

Ontology Aided Modeling of Organic Reaction Mechanisms with Flexible and Fragment Based XML Markup Procedures

Punnaivanam Sankar^{*,†} and Gnanasekaran Aghila[‡]

Department of Chemistry and Computer Science & Engineering and Information Technology,
Pondicherry Engineering College, Puducherry - 605 014, India

Received January 31, 2007

The mechanism models for primary organic reactions encoding the structural fragments undergoing substitution, addition, elimination, and rearrangements are developed. In the proposed models, each and every structural component of mechanistic pathways is represented with flexible and fragment based markup technique in XML syntax. A significant feature of the system is the encoding of the electron movements along with the other components like charges, partial charges, half bonded species, lone pair electrons, free radicals, reaction arrows, etc. needed for a complete representation of reaction mechanism. The rendering of reaction schemes described with the proposed methodology is achieved with a concise XML extension language interoperating with the structure markup. The reaction scheme is visualized as 2D graphics in a browser by converting them into SVG documents enabling the desired layouts normally perceived by the chemists conventionally. An automatic representation of the complex patterns of the reaction mechanism is achieved by reusing the knowledge in chemical ontologies and developing artificial intelligence components in terms of axioms.

INTRODUCTION

Generally, a chemical reaction is described with the graphical representation of individual reactant and product molecules along with the other conditions like solvent, catalyst, temperature, and pressure. The representation of chemical reactions in graphical format involves the depiction of the reactants and the products as 2D graphics using structure editors. The graphical tools to represent reactions such as ChemDraw,¹ ISIS/Draw,² and chemSketch³ are using graphical formats in which reactions can be shown by drawing the molecular structures along with the components like arrows and plus symbols. The relative positioning of the molecular structures according to their role in a chemical reaction is the possible means of representing a chemical reaction in such systems. Some other systems like Reaction SMILES,⁴ SMIRKS,⁵ and REACTOR⁶ use the molecular notations as the basis to represent a chemical reaction. In SMILES and SMIRKS, a reaction is described by representing the molecular species involved in a chemical reaction as reactant, agent, and product in the form of molecular notations separated by ">" symbols, whereas the 'REACTOR' is a virtual reaction processing tool in which the starting compounds are transformed to products according to a set of rules. The input and output for this system includes the molecular representations based on SMILES like mrv,⁶ the connection table based formats^{7,8} such as molfile⁸ and SDfile,⁸ and a markup based format, cml,⁹ developed from Chemical Markup Language¹⁰ (CML). The widely used formats to store the reaction information are the rxnfiles⁸ and RDfiles.⁸ Each rxnfile is composed of molefiles describ-

ing the structural information for the reactants and products of a single reaction. The RDfile⁸ (reaction-data file) is a format to include many reactions as well as molecules together with their associated data. The use of XML¹¹ technology to describe chemical contents resulted in the inclusion of one more format based on the connection table termed as XDfile.⁸ This is an XML-based data format to hold structures or reactions that use any of the connection table based formats. A Java based tool, JChemPaint,^{12,13} is an open-source Java program for drawing 2D chemical structures based on the Chemistry Development Kit.¹⁴ It provides the facility to render and edit chemical reactions with the input files such as molfile, rxnfile, and cmlfile. Since all of these formats to store chemical reaction information are ultimately dependent on a connection table, they are not capable of providing semantically rich reaction representations.

Incorporating the XML technology in storing structures such as XML documents drawn on the screen resulted in the format CDXML.¹⁵ This is an XML encoding of CDX,¹⁶ the native file format of ChemDraw which stores a document as a set of nested objects and properties. The objects include atoms, bonds, fragments, arrows, and text. Properties are the position, color, arrow type, and bond order. The system treats the objects drawn on the screen as molecular fragments composed of atom, bond, text, and other fragments. The attachment point is included for the atoms with nodes. Accordingly, ChemDraw stores the structure drawn on the screen into a series of objects and their associated properties including the automatically generated coordinates in the form of an XML document. However, the pioneering work to describe a chemical reaction in terms of molecular markup is the CMLReact¹⁷ developed as an extension of CML sharing its components with CML components to manage chemical reactions by including the reaction components in

* Corresponding author e-mail: gapspec@yahoo.com.

[†] Department of Chemistry.

[‡] Department of Computer Science & Engineering and Information Technology.

the current CML functionality. It handles the reaction markup through the inclusion of the description of molecular species such as reactant, product, etc., placed in a container element <reaction>. The reaction is represented by describing molecular species behaving as reactants and products and attaching directly within the reaction markup using appropriate tags like <reactant> and <product>, respectively. In an alternate approach, the reaction markup is provided with the links to the corresponding molecular markup available separately. This approach results in a succinct markup eliminating the inclusion of an entire description of molecules inside the reaction markup.

The information components within the reaction in CMLReact are annotated by referring to specific bonds and atoms through their IDs and using mechanism components to describe their role, properties, etc. In a second approach, the mechanism component is used to identify types of bond formed/broken without reference to actual atoms and bonds. Mostly in CMLReact the reaction is represented through some editing tools. Accordingly the coordinates of the structures are fixed through the editors. So the structures can be presented on the screen at the same location from where they were drawn as this presentational information is already fixed into the content markup. Indeed the description of mechanism components in CMLReact is demonstrated early in CMLSnap,¹⁸ developed as an extension of CML to create animated reactions. In this work, the representation of multistep reactions is described by considering each step as a row in a table providing the information about the components involved in the step. CML is used to hold any number of such components including the 2D coordinates. The mechanism is described by creating snapshots using editing tools and edited by changing the bonds, atom positions, and charges. Each and every step contains all the atoms in the system whether or not they are involved in the current step. Finally the mechanism of the reaction is shown as an animated diagram of reaction by transforming the XML into Scalable Vector Graphics¹⁹ (SVG).

The reaction representation in terms of a simple molecular markup is not sufficient for encoding mechanism schemes, because the reaction mechanism formalization needs the semantics to be associated with the molecular fragments undergoing chemical transformations in each and every step of the reaction. So the reaction markup becomes complete only when the entire mechanistic pathways are incorporated into the reaction markup. The mechanism of a chemical reaction is the actual process by which the reactants are converted into products. Accordingly the description of the reaction mechanism requires certain critical types of information about the course of the reaction such as the nature and place of bond cleavage, number of steps involved, relative rates of each step, main product, side product, and so forth. In addition to this the mechanism depiction should include the changes in the electronic structures of molecules as they pass from starting materials into the products. So the formalization of a reaction mechanism is a difficult task, because it requires the encoding of chemical species along with the other features like transition states, intermediate species indicating the half bonded chemical moieties, charges (+, -, δ +, δ -), lone pair electrons, free radicals, reaction arrows (forward, backward, and two way arrows, etc.), and reaction conditions (temperature, pressure, heat, light, catalyst, etc.). Most

importantly the electron movements which are normally shown as a curved double headed arrow to indicate the movement of an electron pair and a single barbed fish hook headed arrow for the single electron movement should also be described along with the mechanism components. The reaction representation becomes meaningful and as complete as possible when these complex features of the mechanism are included in the markup and no such approach is available.

Another aspect of formalizing the reaction mechanism is making the representation automatic. This is needed, because even for a small reaction scheme involving two or three steps the flawless markup coding is difficult in manual attempts. In this situation, a system associated with a knowledge backup is a better choice to make the mechanism representations automatic. This can be achieved by developing systems supported by chemical ontologies comprising the needed knowledge corresponding to the chemical reactions. This knowledge based representation system makes the automatic reaction representation easier. The next aspect associated with the reaction mechanism modeling is the rendering of a mechanism scheme as a graphical representation preferably on a browser. The transformation of the mechanism markup into a graphical display is equally important as that of the markup. A better rendering system for reaction mechanism schemes should be capable of displaying all the complex features of the mechanistic pathways in a conventional format viewed by the chemists. Further, the display system must be developed consistently with the global standards. In the present trend of transforming the WWW into a new wave of the semantic Web,^{20–24} the storage, retrieval, and communication of chemical contents are heading toward new technology completely based on markup procedures. Thus a communication oriented representation methodology based on a markup procedure is gaining importance in both representing and rendering chemical information.^{25–29} The progress of WWW based on XML syntax resulted in the evolution of SVG to describe the two-dimensional graphics in XML to be communicated via the Internet with the advantages of lesser storage and shorter download time. The SVG is a W3C endorsed XML specification capable of capturing rich graphical content with a group of XML elements. It allows three types of graphic objects, namely vector graphical shapes (e.g., paths consisting of straight lines and curves), images, and text. The features of SVG allow the drawings to be interactive and dynamic. The Document Object Model (DOM) of SVG favors the development of sophisticated applications because the DOM provides complete access to all elements, attributes, and properties posing it as a suitable media for the transaction of reaction mechanism schemes in XML syntax.

