

# Model Tool to Describe Chemical Structures in XML Format Utilizing Structural Fragments and Chemical Ontology

Punnaivanam Sankar,<sup>\*,†</sup> Krief Alain,<sup>‡</sup> and Gnanasekaran Aghila<sup>§</sup>

Department of Chemistry, Pondicherry Engineering College, Puducherry-605 014, India, Department of Chemistry, Facultés Universitaires Notre-Dame de la Paix, Namur, B 5000, Belgium, and Department of Computer Science, Pondicherry University, Puducherry-605 014, India

Received February 5, 2010

We have developed a model structure-editing tool, ChemEd, programmed in JAVA, which allows drawing chemical structures on a graphical user interface (GUI) by selecting appropriate structural fragments defined in a fragment library. The terms representing the structural fragments are organized in fragment ontology to provide a conceptual support. ChemEd describes the chemical structure in an XML document (ChemFull) with rich semantics explicitly encoding the details of the chemical bonding, the hybridization status, and the electron environment around each atom. The document can be further processed through suitable algorithms and with the support of external chemical ontologies to generate understandable reports about the functional groups present in the structure and their specific environment.

## INTRODUCTION

ChemEd is the first of a series of tools designed in the context of EnCOrE,<sup>1</sup> a project aimed to create a Web-based encyclopedia of organic chemistry built collaboratively but under a strict editorial board policy. We have identified several tools with original features to achieve this objective and have ranked first the design of a tool able to interoperate between a structure and the perception of a chemist which allows, among others, creativity through chemical synthesis and organization of chemical data.

Structure, which is one of the most important representations in chemistry, has a long history. J. J. Berzelius initiated the method of representing the chemical structures with modern symbols in the early 19th century.<sup>2</sup> He introduced the classical system of chemical symbols and formulas such as C, H, O, CH<sub>3</sub>–, C<sub>6</sub>H<sub>5</sub>–, etc. These symbolic representations are still being used with little modification. In recent years and especially as the consequence of the seminal work of E. J. Corey<sup>3</sup> and S. Rubenstein,<sup>4</sup> chemical structures have been digitalized as graphics, opening completely new capabilities. They can be drawn on a computer screen, using appropriate editors, modified, displayed, retrieved, incorporated in documents, and processed in various formats on a computer with appropriate operating systems. Some editors,<sup>5</sup> such as ChemDraw,<sup>5a</sup> MDL ISIS/Draw,<sup>5b</sup> ACD/ChemSketch,<sup>5c</sup> and MarvinSketch<sup>5d</sup> are commercially available; some such as MDL ISIS/Draw-2.5<sup>5b</sup> ACD/ChemSketch freeware,<sup>5c</sup> and MarvinSketch<sup>5d</sup> can be used freely. Another tool JChemPaint<sup>5e</sup> is a free, open-source editor for two-dimensional (2D) chemical structures. These editors are capable of supporting widely used file formats, such as Connection Table (CTab),<sup>6</sup> MDL Molfile,<sup>7</sup> and SMILES.<sup>8</sup>

Some of them are starting to use Extensible Markup Language (XML)<sup>9</sup> formats, including Chemical Markup Language<sup>10</sup> (CML).

ChemEd is not yet as sophisticated as those editors but is an alternative one which allows one to draw and encode even complex chemical structures (ferrocene, structure 10, Figures 5 and 16) along with the features, such as, for example, supramolecular interactions. ChemEd automatically encodes the information in XML, a semantically rich, machine-readable format, which is then processed to complicated inferences particularly needed for chemical reaction descriptions or to natural language meaningful for the chemist (cyclohexane carboxaldheyde, structure 7, Figures 4 and 31).

The basic concept involves the description of: chemical structures as atoms (including isotope label, hybridization, charges, etc.); orbitals (electronic content, such as lone pair, empty, unpair, etc.); bonds (such as single, double, triple, type of bond, like  $\sigma$ ,  $\pi$ , etc.); groups (halides, hydroxy, carbonyl, carboxyl, acyl, etc.); functional groups (FG) (organometallic, organic halide, aldehyde, and ketone FGs); and structural features (chains, cycles, polycycles, etc.).

The chemical structures are built, as will be described below, from selected fragments consisting of atoms, “electronLink”, and bonds whose XML description, stored in a library “ChemLib”, can be interoperated with external chemical ontologies developed in Protégé.<sup>11</sup>

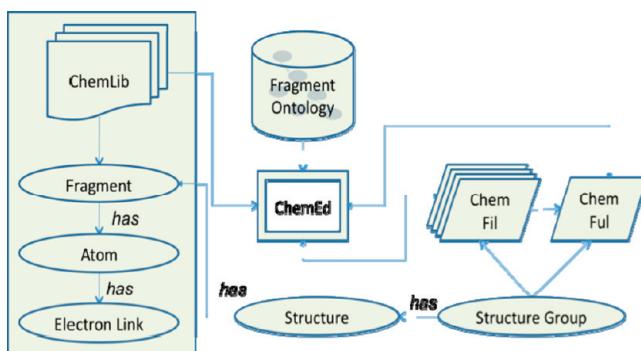
**Architecture of the System.** The structure editor is programmed in JAVA.<sup>12</sup> It has a built in resource ChemLib, a collection of XML documents representing various basic structural fragments needed for structure construction. Each fragment is defined in terms of its constituent atoms as “atom” elements. Every atom in each fragment in turn is defined with an appropriate number of “electronLink” elements to provide semantics about the orbital, the electron status of the orbital, and the bonding information. In the proposed conceptual structures description plan, a chemical structure is considered to be composed of structural frag-

\* Email:gapspec@yahoo.com.

† Pondicherry Engineering College.

‡ Facultés Universitaires Notre-Dame de la Paix.

§ Pondicherry University.



**Figure 1.** Architecture of the structure description system.

ments. Accordingly, the chemical structure is conceptualized as a combination of one or more structural fragments. Some of the basic ones have been identified, described, and organized in fragment ontology. ChemEd works on the basis of the knowledge related to the chemical structure represented in XML format. The editor allows the selection of fragments to build chemical structures on screen. During this construction, the details of the fragments are parsed from ChemLib, and appropriate semantics are added to generate the complete description in ChemFul. The system also generates a concise instruction code, ChemFil, to store the structures in XML format. The ChemFil is a short instruction sequence on the structure construction, which can be read by ChemEd to generate the ChemFul and is able to build the structure on screen. In ChemEd, drawings are captured as a structure group, which may contain one or more structures. So the structure construction makes the fragments into a structure and then wraps the structures as a structure group. This facilitates the description of more than one interacting structure. The architecture of the structure description system is shown in Figure 1.

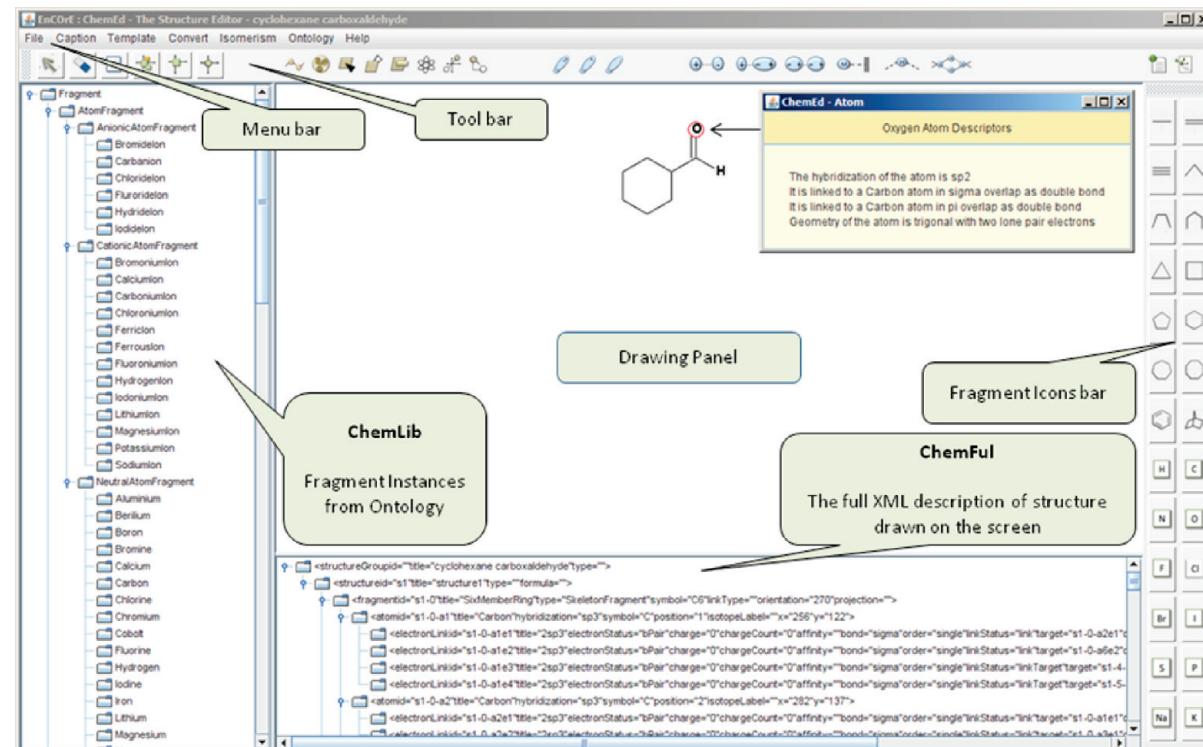
**Graphical User Interface (GUI).** The GUI developed for ChemEd is shown in Figure 2. The toolbar showing the

important functionalities is provided in Figure 3. This software application has the usual capabilities to draw a structure, to select various fragments and templates, to change atoms and bond types, to display the structure, and to create XML as output. Additionally the system has the facility to draw and encode supramolecular interactions such as ion–ion, ion–dipole, dipole–dipole, metal– $\pi$ , hydrogen bonding, and multicenter bonding (Figures 4 and 5). The description of the functionalities and the structure drawing methodology are detailed in the Supporting Information.

**ChemFull and ChemFil.** The vital part of ChemEd is ChemFull which captures in XML the complete structural features of the structures drawn on the screen and facilitates the process to arrive at useful inferences, textual treatment of structure, and interoperability with external ontologies. To generate ChemFull, the system uses the generic information described for the basic fragments in XML. The XML format of any fragment is shown in Figure 6.

The concept plan of a fragment describes the fragments in three semantic levels, viz. “fragment”, “atom”, and “electronLink” levels. This is achieved by identifying appropriate attributes for each level. The semantics in the “fragment” level includes the ‘title’ attribute to specify the fragment name. The ‘type’ attribute is used to indicate the class to which the fragment belongs. The ‘symbol’ attribute denotes the symbol of the fragment, like “C/H/N/OH/CHO/COOH/etc”. The final attribute, named as ‘smiles’, is introduced to hold a string notation representing the fragment. For example, the ‘smiles’ attribute values “C”, “C1CC1”, and “O” are used to represent the fragments “Carbon”, “ThreeMemberRing”, and “HydroxylGroup”, respectively. This attribute is used to generate one of the SMILES strings through appropriate algorithms for the completed structure.

