

In-Silico Chemical Bond Validation Technique based on a Semantic Chemical Structure Markup System

Punnaivanam Sankar
Department of Chemistry
Pondicherry Engineering
College, Puducherry - 605 014,
India

ABSTRACT

A technique to automatically validate the chemical bonding between the atoms during the structure construction on a computer screen using a chemical structure editor is developed. The technique involves the capturing of semantics of the electronLinks between the atoms involved in chemical bonding. An artificial intelligence component is developed to support the system to bring the decisions of valid chemical bonding based on the semantics of the electronLinks involved in bonding.

General Terms

Knowledge Representation, Artificial Intelligence

Keywords

XML, ontology, chemical bonding.

1. INTRODUCTION

In the recent years the Internet has spread widely enforcing the *in-silico* operation on the chemical information such as search, retrieval, communication etc. through appropriate representation and encoding procedures compatible for the internet technology. The influence of Internet system over the storage, retrieval and the communication of chemical structures are significant because the contents in Internet are managed by markup languages. The treatment of chemical structures in a web media needed a markup based representation technique in which the structural information are to be provided as tagged information. As a consequence of this the Chemical Markup Language¹ (CML) has emerged to describe the structural features. The CML is an XML² based markup language capturing the structural information through a concise set of tags with the associated semantics.

The technology of drawing structures on the screen as input resulted in the development of several structure editors. The prominent editors such as Cambridgesoft ChemDraw³, Symyx Draw⁴, ACDs ChemSketch⁵, MarvinSketch⁶, JChemPaint⁷ possesses the functionality to draw, render and to characterize the structures drawn on the screen by linking with suitable databases. These editors are capable of supporting a wide range of file formats. None of the tools describes explicitly the type of bonds connecting atoms in terms of the information related to the electrons involved in bonding. This is because of the absence of appropriate semantics in the representation of chemical structures. So the available tools

are suitable to be supported by databases and not by knowledge bases. To achieve knowledge based tools to describe and process chemical structures, the tool has to be supported by some conceptual knowledge backup. As the knowledge about the chemical bonding is already established, the AI components can be developed through the knowledge representation procedures followed by the implementation in a generic platform. The resultant components are capable of serving as reusable resources in a common platform and to make the applications intelligent enough. Further it is also possible to integrate them with the ontologies to develop and extend applications globally to bring the system open.

The WWW^{8,9} is transforming into a new generation web technology demanding the contents available in the Internet to be processed by the computers. So the next generation communication media is expected to be the evolving versions of web technologies and applications⁸⁻¹⁶ in which the information will be made available in a semantically rich format and accessible by machines for various processes. This emerging trend needs innovative and semantically rich representation formats to handle chemical bonding during structure construction. Accordingly the technique of automatic bond validation is proposed for the structure construction on computer screen with a structure editor, ChemEd¹⁷ capable of describing chemical structures in XML².

2. METHOD

The proposed technique is tested with the tool ChemEd¹⁷, developed by us. The tool allows the structure construction through the selection of appropriate structural fragments defined in the built in Fragment Library of the tool. The ChemEd describes the chemical structure in terms of a structure group, which may contain one or more structures. Each structure in turn is described with the associated structural fragments. The fragments are described with the atoms with which it is composed of. Finally, each atom is associated with the description of suitable number of electronLinks. Thus the ChemEd captures the whole structure description as an XML² document for a convenient process. During the structure construction in ChemEd¹⁷ using the structural fragments, an appropriate bond validation is needed in order to enable the chemical bond formation between two fragments along the atoms according to chemical laws. A

validation technique suitable for a structure description in XML has been developed and discussed.

3. RESULT AND DISCUSSION

The fragment based structure construction involves the establishment of chemical bonding between two atoms of structural fragments. The actual chemical entity involved in chemical bonding¹⁸ is the outermost electrons of the atoms. The electrons are distributed around the atoms in electron orbitals and the orbitals are arranged in different energy levels. Each orbital can accommodate two electrons maximum. So the status of the orbital can be empty, one electron and a pair of electrons. The single electron status and the paired electron status are called as “unpaired” and “lone pair” electron respectively. For a chemical bond between two atoms to be formed, two electrons are needed. After the bond formation the electrons used for bonding between two atoms are shared by both the atoms. These shared electrons in the form of chemical bonding are called as bond pairs. When the bond is formed by sharing of unpaired electron from each atom involved, the bond formed is called as covalent bond. If the two electrons needed for a bond formation is provided by the same atom and shared by both the atoms, the bond formed is a coordinate-covalent bond. It is also called as dative bond. There is another type of bond termed ionic bond forms by the electrostatic attraction. This bond results between two oppositely charged electron sites. The charged electron site is a result of either loss or a gain of electron on the electron site. When an electron site loses an electron it becomes positively charged and the electron site gains the electron becomes negatively charged.

