Journal of Molecular Graphics and Modelling 61 (2015) 30–43

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Encoding of Fundamental Chemical Entities of Organic Reactivity Interest using chemical ontology and XML

Durairaj Vijayasarathi [b], Punnaivanam Sankar [a,*]

[a] Department of Chemistry, Pondicherry Engineering College, Puducherry 605 014, India
[b] Department of Chemistry, Kanchi Mamunivar Centre for Post Graduate Studies, Puducherry 605 008, India

ARTICLE INFO

Article history:
Received 9 December 2014
Received in revised form 16 April 2015
Accepted 9 June 2015
Available online 26 June 2015

Keywords:
Organic reactivity
Chemical structure
Chemical entity
Notation
XML
Chemical ontology

ABSTRACT

Fundamental chemical entities are identified in the context of organic reactivity and classified as appropriate concept classes namely ElectronEntity, AtomEntity, AtomGroupEntity, FunctionalGroupEntity and MolecularEntity. The entity classes and their subclasses are organized into a chemical ontology named "ChemEnt" for the purpose of assertion, restriction and modification of properties through entity relations. Individual instances of entity classes are defined and encoded as a library of chemical entities in XML. The instances of entity classes are distinguished with a unique notation and identification values in order to map them with the ontology definitions. A model GUI named Entity Table is created to view graphical representations of all the entity instances. The detection of chemical entities in chemical structures is achieved through suitable algorithms. The possibility of asserting properties to the entities at different levels and the mechanism of property flow within the hierarchical entity levels is outlined.

1. Introduction

The description of organic reactivity [1–5] by computer based technique requires the identification and encoding of generic sub-structural components of molecular structure. If a molecular structure is considered as a chemical entity, the sub-structural components include any characteristic and meaningful sub-structural units with which the molecular structure is composed. Accordingly, a chemical entity in the context of organic reactivity is to be defined in terms of various sub-structural units of a molecular structure. These sub-structural units can be considered as the fundamental chemical entities representing any distinguishable, stable or unstable chemical species needed to describe the organic reactivity. If such fundamental chemical entities of organic reactivity are identified and encoded properly, it can be used to develop meaningful reactivity prediction system. Subsequently, the system can be extended to arrive at a useful reaction description system as well.

Considering the sub-structural components of a molecule as the fundamental chemical entities of organic reactivity, there are only a few categories to be pursued in this interest. In describing the organic reactivity, the role of characteristic groups [1–8] like carbonyl, hydroxyl, amino, etc. is crucial. The presence of such characteristic chemical groups on a carbon skeleton is normally described with term functional group [7–9]. Hence, the functional group of molecule is an essential chemical entity in detailing the organic reactivity. In order to describe the characteristic chemical groups and the functional groups with the respective atoms, the concept of atom as a chemical entity is needed. Finally, the inclusion of electron as a chemical entity is inevitable because it is involved in chemical bonding.

In a reaction of aldehyde with an alcohol to form acetal through hemi-acetal, the first step is the protonation of the aldehyde in acid catalysed condition. The protonation involves one of the lone pairs of carbonyl oxygen atom in the formation of a coordinate covalent bond with a hydrogen ion [5]. As a result of this bonding, the lone pair of oxygen atom changes into a bond pair and the vacant electron status of hydrogen ion also attains the status of a bond paired electron. The change of lone pair entity into a bond pair entity of the oxygen atom is an electron level transformation. Due to this change, the oxygen atom acquires the property of a bond pair electron in place of the lone pair electron. In detailing the organic reactivity, the change in the properties of electron neighbourhood of an atom is essential and important. In consequence, the consideration of valence electrons of every atom as distinct chemical entities is inevitable. The inclusion of electron as a dis-

* Corresponding author.
E-mail address: gapspec@gmail.com (S. Punnaivanam).