In the present work mechanism models for the common organic reactions are proposed with a flexible markup system capturing each and every component of the mechanistic pathways. Unlike the CMLReact, the developed markup system allows the mechanism components to be described in a fragment based construct enabling the incorporation of needed semantics associated with the structural fragments in which the presentational coordinate system is separated from the markup. Accordingly the formalization of molecular fragments which are substituted or added or eliminated or undergoing a rearrangement becomes easier. So the molecular fragments can be meaningfully processed for desirable

```

<?xml version="1.0"?>
<!DOCTYPE reactionScheme SYSTEM "rxnDTD.dtd">
<!--Reaction mechanism modeling-->
<reactionScheme xmlns="http://www.chemonto-india.info/rxnMkp" id="rs" type="" title="">
  <reaction id="" type="" title="">
    <reactionStep id="">
      <reactantList id="">
        <reactant id="" type="" title="" formula="" />
        <reactant id="" type="" title="" formula="" />
        <relation id="" title="" target="" />
      </reactantList>
      <productList id="">
        <relation id="" title="" target="" />
        <product id="" type="" title="" formula="" />
        <product id="" type="" title="" formula="" />
        <relation id="" title="" target="" />
      </productList>
      <co-participantList>
        <substance id="" type="catalyst" title="" formula="" />
        <substance id="" type="" title="" formula="" />
      </co-participantList>
      <conditionList>
        <condition id="" type="" title="">normal</condition>
        <condition id="" type="" title="">normal</condition>
      </conditionList>
    </reactionStep>
    <mechanism id="" type="" title="">
      <mechanismStep id="" type="" title="">
        <reactantComponentList id="" type="" title="">
          <mechanismComponent id="" type="" title="">
            <structureList id="">
              <baseStructure id="" title="" type="" class="" orientation="" formula="" />
              <jointStructure id="" title="" type="" class="" orientation="" projection="" target=""
                position="" formula="" />
              ...
              <eMoveStructure id="" title="" type="" orientation="" projection="" source=""
                sourcePosition="" target="" targetPosition="" />
            </structureList>
          </mechanismComponent>
          ...
        </reactantComponentList>
        <productComponentList id="" type="" title="">
          ...
        </productComponentList>
      </mechanismStep>
      <mechanismStep id="" type="" title="">
        ...
      </mechanismStep>
    </mechanism>
  </reaction>
</reactionScheme>
    
```

Figure 1. The skeleton structure of a reaction mechanism markup.

layout, searching, mining, etc. Also the proposed system is capable of describing both the chemical as well as the nonchemical features of the mechanism components such as transition states, reaction intermediates, charged species, partial charges, electron movement, reaction signs, arrows, square brackets, etc. The proposed system is developed into a knowledge based representation system capable of representing the mechanism schemes in XML automatically with the support of chemical knowledge represented in the form of ontologies and AI components. The representation in XML followed by their conversion into SVG is a convenient and global technique to store, retrieve, and communicate reaction

schemes over the networks. The combination of XML and SVG is an open standard technique to exchange reaction schemes. This enables the applications developed based on this technology as globally accessible and acceptable with a promising scope for the development of new generation tools assisting the evolving semantic Web and computer aided synthesis design techniques.

GENERAL STRUCTURE OF REACTION MECHANISM MARKUP

The structure of a reaction document containing the mechanism components is designed through careful analysis

```

<mechanismComponent id="McStep1/rctLst/cpt2" type="" title="">
  <relation id="McStep1/rctLst/cpt2/rln" title="combinesWith" target="McStep1/rctLst/cpt1" />
  <structureList id="McStep1/rctLst/cpt2/sl">
    <baseStructure id="McStep1/rctLst/cpt2/Frg-0" title="CAtom" type="base" class="alkyl"
      orientation="0" formula="C" />
    <jointStructure id="McStep1/rctLst/cpt2/Frg-1" title="methyl" type="jAtomGroup"
      class="methyl" orientation="90" projection="above"
      target="McStep1/rctLst/cpt2/Frg-0" position="1" formula="CH3" />
    <jointStructure id="McStep1/rctLst/cpt2/Frg-2" title="HAtom" type="jAtomGroup"
      class="HAtom" orientation="270" projection="above"
      target="McStep1/rctLst/cpt2/Frg-0" position="2" formula="H" />
    <jointStructure id="McStep1/rctLst/cpt2/Frg-3" title="HAtom" type="jAtomGroup"
      class="HAtom" orientation="180" projection="below"
      target="McStep1/rctLst/cpt2/Frg-0" position="3" formula="H" />
    <jointStructure id="McStep1/rctLst/cpt2/Frg-4" title="BrAtom" type="jAtomGroup"
      class="BrAtom" orientation="0" projection="in"
      target="McStep1/rctLst/cpt2/Frg-0" position="4" formula="Br" />
  </structureList>
</mechanismComponent>

```

Figure 2. The XML code describing an alkyl halide molecule involved in an S_N2 mechanism.

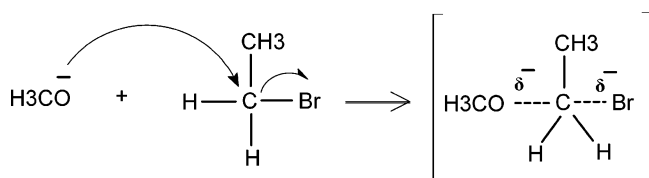


Figure 3. The graphical representation for the formation of a transition state in an S_N2 mechanism.

of the vocabulary available for reaction markup in CML-React¹⁷ and by studying the reaction mechanisms of primary reaction types^{30–32} with respect to the proposed mechanism markup methodology. The reaction mechanisms involved in the primary reaction classes such as the substitution, addition, elimination, and rearrangement reactions are considered for the proposed markup procedure. The study covers the mechanisms for the nucleophilic, electrophilic, and free radical reactions of the above reaction classes. Based on the vocabulary needed to formalize the reaction mechanism, a generic structure of the reaction document containing the mechanism components is developed.

The reaction document starts with the root element <reactionScheme>. The individual reaction information is placed within this element using one or more <reaction> elements. The content of a <reaction> element can be viewed as two parts. The first part provides the basic information about the reaction such as the reaction steps, starting materials, products, and the reaction conditions along with the possible semantics. The second part deals with the description of the mechanism of the reaction. This compartment of the document covers the entire mechanistic pathways describing the reaction in a comprehensive format. The components concerned to the mechanism are placed in a container element <mechanism>. Each mechanism element consists of one or more <mechanismStep> elements. Which in turn is composed with several <mechanismComponent> elements placed inside <reactantComponentList> and <productComponentList> appropriately along with the <relation> elements if needed. The structural features of the mechanism components are described with a unique markup system and enclosed inside an XML element named as

<structureList> placed inside the <mechanismComponent> element. A mechanism component is considered as a portion of the mechanism step which may be a molecule or an ionic species or transition state or an intermediate chemical species. These components are captured with a markup methodology flexible enough to describe the reaction mechanisms as complete as possible. The skeleton markup structure of the reaction mechanism markup is shown in Figure 1.

According to this markup technique each and every component of a mechanism step is described with a list of substructures enclosed in a <structureList> element. The chemical and nonchemical structural features of the mechanism components are described within this element along with appropriate semantics in a fragment based approach. This approach allows the description of any component of a mechanism step with a combination of two types of substructures namely a base structure and a joint structure along with a third structure to describe the electron movements. The base structure is the starting fragment of any component over which the entire structure is to be built. The joint structures are any structural fragment to be attached to the base structure or with other joint structures. In order to encode this information, a set of appropriate XML elements namely <baseStructure>, <jointStructure>, and <eMoveStructure> is defined inside the <structureList> along with the needed semantics in the form of attributes. This skeleton markup frame is found to be suitable to describe any structural feature which is chemical, nonchemical, and a combination of both into a semantic markup describing the reaction mechanism.

Representation of Mechanism Components. Using the markup skeleton developed for the study, the reaction mechanisms of substitution, addition, elimination, and rearrangement reactions are modeled by the system. Subsequently the mechanism schemes represented in XML markup are rendered in a browser as two-dimensional graphics showing all the components of the mechanism as perceived by the chemists conventionally. The code generated for the reaction of an alkyl halide proceeding through a bimolecular mechanism (S_N2) is presented in Figure 2, and the graphics represented by the code are shown in Figure 3.

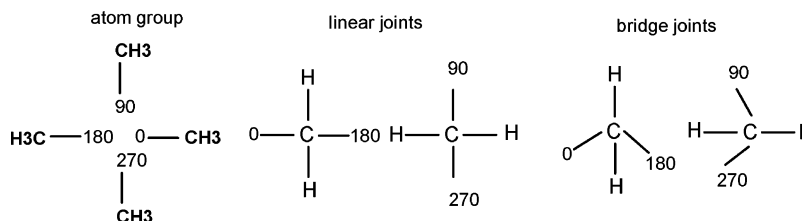


Figure 4. Joint structural fragments with different orientations and arrangements.

```

<mechanismComponent id="McStep1/pdtLst/cpt1" type="" title="">
  <relation id="McStep1/pdtLst/cpt1/RIn" title="irreversibleWith" target="" />
  <structureList id="McStep1/pdtLst/cpt1/sl">
    <baseStructure id="McStep1/pdtLst/cpt1/Frg-0" title="transCAtom" type="base" class="alkyl"
      orientation="0" formula="C" />
    <jointStructure id="McStep1/pdtLst/cpt1/Frg-1" title="methyl" type="jAtomGroup"
      class="methyl" orientation="90" projection="in"
      target="McStep1/pdtLst/cpt1/Frg-0" position="1" formula="CH3" />
    <jointStructure id="McStep1/pdtLst/cpt1/Frg-2" title="HAtom" type="jAtomGroup"
      class="HAtom" orientation="225" projection="below"
      target="McStep1/pdtLst/cpt1/Frg-0" position="2" formula="H" />
    <jointStructure id="McStep1/pdtLst/cpt1/Frg-3" title="HAtom" type="jAtomGroup"
      class="HAtom" orientation="315" projection="above"
      target="McStep1/pdtLst/cpt1/Frg-0" position="3" formula="H" />
    <jointStructure id="McStep1/pdtLst/cpt1/Frg-4" title="transBrAtom" type="jAtomGroup"
      class="BrAtom" orientation="0" projection="in"
      target="McStep1/pdtLst/cpt1/Frg-0" position="4" formula="Br" />
    <jointStructure id="McStep1/pdtLst/cpt1/Frg-5" title="transmethoxide" type="jAtomGroup"
      class="methoxide" orientation="180" projection="in"
      target="McStep1/pdtLst/cpt1/Frg-0" position="5" formula="OCH3" />
    <jointStructure id="McStep1/pdtLst/cpt1/Frg-6" title="partialMinus" type="misc"
      class="partialMinus" orientation="" projection=""
      target="McStep1/pdtLst/cpt1/Frg-0" position="6" formula="" />
    <jointStructure id="McStep1/pdtLst/cpt1/Frg-7" title="partialMinus" type="misc"
      class="partialMinus" orientation="" projection=""
      target="McStep1/pdtLst/cpt1/Frg-0" position="7" formula="" />
    <jointStructure id="McStep1/pdtLst/cpt1/Frg-8" title="Rxnsqbkt" type="misc" class=""
      orientation="0" projection="" target="McStep1/pdtLst/cpt1/Frg-0" position="8"
      formula="" />
    <jointStructure id="McStep1/pdtLst/cpt1/Frg-9" title="Rxnsqbkt" type="misc" class=""
      orientation="180" projection="" target="McStep1/pdtLst/cpt1/Frg-0" position="9"
      formula="" />
  </structureList>
</mechanismComponent>
    
```

Figure 5. The XML code describing the transition state of an S_N2 mechanism.

