



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.

© 2015 Elsevier Inc. All rights reserved.

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 describ-

ing 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).

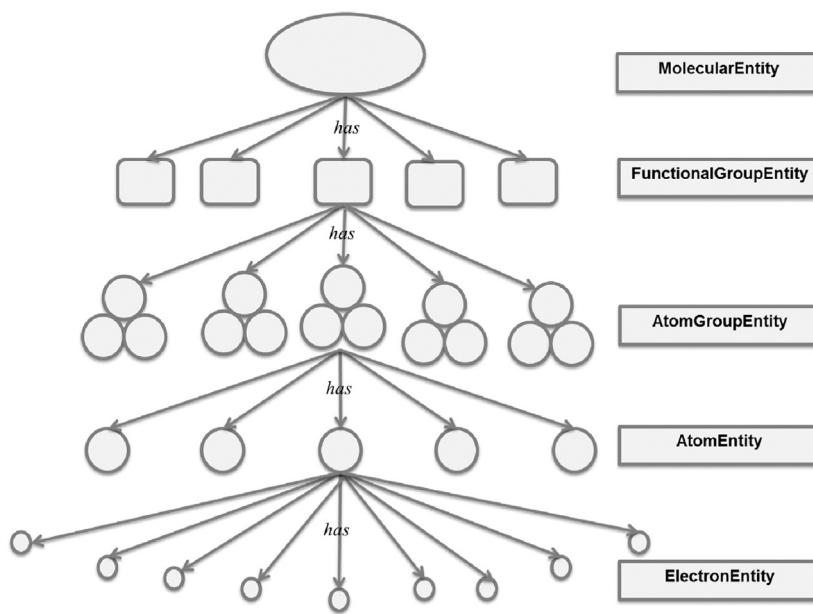


Fig. 1. Fundamental Chemical Entities and their relationship.