During a chemical bonding, the electron sites associated with the atoms involved in bonding experiences significant changes. For example two electron sites with unpaired electrons changes into a bond pair after the chemical bonding. Similarly the charge status and the details of target fragment also experiences changes appropriately. In ChemEd¹⁷ system of structure construction, appropriate semantic are identified to capture the chemical bonding and are associated with the <atom> element using <electronLink/> elements in the mark up. The important semantics concerned to the chemical bonding details is shown below:

```
<electronLink id="" title="" electronStatus="" charge=""  
chargeCount="" affinity="" bond="" order="" target=""  
linkStatus=""/>
```

The ‘id’ attribute is used to hold the unique ‘id’ value generated by ChemEd during chemical bonding. The ‘title’ attribute is to denote the name of the electron site as “1s/2s/2p etc”. The number of electrons present in the electronLink is provided by the attribute ‘electronStatus’ using the values like “empty/uPair/lPair/bPair” for “empty/unpaired/lone paired/bond paired” status respectively. The attributes ‘charge’ and ‘chargeCount’ is used to fix the charge status of the electronLink. The ‘charge’ attribute can hold the values such as “+/-” to indicate the positive / negative charge.

Whereas the ‘chargeCount’ attribute takes up the values like “0/1/2” to bring the number of positive or negative charges on the electronLink. The remaining five attributes are used to provide the semantics about the chemical bonding. The values for ‘affinity’ attribute are “covalent/ionic/dative” indicating the default nature of the electronLink. The ‘bond’ attribute is used for capturing the type of bond such as sigma/pi/aromatic etc. The ‘order’ attribute implies the bond order with values “single/double/triple” for single bond, double bond and triple bond respectively. The ‘target’ attribute take up the value of ‘id’ attribute of another electronLink to which this electronLink is mapped for bonding. Finally the linkStatus holds the detail of role of the electronLink in chemical bonding with values “linkSource/linkTarget”. Accordingly for example a default electronLink markup for an unpaired electron belonging to a HydrogenAtom is shown below:

```
<electronLink id="hydrogen-0-a1e1" title="1s"  
electronStatus="uPair" charge="0" chargeCount="0"  
affinity="covalent" bond="" order="" target=""  
linkStatus=""/>
```

The same electronLink after mapped with the electronLink of another Hydrogen atom through a covalent bond is provided below:

```
<electronLink id="hydrogen-0-a1e1" title="1s"  
electronStatus="bPair" charge="0" chargeCount="0"  
affinity="" bond="sigma" order="single" target="hydrogen-1-  
a1e1" linkStatus="linkSource"/>
```

Whereas the semantics of the target electronLink is shown below:

```
<electronLink id="hydrogen-1-a1e1" title="1s"  
electronStatus="bPair" charge="0" chargeCount="0"  
affinity="" bond="sigma" order="single" target="hydrogen-0-  
a1e1" linkStatus="linkTarget"/>
```

So it is obvious that the semantics of chemical bonding can be captured automatically with the semantics of the source and target electronLinks. Based on the source and target electronLinks involved in chemical bonding an artificial intelligence component is developed with a suitable knowledge representation format arrived from the established knowledge about the chemical bonding. For example if an electronLink with unpaired electron is targeted with another electronLink with unpaired electron, bond validation allows the linkage to connect the two atoms establishing the linkage between the two atoms. On the other hand invalid fragment selections are automatically sensed and rejected by the system. The system will not allow linkage between one electron site with unpaired electron and another with lone pair electrons. Further the type of bond such as covalent, ionic and dative to be established between the concerned atoms is also automatically determined by the AI component providing the intelligence characteristics to the system. The knowledge representation format in terms of a semantic network providing the AI component is presented in the Figure 1.