In the “atom” level semantics, an ‘id’ attribute is provided to identify uniquely the atoms as a1, a2, a3, etc. belonging



**Figure 2.** GUI of ChemEd showing major features.

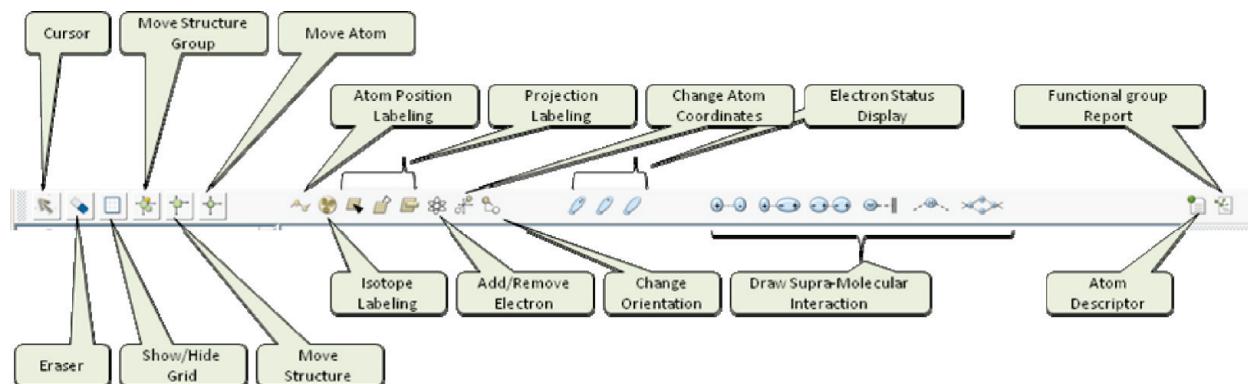


Figure 3. Horizontal tool bar of ChemEd.

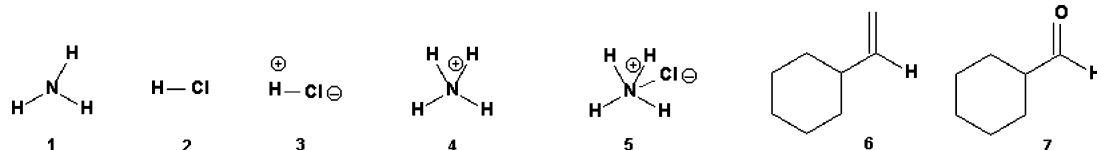


Figure 4. Example structures to describe the structure drawing in ChemEd.

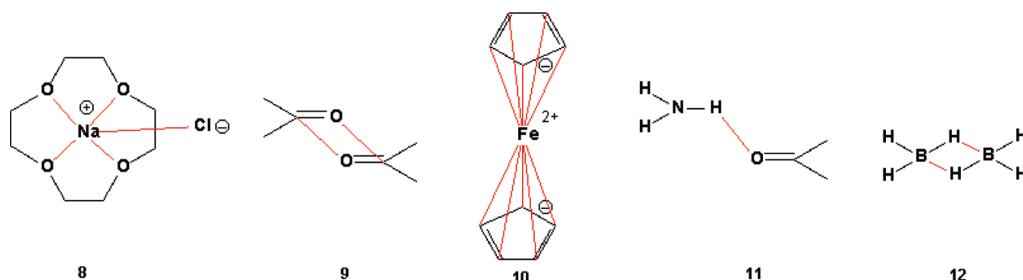


Figure 5. Supra-molecular interactions in structures created in ChemEd.

```

<fragment title="" type="" symbol="smiles="">
    <atom id="" title="" hybridization="" symbol="" position="" isotopeLabel="" orientation="">
        <electronLink id="" title="" electronStatus="" charge="" chargeCount="" affinity="">
            bond="" order="" target="" linkStatus="">
        </electronLink>
    </atom>
</fragment>

```

Figure 6. XML format of fragment.

to a particular fragment. The name and symbol of the atom are indicated with the ‘title’ and ‘symbol’ attributes, respectively. The ‘hybridization’ attribute at this level specifies the hybridization status of the atom, if any. The ‘position’ attribute allows the numbering of atoms in a skeleton. The ‘isotopeLabel’ attribute brings the possibility of labeling the atoms at any stage of structure construction guided by external ontology. The last attribute, ‘orientation’ is a presentational attribute used by the system to create basic cyclic skeleton structures.

The innermost part of the markup provides the semantics about the “electronLink” associated with each atom of the fragment. The id values such as a1e1, a1e2, a2e1, a2e2, etc. associate the electronLink tags with the corresponding atoms. The value of the title attribute provides the name representing an orbital such as s-, p-orbital, etc. with the values of “1s/2s/2px/2py/2pz/etc”. In order to represent the hybridization status, the name of the hybridization is indicated as “2sp3/2sp2/2sp/etc”. The electronic status of the “electronLink”, such as lone pair, unpaired, and empty, is encoded using the values “lPair”, “uPair”, and “empty”, respectively. The bond pair is indicated as “bPair” for the ‘electronStatus’ attribute. The ‘charge’ and ‘chargeCount’ attributes are used

to indicate the charge with the values of “+”, “-”, and “0” for positive, negative, and neutral charges, respectively, with its count. The partial charges normally shown as  $\delta+$  and  $\delta-$  can also be accommodated into the same attribute. In the present study, the values of “p+” and “p-” are used to indicate the partial charges due to some programming and encoding restrictions. Meaningful inferences can be obtained by combining the values of ‘electronStatus’, ‘charge’, and ‘chargeCount’ attributes. For example, a value of empty for ‘electronStatus’ followed by a “+” in the charge and an “1” in the charge attribute can be inferred as the positive charge due to the loss of an unpaired electron. Similarly, the values “lPair”, “-” and “1” for ‘electronStatus’, ‘charge’, and ‘chargeCount’ attributes, respectively, provide the meaning of acquired negative charge due to the gain of one electron.

The next five attributes namely ‘affinity’, ‘bond’, ‘order’, ‘target’, and ‘linkStatus’ are used to provide the semantics of the chemical bonding: (i) the ‘affinity’ attribute holds the values indicating the normal tendency of the “electronLink” toward chemical bonding using the values, such as “covalent/ionic/coordinate” quoting for covalent-, ionic- or dative-bond; (ii) the type of bond is denoted with values, like “sigma/pi-y/pi-z”, to separately encode the  $\sigma$  and  $\pi$  systems; (iii)

```

<structureGroup id="" title="" type="">
  <structure id="" title="" type="" formula="" X="" Y="" caption="" captionX="" captionY="">
    <fragment id="" title="" type="" symbol="" linkType="" orientation="" projection="">
      x1="" y1="" x2="" y2=""
      <atom id="" title="" hybridization="" symbol="" position="" isotopeLabel="" x="" y="">
        <electronLink id="" title="" electronStatus="" charge="" chargeCount="" affinity="" bond="">
          order="" linkStatus="" target="" orientation="" projection="" x1="" y1="" x2="" y2="">/>
      </atom>
    </fragment>
  </structure>
</structureGroup>

```

**Figure 7.** XML template of ChemFul.

the ‘order’ attribute indicates bond order with values “single/double/triple”; (iv) the ‘target’ attribute of the source “electronLink” is used to hold the unique ‘id’ value of the “electronLink” of the targeted atom to represent a chemical bond. At the same time the ‘target’ attribute of the target “electronLink” is filled with the ‘id’ value of the “electronLink” of the source atom; (v) the ‘linkStatus’ attribute specifies whether the link is between two “electronLink” belonging to the same or different fragments. In case of a link within the fragment, the linkStatus holds the value as “link”. If the link is between two different fragments, then the “electronLink” belongs to the target fragment, which takes up the value as “linkTarget”, and that of the source fragment is filled with the value as “linkSource”. The same attribute is used to hold the bridge-head positions in bridged structures with the values of “bridgeTarget” and “bridgeSource”.

During the structure construction, the semantics about the construction are captured and encoded in all three levels. The chemical bonding details are encoded by the system in the “electronLink” elements associated with the corresponding atoms. The chemical bonding between the “electronLink” is captured with the descriptor attributes, viz. affinity, bond, order, target, and linkStatus. These attributes hold the natural bonding tendency of every “electronLink” before it is linked with another fragment. The semantics are added to these attributes to represent the acquired properties of the “electronLink” due to bonding linkages after it is mapped with another “electronLink” to describe a chemical bond.

For example, the sequence of values shown below describes the tendency of an “electronLink” toward the formation of a covalent bond before it is mapped with another: affinity = “covalent” bond = “” order = “” linkStatus = “” target = “”.

The same attributes with the appropriate values, as given below, informs that the link is mapped with another one representing a single bond between the two “electronLink”: affinity = “” bond = “sigma” order = “single” linkStatus = “linkTarget” target = “s1-1-a1e1”.

Also the semantics details the bond between the two links as a  $\sigma$  bond whose order is filled with the value “single” to indicate it is a single bond. The value in the ‘linkStatus’ attribute shows that it is the mapping target for the other fragment to which it is linked during the structure construction. As and when a new fragment is added to the existing group of fragments, the system automatically encodes the bonding details, and finally the complete structural description is generated by the system as an XML document, i.e., ChemFul. The XML format of ChemFul is shown in Figure 7.

The XML format of ChemFul includes the description of individual fragments inside a “structure” element and then

all the “structure” elements in a “structureGroup” element. So, ChemEd generates ChemFul with two additional layers of semantics in terms of “structure” and “structureGroup” along with the details of the fragments. At present, the “structureGroup” layer is provided with ‘id’, ‘title’ and ‘type’ attributes. The structure layer contains ‘id’, ‘title’, ‘type’, ‘formula’, ‘X’, ‘Y’, ‘caption’, ‘captionX’, ‘captionY’ attributes. The ‘formula’ attribute holds the formula of the structure, if needed. In the “structure” layer, the presentational attributes ‘X’ and ‘Y’ are used to render the individual structures in the “structureGroup”. The ‘caption’, ‘captionX’, ‘captionY’ attributes provide the facility to capture the caption, if any of the structures along with its location are on the drawing panel. These layers facilitate the extension of ChemEd to accommodate more attributes to represent the structures as both virtual substances and materials in the future.

ChemEd adds additional semantics about the presentational details in each level of the fragments to encode the rendering information of the fragments on the screen, according to their construction. This is done with attributes: (i) ‘orientation’, ‘projection’, ‘x1’, ‘y1’, ‘x2’, and ‘y2’ to the “electronLink” element; (ii) ‘x’ and ‘y’ for the “atom” element; and (iii) ‘linkType’, ‘orientation’, ‘projection’, ‘x1’, ‘y1’, ‘x2’, and ‘y2’ into the “fragment” element. The ‘linkType’ attribute in the fragment level is used to hold values, such as “direct”, “bridge”, “fuse”, or “spiro”, representing the nature of the link when the fragment is joined with other fragments during the structure construction. The ‘orientation’ attribute provides the description of imaginary orientation of bond links in a two-dimensional (2D) plane. A value from 0 to 360 in an anticlockwise direction is suggested as the possible orientation in which the atoms are oriented with respect to the mapped atom. This indicates a fragment can be rendered with any orientation as specified by the values of this attribute. The orientations are restricted to values, such as 0, 30, 60, 90, 120, 150, 180, 210, 240, 270, 300, and 330 in the present study in order to show the feasibility of the rendering approach. In the future, more possible orientations can be included for rendering a fragment. It is also possible to make the fragments adjust with respect to their surrounding fragments automatically with improved algorithms. The exact coordinates in pixel values at which the fragments are placed on the screen are provided using ‘x1’, ‘y1’, ‘x2’, and ‘y2’ for rendering the chemical structures on the computer screen as 2D graphics. These attributes render the “electronLink” as points with  $x1 = x2$  and  $y1 = y2$ , indicating unmapped free links or mapped bond links showing the chemical bond with the values as  $x1 \neq x2$  and  $y1 \neq y2$ . Accordingly, a point at an atom indicates the presence of an open “electronLink”. A line connecting between two atoms represents a chemical bond.

```

Code 1: "electronLink" with 'uPair' electronStatus of Hydrogen Fragment
<electronLink id="a1e1" title="1s" electronStatus="uPair" charge="0" chargeCount="0" affinity="covalent"
  bond="" order="" target="" linkStatus="" />
Code 2: "electronLink" with 'uPair' electronStatus of Chlorine Fragment
<electronLink id="a1e1" title="2sp3" electronStatus="uPair" charge="0" chargeCount="0"
  affinity="covalent" bond="" order="" target="" linkStatus="" />
Code 3: "electronLink" with 'bPair' electronStatus of Hydrogen Fragment
<electronLink id="s1-0-a1e1" title="1s" electronStatus="bPair" charge="0"
  chargeCount="0" affinity="" bond="sigma" order="single" linkStatus="linkTarget"
  target="s1-1-a1e1" orientation="0" projection="" x1="301" y1="188" x2="331"
  y2="188" />
Code 4: "electronLink" with 'bPair' electronStatus of Chlorine Fragment
<electronLink id="s1-1-a1e1" title="2sp3" electronStatus="bPair" charge="0"
  chargeCount="0" affinity="" bond="sigma" order="single" linkStatus="linkSource"
  target="s1-0-a1e1" orientation="0" projection="" x1="331" y1="188" x2="301"
  y2="188" />