The structure of ethylbromide molecule is described with a base structure and four joint structures as seen from Figure 2. A base structure represents the starting fragment of any chemical structure, and it is default. The semantics of the base structure is provided with the attributes namely 'id', 'title', 'type', 'class', 'orientation', and 'formula'. The 'id' attribute is used to hold the unique ID of the fragment. The fragment's name and type is provided in the 'title' and the 'type' attribute, respectively. The fourth attribute 'class' is used to include the functional class of the fragment if any. For example the fragment "CAtom" is defined as the "alkyl" class, and a "benzene" fragment belongs to the "aryl" class. The 'orientation' attribute provides the direction of the fragment using the numerical values of 0–360 on an anticlockwise basis. Accordingly the orientation values of '0' and '180' for the 'methyl' fragment represents the group as "CH3" and "H3C", respectively, in the case of base structures. The final attribute 'formula' holds the formula

of the structural fragment to be used to determine the molecular formula and weight of a structure constructed with this markup technique.

The <jointStructure> element contains the information for the joint fragments. The attributes 'id', 'title', 'class', and 'formula' are used to provide the same semantics as described for the base structure. The 'orientation' attribute for any joint structure specifies the direction along which the fragment is attached to another structural fragment. Apart from these attributes, the joint structure consists of additional attributes such as 'projection', 'target', and 'position' to specify the location and plane of the attachment. The value of "in" for the joint bromine atom in Figure 2 indicates that the bond connecting carbon and bromine atoms is the reference plane. The methyl group and one of the hydrogen atoms are represented as projecting above this reference plane, and the other hydrogen atom is marked as below the plane projection. This is useful to some extent in describing

```

<mechanismComponent id="McStep1/rctLst/cpt1" type="" title="">
  <structureList id="McStep1/rctLst/cpt1/strLst">
    <baseStructure id="McStep1/rctLst/cpt1/Frg-0" title="methoxide" type="base"
      class="methoxide" orientation="180" formula="OCH3" />
    <jointStructure id="McStep1/rctLst/cpt1/Frg-1" title="minus" type="misc" class="minus"
      orientation="" projection="" target="McStep1/rctLst/cpt1/Frg-0" position="1"
      formula="" />
    <eMoveStructure id="McStep1/rctLst/cpt1/Arc-1" title="IPair" type="clockwise"
      orientation="" projection="" source="McStep1/rctLst/cpt1/Frg-0"
      sourcePosition="1" target="McStep1/rctLst/cpt2/Frg-0" targetPosition="3" />
  </structureList>
</mechanismComponent>

```

Figure 6. The XML code generated for the description of ionic species and a lone pair electron movement.

the groups undergoing inversion and the groups occupying positions during the transition states but not the exact stereo positions. The remaining attributes namely the ‘target’ and ‘position’ play a significant role in the construction of the structure. The ‘target’ attribute specifies the target fragment for the joint structure. This is done by marking the ID value of the target fragment as the value for the target attribute in the joint fragment. Subsequently, the exact location within the target fragment to which the joint fragment is to be attached is indicated by the ‘position’ attribute with numerical values. For this provision, the structural fragments are defined with imaginary points of attachments which may be considered as electronic positions for the bond formation. For example, a ‘CAtom’ fragment is defined with four points of attachments. According to the code, all four joint fragments are represented as located on these four positions as ‘2’ and ‘3’ for the hydrogen atoms and ‘1’ and ‘4’ for the methyl and bromine fragments, respectively. The choice of a different arrangement of the same structural fragments is handled with the ‘type’ attribute. A methyl fragment can be shown as a group of atoms like “CH3” or as atoms connected to bonds. A methylene fragment can be of linear or bridge type. Similarly there are fuse type structures and miscellaneous structures like electron pairs, charges, arrows, square brackets, etc. to be used for the construction of chemical structures. The semantics of this information is incorporated as the value “linear/bridge/fuse/jAtomGroup/misc” for the ‘type’ attribute of joint structures in the structure markup. With making use of these attributes it is possible to describe the structural fragments with different alignments with respect to the other structural fragments resulting in a flexible markup procedure to encode the complex features of the reaction mechanisms. In Figure 4 some joint structures with different orientations and arrangements defined for the present study are provided.

The description of transition state species requires the encoding of half bonded structures, partial charges, square brackets, etc. The code fragment for the transition state structure of the transition state of S_N2 mechanism is shown in Figure 5.

The XML code for the transition state depicts the groups occupying the transitional positions. The incoming nucleophilic unit and the existing bromine atom are indicated with ‘in plane’ projections as half bonded species with the base carbon conveying the transition state condition. The methyl fragment is now occupying the ‘in plane’ projection as a result of the rear side attack of the nucleophile. The values of orientation attribute are changed to indicate appropriate directions of the groups. The partial charges and the square

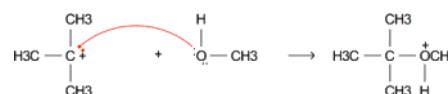


Figure 7. The graphical display obtained for the intermediate species involved in an S_N1 mechanism.

brackets are described as additional joint structures in the code. For the description of normal charges to depict the ionic species, the joint structure construct is used with similar markup procedures. The code showing the representation of ionic species involved in the aliphatic nucleophilic substitution of an alkyl halide is presented in Figure 6 along with the electron movement structure describing the movement of lone pair from the anion to the carbon atom of the alkyl halide.

A significant part of the system is the representation of the electron movements during the chemical reactions. This important feature of the markup describes the changes taking place in the electronic structures during the course of reaction. In the S_N2 mechanism, there are two important electron shifts to be shown. One is the movement of a lone pair from the nucleophilic unit to the carbon center of the alkyl halide. The other shift is a movement of an unpaired electron within the alkyl halide molecule leading to the formation of halide ion. This information is encoded with the element <eMove-structure> with appropriate attributes. The movement of a lone pair and unpaired electrons is indicated with the values of ‘IPair’ and ‘uPair’ for the title attribute, respectively. The clockwise or anticlockwise movement of the electrons is shown with the appropriate values in the ‘type’ attribute. The last four attributes are used to specify the information about the origin and the destination of the electronic arc. The ‘source’ and ‘sourcePosition’ attributes provide the origin point of the electron shift, and the destination is shown with the ‘target’ and ‘targetPosition’ attributes. The markup procedure enables the description of electron movements within the molecule as well as from one component to other making use of the attributes defined. The unimolecular mechanism for the nucleophilic substitution reaction of a tertiary alkyl halide with a weak nucleophile involves the formation of an intermediate chemical species, carbonium ion. Also the scheme requires the description of the lone pair electrons explicitly for the oxygen atom of the weak nucleophile to be transferred to the carbonium carbon center. The graphical display obtained for the description of intermediate structures and the lone pair electron movement from the weak nucleophile to the carbocation are shown in Figure 7, and the corresponding XML code for the structures is shown in Figure 8.

```

<mechanismComponent id="McStep2/rctLst/cpt1" type="" title="">
  <structureList id="McStep2/rctLst/cpt1/sl">
    <baseStructure id="McStep2/rctLst/cpt1/Frg-0" title="carboniumCAtom" type="base"
      class="alkyl" orientation="0" formula="C" />
    <jointStructure id="McStep2/rctLst/cpt1/Frg-1" title="methyl" type="jAtomGroup"
      class="methyl" orientation="90" projection="above"
      target="McStep2/rctLst/cpt1/Frg-0" position="1" formula="CH3" />
    <jointStructure id="McStep2/rctLst/cpt1/Frg-2" title="methyl" type="jAtomGroup"
      class="methyl" orientation="270" projection="above"
      target="McStep2/rctLst/cpt1/Frg-0" position="2" formula="CH3" />
    <jointStructure id="McStep2/rctLst/cpt1/Frg-3" title="methyl" type="jAtomGroup"
      class="methyl" orientation="180" projection="below"
      target="McStep2/rctLst/cpt1/Frg-0" position="3" formula="CH3" />
  </structureList>
</mechanismComponent>
<mechanismComponent id="McStep2/rctLst/cpt2" type="" title="">
  <structureList id="McStep2/rctLst/cpt2/strLst">
    <baseStructure id="McStep2/rctLst/cpt2/Frg-0" title="methoxide" type="base"
      class="methoxide" orientation="0" formula="OCH3" />
    <jointStructure id="McStep2/rctLst/cpt2/Frg-1" title="HAtom" type="jAtomGroup"
      class="HAtom" orientation="90" projection="" target="McStep2/rctLst/cpt2/Frg-0"
      position="1" formula="H" />
    <jointStructure id="McStep2/rctLst/cpt2/Frg-2" title="ePair" type="misc" class="misc"
      orientation="90" projection="" target="McStep2/rctLst/cpt2/Frg-0" position="3"
      formula="" />
    <jointStructure id="McStep2/rctLst/cpt2/Frg-3" title="ePair" type="misc" class="misc"
      orientation="0" projection="" target="McStep2/rctLst/cpt2/Frg-0" position="2"
      formula="" />
    <eMoveStructure id="McStep2/rctLst/cpt2/Arc-1" title="IPair" type="anticlockwise"
      orientation="" projection="" source="McStep2/rctLst/cpt1/Frg-0"
      sourcePosition="3" target="McStep2/rctLst/cpt1/Frg-0" targetPosition="4" />
  </structureList>
</mechanismComponent>
    
```

Figure 8. The XML code generated for the description of carbonium ion and lone pair electrons and their movement involved in an S_N1 mechanism.

The description of the aromatic system differs from that of the alkyl system because in aromatic systems the presence of a substituent either activates or deactivates the ring. As a result of this, electron-rich or -deficient situations at *ortho*-, *meta*-, and *para*-positions are formed depending on the nature of the substituent. The presence of an electron withdrawing or a donating substituent in the ring influences the reaction course. Also the mechanism of the aromatic reaction involves the movement of electrons within the ring. So the reaction involving aromatic rings needs the description of positions within the ring to attach semantics for a reaction representation. The encoding of an aromatic ring with two substituents and the electron movements during a nucleophilic substitution reaction is shown in Figure 9. The complete display obtained for the reaction is presented in Figure 10.