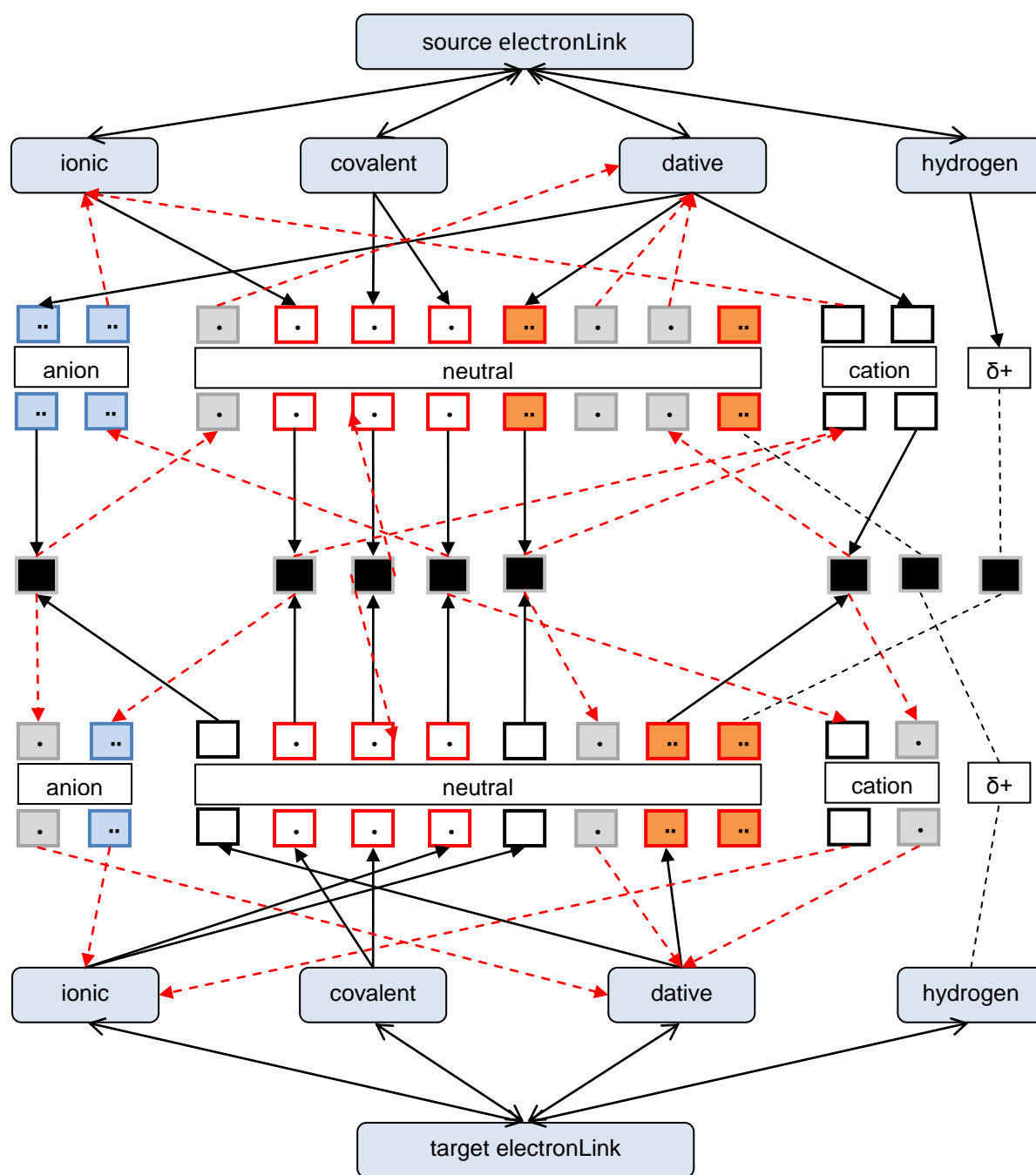


Fig 1: The semantic network for bond validation

The knowledge described in the semantic network for bond validation is working on three aspects about the electronLinks involved in bond mapping. The first one is the normal tendency of the atoms to form a covalent or ionic or dative bond. This information is present in the 'affinity' attribute of every electronLink. The second one is the charge status of the electronLink such as neutral or anionic or cationic or partial as referenced from the 'charge' attribute. The electronic status is the third consideration in which the electronLinks with different electronic status. This is available with the 'electronStatus' attribute. As the semantics are already available in the markup, the decision on the valid bonding and