```

**Figure 8.** Changes in “electronLink” during the construction of molecular hydrogen chloride structure (see also 2 in Figure 4).

```

<?xml version="1.0"?>
<structureGroup id="" title="molecular hydrogen chloride" type="">
  <structure id="s1" title="structure1" type="" formula="" X="0" Y="0" caption="" captionX="0"
    captionY="0">
    <fragment id="s1-0" title="Hydrogen" type="AtomFragment" symbol="H" linkType=""
      orientation="" projection="" x1="0" y1="0" x2="0" y2="0">
      <atom id="s1-0-a1" title="Hydrogen" hybridization="" symbol="H" position=""
        isotopeLabel="" x="301" y="188">
        <electronLink id="s1-0-a1e1" title="1s" electronStatus="bPair" charge="0"
          chargeCount="0" affinity="" bond="sigma" order="single" linkStatus="linkTarget"
          target="s1-1-a1e1" orientation="0" projection="" x1="301" y1="188" x2="331"
          y2="188" />
      </atom>
    </fragment>
    <fragment id="s1-1" title="Chlorine" type="AtomFragment" symbol="Cl" linkType="direct"
      orientation="0" projection="" x1="301" y1="188" x2="331" y2="188">
      <atom id="s1-1-a1" title="Chlorine" hybridization="sp3" symbol="Cl" position=""
        isotopeLabel="" x="331" y="188">
        <electronLink id="s1-1-a1e1" title="2sp3" electronStatus="bPair" charge="0"
          chargeCount="0" affinity="" bond="sigma" order="single" linkStatus="linkSource"
          target="s1-0-a1e1" orientation="0" projection="" x1="331" y1="188" x2="301"
          y2="188" />
        <electronLink id="s1-1-a1e2" title="2sp3" electronStatus="IPair" charge="0"
          chargeCount="0" affinity="coordinate" bond="" order="" linkStatus="" target=""
          orientation="" projection="" x1="331" y1="188" x2="331" y2="188" />
        <electronLink id="s1-1-a1e3" title="2sp3" electronStatus="IPair" charge="0"
          chargeCount="0" affinity="coordinate" bond="" order="" linkStatus="" target=""
          orientation="" projection="" x1="331" y1="188" x2="331" y2="188" />
        <electronLink id="s1-1-a1e4" title="2sp3" electronStatus="IPair" charge="0"
          chargeCount="0" affinity="coordinate" bond="" order="" linkStatus="" target=""
          orientation="" projection="" x1="331" y1="188" x2="331" y2="188" />
      </atom>
    </fragment>
  </structure>
</structureGroup>

```

**Figure 9.** ChemFul for the molecular hydrogen-chloride structure (see also 2 in Figure 4).

In order to bring more insight in the semantics of bond mapping done by ChemEd, we describe and comment on the XML code generated for hydrogen chloride (molecular and ionic description), ammonia, ammonium chloride, ferrocene, and diborane.

In the construction of molecular hydrogen chloride, the “electronLink” with the ‘uPair’ electronStatus for the fragments “Hydrogen” (Code 1, Figure 8) and “Chlorine” (Code 2, Figure 8) are changed into “electronLink” with “bPair” for ‘electronStatus’ of “Hydrogen” (Code 3, Figure 8) and “Chlorine” (Code 4, Figure 8) fragments after they are joined together to construct the molecular hydrogen chloride structure.

Before joining the two fragments, the semantics of “electronLink” in both of the “Hydrogen” and “Chlorine” fragments indicate that the “electronLink” is eligible for a covalent-bond formation according to the values of “covalent” in the ‘affinity’ attribute and “null” for the ‘bond’, ‘order’, ‘linkStatus’, and ‘target’ attributes. The same attributes are filled with the values like “”, “sigma”, and “single” for ‘affinity’, ‘bond’, and ‘order’ attributes, respec-

tively, indicating the bond mapping between the “Hydrogen” and “Chlorine” fragments as a single bond. The null value in the ‘affinity’ attribute does not allow further bonding on the “electronLink”. The ‘linkStatus’ of hydrogen and chlorine is filled with “linkTarget” and “linkSource”, respectively, denoting that the source “Chlorine” fragment is fixed onto the target “Hydrogen” fragment. The system maps this bond link additionally with the ‘target’ attribute to hold ‘id’ values of the fragments mutually. The ChemFul code generated for molecular hydrogen chloride is shown in Figure 9 (see also 2 in Figure 4).

In the case of the construction of ionic hydrogen chloride (3, Figure 4), the ionic fragments, “HydrogenIon” and “ChlorideIon” are placed on the screen as two separate structures, and they are joined with the “Ion-Ion Interaction” tool by dragging a line between the hydrogen and chlorine atoms. This activity encodes the line as an ion–ion interaction with an “interaction” element in ChemFul. This is a mapping between the “electronLink” of source (“HydrogenIon”) and target (“ChlorideIon”) objects on the screen. So the ‘id’ value of the “electronLink” is entered in the respective “sourceId”

```

<?xml version="1.0"?>
<structureGroup id="" title="ionic hydrogen chloride" type="">
  <structure id="s1" title="structure1" type="" formula="" X="0" Y="0" caption="" captionX="0" captionY="0">
    <fragment id="s1-0" title="Hydrogenlon" type="AtomFragment" symbol="H+" linkType=""
      orientation="" projection="" x1="0" y1="0" x2="0" y2="0">
      <atom id="s1-0-a1" title="Hydrogen" hybridization="" symbol="H" position=""
        isotopeLabel="" x="221" y="210">
        <electronLink id="s1-0-a1e1" title="1s" electronStatus="empty" charge="+" chargeCount="1" affinity="ionic" bond="" order="" linkStatus="" target="" orientation="" projection="" x1="221" y1="210" x2="221" y2="210"/>
      </atom>
    </fragment>
  </structure>
  <structure id="s2" title="structure2" type="" formula="" X="35" Y="-1" caption="" captionX="-35" captionY="1">
    <fragment id="s2-0" title="Chloridelon" type="AtomFragment" symbol="Cl-" linkType=""
      orientation="" projection="" x1="0" y1="0" x2="0" y2="0">
      <atom id="s2-0-a1" title="Chlorine" hybridization="sp3" symbol="Cl" position=""
        isotopeLabel="" x="250" y="212">
        <electronLink id="s2-0-a1e1" title="2sp3" electronStatus="IPair" charge="-" chargeCount="1" affinity="ionic" bond="" order="" linkStatus="" target="" orientation="" projection="" x1="250" y1="212" x2="256" y2="209"/>
        <electronLink id="s2-0-a1e2" title="2sp3" electronStatus="IPair" charge="0" chargeCount="0" affinity="coordinate" bond="" order="" linkStatus="" target="" orientation="" projection="" x1="250" y1="212" x2="256" y2="209"/>
        <electronLink id="s2-0-a1e3" title="2sp3" electronStatus="IPair" charge="0" chargeCount="0" affinity="coordinate" bond="" order="" linkStatus="" target="" orientation="" projection="" x1="250" y1="212" x2="256" y2="209"/>
        <electronLink id="s2-0-a1e4" title="2sp3" electronStatus="IPair" charge="0" chargeCount="0" affinity="coordinate" bond="" order="" linkStatus="" target="" orientation="" projection="" x1="250" y1="212" x2="256" y2="209"/>
      </atom>
    </fragment>
  </structure>
  <interaction title="ionic" type="ion-ion" sourceObject="Hydrogenlon" targetObject="Chloridelon" sourceld="s1-0-a1e1" targetId="s2-0-a1e1" x1="221" y1="210" x2="250" y2="212"/>
</structureGroup>

```

**Figure 10.** ChemFull for the ionic hydrogen-chloride structure (see also 3 in Figure 4).

```

<?xml version="1.0"?>
<structureGroup id="" title="ionic hydrogen chloride" type="">
  <structure id="s1" title="structure1" type="" formula="" X="0" Y="0" caption="" captionX="0" captionY="0">
    <fragment id="s1-0" title="Hydrogenlon" type="AtomFragment" link="" orientation=""
      target1="" target2="" bgX1="" bgY1="" bgX2="" bgY2="">
    </structure>
  <structure id="s2" title="structure2" type="" formula="" X="35" Y="-1" caption="" captionX="-35" captionY="1">
    <fragment id="s2-0" title="Chloridelon" type="AtomFragment" link="" orientation=""
      target1="" target2="" bgX1="" bgY1="" bgX2="" bgY2="">
    </structure>
    <interaction title="ionic" type="ion-ion" sourceObject="Hydrogenlon" targetObject="Chloridelon" sourceld="s1-0-a1e1" targetId="s2-0-a1e1"/>
</structureGroup>

```

**Figure 11.** ChemFil for the ionic hydrogen-chloride structure (see also 3 in Figure 4).

and “targetId” along with the exact coordinate values for the rendering purpose, as shown in Figure 10.

The ChemEd generates the ChemFil code consisting of the instruction on how the structure is constructed with the fragments, the structural modifications, and the interactions drawn, etc., while the structure is saved as an XML file. The ChemFil code generated for the structure of ionic hydrogen chloride (see also 3 in Figure 4) is shown in Figure 11. ChemFil captures the structure-building sequence with “fragment” elements, using the appropriate attributes identified. The attributes ‘id’, ‘title’, and ‘type’ are used to capture the id, title, and type of fragments. The ‘link’ attribute captures the information on how the fragment is joined with its target. The direction along with the fragment are joined and captured in the ‘orientation’ attribute. The unique id value of target “electronLink” is recorded in the ‘target1’ attribute. In case of a bridge- or fuse-type fragment, the second id value is stored in the ‘target2’ attribute. The ‘bgX1’, ‘bgY1’, ‘bgX2’, and ‘bgY2’ attributes are exclusively used to store the relative

coordinate values of bridge fragment atoms. ChemEd uses this document to store and open the stored structures on the screen.

Next we disclose the changes in the XML part generated for the “Nitrogen” fragment in ammonium chloride (see also 5 in Figure 4), involving ammonia (see also 1 in Figures 4 and 12) and then the ammonium ion (see also 4 in Figures 4 and 13) along with the ion–ion interaction encoding in the ammonium chloride structure group (see also 5 in Figures 4 and 14).