It is found that convincing results are obtained when the principles of the proposed markup procedure are applied to the mechanisms of the primary organic reactions involving aliphatic, aromatic, and acyclic compounds. The important reaction mechanisms considered for the representation with the system is provided in Table 1

KNOWLEDGE MANAGEMENT FOR AUTOMATIC REPRESENTATION OF REACTION MECHANISM

The Chemical Ontological Support System³³ (COSS) developed for integrating the chemical knowledge to support the automatic representation of chemical reactions includes the chemical ontologies namely the reaction, reagent, and

compound ontologies. This knowledge base is strengthened with the addition of artificial intelligence (AI) components in terms of axiom definitions and relations to draw valid inferences with regard to identification of functional groups, fixing the preferable mechanism, alkyl group transformations during chemical reactions, etc. So, the COSS in its present condition is capable of serving as a supporting system to enable the representation of reaction mechanisms for common organic reactions automatically. The architecture of the system is shown in Figure 11.

Materials and Methods. The basic input needed for the system is the fragment based markup of the reactant. The COSS then identifies automatically the various functional groups present in the input structure and lists the eligible reaction classes the structure can undergo as well as the appropriate reagents for selection. A decision on the reaction class and the reagent then triggers a series of events which includes the identification of probable mechanism path, possible product classes, and building the information about reactants, products, and group transformations taking place if any. Based on the data generated through COSS, the reaction document is then written in XML through predefined templates capturing the entire features of the reaction including the mechanism components. The reaction markup then linked to the fragments library containing the coordinates of the fragments for the display. The reaction scheme is then rendered on the screen using a panel or a picture box

```

<mechanismComponent id="McStep1/rctLst/cpt1" type="" title="">
  <structureList id="McStep1/rctLst/cpt1/sl">
    <baseStructure id="McStep1/rctLst/cpt1/Frg-0" title="benzene" type="base"
      class="aryl" orientation="270" formula="C6H6" />
    <jointStructure id="McStep1/rctLst/cpt1/Frg-1" title="ClAtom" type="jAtomGroup"
      class="ClAtom" orientation="90" projection="in"
      target="McStep1/rctLst/cpt1/Frg-0" position="1" formula="Cl" />
    <jointStructure id="McStep1/rctLst/cpt1/Frg-2" title="HAtom" type="jAtomGroup"
      class="HAtom" orientation="45" projection="in"
      target="McStep1/rctLst/cpt1/Frg-0" position="2" formula="H" />
    <jointStructure id="McStep1/rctLst/cpt1/Frg-3" title="HAtom" type="jAtomGroup"
      class="HAtom" orientation="315" projection="in"
      target="McStep1/rctLst/cpt1/Frg-0" position="3" formula="H" />
    <jointStructure id="McStep1/rctLst/cpt1/Frg-4" title="nitro" type="jAtomGroup"
      class="nitro" orientation="270" projection="in"
      target="McStep1/rctLst/cpt1/Frg-0" position="4" formula="NO2" />
    <jointStructure id="McStep1/rctLst/cpt1/Frg-5" title="HAtom" type="jAtomGroup"
      class="HAtom" orientation="225" projection="in"
      target="McStep1/rctLst/cpt1/Frg-0" position="5" formula="H" />
    <jointStructure id="McStep1/rctLst/cpt1/Frg-6" title="HAtom" type="jAtomGroup"
      class="HAtom" orientation="135" projection="in"
      target="McStep1/rctLst/cpt1/Frg-0" position="6" formula="H" />
    <eMoveStructure id="McStep1/rctLst/cpt1/Arc-1" title="IPair" type="anticlockwise"
      orientation="" projection="" source="McStep1/rctLst/cpt1/Frg-0"
      sourcePosition="1" target="McStep1/rctLst/cpt1/Frg-0"
      targetPosition="5" />
    <eMoveStructure id="McStep1/rctLst/cpt1/Arc-2" title="IPair" type="anticlockwise"
      orientation="" projection="" source="McStep1/rctLst/cpt1/Frg-0"
      sourcePosition="5" target="McStep1/rctLst/cpt1/Frg-0"
      targetPosition="4" />
  </structureList>
</mechanismComponent>

```

Figure 9. The XML code describing an aromatic ring involved in nucleophilic substitution.

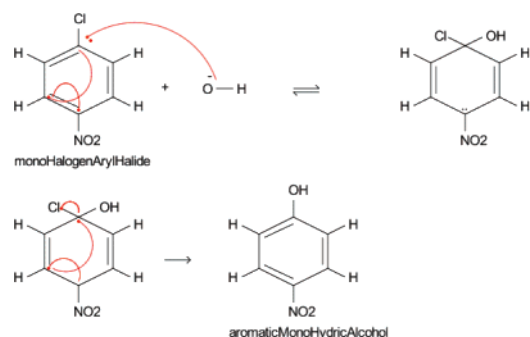


Figure 10. The rendering of the addition–elimination mechanism of an aromatic nucleophilic substitution reaction in a browser.

as a standalone system. A default SVG document is saved into the desired location for the browser display.

Knowledge Decision Support. The knowledge decision support for the present work is provided by the axioms and binary relations interoperating with the chemical ontologies.³³ The AI components in the form of axioms are defined and implemented in XML to handle the situations like the identification of functional groups, decision on the preferable mechanism, probable product classes, and the alkyl group transformations during chemical reactions. This approach is needed especially to make the system behave intelligently. In knowledge engineering in terms of ontology, there is a provision to attach restrictions on any concept class named as axioms. The axioms provide additional semantics about the concept class which cannot be defined normally. In a way the axiom definitions can be considered as artificial

Table 1. Reaction Classes and Mechanisms Considered for the Present Study

nucleophilic substitution at saturated carbon (S_N1 and S_N2 mechanisms)
nucleophilic substitution at aromatic carbon – addition–elimination and elimination–addition (benzyne) mechanisms
mechanisms of various nucleophilic acyl substitution reactions
electrophilic substitution at saturated carbon
electrophilic substitution at aromatic carbon
acid/base catalyzed nucleophilic addition at carbonyl carbon
acid catalyzed and uncatalyzed electrophilic addition on alkenes
radical substitution at alkane
radical addition at alkene
β -elimination reaction at saturated alkyl halides (E1 and E2) and alcohols
rearrangements in S_N1 and E1 mechanisms

intelligence components with which critical inferences about the reaction mechanisms are derived.

The first type of axiom designed for the present study is the axiom-functionalGroup to identify the functional groups present in the structure markup based on fragments. Each fragment in the markup, whether it is a base or a joint structure, is incorporated with an attribute named ‘class’ containing the functional class of the fragment such as “alkyl”, “aryl”, “HAtom”, “carbonyl”, “hydroxide”, etc. when the fragments are integrated to build the structure of a molecule, the system identifies the locations of each and every fragment and builds the combination of classes. For example if a ‘BrAtom’ fragment is attached to a base carbon atom, the system builds class combination as “alkyl@BrAtom”. Similarly a ‘HAtom’ fragment joined to a carbonyl fragment which in turn attached to a benzene ring, and then the

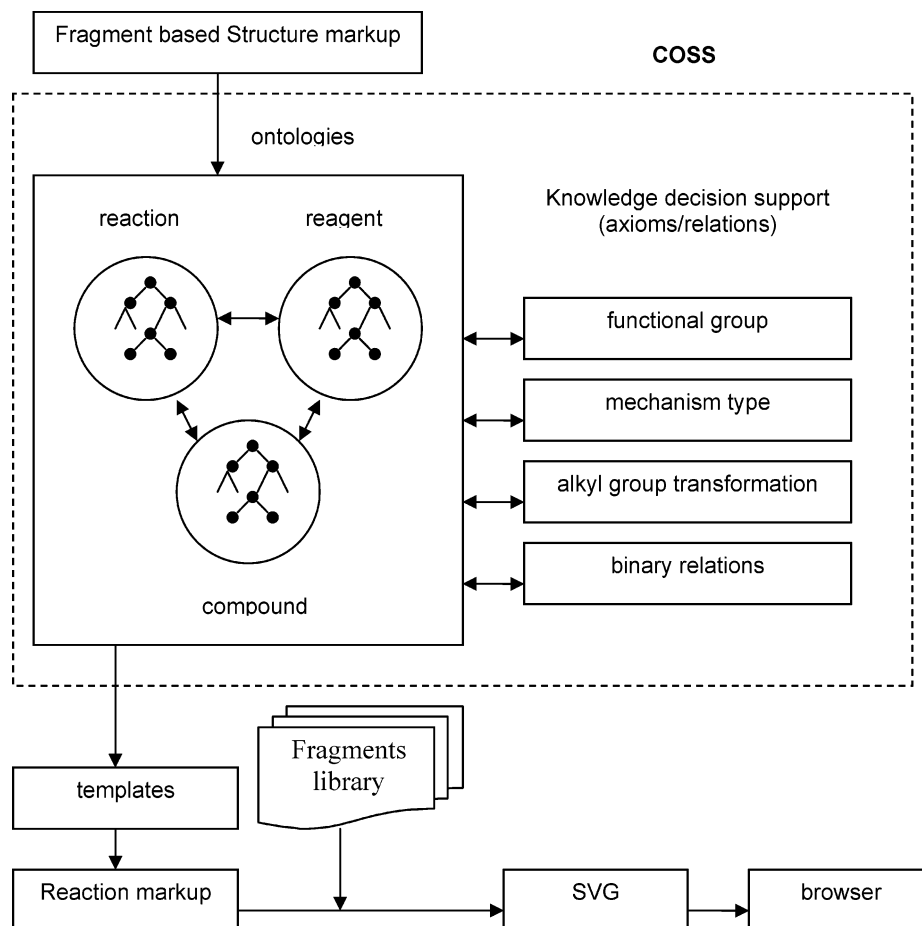


Figure 11. The architecture of the automatic mechanism representation system.

combination class generated by the system is “aryl@carbonyl-HAtom”. Subsequently, the functional class combinations are linked to the axiom-functionalGroup to identify the valid functional moieties in the structure markup. Accordingly the system provides the result as “monoHalogenAlkylHalide” for the earlier and “aromaticAldehyde” for the later. The knowledge representation aiding this identification of functional groups from the structural fragments is shown in Figure 12. The XML implementation of the same is presented in Figure 13.