the type of bonding is achieved by developing axioms. According to the chemical bonding description shown in the Figure 1, the two electronLinks involved in the bond mapping are designated as source and target electronLinks. The source electronLink is the one belonging to an atom being linked to another atom containing the target electronLink. Both source and target electronLinks are related to the possible charge and electronic status to indicate various types of chemical bonding. The solid arrow head connectors leading from source and target electronLink and ends with a solid square box signify a valid chemical bonding. The resultant features of the respective electronLink after the establishment of the

bond mapping is shown by dashed arrow heads. This is indicated in a reverse direction so that the dashed arrows start from the solid square boxes and ends with the respective electronLinks providing the modified charge and electronic status of the respective electronLink. The knowledge representation for the bond validation is implemented in XML² in the form of axioms^{15,16}. The sample code for the axioms corresponding to the formation of a covalent, ionic and a dative bond between the source and target electronLinks are shown below:

```
<axiom bondId="covalent" sEStatus="uPair"
sChargeType="0" sBondType="covalent" tEStatus="uPair"
tChargeType="0" tBondType="covalent" rsEStatus="bPair"
rsChargeType="0" rtEStatus="bPair" rtChargeType="0"/>
```

```
<axiom bondId="ionic" sEStatus="uPair" sChargeType="0"
sBondType="ionic" tEStatus="uPair" tChargeType="0"
tBondType="covalent" rsEStatus="empty" rsChargeType="+"
rtEStatus="iPair" rtChargeType="-"/>
```

```
<axiom bondId="dative" sEStatus="iPair" sChargeType="-"
sBondType="ionic" tEStatus="empty" tChargeType="0"
tBondType="dative" rsEStatus="bPair" rsChargeType="0"
rtEStatus="bPair" rtChargeType="-"/>
```

The axiom constructed for the bond validation contains the attributes, 'sEStatus', 'sChargeType', and 'sBondType' to hold the electronic status, charge status and the bonding tendency of the source electronLink. Similarly the attributes 'tEStatus', 'tChargeType', and 'tBondType' are defined for the target electronLink. In order to describe the resultant status of the electronLinks after the bonding, the attributes 'rsEStatus', 'rsChargeType', 'rtEStatus' and 'rtChargeType' are used. The type of bond formed after the electronLink linkage is shown with the 'bondId' attribute. Based on this implementation a covalent bond between a neutral source electronLink with unpaired electron having a tendency to form a covalent bond is valid when this source electronLink is linked with a target electronLink with similar electronLink status. Consequently the resultant change is only in the electronic status. After bonding, the unpaired electrons of both electronLinks become bond pairs retaining the charge status as neutral. If the bonding tendency of the source electronLink is ionic and the charge status is neutral, when linked with the target electronLink with the covalent bonding tendency possessing neutral charge, the resultant bond is ionic. Subsequently, the source electronLink acquires a positive charge by losing the electron to the target electronLink making it an ion pair with a negative charge. Similarly the dative bond formation between a negatively charged ionic species and the empty dative electronLinks of metals are described in the third piece of the code fragment. The role of AI component developed in building the structures is explained with some simple structures constructed with the system.

A simplest case of building hydrogen molecule involves two hydrogen atoms. The semantics of the 1s electronLinks of both atoms provided by the semantics are shown below:

The source hydrogen atom electronLink

```
<electronLink id="hydrogen-0-a1-o1" title="1s"
electronStatus="uPair" charge="0" chargeCount="0"
affinity="covalent" bond="" order="" target=""
linkStatus=""/>
```

The target hydrogen atom electronLink

```
<electronLink id="hydrogen-0-a1-o1" title="1s"
electronStatus="uPair" charge="0" chargeCount="0"
affinity="covalent" bond="" order="" target=""
linkStatus=""/>
```

In the above bond mapping, the source and the target electronLink details of both hydrogen atoms fit into the first axiom definition shown above. So the system allows the formation of a bond linkage between two hydrogen fragments and it identifies the chemical bond formed is a covalent bond. The second axiom describes the bond mapping between sodium atom and chlorine atom to form an ionic bond between them.