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.

```

<electronEntityList>
    <electronEntity id="EE-001" title="(00)lonePairEE" type="" charge="0" chargeCount="0" notation="(00)lp">
        <electronLink id="" title="lonePair" type="" electronStatus="IPair" charge="0" chargeCount="0" affinity="" bond="" order="" linkStatus="" target=""
    orientation="" projection="" priority="" gcCode="" notation="(00)lp" partCharge="" partChargeVal="" x1="" y1="" x2="" y2="" />
    </electronEntity>
    <electronEntity id="EE-002" title="(-1)lonePairEE" type="ion" charge="-" chargeCount="1" notation="(-1)lp">
        <electronLink id="" title="lonePair" type="" electronStatus="IPair" charge="-" chargeCount="1" affinity="" bond="" order="" linkStatus="" target=""
    orientation="" projection="" priority="" gcCode="" notation="(-1)lp" partCharge="" partChargeVal="" x1="" y1="" x2="" y2="" />
    </electronEntity>
    <electronEntity id="EE-003" title="(00)unPairEE" type="radical" charge="0" chargeCount="0" notation="(00)up">
        <electronLink id="" title="unPair" type="" electronStatus="uPair" charge="0" chargeCount="0" affinity="" bond="" order="" linkStatus="" target=""
    orientation="" projection="" priority="" gcCode="" notation="(00)up" partCharge="" partChargeVal="" x1="" y1="" x2="" y2="" />
    </electronEntity>
    <electronEntity id="EE-004" title="(-1)unPairEE" type="radicalIon" charge="-" chargeCount="1" notation="(-1)up">
        <electronLink id="" title="unPair" type="" electronStatus="uPair" charge="-" chargeCount="1" affinity="" bond="" order="" linkStatus="" target=""
    orientation="" projection="" priority="" gcCode="" notation="(-1)up" partCharge="" partChargeVal="" x1="" y1="" x2="" y2="" />
    </electronEntity>
    <electronEntity id="EE-005" title="(+1)unPairEE" type="radicalIon" charge="+" chargeCount="1" notation="(+1)up">
        <electronLink id="" title="unPair" type="" electronStatus="uPair" charge="+" chargeCount="1" affinity="" bond="" order="" linkStatus="" target=""
    orientation="" projection="" priority="" gcCode="" notation="(+1)up" partCharge="" partChargeVal="" x1="" y1="" x2="" y2="" />
    </electronEntity>
    <electronEntity id="EE-006" title="(00)hookPairEE" type="hook" charge="0" chargeCount="0" notation="(00)hp">
        <electronLink id="" title="bondPair" type="" electronStatus="bPair" charge="0" chargeCount="0" affinity="" bond="" order="" linkStatus="" target=""
    orientation="" projection="" priority="" gcCode="" notation="(00)hp" partCharge="" partChargeVal="" x1="" y1="" x2="" y2="" />
    </electronEntity>
    <electronEntity id="EE-007" title="(+1)hookPairEE" type="hook" charge="+" chargeCount="1" notation="(+1)hp">
        <electronLink id="" title="bondPair" type="" electronStatus="bPair" charge="+" chargeCount="1" affinity="" bond="" order="" linkStatus="" target=""
    orientation="" projection="" priority="" gcCode="" notation="(+1)hp" partCharge="" partChargeVal="" x1="" y1="" x2="" y2="" />
    </electronEntity>
    <electronEntity id="EE-008" title="(-1)hookPairEE" type="hook" charge="-" chargeCount="1" notation="(-1)hp">
        <electronLink id="" title="bondPair" type="" electronStatus="bPair" charge="-" chargeCount="1" affinity="" bond="" order="" linkStatus="" target=""
    orientation="" projection="" priority="" gcCode="" notation="(-1)hp" partCharge="" partChargeVal="" x1="" y1="" x2="" y2="" />
    </electronEntity>
    <electronEntity id="EE-009" title="(+δ)hookPairEE" type="hook" charge="+" chargeCount="δ" notation="(+δ)hp">
        <electronLink id="" title="bondPair" type="" electronStatus="bPair" charge="+" chargeCount="δ" affinity="" bond="" order="" linkStatus="" target=""
    orientation="" projection="" priority="" gcCode="" notation="(+δ)hp" partCharge="" partChargeVal="" x1="" y1="" x2="" y2="" />
    </electronEntity>
    <electronEntity id="EE-010" title="(-δ)hookPairEE" type="hook" charge="-" chargeCount="δ" notation="(-δ)hp">
        <electronLink id="" title="bondPair" type="" electronStatus="bPair" charge="-" chargeCount="δ" affinity="" bond="" order="" linkStatus="" target=""
    orientation="" projection="" priority="" gcCode="" notation="(-δ)hp" partCharge="" partChargeVal="" x1="" y1="" x2="" y2="" />
    </electronEntity>
    <electronEntity id="EE-011" title="(00)vacantPairEE" type="" charge="0" chargeCount="0" notation="(00)vp">
        <electronLink id="" title="vacantPair" type="" electronStatus="vacant" charge="0" chargeCount="0" affinity="" bond="" order="" linkStatus="" target=""
    orientation="" projection="" priority="" gcCode="" notation="(00)vp" partCharge="" partChargeVal="" x1="" y1="" x2="" y2="" />
    </electronEntity>
    <electronEntity id="EE-012" title="(+1)vacantPairEE" type="ion" charge="+" chargeCount="1" notation="(+1)vp">
        <electronLink id="" title="vacantPair" type="" electronStatus="vacant" charge="+" chargeCount="1" affinity="" bond="" order="" linkStatus="" target=""
    orientation="" projection="" priority="" gcCode="" notation="(+1)vp" partCharge="" partChargeVal="" x1="" y1="" x2="" y2="" />
    </electronEntity>
</electronEntityList>

```

Fig. 3. XML encoding of ElectronEntities.

semantic representation format for describing chemical structures in terms of atoms with electron neighborhood is developed along with a suitable structure editor, “ChemEd” [15,17]. In line with this approach, a generic conceptual model of functional groups [18] is achieved with the support of chemical ontology. Based on this model the ChemEd is used to define, encode and detect functional groups [18]. There is no clear and formal approach exclusively for the description of organic reactivity in terms of chemical entities. We report a methodology to encode and detect Fundamental Chemical Entities of Organic Reactivity Interest. The possibility of associating properties and detecting them on chemical structure is outlined in this study. A chemical ontology is developed for the purpose of property assertion, inheritance and restriction, etc. on the chemical entities. The proposed approach may be suitable to develop applications on predicting organic reactivity in a semantic framework to provide precise information. The methodology reported is expected to be suitable for the retrieval of precise information when compared to the reaction database searches.

2. Chemical entity classification

The fundamental chemical entities identified to describe the organic reactivity are classified into appropriate concept categories viz. ElectronEntity (EE), AtomEntity (AE), AtomGroupEntity

(AGE), FunctionalGroupEntity (FGE) and MolecularEntity (ME). These entities are organized into a concept taxonomy and structured as a chemical ontology named as “ChemEnt”. The concepts and their hierarchical relationships defined in ChemEnt are shown in Fig. 1.

According to the concept classification of chemical entities, the top level concept in the hierarchy is the MolecularEntity. A MolecularEntity represents the structural composition of smaller to bigger molecules like water, ethanol, acetic acid, carbohydrates, amino acids, etc. The next entity class in the hierarchy is the FunctionalGroupEntity. It is related to its parent concept, MolecularEntity with a ‘has’ relationship. Accordingly, a ME class can be related with one or more FGE in order to describe the molecular entity with appropriate properties of respective functional group entities. The functional group entity class in turn is related to one or more AtomGroupEntity concepts with ‘has’ relationships. The concept of AGE represents the structure of characteristic groups like hydroxyl group, carbonyl group, formyl group, etc. Each AGE is related to the concept of one or more AtomEntity concepts which in turn is related to one or more ElectronEntity concepts through ‘has’ relationship.