The decision on the favorable mechanism with respect to a given set of conditions like substrate type, reagent strength, etc., is a complex situation. For example in nucleophilic substitution on a saturated carbon atom and in β -elimination at saturated alkanes the reaction may undergo either through unimolecular or bimolecular mechanisms depending on the reaction conditions. The second axiom definition handles this situation. The problem of deciding the correct mechanism is sorted out with an axiom named axiom-preferredMechanism. The related knowledge and their interrelations are shown in Figure 14. A portion of the XML implementation of this axiom is shown in Figure 15. The preferred mechanism either S_N1 or S_N2 in the case of the nucleophilic substitution reaction and E1 or E2 for β -elimination is mainly decided by the substrate type and the reagent strength. The polarity of the solvent is an enhancing factor for the respective mechanisms. So the definitions are incorporated

with the values corresponding to the substrate type, reagent strength, and solvent polarity. In order to make the definitions generic to be applicable for any other reaction conditions, the definitions are included with needed knowledge such as temperatureRange, pressureRange, catalystType, substrateType, substituentType, reagentStrength, and solventPolarity. A value of ‘null’ for these conditions implies that the corresponding conditions are not considered for the decision making.

The attribute ‘substituentType’ in combination with ‘substrateType’ is used in drawing useful inferences in the reactions involving aromatic substrates. The presence of substituents in *ortho*-, *meta*-, and *para*-positions and their influence on the reaction course is handled with the same axiom-preferredMechanism definitions. Depending on the electron donating or withdrawing capacity of the substituents, the ring gets activated or deactivated influencing the reaction path and course significantly. For example the presence of electron withdrawing substituents in *ortho*- and *para*-positions in halobenzene activates the ring toward nucleophilic substitution reaction. The reaction proceeds through the addition–elimination mechanism. In unactivated halobenzenes, the substitution by a nucleophile is proceeding through the ‘benzyne mechanism’. The encoding of the knowledge related to this issue is shown in Figure 16.

Another situation which requires this sort of knowledge based approach arises in the β -elimination reactions. In these reactions two groups on the adjacent carbons in an alkyl

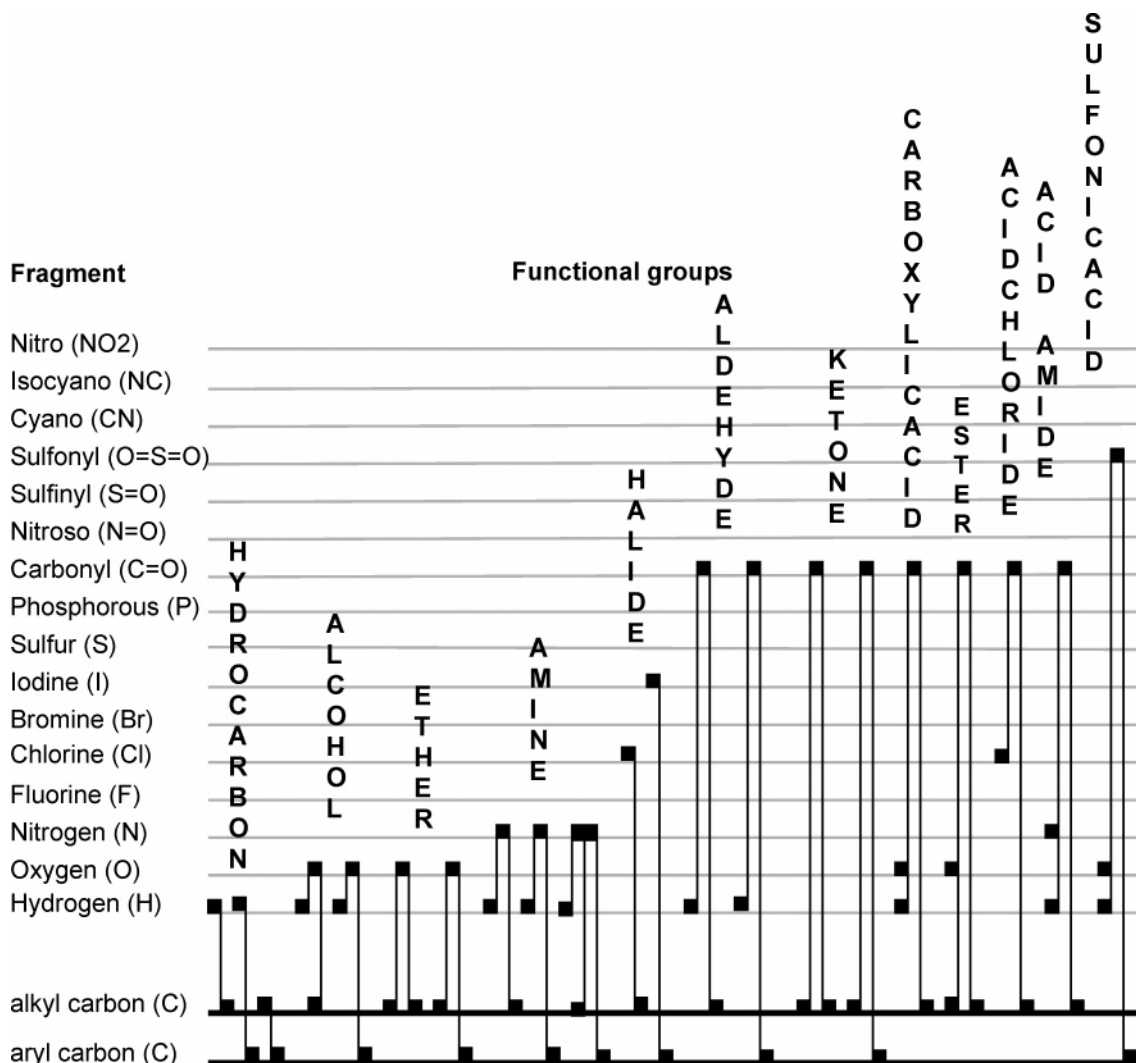


Figure 12. The knowledge representation for a few functional groups and the associated structural fragments.

```
<?xml version="1.0"?>
<!DOCTYPE axiom-functionalGroup SYSTEM "rxnontoDTD.dtd" >
< axiom-functionalGroup xmlns = "http://www.chemonto-india.info" id="" title="" type="">
  <fngrpSet id="alkyl@HAtom" type="saturated" title="alkane" />
  <fngrpSet id="alkyl@methyl" type="saturated" title="alkane" />
  <fngrpSet id="alkenyl@CAtom" type="unsaturated" title="alkene" />
  <fngrpSet id="alkenyl@HAtom" type="unsaturated" title="alkene" />
  <fngrpSet id="alkyl@OAtom-HAtom" type="saturated" title="aliphaticMonoHydricAlcohol" />
  <fngrpSet id="aryl@OAtom-HAtom" type="unsaturated" title="aromaticAlcohol" />
  <fngrpSet id="alkyl@OAtom-alkyl" type="saturated" title="aliphaticEther" />
  <fngrpSet id="alkyl@hydroxide" type="saturated" title="aliphaticMonoHydricAlcohol"/>
  <fngrpSet id="alkyl@BrAtom" type="saturated" title="monoHalogenAlkylHalide" />
  <fngrpSet id="alkyl@carbonyl-OAtom" type="saturated" title="aliphaticCarboxylicacid"/>
  <fngrpSet id="alkyl@carbonyl-alkyl" type="saturated" title="aliphaticKetone" />
  <fngrpSet id="alkyl@carbonyl-HAtom" type="saturated" title="aliphaticAldehyde"/>
  <fngrpSet id="aryl@carbonyl-HAtom" type="saturated" title="aromaticAldehyde"/>
  <fngrpSet id="alkyl@carbonyl-ClAtom" type="saturated" title="aliphaticAcidhalide"/>
  ...
</ axiom-functionalGroup >
```

Figure 13. The XML implementation of the axiom-functionalGroup.

system is removed, and a double bond is formed between the carbon atoms. Since one of the groups attached to the carbon atom is removed completely, there is a change from one alkyl group to another. Complete knowledge about the alkyl groups and their interrelations is needed to encode the

group transformations. For this purpose an alkyl group treelike diagram is developed to represent the knowledge about the alkyl groups and their relationship with each other. The diagram showing the branching and their relations among the alkyl groups is presented in Figure 17, and the