Figure 12 reports the XML part for the nitrogen in ammonia built by attaching three “Hydrogen” fragments to the “Nitrogen” fragment defined with three “electronLink” having an “uPair” “electronStatus” and with a fourth one with “IPair” “electronStatus”. The three “electronLink” with “uPair” “electronStatus” are used to create three nitrogen–hydrogen single bonds.

Figure 13 reports the XML part for the nitrogen in the ammonium ion built by attaching a “HydrogenIon” fragment to the “Nitrogen” fragment of the ammonia structure. This

```

<fragment id="s1-0" title="Nitrogen" type="AtomFragment" symbol="N" linkType=""
orientation="" projection="" x1="336" y1="225" x2="336" y2="225">
<atom id="s1-0-a1" title="Nitrogen" hybridization="sp3" symbol="N" position=""
isotopeLabel="" x="336" y="225">
<electronLink id="s1-0-a1e1" title="2sp3" electronStatus="bPair" charge="0"
chargeCount="0" affinity="" bond="sigma" order="single" linkStatus="linkTarget"
target="s1-1-a1e1" orientation="60" projection="" x1="336" y1="225" x2="351"
y2="199"/>
<electronLink id="s1-0-a1e2" title="2sp3" electronStatus="bPair" charge="0"
chargeCount="0" affinity="" bond="sigma" order="single" linkStatus="linkTarget"
target="s1-2-a1e1" orientation="210" projection="" x1="336" y1="225" x2="310"
y2="240"/>
<electronLink id="s1-0-a1e3" title="2sp3" electronStatus="bPair" charge="0"
chargeCount="0" affinity="" bond="sigma" order="single" linkStatus="linkTarget"
target="s1-3-a1e1" orientation="330" projection="" x1="336" y1="225" x2="362"
y2="240"/>
<electronLink id="s1-0-a1e4" title="2sp3" electronStatus="IPair" charge="0"
chargeCount="0" affinity="coordinate" bond="" order="" linkStatus="" target=""
orientation="" projection="" x1="336" y1="225" x2="336" y2="225"/>
</atom>
</fragment>

```

**Figure 12.** The “electronLink” for the nitrogen atom of the “Nitrogen” Fragment in the ammonia structure (see also 1 in Figure 4).

```

<fragment id="s1-0" title="Nitrogen" type="AtomFragment" symbol="N" linkType=""
orientation="" projection="" x1="336" y1="225" x2="336" y2="225">
<atom id="s1-0-a1" title="Nitrogen" hybridization="sp3" symbol="N" position=""
isotopeLabel="" x="336" y="225">
<electronLink id="s1-0-a1e1" title="2sp3" electronStatus="bPair" charge="0"
chargeCount="0" affinity="" bond="sigma" order="single" linkStatus="linkTarget"
target="s1-1-a1e1" orientation="60" projection="" x1="336" y1="225" x2="351"
y2="199"/>
<electronLink id="s1-0-a1e2" title="2sp3" electronStatus="bPair" charge="0"
chargeCount="0" affinity="" bond="sigma" order="single" linkStatus="linkTarget"
target="s1-2-a1e1" orientation="210" projection="" x1="336" y1="225" x2="310"
y2="240"/>
<electronLink id="s1-0-a1e3" title="2sp3" electronStatus="bPair" charge="0"
chargeCount="0" affinity="" bond="sigma" order="single" linkStatus="linkTarget"
target="s1-3-a1e1" orientation="330" projection="" x1="336" y1="225" x2="362"
y2="240"/>
<electronLink id="s1-0-a1e4" title="2sp3" electronStatus="bPair" charge="0"
chargeCount="1" affinity="ionic" bond="dative" order="1" linkStatus="linkTarget"
target="s1-4-a1e1" orientation="120" projection="" x1="336" y1="225" x2="321"
y2="199"/>
</atom>
</fragment>

```

**Figure 13.** The “electronLink” for the “Nitrogen” Fragment in the ammonium ion structure (see also 4 in Figure 4).

```

<interaction title="ionic" type="ion-ion" sourceObject="Nitrogen" targetObject="Chloridelon"
sourceId="s1-0-a1e4" targetId="s2-0-a1e1" x1="336" y1="225" x2="365" y2="210"/>

```

**Figure 14.** Ion–ion interaction encoding in ammonium chloride structure group (5, Figure 4).

utilizes the fourth “electronLink” with “IPair” ‘electronStatus’ to create a dative bond with the “HydrogenIon” fragment, leaving a positive charge as can be seen in the highlighted portion in Figure 13.

Placing a “ChlorideIon” fragment near the ammonium-ion structure (see also 4 in Figure 4) creates the chloride ion as a second structure. The ion–ion interaction between the ammonium and chloride ions is drawn with the “Ion–Ion Interaction” tool by dragging a line from the nitrogen atom of the ammonium ion to the chlorine atom of the chloride ion, creating ammonium chloride (see also 5 in

Figure 4). Figure 14 reports the XML part representing the ion–ion interaction in ammonium chloride.

We have described already that the cyclohexane carboxaldehyde 7 (Figure 4) is built from 6 (Figure 4) by the replacement of a carbon atom by an oxygen atom. This change is reflected in the ChemFil encoding of 7 with a “modification” element (highlighted portion of Figure 15). The “modification” element has a ‘title’ attribute to encode the nature of modification. The ‘sourceObject’ and ‘targetObject’ specifies the source and target objects. The ‘targetId1’ attribute is used to hold the ‘id’ values of the target atom

```

<structureGroup id="" title="cyclohexane carboxaldehyde" type="">
<structure id="s1" title="structure1" type="" formula="" X="0" Y="0" caption="" captionX="0"
captionY="0">
<fragment id="s1-0" title="SixMemberRing" type="SkeletonFragment" link="" orientation="270"
target1="" target2="" bgX1="" bgY1="" bgX2="" bgY2="" />
<fragment id="s1-1" title="SingleBond" type="SkeletonFragment" link="direct"
orientation="30" target1="s1-0-a2e3" target2="" bgX1="" bgY1="" bgX2="" bgY2="" />
<fragment id="s1-2" title="DoubleBond" type="SkeletonFragment" link="direct"
orientation="90" target1="s1-1-a1e2" target2="" bgX1="" bgY1="" bgX2="" bgY2="" />
<fragment id="s1-3" title="Hydrogen" type="AtomFragment" link="direct" orientation="330"
target1="s1-1-a1e4" target2="" bgX1="" bgY1="" bgX2="" bgY2="" />
<modification title="Atom" sourceObject="Oxygen" targetObject="Carbon" targetId1="s1-2-a1"
targetId2="" x1="0" y1="0"/>
</structure>
</structureGroup>

```

**Figure 15.** ChemFil code instruction to build the structure of cyclohexane carboxaldehyde (see also 7 in Figure 4).

```

<structureGroup id="" title="ferrocene" type="">
<structure id="s1" title="structure1" type="" formula="" X="0" Y="0">
<fragment id="s1-0" title="FerrousIon" type="AtomFragment" link="" orientation="" target1="" target2="" bgX1="" bgY1="" bgX2="" bgY2="">
</structure>
<structure id="s2" title="structure2" type="" formula="" X="5" Y="-65">
<fragment id="s2-0" title="CyclopentadienylAnionRing" type="TemplateFragment" link="" orientation="90" target1="" target2="" bgX1="" bgY1="" bgX2="" bgY2="">
</structure>
<structure id="s3" title="structure3" type="" formula="" X="4" Y="55">
<fragment id="s3-0" title="CyclopentadienylAnionRing" type="TemplateFragment" link="" orientation="270" target1="" target2="" bgX1="" bgY1="" bgX2="" bgY2="">
</structure>
<interaction title="coordinate" type="metal-pi" sourceObject="Iron" targetObject="Carbon" sourcelId="s1-0-a1e4" targetId="s2-0-a1e4"/>
<interaction title="coordinate" type="metal-pi" sourceObject="Iron" targetObject="Carbon" sourcelId="s1-0-a1e4" targetId="s2-0-a1e4"/>
<interaction title="coordinate" type="metal-pi" sourceObject="Iron" targetObject="Carbon" sourcelId="s1-0-a1e5" targetId="s2-0-a2e4"/>
<interaction title="coordinate" type="metal-pi" sourceObject="Iron" targetObject="Carbon" sourcelId="s1-0-a1e5" targetId="s2-0-a3e4"/>
<interaction title="coordinate" type="metal-pi" sourceObject="Iron" targetObject="Carbon" sourcelId="s1-0-a1e6" targetId="s2-0-a4e4"/>
<interaction title="coordinate" type="metal-pi" sourceObject="Iron" targetObject="Carbon" sourcelId="s1-0-a1e6" targetId="s2-0-a5e4"/>
<interaction title="coordinate" type="metal-pi" sourceObject="Iron" targetObject="Carbon" sourcelId="s1-0-a1e7" targetId="s3-0-a1e4"/>
<interaction title="coordinate" type="metal-pi" sourceObject="Iron" targetObject="Carbon" sourcelId="s1-0-a1e7" targetId="s3-0-a1e4"/>
<interaction title="coordinate" type="metal-pi" sourceObject="Iron" targetObject="Carbon" sourcelId="s1-0-a1e8" targetId="s3-0-a2e4"/>
<interaction title="coordinate" type="metal-pi" sourceObject="Iron" targetObject="Carbon" sourcelId="s1-0-a1e8" targetId="s3-0-a3e4"/>
<interaction title="coordinate" type="metal-pi" sourceObject="Iron" targetObject="Carbon" sourcelId="s1-0-a1e9" targetId="s3-0-a4e4"/>
<interaction title="coordinate" type="metal-pi" sourceObject="Iron" targetObject="Carbon" sourcelId="s1-0-a1e9" targetId="s3-0-a5e4"/>
</structureGroup>

```

**Figure 16.** The ChemFil code generated by ChemEd for Ferrocene structure (see also **10** in Figure 5).

replaced. While structure is saved in the disk, ChemEd adds “Hydrogen” fragments to all the unutilized “electronLink” with “uPair” ‘electronStatus’ eligible for forming a  $\sigma$  bond. The ChemFull encodes all the “Hydrogen” fragments whether it is deliberately drawn or added by the system. However, ChemFil encodes only those “Hydrogen” fragments which are drawn by the user.

The ferrocene structure (**10**) is constructed with one “FerrousIon” and two “CyclopentadienylAnionRing” fragments. ChemEd allows the drawing of metal– $\pi$  interactions between the Fe(II) ion and the cyclopentadienyl ring by selecting the appropriate tool. The “FerrousIon” fragment is defined with 9 “electronLink” (Figure 20) to represent a total of 18 electron noble electron configuration at the central metal. The first three “electronLink” have “lPair” ‘electronStatus’ representing three filled 3d orbitals. The remaining six ‘electronLink’ are defined with “empty” ‘electronStatus’ representing the availability of six vacant metal orbitals describing the metal– $\pi$  interactions. The ferrocene structure is described<sup>13</sup> as a  $\pi$ -complex of Fe(II) having two electrons less with two cyclopentadienyl anions contributing six electrons each. The interaction mapping is done between the “electronLink” of five carbon atoms (“pi” status in the “bond” attribute involving six electrons) in the “CyclopentadienylAnionRing” fragment and six “electronLink” of iron atom (“empty” in the ‘electronStatus’ attribute and “” status in the ‘bond’ attribute) in the “FerrousIon” fragment. The ChemFil code for the structure of ferrocene (**10** in Figures 5 and 16) shows the involvement of the last 6 “electronLink” on the Ferrous (II) ion (**15** in Figure 18), describing 12

metal– $\pi$  interactions, 6 form each cyclopentadienyl anion structure, as can be seen from the ‘sourceId’ and ‘targetId’ attributes.