http://dx.doi.org/10.1016/j.jmgm.2015.06.001
1093-3263/

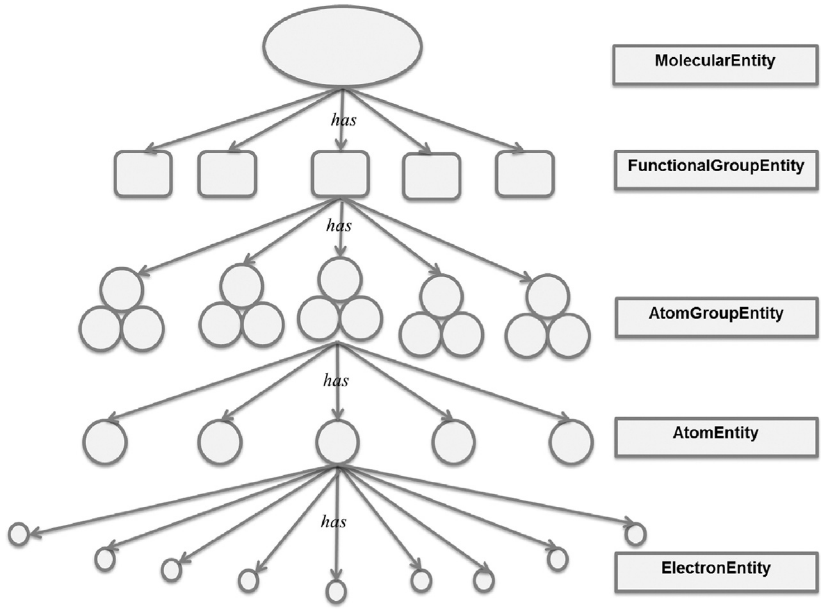

Fig. 1. Fundamental Chemical Entities and their relationship.

tinct chemical entity brings the possibility encoding the electron entities with appropriate properties needed for the description of organic reactivity.

The changes in the electron neighbourhood impart specific reactivity properties of an atom in the chemical structure. For example, in the structural change of an ammonia molecule into an ammonium ion, the positive charge arising on the Nitrogen atom is susceptible to non-covalent interactions. Depending on the number and availability of bonded electrons, non-bonded electrons, lone pairs and vacant electron status, the manifestation of atom differs in the context of organic reactivity. Accordingly, different species of same atom adopt different chemical properties depending on its electron environment. The chemical environment and the properties of carbon atom in the structure of methane, ethylene and acetylene are different. Similarly the characteristics of nitrogen atom in primary, secondary and tertiary amine structures vary. It is imperative to consider every atom species with distinct electron environment as an instance of the same atom entity.

The approach of identifying basic structural components in terms of conceptual chemical entities is useful in associating respective properties pertaining to the organic reactivity at different entity levels. Further, the concept of chemical entities, entity instances and their relationships can be organized into a chemical ontology. The resultant chemical ontology is a suitable media for the property inheritance from one level into other. Chemical ontologies [10–20] are intended to define and describe the chemical concepts on a machine understandable common language like Web Ontology Language [21] (OWL). The ontological descriptions are shared across the community as a generic knowledge resource. Several reports are available on the use of chemical ontology for the purpose of chemical knowledge exchange in a common media. A notable one is the Chemical Entities of Biological Interest (ChEBI) [22]. It is an ontological database comprising huge number of chemical entities in relation to their specific biological properties. ChEBI is a freely available dictionary of small molecular entities and their structures in standard structure representations. Glycomics ontology [23] (GlycO) provides extensive knowledge and semantically rich descriptions of carbohydrate structure, glycan binding relationships, glycan biosynthetic pathways and the developmental biology of stem cells. A similar database of bioactive molecules with properties related to drug activities is developed in the name of ChEMBL [24] and maintained by European Bioinformatics Institute.

The term chemical entity is generally used to denote molecular entity referring to the chemical name of molecules. IUPAC Gold Book [8] treats the term chemical entity to specify the representation of any constitutionally or isotopically distinct chemical species. The species include atom, ion, ion pair, radical, radical ion, molecule, conformer, complex, etc., classifiable as a separately distinguishable entity. A recent study reports a methodology of Chemical Entity Semantic Specification (CHESS) in order to provide cross-domain knowledge integration [25]. In this study, the components of molecular structures like atoms, bonds and functional groups are considered as specific chemical entities except electrons. This system is a RDF based chemical information specification system backed by a chemical ontology, CHENINF [13]. The focus of the system is on encoding of semantic chemical information for facile chemical knowledge integration across web media.

We have reported chemical ontologies developed in relevance to the chemical reactions with a focus of arriving methodologies for representing the reactions [14] and their mechanisms [15]. A

```
<atom id=" " title=" " type="" hybridization="" symbol=" " charge="" chargeCount="" position=""
          isotopeLabel="" electroNegativity="" oxdnState="" block=" " gcCode="" notation=""
          partCharge="" partChargeVal="" x="" y="">

          <electronLink id="" title="" type="" electronStatus="" charge="" chargeCount="" affinity=""
                   bond=""  order="" linkStatus="" target="" orientation="" projection="" priority=""
                   gcCode=" " notation="" partCharge="" partChargeVal="" x1="" y1="" x2="" y2=""/>
```

Fig. 2. Description template of atom and electronLink elements in XML.