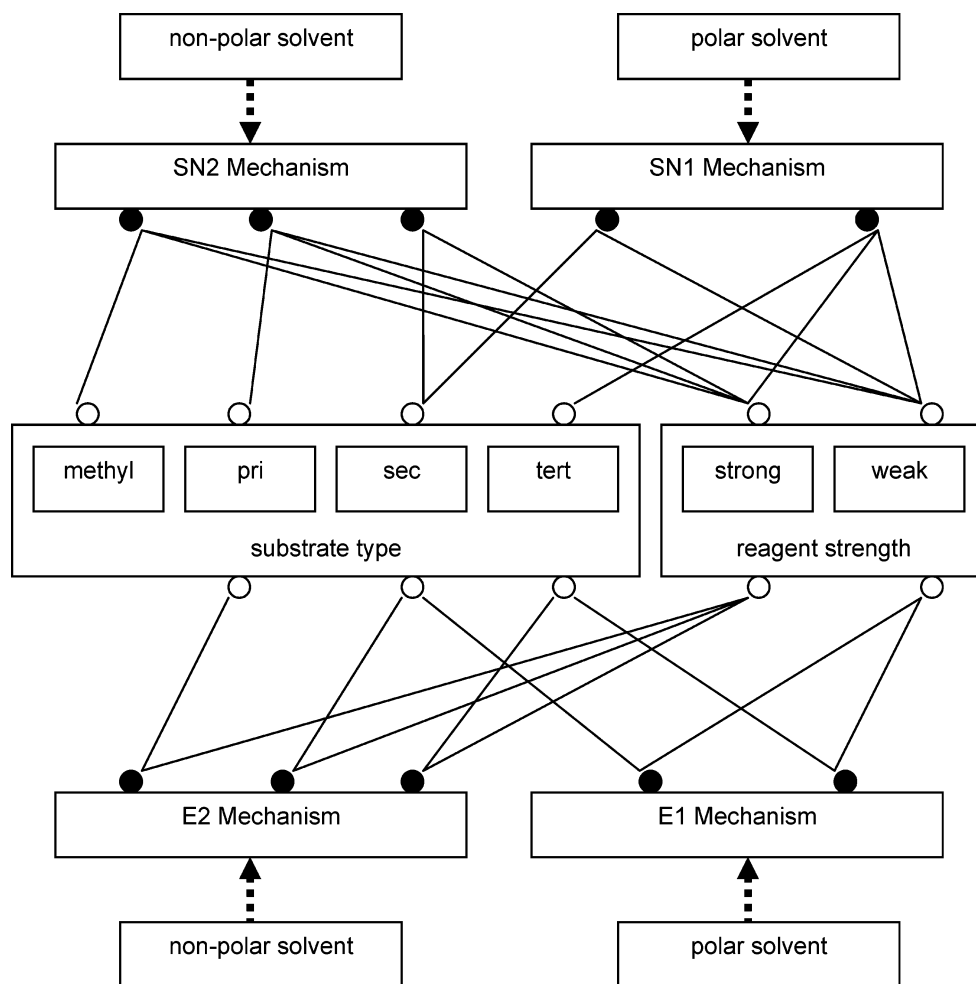


Figure 14. Knowledge decision diagram showing the relationships between the reaction mechanisms and the reaction conditions for nucleophilic substitution and β -elimination reactions.

```

<axiom-preferredMechanism id="sub/nuc/saC-ind-apm" type="mechanism"
  title="preferred mechanism" >
  <mechanismConditionSet id="sub/nuc/saC-ind-apm-001" type="mechanism"
    title="SN2" temperatureRange="" pressureRange="" catalystType=""
    substrateType="methyl" substituentType="" reagentStrength="strong"
    solventPolarity="non-polar" />
  <mechanismConditionSet id="sub/nuc/saC-ind-apm-002" type="mechanism"
    title="SN2" temperatureRange="" pressureRange="" catalystType=""
    substrateType="pri" substituentType="" reagentStrength="strong"
    solventPolarity="non-polar" />
  <mechanismConditionSet id="sub/nuc/saC-ind-apm-003" type="mechanism"
    title="SN1" temperatureRange="" pressureRange="" catalystType=""
    substrateType="tert" substituentType="" reagentStrength="strong"
    solventPolarity="polar" />
  ...
</axiom-preferredMechanism>
    
```

Figure 15. Encoding of axiom-preferredMechanism for an aliphatic nucleophilic substitution reaction.

sample encoding of the knowledge representation is shown in Figure 18.

According to the branching diagram of the alkyl groups, the methyl carbon is considered as the hierarchically lowest level to which three hydrogen atoms are connected. The next level linear alkyl groups such as ethyl, n-propyl, n-butyl, and n-pentyl are represented by connecting methylene groups step by step over the methyl carbon. A carbon designated

as 'isopropyl' is linked to a hydrogen atom and two methyl carbons. Subsequently, the isobutyl group is obtained by linking the isopropyl with a methylene carbon. A tertiary butyl carbon is shown to be connected to three methyl carbon atoms. This approach considers each of the alkyl fragments as a combination of three subalkyl fragments. The encoding of this arrangement of alkyl groups is pursued with the <group-TransferSet> element along with the appropriate attributes

```

<axiom-preferredMechanism id="sub/nuc/ArC-ind" type="Bimolecular Displacement"
  title="preferred mechanism" >
  <mechanismConditionSet id="sub/nuc/ArC-ind-001" type="Addition-Elimination"
    title="arylAE" temperatureRange="" pressureRange="" catalystType=""
    substrateType="12" substituentType="activating" reagentStrength="strong"
    solventPolarity="non-polar" />
  <mechanismConditionSet id="sub/nuc/ArC-ind-002" type="Addition-Elimination"
    title="arylAE" temperatureRange="" pressureRange="" catalystType=""
    substrateType="14" substituentType="activating" reagentStrength="strong"
    solventPolarity="non-polar" />
  ...
  <mechanismConditionSet id="sub/nuc/ArC-ind-005" type="mechanism" title="Benzyne"
    temperatureRange="" pressureRange="" catalystType=""
    substrateType="1" substituentType="" reagentStrength="strong"
    solventPolarity="polar" />
  <mechanismConditionSet id="sub/nuc/ArC-ind-006" type="mechanism" title="Benzyne"
    temperatureRange="" pressureRange="" catalystType=""
    substrateType="12" substituentType="deActivating" reagentStrength="strong"
    solventPolarity="polar" />
  <mechanismConditionSet id="sub/nuc/ArC-ind-007" type="mechanism" title="Benzyne"
    temperatureRange="" pressureRange="" catalystType=""
    substrateType="14" substituentType="deActivating" reagentStrength="strong"
    solventPolarity="polar" />
  ...
</axiom-preferredMechanism>

```

Figure 16. Encoding of axiom-preferredMechanism for an aromatic nucleophilic substitution reaction.

namely 'sourceGroup', 'targetGroup1', 'targetGroup2', and 'targetGroup3'. Accordingly, a source group like the isopropyl group can be split into three target groups namely 'hydrogen', 'methyl', and 'methyl'. This information can be utilized to arrive at the group transformation taking place during β -elimination reactions. Elimination of the HBr molecule from 2-bromopentane results in *trans*-2-pentene. In this reaction, the bromine atom from the second carbon and a hydrogen atom from the third carbon are eliminated to form the reaction product. During this process the third carbon holding two hydrogen atoms and an ethyl group (n-propyl) is transformed into an alkene carbon with only an ethyl and a hydrogen attachment. The alkyl group transformation can be fixed with axiom-groupTransformation to determine the alkyl group changes arising due to β -elimination on an alkyl system. A reverse situation arises in the electrophilic and free radical addition in alkenes. This situation is prevailing in the addition of HBr to alkenes, in electrophilic conditions. The addition of HBr to 2-methyl-2-butene results in the product of 2-methyl-2-bromobutane. This indicates that the hydrogen atom of the reagent is added to the carbon atom holding one hydrogen and one methyl group already. This makes the carbon into an ethyl group carbon. A similar group transformation is taking place in the case of acid-catalyzed hydration of 2-methyl-2butene too. These situations are also handled with the same axiom-groupTransformation definitions.

RENDERING SYSTEM

The reaction representation becomes complete only when it is associated with a suitable display system to view the mechanism schemes in a graphical format. In the present study the mechanism schemes are represented with a unique markup procedure so as to enable the description of all the components of a chemical reaction including the mechanistic

path ways. Accordingly a display technique which can go hand-in-hand with the structure markup procedure is felt as suitable for rendering the reaction schemes in a browser. So an appropriate rendering system is also developed to display the reaction mechanisms as 2D graphics in a way the chemists view the reaction schemes conventionally. The architecture of the rendering system is shown in Figure 19.

The first component of the display system is the reaction mechanism markup generated automatically by the COSS. This is an XML document and forms as the input for the rendering system. The descriptions made in the input XML file are then converted into an SVG file in which the contents of the mechanism markup are captured as 2D graphics. This is done by linking the input data with the second component of the system termed as structureXY. The structureXY system is a cluster of XML documents describing the *x*, *y* coordinate values of various structural fragments needed for the construction of reaction mechanism components. According to this system the chemical and nonchemical structural fragments of mechanism components are viewed primarily as lines, texts, and points. In order to describe this information, a simple XML extension language is developed to markup the *x*, *y* coordinates of the representative structural fragments. The rendering language is constructed with only three XML elements namely <lineXY>, <textXY>, and <pointXY> enclosed in a container element <structureXY>. Thus the construction of the structureXY system involves the identification of structural pieces in terms of points, lines, and texts followed by assigning the *x*, *y* coordinate values relative to a reference point within the structure. In order to understand the construction of the structureXY system, the markup code for a few structural fragments belonging to the base, joint structures are shown in Figure 20.