The multicenter bond<sup>13</sup> in diborane<sup>13a,b</sup> structure **12** is drawn using the “Multi Centered Bonding Interaction” tool. The “Boron” fragment is defined with four “electronLink”. The first three “electronLink” represent three 2sp<sup>2</sup> orbitals with “uPair” ‘electronStatus’. These three are used to describe three boron–hydrogen single bonds. The fourth one is defined with an “empty” ‘electronStatus’, representing the 2pz orbital. The multicentered bonding is encoded as an “interaction” of a hydrogen atom of boron–hydrogen bond to the other “Boron” Fragment as shown in Figure 17.

**ChemLib, The Fragment Library.** ChemLib contains the generic description of the basic fragments that have been identified to build a large variety of different structures and template fragments. ChemLib is a built-in component of ChemEd. The system can be extended to incorporate external template fragment libraries by creating and storing user-defined fragments, making use of the basic fragments.

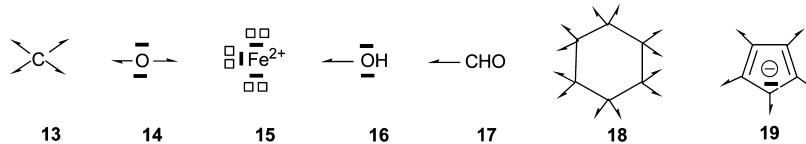
The basic fragments are created manually as XML documents with appropriate semantics in each level. They are stored as separate XML files and packed into four basic classes of fragments (“AtomFragment”, “AtomGroupFragment”, “SkeletonFragment”, and “TemplateFragment”). Some representative fragments: “AtomFragment”, such as “Carbon” (**13**, Figure 18), “Oxygen” (**14**, Figure 18), and “Ferrous(II)Ion” (**15**, Figure 18); AtomGroupFragment”, such as “HydroxyGroup” (**16**, Figure 18) and “FormylGroup” (**17**, Figure 18); “SkeletonFragment”, such as “SixMemberRing” (**18**, Figure 18); and “TemplateFragment”, such as “Cyclo-

```

<structureGroup id="" title="diborane" type="">
  <structure id="s1" title="structure1" type="" formula="" X="0" Y="0" caption="" captionX="0"
    captionY="0">
    <fragment id="s1-0" title="Boron" type="AtomFragment" link="" orientation="" target1=""
      target2="" bgX1="" bgY1="" bgX2="" bgY2=""/>
    <fragment id="s1-1" title="Hydrogen" type="AtomFragment" link="direct" orientation="30"
      target1="s1-0-a1e1" target2="" bgX1="" bgY1="" bgX2="" bgY2=""/>
    <fragment id="s1-2" title="Hydrogen" type="AtomFragment" link="direct" orientation="150"
      target1="s1-0-a1e2" target2="" bgX1="" bgY1="" bgX2="" bgY2=""/>
    <fragment id="s1-3" title="Hydrogen" type="AtomFragment" link="direct" orientation="210"
      target1="s1-0-a1e3" target2="" bgX1="" bgY1="" bgX2="" bgY2=""/>
  </structure>
  <structure id="s2" title="structure2" type="" formula="" X="106" Y="0" caption=""
    captionX="-159" captionY="-1">
    <fragment id="s2-0" title="Boron" type="AtomFragment" link="" orientation="" target1=""
      target2="" bgX1="" bgY1="" bgX2="" bgY2=""/>
    <fragment id="s2-1" title="Hydrogen" type="AtomFragment" link="direct" orientation="210"
      target1="s2-0-a1e1" target2="" bgX1="" bgY1="" bgX2="" bgY2=""/>
    <fragment id="s2-2" title="Hydrogen" type="AtomFragment" link="direct" orientation="30"
      target1="s2-0-a1e2" target2="" bgX1="" bgY1="" bgX2="" bgY2=""/>
    <fragment id="s2-3" title="Hydrogen" type="AtomFragment" link="direct" orientation="330"
      target1="s2-0-a1e3" target2="" bgX1="" bgY1="" bgX2="" bgY2=""/>
  </structure>
  <interaction title="sigma" type="multi" sourceObject="Hydrogen" targetObject="Boron"
    sourceId="s1-1-a1e1" targetId="s2-0-a1e1"/>
  <interaction title="sigma" type="multi" sourceObject="Hydrogen" targetObject="Boron"
    sourceId="s2-1-a1e1" targetId="s1-0-a1e1"/>
</structureGroup>

```

**Figure 17.** ChemFil code generated by ChemEd for Diborane structure (see also **12** in Figure 5).



ElectronLink-Unpair: —; ElectronLink-LonePair: -; ElectronLink-Empty □; Bond: —

**Figure 18.** Representative fragments from ChemLib, **13**: “Carbon”, **14**: “Oxygen”, **15**: “Ferrous(II)Ion”, **16**: “HydroxyGroup”, **17**: “FormylGroup”, **18**: “SixMemberRing”, and **19**: “CyclopentadienylAnion”.

```

<fragment title="Carbon" type="AtomFragment" symbol="C" smiles="C">
  <atom id="a1" title="Carbon" hybridization="sp3" symbol="C" position="" isotopeLabel=""
    orientation="">
    <electronLink id="a1e1" title="2sp3" electronStatus="uPair" charge="0" chargeCount="0"
      affinity="covalent" bond="" order="" target="" linkStatus="">
    <electronLink id="a1e2" title="2sp3" electronStatus="uPair" charge="0" chargeCount="0"
      affinity="covalent" bond="" order="" target="" linkStatus="">
    <electronLink id="a1e3" title="2sp3" electronStatus="uPair" charge="0" chargeCount="0"
      affinity="covalent" bond="" order="" target="" linkStatus="">
    <electronLink id="a1e4" title="2sp3" electronStatus="uPair" charge="0" chargeCount="0"
      affinity="covalent" bond="" order="" target="" linkStatus="">
  </atom>
</fragment>

```

**Figure 19.** XML code for the “Carbon” fragment (see also **13** in Figure 18).

```

<fragment title="FerrousIion" type="AtomFragment" symbol="Fe2+" smiles="[Fe2+]>
  <atom id="a1" title="Iron" hybridization="d2sp3" symbol="Fe" position="" isotopeLabel=""
    orientation="">
    <electronLink id="a1e1" title="3d" electronStatus="IPair" charge="0" chargeCount="0"
      affinity="coordinate" bond="" order="" target="" linkStatus="">
    <electronLink id="a1e2" title="3d" electronStatus="IPair" charge="0" chargeCount="0"
      affinity="coordinate" bond="" order="" target="" linkStatus="">
    <electronLink id="a1e3" title="3d" electronStatus="IPair" charge="0" chargeCount="0"
      affinity="coordinate" bond="" order="" target="" linkStatus="">
    <electronLink id="a1e4" title="d2sp3" electronStatus="empty" charge="+" chargeCount="2"
      affinity="ionic" bond="" order="" target="" linkStatus="">
    <electronLink id="a1e5" title="d2sp3" electronStatus="empty" charge="0" chargeCount="0"
      affinity="coordinate" bond="" order="" target="" linkStatus="">
    <electronLink id="a1e6" title="d2sp3" electronStatus="empty" charge="0" chargeCount="0"
      affinity="coordinate" bond="" order="" target="" linkStatus="">
    <electronLink id="a1e7" title="d2sp3" electronStatus="empty" charge="0" chargeCount="0"
      affinity="coordinate" bond="" order="" target="" linkStatus="">
    <electronLink id="a1e8" title="d2sp3" electronStatus="empty" charge="0" chargeCount="0"
      affinity="coordinate" bond="" order="" target="" linkStatus="">
    <electronLink id="a1e9" title="d2sp3" electronStatus="empty" charge="0" chargeCount="0"
      affinity="coordinate" bond="" order="" target="" linkStatus="">
  </atom>
</fragment>

```

**Figure 20.** XML code for the “Ferrous (II) ion” fragment (see also **15** in Figure 18).

pentadienylAnion” (**19**, Figure 18) are shown in Figure 18. The Figures 19–22 disclose the XML code for the fragments

representing “Carbon”, “Ferrous (II) ion”, “FormylGroup”, and “SixMemberRing” fragments.

```

<fragment title="FormylGroup" type="AtomGroupFragment" symbol="CHO" smiles="C=O">
  <atom id="a1" title="Carbon" hybridization="sp3" symbol="C" position="" isotopeLabel=""
    orientation="">
    <electronLink id="a1e1" title="2sp3" electronStatus="uPair" charge="0" chargeCount="0"
      affinity="covalent" bond="" order="" target="" linkStatus="" />
  </atom>
</fragment>

