><bondmod>D</bondmod>
<substval>O</substval></subst></subslist>
</XyM>
```

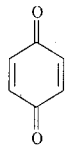


Figure 33. Synonymous XyMMLs for drawing 1,4-benzoquinone.

a bond list, an atom list (vacant); the cyclohexanev skeleton takes a bond list but no atom list; and the bzdrv skeleton takes a bond pattern but no atom list. As found easily, such skeleton-name synonyms of a XyMML provide a nonunique but unambiguous nature with respect to a structural diagram as well as a structure.[33]

9.4. Reaction Schemes. In a traditional way of literal communication, a reaction scheme containing several structural formulas has been treated as image data produced by an interactive structure editor. The involved formulas cannot be treated separately, since the data of a specific formula cannot be taken out from the reaction scheme. On the other hand, a reaction scheme given by the XyMML has the data of each specific formula by virtue of a XyM element.

This feature of the XyMML has a merit in realizing effective cross-reference. As long as structural data are communicated in image, cross-reference concerning formula numbers has been incomplete. Thus, a number given to each formula cannot be called distinctly, since there is no reference key to be called. On the other hand, the XyMML proposed in this article supports cross-reference, where a reference key (e.g. "PhH") is given to each formula, as found in Scheme 1:

```
<XyM id="PhH" x="400" y="750" xshift="220" yshift="200">
```

The formula number (**1**) can be called by using the XyMref element, i.e., ⟨XyMref refid="PhH"⟩. As clarified in the next section, this feature comes from the XyMTeX system for typesetting chemical structures, which has provided us with a tool for cross-reference concerning every structure.[22]

10. IMPLEMENTATION

10.1. Processing Steps of XyMML. In the present paper, we have developed a method for printing structural formulas as the first step of applying the XyMML system to electronic communication and documentation (Figure 34). The XyMML system has been implemented as a DTD (document type definition) named xymartcl.dtd, a translation file named bcsj_ta.rep, and a class file named xymartcl.cls for LaTeX2e processing. The XyMTeX packages reported previously[24] have been used during the LaTeX2e processing.[34] It should be added here that the manuscript of this paper has been

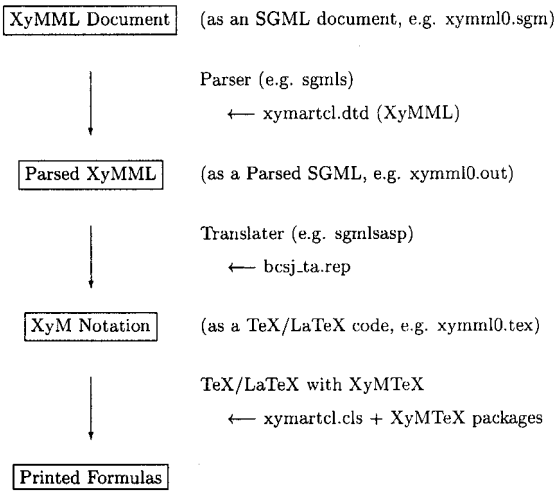


Figure 34. XyMML for printing structural formulas.

prepared by means of XyMML according to the process shown in Figure 34.

10.2. Document Type Definition for XyMML. The main part of the implementation of the XyMML system is the DTD named xymartcl.dtd, in which the elements described in this paper are defined.[35] For example, the XyM element has been defined as follows:

```
<!ELEMENT XyM - -
(XyMname,XyMnmsfx?,XyMbody?,
skbndlst?,(bondlist|bndpatrn)?,atomlist?,subslist,omitlist?)
--XyM Notation -->
```

The structure of this definition is illustrated by Figure 1. The attributes for the XyM element are concerned with cross-reference, the size of the resulting formula, and the adjustment of *x*,*y*-coordinates, i.e.,

```
<!ATTLIST XyM id ID #IMPLIED
s CDATA #IMPLIED
x NUTOKEN #IMPLIED
y NUTOKEN #IMPLIED
xshift NMTOKEN #IMPLIED
yshift NMTOKEN #IMPLIED
-- id = xym compound reference -->
```

The other elements described in the preceding sections have been defined in similar ways. Their structures are illustrated in Figures 17, 20, 23, 25, 27, 29, and 31.

The DTD (xymartcl.dtd) has been used to parse SGML documents containing structural formulas based on XyMML, where sgmls[36] has been used as an SGML parser. For example, an SGML document named xymml0.sgm (the manuscript of this article) has been treated by typing the following command into a command line on a display:

```
a:\sgml\sgmls eucjpn1.dec xymartcl.dtd xymml0.sgm > xymml0.out
```

where the file eucjpn1.dec is an SGML declaration[31] and the file xymartcl.dtd is the DTD file defining XyMML. Thereby, we have obtained an intermediate file with full tags (e.g. xymml0.out).

10.3. Translation of XyMML to XyM Notation. The intermediate file with full tags (e.g. xymml0.out) has been treated by an SGML translator such as sgmlsasp.[36]

```
a:\sgml\sgmlsasp bcsj_ta.rep < xymml0.out > xymml0.tex
```

where the file bcsj_ta.rep is a translation file producing a TeX file of the *Bull. Chem. Soc. Jpn.* or *J. Chem. Inf. Comput. Sci.* format. Thereby, we obtain a TeX file (e.g. xymml0.tex) suitable to TeX/LaTeX processing.

For example, the XyMML shown in Figure 13 is translated into the following TeX/LaTeX code:

```
\begingroup
\edef\xymrefa{PHCL}\edef\xymrefb{}
\edef\kkk{400}
\begin{tabular}{c}
\ifx\kkk\empty\else
\begin{picture}(400,750)(220,200)\fi
\bzdrv{{1}==Cl;}
\ifx\kkk\empty\else
\end{picture}\fi
\\[-\baselineskip]
\ifx\xymrefa\empty\else{\bf \compd\xymrefb
\expandafter\label{\xymrefa}}\fi
\end{tabular}\endgroup%%%
```

which contains a XyM notation,