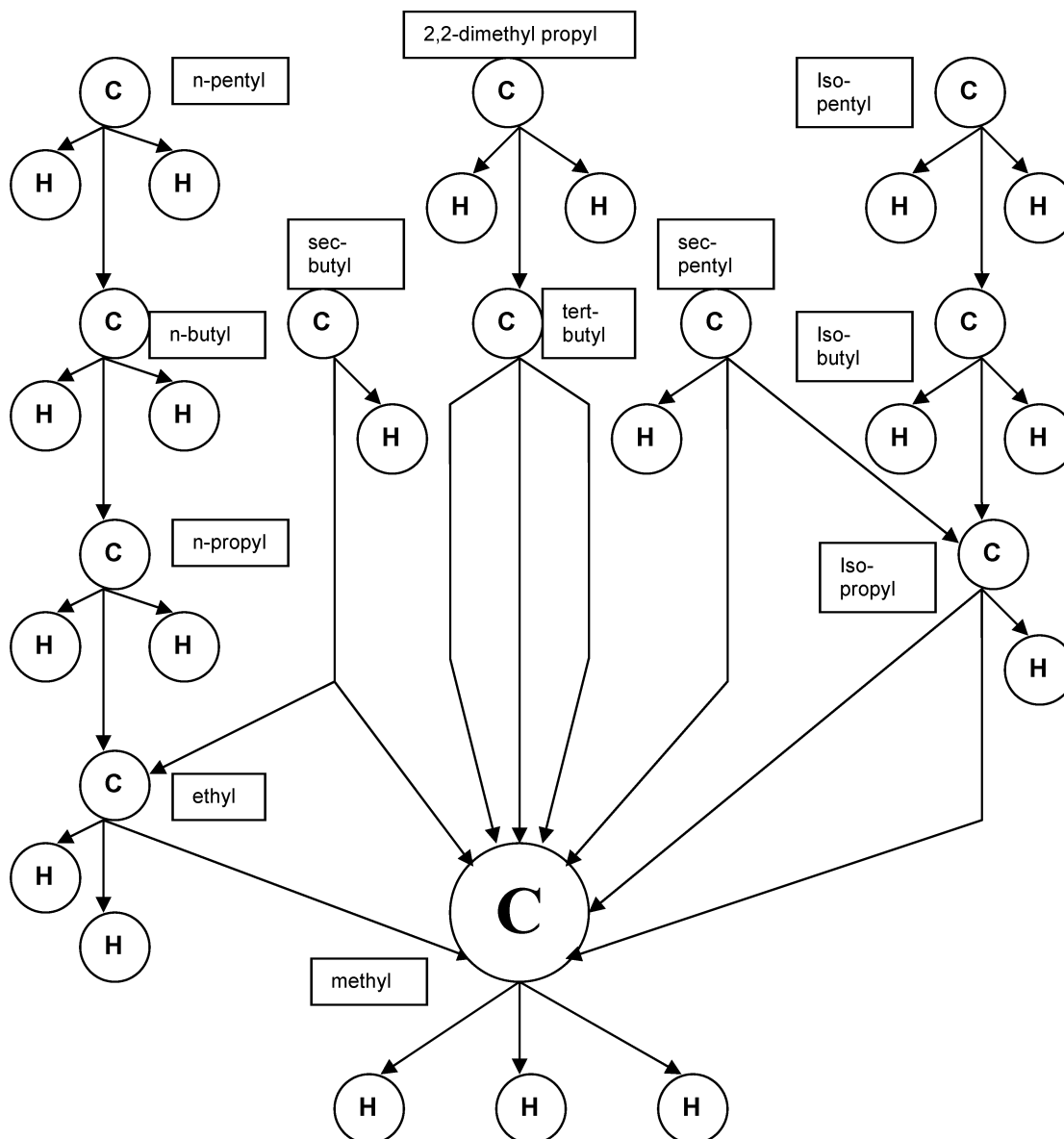


Figure 17. The knowledge representation of branching and their relations among the alkyl groups.

```

<axiom-groupTransformation id=" " type="" title="" >
  <groupTransferSet id=" 001" type="" title=""
    sourceGroup="methyl" targetGroup1="hydrogen" targetGroup2="hydrogen"
    targetGroup3="hydrogen"/>
  <groupTransferSet id=" 002" type="" title=""
    sourceGroup="ethyl" targetGroup1="methyl" targetGroup2="hydrogen"
    targetGroup3="hydrogen"/>
  <groupTransferSet id="003" type="" title=""
    sourceGroup="n-propyl" targetGroup1="ethyl" targetGroup2="hydrogen"
    targetGroup3="hydrogen"/>
  <groupTransferSet id=" 004" type="" title=""
    sourceGroup="n-butyl" targetGroup1="n-propyl" targetGroup2="hydrogen"
    targetGroup3="hydrogen"/>
  <groupTransferSet id=" 005" type="" title=""
    sourceGroup="iso-propyl" targetGroup1="methyl" targetGroup2="methyl"
    targetGroup3="hydrogen"/>
  ...
</axiom-groupTransformation>
    
```

Figure 18. Encoding of axiom-groupTransformation.

Any line segment in the structural fragment is described with the <lineXY> element. The x, y coordinate positions

for the line segment are provided with four attributes namely x1, y1, x2, and y2. The semantics of the line are incorporated

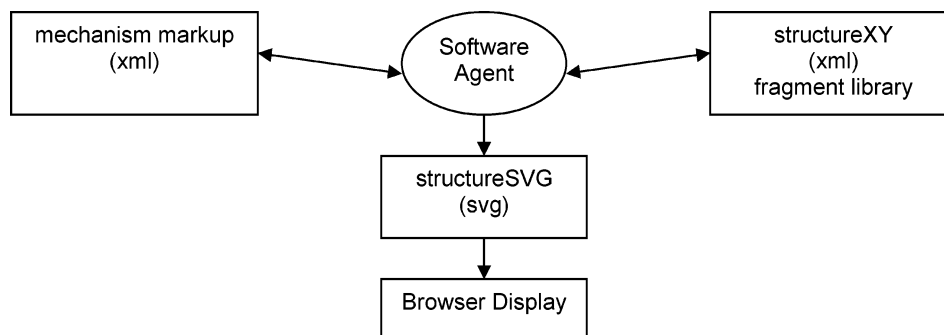


Figure 19. The architecture of a rendering system.

```

<structureXY id="CAAtom0" title="carbon" type="alkyl" formula="C">
  <textXY id="CAAtom0Tx-1" title="C" type="atom" x1="-6" y1="6"/>
  <pointXY id="CAAtom0Pt-1" type="uPair" x1="0" y1="-10" />
  <pointXY id="CAAtom0Pt-2" type="uPair" x1="0" y1="10" />
  <pointXY id="CAAtom0Pt-3" type="uPair" x1="-10" y1="0"/>
  <pointXY id="CAAtom0Pt-4" type="uPair" x1="10" y1="0" />
</structureXY>
<structureXY id="carbonyl0" title="carbonyl" type="carbonyl" formula="CO">
  <lineXY id="carbonyl0Ln-1" type="sglBond" x1="0" y1="0" x2="20" y2="0" />
  <lineXY id="carbonyl0Ln-2" type="dblBond" x1="28" y1="-10" x2="28" y2="-30" />
  <lineXY id="carbonyl0Ln-3" type="dblBond" x1="32" y1="-10" x2="32" y2="-30" />
  <textXY id="carbonyl0Tx-1" type="atom" title="O" x1="24" y1="-36" />
  <textXY id="carbonyl0Tx-2" type="atom" title="C" x1="24" y1="6" />
  <pointXY id="carbonyl0Pt-1" type="bPair" x1="30" y1="-10"/>
  <pointXY id="carbonyl0Pt-2" type="bPair" x1="30" y1="10" />
  <pointXY id="carbonyl0Pt-3" type="bPair" x1="20" y1="0" />
  <pointXY id="carbonyl0Pt-4" type="uPair" x1="40" y1="0" />
</structureXY>
<structureXY id="BrAtom180" title="" type="BrAtom" formula="Br">
  <lineXY id="BrAtom180Ln-1" type="sglBond" x1="0" y1="0" x2="-20" y2="0" />
  <textXY id="BrAtom180Tx-1" type="atom" title="Br" x1="-36" y1="6" />
</structureXY>
<structureXY id="partialMinus" title="" type="">
  <textXY id="partialMinusTx-1" type="misc" title="δ" x1="0" y1="0" />
  <textXY id="partialMinusTx-2" type="misc" title="-" x1="10" y1="-10" />
</structureXY>
<structureXY id="twArrow0" title="" type="">
  <lineXY id="twArrow0Ln-1" type="misc" x1="0" y1="0" x2="30" y2="0" />
  <lineXY id="twArrow0Ln-2" type="misc" x1="26" y1="-4" x2="30" y2="0" />
  <lineXY id="twArrow0Ln-3" type="misc" x1="0" y1="4" x2="30" y2="4" />
  <lineXY id="twArrow0Ln-4" type="misc" x1="4" y1="8" x2="0" y2="4" />
</structureXY>
  
```

Figure 20. The markup code examples of base, joint, and miscellaneous structures in the structureXY library.

in the 'type' attribute using the values such as 'sglBond', 'dblBond', 'tplBond', and 'misc'. Accordingly, the type attribute indicates a line segment as a single bond or double bond or triple bond or a miscellaneous line. The <textXY> and <pointXY> elements are associated with only x1 and y1 values sufficient to hold the x, y coordinate values. The 'type' attribute in <textXY> element holds the meaning like atom and other miscellaneous text (R, n, etc.) with the values such as 'atom' and 'misc', respectively. Similarly, the value of the 'type' attribute in <pointXY> is used to indicate whether the point is the location of a lone pair, bond pair, unpaired electron, and miscellaneous points using the values of 'lPair', 'bPair', 'uPair', and 'misc', respectively. The <textXY> element is provided with an additional attribute, 'title', to hold the atom symbols like C, H, O, N, S, F, Cl,

Br, I, etc. All three elements are provided with an 'id' attribute to contain a unique identification value. The container for the whole markup of any structural fragment is the <structureXY> which in turn is placed in <structureXYList> as the root element in an XML document.

According to this markup methodology, the features of a structural fragment are marked simply with <lineXY>, <textXY>, and <pointXY> elements. A point within the structural unit is identified as the origin point and is assigned with a value of '0' for both x1 and y1 attributes. This point is considered as the reference point, with respect to this the line segments, text positions, and other substituent points in the structure are appropriately assigned the values for x1, y1, x2, and y2, respectively. The structural fragment assigned with appropriate coordinate values forms as a reusable

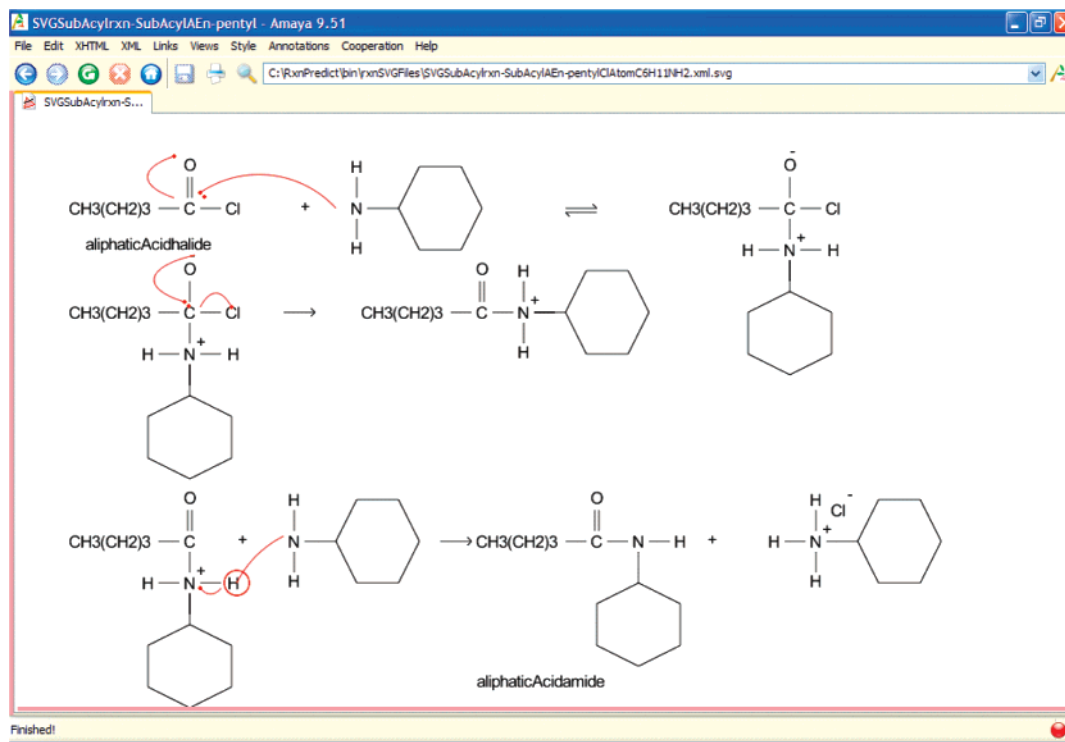


Figure 21. The layout showing the addition–elimination mechanism of a nucleophilic acyl substitution.

structural fragment and is stored in the structureXY library. These reusable components are used for the construction of simple to complex mechanism components for any number of times. The structureXY library consists of the structural fragments belonging to the baseStructure, jointStructures, and miscellaneous structures as separate XML documents. The XML documents thus obtained are collectively considered as the structureXY system.