```

**Figure 21.** XML code for the “FormylGroup” fragment (see also 17 in Figure 18).

```

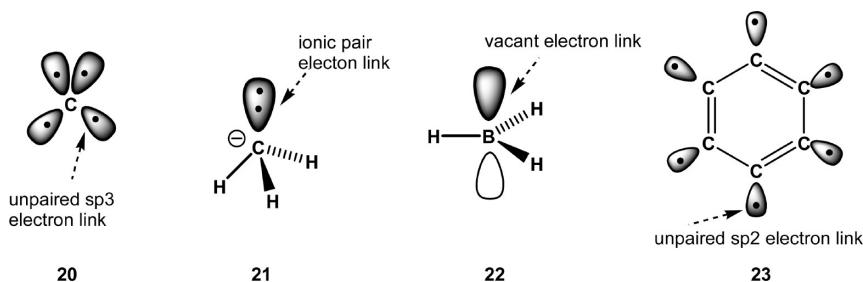
<fragment title="SixMemberRing" type="SkeletonFragment" symbol="C6" smiles="C1CCCCC1">
  <atom id="a1" title="Carbon" hybridization="sp3" symbol="C" position="1" isotopeLabel=""
    orientation="0">
    <electronLink id="a1e1" title="2sp3" electronStatus="bPair" charge="0" chargeCount="0"
      affinity="" bond="sigma" order="single" target="a2e1" linkStatus="link" />
    <electronLink id="a1e2" title="2sp3" electronStatus="bPair" charge="0" chargeCount="0"
      affinity="" bond="sigma" order="single" target="a2e2" linkStatus="link" />
    <electronLink id="a1e3" title="2sp3" electronStatus="uPair" charge="0" chargeCount="0"
      affinity="covalent" bond="" order="" target="" linkStatus="" />
    <electronLink id="a1e4" title="2sp3" electronStatus="uPair" charge="0" chargeCount="0"
      affinity="covalent" bond="" order="" target="" linkStatus="" />
  </atom>
  <atom id="a2" title="Carbon" hybridization="sp3" symbol="C" position="2" isotopeLabel=""
    orientation="60">
    <electronLink id="a2e1" title="2sp3" electronStatus="bPair" charge="0" chargeCount="0"
      affinity="" bond="sigma" order="single" target="a1e1" linkStatus="link" />
    <electronLink id="a2e2" title="2sp3" electronStatus="bPair" charge="0" chargeCount="0"
      affinity="" bond="sigma" order="single" target="a3e1" linkStatus="link" />
    <electronLink id="a2e3" title="2sp3" electronStatus="uPair" charge="0" chargeCount="0"
      affinity="covalent" bond="" order="" target="" linkStatus="" />
    <electronLink id="a2e4" title="2sp3" electronStatus="uPair" charge="0" chargeCount="0"
      affinity="covalent" bond="" order="" target="" linkStatus="" />
  </atom>
  <atom id="a3" title="Carbon" hybridization="sp3" symbol="C" position="3" isotopeLabel=""
    orientation="0">
    <electronLink id="a3e1" title="2sp3" electronStatus="bPair" charge="0" chargeCount="0"
      affinity="" bond="sigma" order="single" target="a2e2" linkStatus="link" />
    <electronLink id="a3e2" title="2sp3" electronStatus="bPair" charge="0" chargeCount="0"
      affinity="" bond="sigma" order="single" target="a4e1" linkStatus="link" />
    <electronLink id="a3e3" title="2sp3" electronStatus="uPair" charge="0" chargeCount="0"
      affinity="covalent" bond="" order="" target="" linkStatus="" />
    <electronLink id="a3e4" title="2sp3" electronStatus="uPair" charge="0" chargeCount="0"
      affinity="covalent" bond="" order="" target="" linkStatus="" />
  </atom>
  <atom id="a4" title="Carbon" hybridization="sp3" symbol="C" position="4" isotopeLabel=""
    orientation="300">
    <electronLink id="a4e1" title="2sp3" electronStatus="bPair" charge="0" chargeCount="0"
      affinity="" bond="sigma" order="single" target="a3e2" linkStatus="link" />
    <electronLink id="a4e2" title="2sp3" electronStatus="bPair" charge="0" chargeCount="0"
      affinity="" bond="sigma" order="single" target="a5e1" linkStatus="link" />
    <electronLink id="a4e3" title="2sp3" electronStatus="uPair" charge="0" chargeCount="0"
      affinity="covalent" bond="" order="" target="" linkStatus="" />
    <electronLink id="a4e4" title="2sp3" electronStatus="uPair" charge="0" chargeCount="0"
      affinity="covalent" bond="" order="" target="" linkStatus="" />
  </atom>
  <atom id="a5" title="Carbon" hybridization="sp3" symbol="C" position="5" isotopeLabel=""
    orientation="240">
    <electronLink id="a5e1" title="2sp3" electronStatus="bPair" charge="0" chargeCount="0"
      affinity="" bond="sigma" order="single" target="a4e2" linkStatus="link" />
    <electronLink id="a5e2" title="2sp3" electronStatus="bPair" charge="0" chargeCount="0"
      affinity="" bond="sigma" order="single" target="a6e1" linkStatus="link" />
    <electronLink id="a5e3" title="2sp3" electronStatus="uPair" charge="0" chargeCount="0"
      affinity="covalent" bond="" order="" target="" linkStatus="" />
    <electronLink id="a5e4" title="2sp3" electronStatus="uPair" charge="0" chargeCount="0"
      affinity="covalent" bond="" order="" target="" linkStatus="" />
  </atom>
  <atom id="a6" title="Carbon" hybridization="sp3" symbol="C" position="6" isotopeLabel=""
    orientation="180">
    <electronLink id="a6e1" title="2sp3" electronStatus="bPair" charge="0" chargeCount="0"
      affinity="" bond="sigma" order="single" target="a5e2" linkStatus="link" />
    <electronLink id="a6e2" title="2sp3" electronStatus="bPair" charge="0" chargeCount="0"
      affinity="" bond="sigma" order="single" target="a1e2" linkStatus="link" />
    <electronLink id="a6e3" title="2sp3" electronStatus="uPair" charge="0" chargeCount="0"
      affinity="covalent" bond="" order="" target="" linkStatus="" />
    <electronLink id="a6e4" title="2sp3" electronStatus="uPair" charge="0" chargeCount="0"
      affinity="covalent" bond="" order="" target="" linkStatus="" />
  </atom>
</fragment>

```

**Figure 22.** XML code for the fragment “SixMemberRing” fragment (see also 18 in Figure 18).

The inclusion of semantics about the “electronLink” of each atom in the structure description provides the meaning of the electronic status on the orbital, such as empty, lone pair, and unpaired orbitals and bond as bond paired. For example, a sp<sub>3</sub> carbon is described as a carbon atom with

four “electronLink” tags to represent four sp<sub>3</sub>-hybridized orbitals with an unpaired electron status in each link (20, Figure 23). Similarly, the methyl carbanion is described with four “electronLink” involving three carbon–hydrogen single bonds, one lone pair electron status, and a negative charge



**Figure 23.** Electron status associated with “electronLink” of few fragments.

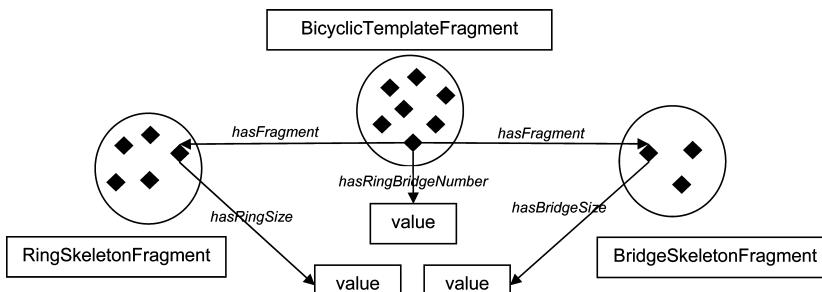


**Figure 24.** Concept graph of fragment ontology generated by the Protégé.

(21, Figure 23). In borane (22, Figure 23), the boron atom is described with four “electronLink”. The first three are utilized for three boron–hydrogen single bonds, and the last one is a vacant electron link described with “empty” ‘electronStatus’. The carbon skeleton present in the benzene ring (23, Figure 23) is described with six carbon atoms (sp<sup>2</sup>

hybridized) forming the aromatic ring and each atom bearing one “electronLink” with “uPair” electron status for further bonding.

**Fragment Ontology.** Since the conceptual content of chemical structure is identified in terms of structural fragments and their constituents, the vocabulary related to various



**Figure 25.** Relationships in a bridged bicyclic ring system.

Code 1: “RingSkeletonFragment” and “BridgeSkeletonFragment” are related to “BicyclicTemplateFragment”

```

<SubClassOf>
<Class URI="&Fragment ontology;BicyclicTemplateFragment">
<ObjectIntersectionOf>
<ObjectSomeValuesFrom>
<ObjectProperty URI="&Fragment ontology;hasFragment"/>
<Class URI="&Fragment ontology;BridgeSkeletonFragment"/>
</ObjectSomeValuesFrom>
<ObjectSomeValuesFrom>
<ObjectProperty URI="&Fragment ontology;hasFragment"/>
<Class URI="&Fragment ontology;RingSkeletonFragment"/>
</ObjectSomeValuesFrom>
</ObjectIntersectionOf>
</SubClassOf>
  
```

Code 2: “hasRingSize” relation in “SixMemberRing”

```

<DataPropertyAssertion>
<DataProperty URI="&Fragment ontology;hasRingSize"/>
<Individual URI="&Fragment ontology;SixMemberRing"/>
<Constant datatypeURI="&xsd;integer">6</Constant>
</DataPropertyAssertion>
  
```

Code 3: “hasBridgeSize” relation in “OneMemberBridge”

```

<DataPropertyAssertion>
<DataProperty URI="&Fragment ontology;hasBridgeSize"/>
<Individual URI="&Fragment ontology;OneMemberBridge"/>
<Constant datatypeURI="&xsd;integer">1</Constant>
</DataPropertyAssertion>
  
```

Code 4: “Bicyclo221HeptaneRing” associated with “hasRingBridgeNumber” relations

```

<DataPropertyAssertion>
<DataProperty URI="&Fragment ontology;hasRingBridgeNumber"/>
<Individual URI="&Fragment ontology;Bicyclo221HeptaneRing"/>
<Constant datatypeURI="&xsd;string">6-1</Constant>
</DataPropertyAssertion>
<DataPropertyAssertion>
<DataProperty URI="&Fragment ontology;hasRingBridgeNumber"/>
<Individual URI="&Fragment ontology;Bicyclo221HeptaneRing"/>
<Constant datatypeURI="&xsd;string">5-2</Constant>
</DataPropertyAssertion>
  
```

**Figure 26.** The OWL code snippet generated through Protégé<sup>11</sup> for “subClassOf” definition for “BicyclicTemplateFragment” and the “dataTypeAssertions” for “SixMemberRing”, “OneMemberBridge”.

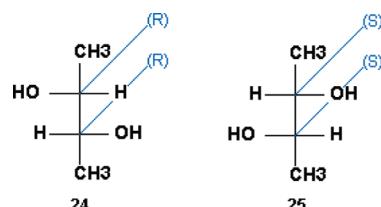
structural fragments is organized in fragment ontology. The top-level concept, such as Fragment, Atom, and “electronLink”, is related with *has* relationship transitively. The ontology extends with the classification of various fragments as subclasses of a fragment. The ontology is created with the state-of-the-art ontology editing tool Protégé version 4.0<sup>11</sup> and implemented in a W3C-approved ontology language, Web Ontology Language (OWL2.0).<sup>11c</sup> The concept graph obtained for the subclass hierarchy of Fragments is shown in the Figure 24.

According to the concept taxonomy, the concept ‘fragment’ is classified into four subclasses, namely “AtomFragment”, “AtomGroupFragment”, “SkeletonFragment”, and “TemplateFragment”: (i) The “AtomFragment” is a structural portion consisting of only one atom. It is further classified as “NeutralAtomFragment”, “AnionicAtomFragment”, and “CationicAtomFragment”. The first one includes the individual instances representing a single atom fragment, such as C, H, O, N, Cl, Br, I, S, P, etc. The second class includes the instances representing the negatively charged atom fragments, like C<sup>-</sup>, H<sup>-</sup>, F<sup>-</sup>, Cl<sup>-</sup>, Br<sup>-</sup>, and I<sup>-</sup>, and the third one is to describe the positively charged atom fragments,

such as C<sup>+</sup>, H<sup>+</sup>, Cl<sup>+</sup>, Br<sup>+</sup>, I<sup>+</sup>, Na<sup>+</sup>, K<sup>+</sup>, Mg<sup>2+</sup>, Fe<sup>2+</sup>, Fe<sup>3+</sup>, etc.; (ii) The “AtomGroupFragment” represents a group of atoms describing meaningful chemical groups, such as carbonyl, formyl, hydroxyl, etc. It is also classified as anionic, cationic, and neutral atom groups representing condensed structures of atom groups, such as OH<sup>-</sup>, COO<sup>-</sup>, NH<sub>2</sub><sup>-</sup>, CN<sup>-</sup>; NH<sub>4</sub><sup>+</sup>, OH<sub>3</sub><sup>+</sup>; OH, CHO, CO, COOH, CH<sub>3</sub>, C<sub>2</sub>H<sub>5</sub>, C<sub>3</sub>H<sub>7</sub>, etc., which can include molecular “AtomGroupFragment”, such as BH<sub>3</sub>, NH<sub>3</sub>, CH<sub>3</sub>OH, (C<sub>2</sub>H<sub>5</sub>)<sub>2</sub>O, and CH<sub>3</sub>COCH<sub>3</sub> particularly useful for describing complexes; and (iii) The “SkeletonFragment” represents the structural fragments with more than one atom connected with bonds. The default atom in the SkeletonFragment is carbon, and the skeleton includes both acyclic and cyclic systems like single and double bonds, three- and six-member rings, etc. Accordingly, the concepts are further classified in more specific categories, resulting in concept taxonomy with a reasonable granularity. The concept tree ends with concepts associating the individual instances.