4. CONCLUSION

The conceptual description of chemical structures in terms of structural fragments, constituent atoms and the associated electronLinks, result in a semantically rich structure markup suitable for the bond mapping procedure during structure construction in ChemEd¹⁷. The technique of automatically validating the chemical bond mapping during structure construction provides the artificial intelligence perspective to the structure editor. The semantic markup construct in XML² and the possibility of interoperability with the other XML² based markup languages makes the proposed technique suitable for the evolving trends of WWW^{8,9}.

5. ACKNOWLEDGMENTS

The author acknowledges the Department of Science & Technology, New Delhi, India for funding under SERC Scheme, File No. SR/S1/OC-86/2009.

6. REFERENCES

- [1] Murray-Rust, P., and Rzepa, H. S. 1999. Chemical Markup Language and XML Part I. Basic Principles. *Journal of Chemical Information and Computer Science*. (39), 928-942.
- [2] W3C Extensible Markup Language (XML). <http://www.w3.org/XML>. accessed November, 10, 2011.
- [3] CambridgeSoft Life Science Enterprise Solutions, Desktop software chemDraw: <http://www.cambridgesoft.com/software/ChemDraw/>. accessed November, 10, 2011.

- [4] MDL ISIS Draw 2.5 <http://mdl-isis-draw.software.informer.com/2.5/>. accessed November, 10, 2011.
- [5] ACD/Labs Chemical Drawing & Nomenclature: http://www.acdlabs.com/products/draw_nom/. accessed November, 10, 2011.
- [6] ChemAxon, Marvin: <http://www.chemaxon.com/marvin/index.html>. accessed November, 10, 2011.
- [7] JChemPaint: <http://sourceforge.net/apps/mediawiki/cdk/index.php?title=JChemPaint>. accessed November, 10, 2011.
- [8] Berners-Lee, T., Hendler, J., and Lassila, O. 2001. The Semantic web. *Scientific American*. (284), 34-43.
- [9] W3C Semantic Web Activity. <http://www.w3.org/2001/sw/>. accessed November, 10, 2011.
- [10] Murray-Rust, P., and Rzepa, H. S. Chemical Markup, XML and the World-Wide Web. 2 Information Objects and the CML DOM. 2001. *Journal of Chemical Information and Modeling*. (41), 1113-1123.
- [11] Murray-Rust, P., Rzepa, H. S., and Wright, M. Development of chemical markup language (CML) as a system for handling complex chemical content. 2001. *New Journal of Chemistry*. (25), 618-634.
- [12] Taylor, K.R., Gledhill, R. J., Essex, J.W., Frey, J.G., Harris, S.W., and De Roure, D.C. Bringing Chemical Data onto the Semantic Web. 2006. *Journal of Chemical Information and Modeling*. (46), 939-952.
- [13] Gkoutos, G. V., Murray-Rust, P., and Rzepa, H. S.; Wright, M. Chemical Markup, XML and the World-Wide Web, Part III: towards a signed semantic Chemical Web of Trust. 2001. *Journal of Chemical Information and Modeling*. (41), 1124-1130.
- [14] Murray-Rust, P. Chemistry for All. 2008. *Nature*. (451), 648-651.
- [15] Sankar, P., and Aghila, G. Design and Development of Chemical Ontologies for Reaction Representation. 2006. *Journal of Chemical Information and Modeling*. 46, 2355-2368.
- [16] Sankar, P., and Aghila, G. Ontology Aided Modeling of Organic Reaction Mechanisms with Flexible and Fragment Based XML Markup Procedures. 2007. *Journal of Chemical Information and Modeling*. (47), 1747-1762.
- [17] Sankar, P., Krief, Alain., and Aghila, G. Model Tool to Describe Chemical Structures in XML Format Utilizing Structural Fragments and Chemical Ontology. 2010. *Journal of Chemical Information and Modeling*. (50), 755-770.
- [18] Wade, L. G., Jr. In *Organic Chemistry*, 5th ed.; Pearson Education (Singapore) Pte, Ltd.: New Delhi, India, 2004; Chapters 1, and 2.