The construction of the structure of a mechanism component is a process of writing an SVG document describing the component in a graphical format as directed by the associated structural markup. A software agent integrates the structure markup and the structureXY system to transform the structural information in the markup into a graphical format utilizing the x , y coordinates of the appropriate structural fragments. The resultant SVG document consists of the structural features of the mechanism component described as 2D graphics which can be viewed in a browser³⁴ compatible with SVG. The advantage of this method is that the graphical components of a mechanism can be rendered on the screen at any desired location simply by incrementing the x , y coordinate values appropriately. The system places any structural fragment or a constructed component in a position corresponding to $x = "0"$ and $y = "0"$ by default. In the present study of representing the chemical reactions, the flexibility of the system and its advantage is felt in the layout of the mechanism schemes. Each and every component of mechanism steps are converted into individual SVG documents. Then they are combined into a single SVG document in which the components are appropriately incremented to appear in the desired locations on the screen.

The movements of electrons are represented with elliptical arcs using the `<path>` and `<circle>` elements of SVG. The path element contains a `d = "path data"` attribute to define the coordinate values and other data needed for drawing an

elliptical arc. The contents of the path data attribute starts with a 'moveto' command 'M' to specify the origin point of the arc in terms of x , y coordinate values. The path data ends with another x , y data to indicate the destination point of the arc. In between these two data, the values needed to specify the size of the arc (radius along x and y axes) and the orientation (clockwise or anticlockwise) of the arc are placed. As the markup code for the electronic arc is defined with the source and target locations, the x , y coordinate values for the corresponding points on the specified structural fragments are identified by the system, and the values for the path data attributes are fed during the writing of SVG. The circle element is used to represent the unpaired and lone pair electrons at the tip of the arc. A single filled circle indicates the movement of single unpaired electron. The lone pair electron movement is shown with a pair of circles at the destination point of the arc. The layout obtained on a browser for the reaction of *n*-hexanoylchloride with cyclohexylamine undergoing addition–elimination mechanism is shown in Figure 21.

CONCLUSION

The development of generic models for the reaction mechanisms for common organic reactions exhibits the usefulness of the chemical ontologies in constructing intelligent applications in the domain of chemistry. The proposed knowledge based representation system is a good platform to handle the complex patterns of reaction mechanisms. Also the system provides a means to incorporate the artificial intelligence components either in the knowledge level or inside the applications to derive meaningful inferences for the prediction of reactions and their mechanisms. The descriptions of the special features of the mechanism components such as arrows, symbols, square brackets, and the electronic movements brings a completeness in repre-

senting the chemical reactions. The flexible XML based markup coding procedure developed for the description of reaction mechanisms and to render the equivalent graphics are easily extendable and can be integrated with other applications. The possibility to store, retrieve, and communicate the reactions and their mechanism in a W3C endorsed media like SVG is a global mode of exchanging the chemical reactions between systems via the Internet.

ACKNOWLEDGMENT

This research work was supported by the Research Grant from All India Council for Technical Education (AICTE), New Delhi, India under Research Promotion Scheme (RPS) 2004-2006 F.No. 8022/RID/NPROJ/RPS-109/2003-04.

Supporting Information Available: DTD for the reaction document, descriptions of XML elements used in the composition of reaction document, and the sample XML and SVG codes generated automatically by the system and browser display for a few mechanisms. This material is available free of charge via the Internet at <http://pubs.acs.org>.

REFERENCES AND NOTES

- CambridgeSoft chemDraw. <http://www.cambridgesoft.com/software/ChemDraw/> (accessed Oct 20, 2006).
- MDL ISIS/Draw. <http://www.mdli.com/index.jsp> (accessed Oct 20, 2006).
- ACD/ChemSketch. [chemsketch/](http://www.acdlabs.com/chemsketch/) (accessed Oct 20, 2006).
- SMILES — A Simplified Chemical Language. <http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html> (accessed Oct 20, 2006).
- ReactionSMILES and SMIRKS. <http://www.daylight.com/meetings/summerschool01/course/basics/smirks.html> (accessed Oct 20, 2006).
- Reactor. <http://www.chemaxon.com/jchem/doc/user/Reactor.html> (accessed Oct 20, 2006).
- 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. <http://www.mdli.com/downloads/public/ctfile/ctfile.pdf> (accessed May 13, 2007).
- Chemical Markup Language (CML). <http://www.xml-cml.org> (accessed May 13, 2007).
- Murray-Rust, P.; Rzepa, H. S. Chemical Markup Language and XML Part I Basic Principles. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 928–942.
- Extensible Markup Language (XML). <http://www.w3.org/XML> (accessed May 13, 2007).
- Krause, S.; Willighagen, E.; Steinbeck, C. JChemPaint - Using the Collaborative Forces of the Internet to Develop a Free Editor for 2D Chemical Structures. *Molecules* **2000**, 5, 93–98.
- Introduction to JChemPaint. <http://almost.cubic.uni-koeln.de/cdk/jcp> (accessed Oct 20, 2006).
- Steinbeck, C.; Han, Y. Q.; Kuhn, S.; Horlacher, O.; Luttmann, E.; Willighagen, E. The Chemistry Development Kit (CDK): An Open-Source Java Library for Chemo- and Bioinformatics. *J. Chem. Inf. Comput. Sci.* **2003**, 43, 493–500.
- CDX Format Specification: The CDXML text-based file format. <http://www.cambridgesoft.com/services/documentation/sdk/chemdraw/cdx/IntroC DXML.htm> (accessed May 13, 2007).
- CDX Format Specification: General. <http://www.cambridgesoft.com/services/documentation/sdk/chemdraw/cdx/General.htm> (accessed May 13, 2007).
- Holliday, G. L.; Murray-Rust, P.; Rzepa, H. S. Chemical Markup, XML, and the World Wide Web. 6. CMLReact, an XML Vocabulary for Chemical Reactions. *J. Chem. Inf. Model.* **2006**, 46, 145–157.
- Holliday, G. L.; Mitchell, J. B. O.; Murray-Rust, P. CMLSnap. A novel method of reaction representation. *Internet J. Chem.* **2004**, 7, Article 4.
- W3C — Scalable Vector Graphics (SVG) XML Graphics for the Web. <http://www.w3.org/Graphics/SVG/> (accessed Oct 20, 2006).
- Murray-Rust, P.; Rzepa, H. S. Chemical Markup, XML and the World-Wide Web. 2. Information Objects and the CML DOM. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 1113–1123.
- Murray-Rust, P.; Rzepa, H. S.; Wright, M. Development of chemical markup language (CML) as a system for handling complex chemical content. *New J. Chem.* **2001**, 25, 618–634.
- Taylor, K. R.; Gledhill, R. J.; Essex, J. W.; Frey, J. G.; Harris, S. W.; De Roure, D. C. Bringing Chemical Data onto the Semantic Web. *J. Chem. Inf. Model.* **2006**, 46, 939–952.
- Gkoutos, G. V.; Murray-Rust, P.; Rzepa, H. S.; Wright, M. Chemical Markup, XML and the World-Wide Web. 3. Towards a Signed Semantic Chemical Web of Trust. *J. Chem. Inf. Comput. Sci.* **2001**, 41, 1124–1130.
- W3C Semantic Web. <http://www.w3.org/2001/sw/> (accessed Oct 20, 2006).
- Fujita, S.; Tanaka, N. XyM Notation for Electronic Communication of Organic Chemical Structures. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 903–914.
- Fujita, S. XyM Markup Language (XyMML) for Electronic Communication of Chemical Documents Containing Structural Formulas and Reaction Schemes. *J. Chem. Inf. Comput. Sci.* **1999**, 39, 915–927.
- Tanaka, N.; Ishimaru, T.; Fujita, S. WWW (World Wide Web) Communication and Publishing of Structural Formulas by XyMML (XyM Markup Language). *J. Comput.-Aided Chem.* **2002**, 3, 81–89.
- Tanaka, N.; Fujita, S. XyMJava System for World Wide Web Communication of Organic Chemical Structures. *J. Comput.-Aided Chem.* **2002**, 3, 37–47.
- Ito, I.; Tanaka, N.; Fujita, S. Development and Application of XyM2Mol System for Converting Structural Data by XyM Notation into Connection Tables. *J. Comput. Chem. Jpn.* **2005**, 4, 79–88.
- Ingold, C. K. In *Structure and Mechanism in Organic Chemistry*, 2nd ed.; CBS Publishers & Distributors: New Delhi, India, 1994; Chapters 5–15.
- Bruckner, R. In *Advanced Organic Chemistry — Reaction Mechanisms*; Academic Press An Imprint of Elsevier: New Delhi, 2005; Chapters 1–6.
- Wade, L. G., Jr. In *Organic Chemistry*, 5th ed.; Pearson Education (Singapore) Pte, Ltd.: New Delhi, India, 2004; Chapters 4, 6, 8, 17, 18, 20, and 21.
- Sankar, P.; Aghila, G. Design and Development of Chemical Ontologies for Reaction Representation. *J. Chem. Inf. Model.* **2006**, 46, 2355–2368.
- Welcome to Amaya W3C's Editor/Browser. <http://www.w3.org/Amaya/Amaya 9.3>. <http://www.w3.org/Amaya/User/New.html> (accessed Oct 20, 2006).

CI700043U