The “SkeletonFragment” concept is included in the ontology to bring the possibility of line segment drawing without showing the hydrogen atoms. A skeleton fragment is a structural portion with one or more chemical bonds connecting carbon atoms by default. This concept includes both open and closed skeletons, i.e., the cyclic and acyclic skeletons. There are five classes of skeleton fragments identified, viz. “BondSkeletonFragment”, “RingSkeletonFragment”, BridgeSkeletonFragment, FuseSkeletonFragment, and “SpiralSkeletonFragment”. The “BondSkeletonFragment” includes the individual instances, such as “SingleBond”, “DoubleBond”, and “TripleBond”. The “RingSkeletonFragment” is defined to include the instances like “Three-MemberRing”, “FourMemberRing”, “FiveMemberRing”, “Six-MemberRing”, “SevenMemberRing”, and “EightMemberRing”. Similarly to include the fuse rings, the concept of “Fuse-SkeletonFragment” is included with the instances “Three-MemberFuseRing”, “FourMemberFuseRing”, “FiveMember-FuseRing”, “SixMemberFuseRing”, “SevenMemberFuseRing”, and “EightMemberFuseRing”. The “SpiralSkeletonFragment” includes “ThreeMemberSpiroRing”, “FourMemberSpiroRing”, “FiveMemberSpiroRing”, and “SixMemberSpiroRing”. The bridge structures are described with a ring associated with a bridge skeleton. For this purpose, “OneMemberBridge”, “Two-MemberBridge”, etc. are defined as the instances of “Bridge-SkeletonFragment” at present.

The “TemplateFragment” is to describe fragments derived from the first three fragment types. For example, a bicyclic template fragment like “Bicyclo221HeptaneRing” can be constructed with a suitable “RingSkeletonFragment” and one



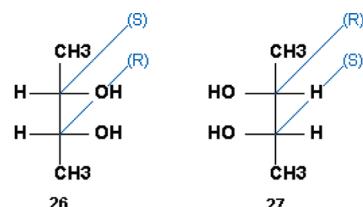
**Figure 27.** The ChemEd configuration around chiral carbons with *R/S* notation.

of the “BridgeSkeletonFragment” fragments. Similarly, a polycyclic template fragment representing a steroid skeleton can be constructed by fusing the ring skeleton fragments appropriately. Different conformations of same cyclic structures like chair and boat forms of cyclohexane ring systems can be created and added into the ChemLib in the “Template-SkeletonFragment” category for structure construction.

In Fragment ontology, the bicyclic template ring system is defined as a combination of ring and bridge skeletons. The concepts and their relationships are shown in Figure 25. Accordingly, each member of “BicyclicTemplateFragment” class is related to a member of both the “RingSkeleton-Fragment” and the “BridgeSkeletonFragment” through a “hasFragment” relationship (Code 1, Figure 26). The exact class member of any “BicyclicTemplateFragment” can be inferred from the values of data-type assertions on the instances belonging to the classes.

For example, the system can infer the appropriate bicyclic ring system through the data-type property assertion defined for the “BicyclicTemplateFragment” members. Here generically represented knowledge related to ring systems provides useful inferences through the relationships of the concepts. The “RingSkeletonFragment” has the “hasRingSize” relation with the individual instance holding the value of the appropriate ring size (Code 2, Figure 26). Similarly the “BridgeSkeletonFragment” is associated with the “has-BridgeSize” relation to hold the value of bridge size (Code 3, Figure 26). Since these two concepts, “RingSkeleton-Fragment” and “BridgeSkeletonFragment” are related to the “BicyclicTemplateFragment” associated with a relation “has-RingBridgeNumber”, allowing the possible combinations of ring and bridge defined inside the ontology (Code 4, Figure 26). Accordingly, if a bicyclic ring system is constructed with a “SixMemberRing” and “OneMemberBridge” or “FiveMemberRing” and “TwoMemberBridge” skeletons, the ChemEd refers to the ontology to fix the bicyclic ring system as a “Bicyclo221HeptaneRing” skeleton. The OWL code snippet generated through Protégé<sup>11</sup> from the fragment ontology is shown in Figure 26.

**Processing of ChemFul.** The markup features of ChemFul provide deeper semantics associated with the chemical structure. Valid inferences about the chemical bonding can be obtained through this approach. The explicit encoding of  $\sigma$  and  $\pi$  bonds, “electronLink” with lone pair electrons, and empty links, poses a new way of structure construction and description. The information available in the ChemFul can be processed appropriately to generate many details about the structure. We describe below our preliminary results to detect the stereochemistry of chiral centers<sup>14</sup> and provide the notations according to both Cahn–Ingold–Prelog<sup>14a,b</sup> and IUPAC rules.<sup>14c</sup> Further, we show the use of ChemFul to generate some atom descriptors and functional group reports with the support of external ontologies developed by us in Protégé<sup>11</sup> for this demonstration.



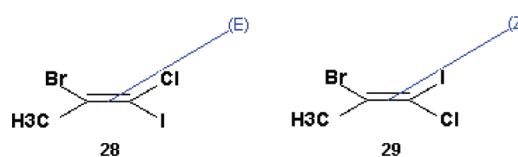
**Top Carbon      Bottom Carbon**

**Figure 28.** Orientation of groups around top and bottom chiral carbons in 24.

**Processing of ChemFul to Infer Stereochemistry in Structures.** The configuration around chiral center in terms of *R/S* notation detected by ChemEd for structures 24–27 containing two adjacent chiral carbons is shown in Figure 27.

The stereochemistry around a carbon atom is obtained by identifying the four atoms attached to a sp<sup>3</sup> carbon along with their orientation. Subsequently the groups or atoms linked to these four atoms are also captured. The configuration in terms of *R/S* notation is attempted with the Fischer projection rule<sup>15</sup> by drawing the groups around carbon atom with the orientations of 0, 90, 180, and 270, as shown in Figure 28. The atom groups along these orientations around a sp<sup>3</sup> carbon are checked for the chirality. In case of a chiral carbon, the groups are arranged in sequence starting with the highest to lowest priority, according to the priority rule. The corresponding orientations are also sequenced respectively. From this sequence the *R/S* notation of the chiral carbon is obtained by a rearranging the orientations of the groups. For example, the priority sequence of groups along with their orientation sequence for the top chiral carbon (priority sequence: HO;C(OH)(H)(CH<sub>3</sub>);CH<sub>3</sub>;H and orientation sequence: 180;270;90;0) and the bottom chiral carbon (priority sequence: HO;C(OH)(H)(CH<sub>3</sub>);CH<sub>3</sub>;H and orientation sequence: 0;90;270;180) in 24 is shown in Figure 28.

The configuration of the structure is obtained by an interchange of the orientations of lowest and highest priority groups, followed by the elimination of the fourth orientation from the resultant sequence after interchange. This operation sets the groups to visualize in such a way as to arrange the orientations of the groups starting from highest to lowest priority, either in clock- or anticlockwise directions, to be compared with the orientation clock in Figure 28. The resultant direction according to orientation clock allows



**Figure 29.** ChemEd geometrical isomerism with *E/Z* notation.

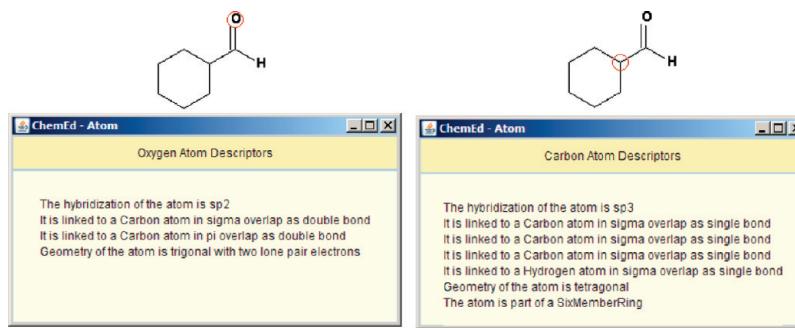


Figure 30. Atom descriptors generated from ChemFul.

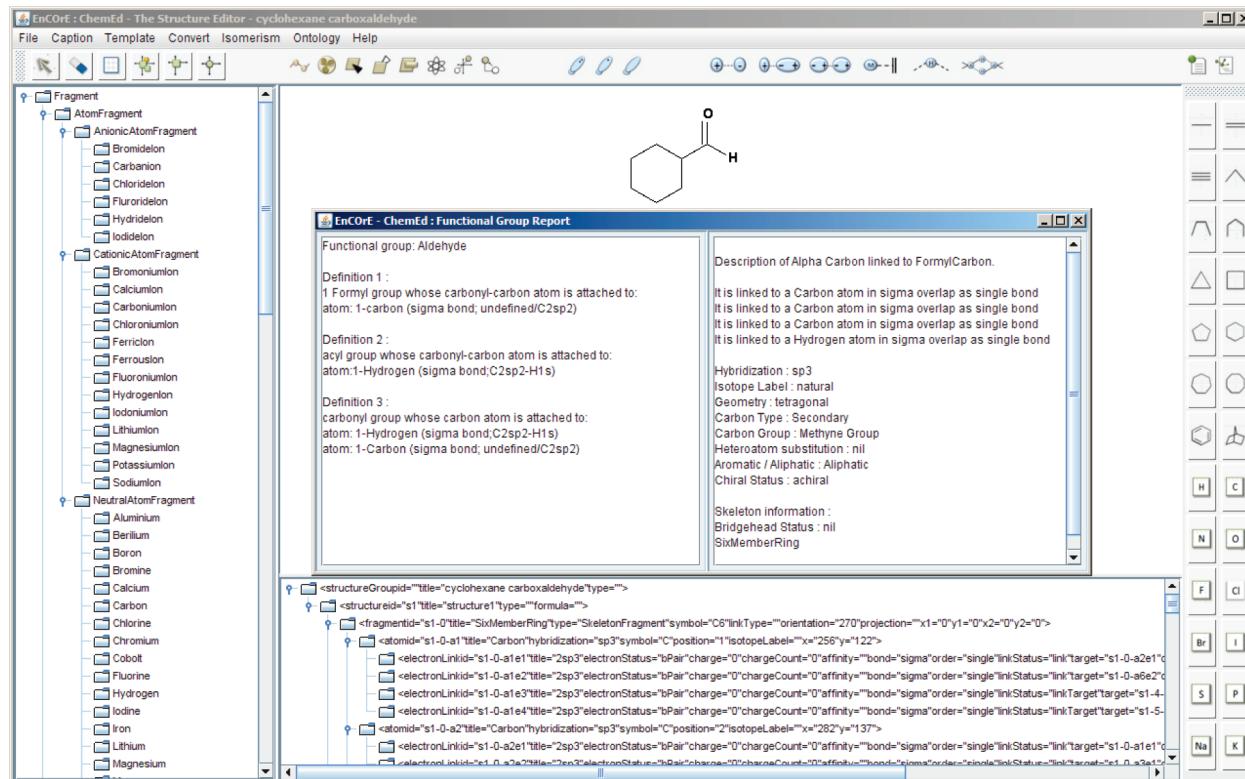


Figure 31. The functional group details captured by ChemEd for Cyclohexane carboxaldehyde 7 structure.

the assignment of configuration of the chiral centers. In the present case, the sequence for both the chiral centers follows a clockwise order in the orientation clock, and so the configurations for both top (0;270;90) and bottom (180;90;270) chiral carbons are **R**.