The hierarchical entity classification favours the encapsulation of necessary attributes or properties with respective entity class and relating them with the next higher or lower level entity classes. Organizing the entity concepts into a chemical ontology provides

Table 1
Details of various ElectronEntities.

S.No	Entity Name	Graphical Representation	Entity Notation	Meaning
01	(00)lonePairEE	(..)	(00)lp	EE representing a lone pair of electron with neutral charge
02	(-1)lonePairEE	(..)	(-1)lp	EE representing a lone pair of electron with a negative charge indicating the gain of a single electron
03	(00)unPairEE	(•)	(00)up	EE representing an unpaired electron with neutral charge indicating a radical
04	(-1)unPairEE	(•)	(-1)up	EE representing an unpaired electron with a negative charge indicating the gain of a single electron from vacant status
05	(+1)unPairEE	(+•)	(+1)up	EE representing an unpaired electron with a positive charge indicating the loss of a single electron from lone pair status
06	(00)hookPairEE	(○)	(00)hp	EE representing an electron hook with neutral charge indicating the resultant EE of hypothetical homolytic cleavage of a covalent bond
07	(+1)hookPairEE	(+○)	(+1)hp	EE representing an unpaired electron hook with positive charge indicating the resultant EE of hypothetical homolytic cleavage of a coordinate covalent bond
08	(-1)hookPairEE	(-○)	(-1)hp	EE representing an unpaired electron hook with negative charge indicating the resultant EE of hypothetical homolytic cleavage of a coordinate covalent bond
09	(+δ)hookPairEE	(+○δ)	(+δ)hp	EE representing an unpaired electron hook with partial positive charge (+δ) indicating the resultant EE of hypothetical homolytic cleavage of a polar covalent bond and the atom associated with it is lesser electronegativity along bond.
10	(-δ)hookPairEE	(-○δ)	(-δ)hp	EE representing an unpaired electron hook with partial negative charge (-δ) indicating the resultant EE of hypothetical homolytic cleavage of a polar covalent bond and the atom associated with it is higher electronegativity along bond.
11	(00)vacantPairEE	(○)	(00)vp	EE representing a vacant electron position with neutral charge
12	(+1)vacantPairEE	(+○)	(+1)vp	EE representing a vacant position of electron with a positive charge indicating the loss of a single electron from unPair neutral EE

the possibility of defining individual instances of a concept class. The ontology can also be used for the assertion of data properties to the individual instances and object properties between the concept classes. Hence, the individual instances of chemical entity classes can be defined in ontology and asserted with the respective reactivity oriented properties. The asserted properties can be detected on a chemical structure along with the inherited properties with the support of a chemical ontology. The ontology support is useful in distinguishing different atomic environment in chemical structures. For example, it is possible to infer the difference between the properties of the nitrogen atom in the structure of ammonia molecule and ammonium ion. The reactivity oriented behaviour of carbonyl group in the structure of functional groups like aldehyde, ketone, carboxylic acid, etc. can be distinguished automatically. This kind of approach can be utilized in developing a reasonable reaction prediction methodology resulting in knowledge based system rather than a database oriented one.

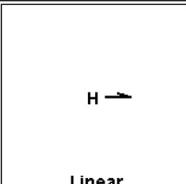
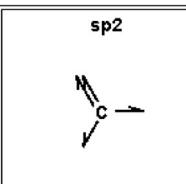
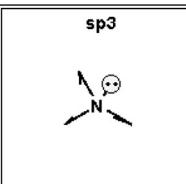
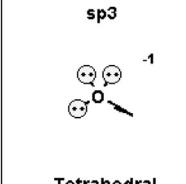
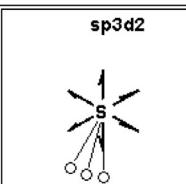
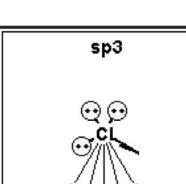
A prerequisite for the property assertion and detection is a generic encoding system enabling the encoding of every individual instances of entity classes viz. EE, AE, AGE, FGE and ME. The encoding system is achieved through a suitable structure representation system developed earlier [17]. The structure representation system describes the chemical composition of a molecular structure in terms of atoms along with the necessary and sufficient number of outer electrons using `<electronLink>` elements explicitly in XML [26]. Every `<electronLink>` element is defined with the necessary semantics about the valence electrons. The utility of this encoding technique is demonstrated through the development of a chemical structure editor (ChemEd) working on the basis of the XML descriptions [17]. Subsequently, the structure editor is used to define and detect organic functional groups in chemical structures [18]. The same representation system is used for the proposed system of encoding the fundamental entities of reactivity interest where the explicit representation of electron neighbourhood is essential. The template of the structure description in XML for generic atom associated with one `electronLink` XML element is shown in Fig. 2.