```
\bzdrv{{1}==Cl;}
```

for chlorobenzene. The code generates the structure of chlorobenzene (**5**) shown in Figure 13, when processed by LaTeX2e loaded with the XyMTeX system.[37] The TeX file (e.g. xymml0.tex) calls the class file (xymartcl.cls) and XyMTeX package files during LaTeX2e processing. All of the figures and schemes shown above have been prepared in light of the XyMML proposed in the present paper. Thus, the manuscript of this paper itself is a sample of the application of the XyMML system.

11. CONCLUSION

The XyM markup language (XyMML) is proposed as a tool for the electronic communication of chemical structural formulas. The XyMML system has the following characteristics advantageous to electronic communication: (1) A XyMML has a nonunique nature so that synonymous XyMMLs are available to generate a sole structure. The synonymous XyMMLs are discussed in terms of derivation synonyms, skeleton-suffix synonyms, skeleton-name synonyms, and skeleton-selection synonyms. (2) A XyMML is written as an ASCII text. (3) A XyMML is readable and understandable by computer as well as by human. (4) A XyMML contains a procedure for linking structural units. (5) A XyMML contains stereochemical information not in the form of layout data but in the form of logical designation. (6) XyMMLs can be collected to give a reaction scheme, in which cross-references to every XyMML are available by using the technique of SGML attributes. The XyMML system has been implemented as a DTD (document type definition) for an SGML parsing, a translation file for translating SGML to LaTeX2e, and a class file for LaTeX2e processing.

REFERENCES AND NOTES

(1) ISO 8879-1996 (JIS X 4151-1992) Information processing, Text and office systems, Standard Generalized Markup Language (SGML).
(2) van Hervijnen, E. *Practical SGML*; Kluwer Academic: Dordrecht, The Netherlands, 1990.
(3) Bryan, M. *SGML. An Author's Guide to the Standard Generalized Markup Language*; Addison-Wesley: Reading, MA, 1988.
(4) Itoh, T. *Kagaku to Kogyo (Tokyo)* **1993**, *46*, 66.
(5) Knuth, D. E. *The TeXbook*; Addison-Wesley: Reading, MA, 1984.
(6) Lamport, L. *LaTeX. A Document Preparation System*, 2nd ed.; Addison-Wesley: Reading, MA, 1994.
(7) Garson, L. ACS' SGML Activities, http://www.oasis-open.org/cover/garson0054.html, 1994.

(8) ISO 12083-1994 (JIS X 0804-1996) Information and documentation, Electronic manuscript preparation and markup.
(9) Bachrach, S. M., Ed. *The Internet: A Guide for Chemists*; American Chemical Society: Washington, D.C., 1996.
(10) Wiggins, G. Chemistry on the Internet: The Library on Your Computer. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 956−965.
(11) Murray-Rust, P.; Rzepa, J. S.; Whittaker, B. J. The World Wide Web as a Chemical Information Tool. *Chem. Soc. Rev.* **1997**, *27*, 1−10.
(12) Rzepa, H. S.; Tonge, A. P. VChemLab: A Virtual Chemisty Laboratory. The Storage, Retrieval, and Display of Chemical Information Using Standard Internet Tools. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1048−1053.
(13) Extensible Markup Language (XML), http://www.w3.org/XML.
(14) CSJ's home page, http://wwwsoc.nacsis.ac.jp/csj/index.html.
(15) Other strong approaches have been discussed in the introduction of our preceding paper.[24]
(16) Raggett, D. W3C Activity: Math, http://www.w3.org/pub/WWW/TR/PR-math/.
(17) Murray-Rust, P.; Rzepa, H. S.; Leach, C. CML−Chemical Markup Language. Presented as a Poster at the 210th ACS Meeting in Chicago, 1995; Abstract 40 (http://www.chi.ic.ac.uk/cml/).
(18) Fujita, S. XyMTeX. A macro package for typesetting chemical structural formulas. *On-line Manual for XyMTeX*, Version 1.00; CTAN: ftp://ctan.tug.org/tex-archive/macros/latex/contrib/other/xymtex/, 1993.
(19) Fujita, S. Typesetting structural formulae with the text formatter TeX/LaTeX. *Comput. Chem.* **1994**, *18*, 109−116.
(20) Fujita, S. XyMTeX for drawing chemical structural formulas. *TUGboats* **1995**, *16*, 80−88.
(21) Fujita, S. XyMTeX. A macro package set for typesetting chemical structural formulas. *On-line Manual for XyMTeX*, Version 1.01; 1996. This can be obtained from the author's cite (see ref 37) and from the CD-ROM attached to ref 22.
(22) Fujita, S. *XyMTeX. Typesetting Chemical Structural Formulas*; Addison-Wesley Japan: Tokyo, 1997.
(23) Ogura, T.; Nishimoto, M.; Hatano, Y. Representation of Mathematical Formula, Chemical Formula and Structral Formula on the WWW Browser. In *The 20th Symposium on Chemical Information and Computer Science, Kumamoto*; Chemical Society of Japan: Tokyo, Japan, 1997; p 21IP15.
(24) Fujita, S.; Tanaka, N. XyM Notation for Electronic Communication of Organic Chemical Structures. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 903−914.
(25) Lafore, R. *Data Structure and Algorithms in Java*; Waite Group: 1998; Japanese Translation, Softbank: Tokyo, 1999.
(26) Active X Whitepaper: http://www.microsoft.co.jp/internet/activex/index.htm.
(27) The XyM is an uppercase form of *χυμ*, where "y" is used in place of Y of special layout ($X^{Y}M$). The pronunciation of XyM is recommended to be "khým", in which the "kh" sound may be a Russian "kh" or more simply an English "k" and the symbol "y" is expected to be pronounced like a German "ü". Note that the word "chemistry" stems from an Arabian root "alchemy", which is, in turn, considered to come from Greek, *χυμείᾱ*.
(28) Davis, C. H.; Rush, J. E. *Information Retrieval and Documentation in Chemistry*; Greenwood: Westport, CT, 1974.
(29) Gould, E. S. *Mechanism and Structure in Organic Chemistry*; Henry Holt: New York, 1959; p 643.
(30) Fujita, S. Synthesis and reactions of *o*-benzoquinone monosulfonimides. *J. Org. Chem.* **1983**, *48*, 177−183, and references cited therein.
(31) The parsing of the XyMML should be done after the SGML declaration containing a line "QUANTITY SGMLREF TAGLVL 50" to expand the limit of the maximum number of opening tags, since the default value of TAGLVL (=24) is short in giving a sufficient result.
(32) Fujita, S. Canonical numbering and coding of imaginary transition structures. A novel approach to the linear coding of individual organic reactions. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 128−137. Several delimiters have been changed in the IUPAC name.
(33) This discussion is quite subtle, since the diagrams produced by a sixheterov, a cyclohexanev, and a bzdrv skeleton are different in their conceptual levels. For example, the diagram produced by a cyclohexanev takes no account of vertices, while the diagram produced by a sixheterov considers that the vertices have carbons but no heteroatoms. Hence, we may consider that the three diagrams have the same appearance but can be regarded as being conceptually different. This discussion may conclude that a XyMML is unique and unambiguous to a structural diagram with consideration of any types of synonyms. It should be noted, however, that a XyMML has a nonunique and unambiguous nature with respect to a structure (not a diagram).
(34) Goossens, M.; Mittelbach, F.; Samarin, A. *The LaTeX Companion*; Addison-Wesley: Reading, MA, 1994.
(35) The DTD and related files will be obtained from the author's site, http://www.chem.kit.ac.jp/fujita/fujitas/fujita.html.
(36) The parser "sgmls" and the translator "sgmlsasp" have been developed by J. Clark and converted to accommodate Japanese characters by S. Tsuchiya. These tools have been obtained from NIFTY-serve, FPRINT forum, No. 135 (36SET.LZH).
(37) The XyMTeX system (version 2.00) will be obtained from the author's site, http://www.chem.kit.ac.jp/fujita/fujitas/fujita.html.

CI990029Q