The priority sequence can also be used to decide the geometrical isomers in terms of *E/Z* notation. For this purpose, the groups around the sp<sup>2</sup> carbon linked to another sp<sup>2</sup> carbon with a double-bond link are sequenced and compared. The presence of higher priority groups on two carbons in the same or opposite orientations can be checked to detect the *E/Z* notation. An example output of ChemEd for *E/Z* notation is presented in Figure 29.

**Generation of Atom Descriptors from ChemFul and External Ontology.** The textual data available in ChemFul on the six levels can be processed independently or jointly, depending on the chemical applications. The markup provides the information regarding the locations of lone pair orbitals, multiple bonds, free radicals, and charges straightforwardly. The structural details captured in ChemFul can be processed to generate the atom descriptors, such as hybridization, bonding with other atoms, geometry, and

location on a skeleton, as shown in the Figure 30. The properties related to a six-member ring are fetched from the ontology.

**Detection of Functional Group Related Information from ChemFul and External Ontology.** Another significant feature of ChemEd is the detection of characteristic organic groups present in the structure. The ChemEd also is capable of providing the locations of the functional groups along with the structural environments of each functional group. This facilitates the interoperability with external ontology to bring information relating to various functional groups, structural skeletons, and characteristic chemical groups described in external ontologies. We have demonstrated the design and development of chemical ontologies for the reaction representation<sup>16</sup> and subsequently the support of chemical ontologies for a fragment-based structure markup system<sup>17</sup> in our earlier work. The structural details in terms of functional groups generated for cyclohexane carboxaldehyde structure 7 by ChemEd are shown in Figure 31. The left part of the report involves the generic description of the aldehyde functional group (C—C(=O)H), whereas the right part describes the specific features of the aldehyde functional

group and especially the description of the carbon  $\alpha$  to the carbonyl group in the context of cyclohexane carboxaldehyde 7.

Note that the generic description of the aldehyde functional group has been achieved to allow the description for example of: (i) Bayer Villiger reaction leading to cyclohexylformate (definition 1) or cyclohexane carboxylic acid (definition 2) and (ii) Grignard reaction of methyl magnesium bromide after hydrolysis to 1-cyclohexylethanol (definition 3).

## CONCLUSION

We have demonstrated the functionalities of ChemEd. We have shown that ChemEd is able to provide meaningful reports for a chemist through machine readable and human understandable structure descriptions in XML for chemical structures drawn on screen. We have included some original tools, which allow a chemist to encode his personal perceptions, like supramolecular interactions. So the functionalities to encode long-range steric or electronic interactions can be added easily. The generation of reports meaningful for chemists will allow the development of interactive applications allowing chemists to encode and process their perceptions about the structure in a collaborative manner. We hope to make clear to the reader that our objective is not to add one more chemical editor to existing ones but to show the possibility for new developments. Our ongoing work will be pursued using a selected community of chemists and communication tools we have used or built in the context of the E-LeGI project<sup>1d</sup> supported by the European Union. Our results will be reported in due course. The semantic markup is suitable to develop algorithms for functional group interchanges and modifications etc. required for a meaningful reaction description. The semantics at various levels will allow the possibility to create JAVA objects associated with suitable properties representing virtual substances and materials to simulate reactions virtually. The markup code is suitable for the conversion into the existing formats, such as the Connection Table, Molfile, CML, SMILES, InChI, etc. with suitable algorithms. ChemEd can be suitably integrated with other applications to build open and shared applications working with a common and approved vocabulary, making use of chemical ontologies.

## ACKNOWLEDGMENT

The authors acknowledge the support of the Belgian Science Foundation (FNRS), especially for the creation of the Fragment Library (A. Krief) and the Facultés Universitaires Notre-Dame de la Paix (Namur, Belgium). The authors acknowledge Prof. Mark A. Musen and Dr. Tania Tudorache of the Biomedical Informatic Research Center (BMIR), Stanford University, for their support in learning Protégé and developing OWL ontologies in Protégé. The authors acknowledge the Department of Science & Technology, India for funding under SERC Scheme (File No. SR/S1/OC-86/2009).

**Supporting Information Available:** The description of the functionalities and the structure drawing methodology of ChemEd is detailed in a word document. The document also contains the screen captures depicting: (i) one of the SMILES string equivalents of the cyclohexane carboxalde-

hyde structure generated by ChemEd; (ii) isotope labeling with the support of external ontology; (iii) atom position numbering in a SixMemberRing; (iv) projection labeling; (v) “Add/Remove Electron” tool to create a carbanion; (vi) “Change Atom Coordinates” to convert a “SixMemberRing” into a “CyclohexaneChairRing” template; (vii) use of “Caption” menu to incorporate captions to structures; (viii) the template viewer panel of ChemEd are provided. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## REFERENCES AND NOTES

- (a) Fleming, I.; Krief, A. unpublished discussion, August 28 1999 as the result of several collaborative sessions supported by Sanofi-Synthelabo company and held in their headquarters, Gentilly, France, (Prof. Castro, B. Sanofi-Synthelabo, France; Fleming, I. Cambridge University, United Kingdom; Gladysz, J. Erlangen University, Germany; Laurencio, C. University of Montpellier II, France; Krief, A. FUNDP, Namur, Belgium; and Laval, A.-M. FUNDP, Namur, Belgium. (b) <http://www.fundp.ac.be/facultes/sciences/departements/chimie/recherche/centres/cos/encore/presentation.html>. Accessed April 6, 2010. (c) This work has been supported in part by the Belgian Science Foundation (FNRS) and the European Union as the integrated project E-LeGI<sup>1d</sup>, (d) <http://www.elegi.org/>. See also <http://www.ist-world.org/ProjectDetails.aspx?ProjectId=12c84b8657ea4a089788938f936f3b96>. Accessed April 6, 2010.
- (a) Cooke, H. A historical study of structures for communication of organic chemistry information prior to 1950. *Org. Biomol. Chem.* **2004**, *2*, 3179–3191. (b) Brecher, J. Graphical representation standards for chemical structure diagrams (IUPAC Recommendations 2008). *Pure Appl. Chem.* **2008**, *80*, 277–410, doi:10.1351/pac200880020277.
- (a) Corey, E. J. General methods for the construction of complex molecules. *Pure Appl. Chem.* **1967**, *14*, 19. (b) Corey, E. J. Centenary lecture. Computer-assisted analysis of complex synthetic problems. *Q. Rev., Chem. Soc.* **1971**, *25*, 455.
- Rubenstein, S. D. In *Electronic documents in chemistry, from ChemDraw 1.0 to present*; 228th American Chemical Society National Meeting, Philadelphia, PA, August 22–26/2004; American Chemical Society: Washington, DC; AN 2004:656510 CAPLUS.
- (a) *chemDraw*, Desktop software; CambridgeSoft Life Science Enterprise Solutions: Cambridge, MA; <http://www.cambridgesoft.com/software/ChemDraw/>. Accessed April 6, 2010. (b) *Symyx*, Transition to Symyx Draw; Symyx Solutions, Inc.: Santa Clara, CA; <http://www.symyx.com/micro/draw31/index.html>. Accessed April 6, 2010; and *mdl isis draw* 2.5; Symyx Solutions, Inc.: Santa Clara, CA; <http://software.informer.com/getfree>. Accessed April 6, 2010; (c) *ACD/Labs Chemical Drawing and Nomenclature*; Advanced Chemistry Development: Toronto, Ontario, Canada; [http://www.acdlabs.com/products/draw\\_nom/](http://www.acdlabs.com/products/draw_nom/). Accessed April 6, 2010; (d) *Marvin*; ChemAxon Kft.: Budapest, 1037 Hungary; <http://www.chemaxon.com/marvin/index.html>. Accessed April 6, 2010; (e) *JChemPaint*; Geeknet: Mountain View, CA; <http://sourceforge.net/apps/mediawiki/cdk/index.php?title=JChemPaint>. Accessed April 6, 2010.
- Dalby, A.; Nourse, J. G.; Douglas Hounshell, W.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J. Description of Several Chemical Structure File Formats Used by Computer Programs Developed at Molecular Design Limited. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 244–255.
- CTFile Formats*; Symyx Solutions, Inc.: Santa Clara, CA; <http://www.mdl.com/downloads/public/ctfile/ctfile.pdf>. Accessed April 6, 2010.
- Weininger, D. SMILES a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28* (1), 31–36. (b) Weininger, D.; Weininger, A.; Weininger, J. Algorithm for generation of unique SMILES notation. *J. Chem. Inf. Comput. Sci.* **1989**, *29* (2), 97–101. (c) Weininger, D. Graphical depiction of chemical structures. *J. Chem. Inf. Comput. Sci.* **1990**, *30* (3), 237–243. (d) *DAYLIGHT SMILES*, A Simplified Chemical Language; Daylight Chemical Information Systems, Inc.: Aliso Viejo, CA; <http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html>. Accessed April 6, 2010.
- W3C Extensible Markup Language (XML); W3C/ERCIM: Sophia-Antipolis, France; <http://www.w3.org/XML>. Accessed April 6, 2010.
- (a) Murray-Rust, P.; Rzepa, H. S. Chemical Markup Language and XML Part I. Basic Principles. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 928–942.
- (a) *Protégé*; Stanford Center for Biomedical Informatics Research: Stanford, CA; <http://protege.stanford.edu/>. Accessed April 6, 2010;

- (b) Protégé, *WikiHomePage*; Stanford Center for Biomedical Informatics Research; Stanford, CA; <http://protege.cim3.net/cgi-bin/wiki.pl/>. Accessed April 6, 2010. (c) W3C, OWL Working Group; W3C/ERCIM: Sophia-Antipolis, France; [http://www.w3.org/2007/OWL/wiki/OWL\\_Working\\_Group](http://www.w3.org/2007/OWL/wiki/OWL_Working_Group). Accessed April 6, 2010.
- (12) JAVA SE 6; Oracle Corporation: Redwood Shores, CA; <http://java.sun.com/javase/6/>. Accessed April 6, 2010.
- (13) (a) Bauerschmidt, S.; Gasteiger, J. Overcoming the Limitations of a Connection Table Description: A Universal Representation of Chemical Species. *J. Chem. Inf. Model.* **1997**, *37*, 705–714. (b) Dietz, A. Yet Another Representation of Molecular Structure. *J. Chem. Inf. Model.* **1995**, *35*, 787–802. (c) Clayden, J.; Greeves, N.; Warren, S.; Wothers, P. Organometallic chemistry. In *Organic Chemistry*. Oxford University Press: New York, 2006; pp 1311–1343.
- (14) (a) Cahn, R. S.; Ingold, C. K.; Prelog, V. The Specification of Asymmetric Configuration in Organic Chemistry. *Experientia* **1956**, *12*, 81–124. (b) Cahn, R. S.; Ingold, C. K.; Prelog, V. Specification of Molecular Chirality. *Angew. Chem.* **1966**, *78*, 413. (c) IUPAC Rules for the Nomenclature of Organic Chemistry, Section E: Stereochemistry (Recommendations); International Union of Pure and Applied Chemistry: Research Triangle Park, NC; 1974.
- (15) Fischer, E. Ueber die Configuration des Traubenzuckers und seiner Isomeren. II. *Ber. Dtsch. Chem. Ges.* **1891**, *24*, 2683–2687.
- (16) Sankar, P.; Aghila, G. Design and Development of Chemical Ontologies for Reaction Representation. *J. Chem. Inf. Model.* **2006**, *46*, 2355–2368.
- (17) Sankar, P.; Aghila, G. Ontology Aided Modeling of Organic Reaction Mechanisms with Flexible and Fragment Based XML Markup Procedures. *J. Chem. Inf. Model.* **2007**, *47*, 1747–1762.

CI100052B