3. Encoding of ElectronEntity

Twelve instances for the concept class of ElectronEntity are identified and encoded for this system at present. Each of the EE is provided with a name and a brief notation for unique identification. The name of entity is to know explicitly the electron status of the electron entity along with the charge associated to it. For example, a neutral lone pair electron status is indicated with a name string “(00) lonePairEE”. The corresponding notation string for the entity is “(00)lp”. The term within the brackets indicates the charge associated with the electron entity. Neutral, positive and negative charges are indicated with “0”, “+” or “-”, respectively, followed by their count with the numerals like 0, 1, 2, 3, etc. Accordingly, the strings “(+1)”, “(-1)” indicate one positive and one negative charge, respectively. Partial positive and partial negative charges are indicated with “(+δ)” and “(-δ)”, respectively. A string “(00)” means that the charge is zero. Similarly a neutral unpaired electron status is represented with the name “(00) unPairEE” and its notation as “(00) up”. Vacant electron status is represented with the terms “vacantPair” and “vp” for name and notation, respectively. Accordingly the “(00) vacantPairEE” and “(00) vp” represents the name and notation string, respectively, for a neutral vacant electron status entity. A concept of hookPair is used to denote the status of an electron covalently bonded to another atom. The hookPair status is required for representing status of an electron entity arising as a result of hypothetical homolytic cleavage of a covalent bond. The notation string for this status is indicated as “hp”. Accordingly, the

Table 2

Details of some AtomEntities.

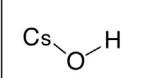
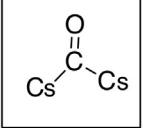
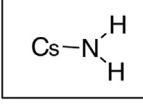
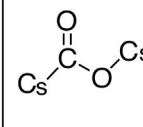
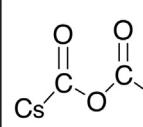
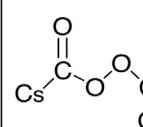
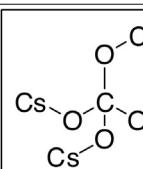
SN O	Entity Class	Structure Representation	Notation	Meaning
1	HydrogenAE	 <p>Linear</p>	H(00)(1sh)	Hydrogen neutral atom entity has one single hookPair
2	CarbonAE	 <p>TrigonalPlanar</p>	C(00)(1dh;2sh)	Carbon neutral atom entity has 1 double hookPair and 2 single hookPairs
3	NitrogenAE	 <p>Tetrahedral</p>	N(00)(1lp;3sh)	Nitrogen neutral atom entity has one lone pair electronLinks and 3 single hookPairs
4	OxygenAE	 <p>Tetrahedral</p>	O(-1)(3lp;1sh)	A negatively charged Oxygen atom entity with 3 lone pair electronLinks and 1 single hookPair
5	SulfurAE	 <p>Octahedral</p>	S(00)(3vp;6sh)	Sulfur neutral atom entity has 6 single hookPair's and 3 vacant pair electronLinks representing 3 empty d-orbitals
6	ChlorineAE	 <p>Tetrahedral</p>	Cl(00)(3lp;5vp;1sh)	Chlorine neutral atom entity has 3 lone pair electronLinks and one single hookPair with 5 d-orbital vacant position electronLinks

notations "(00) hp", "(+1) hp", "(-1) hp" represent hookPair electron status with neutral charge, a positive and a negative charge associated with it, respectively. The details of the electron entities along with the graphical representations and the meaning are shown in [Table 1](#).

The encoding of an ElectronEntity concept is achieved with the description of a <electronEntity> XML element. The semantics of an electron entity is provided with the attributes namely id, title, type, charge, chargeCount and notation to signify meaningful outer electron status for an atom entity. The XML encoding of all the EEs

Table 3

Some representatives of FunctionalGroupEntities.

S.N o	Entity Name	Structure Representation	Notation
1	Alcohol-FGE		O[(sb@C;sb@H)]
2	Ketone-FGE		C[(db@O;sb@C;sb@C)]
3	PrimaryAmine-FGE		N[(sb@C;sb@H;sb@H)]
4	CarboxylicEster-FGE		C[(db@O;sb@O;sb@C)]O[(sb@C;sb@C)]
5	Anhydride-FGE		C[(db@O;sb@O;sb@C)]O[(sb@C;sb@C)]C[(db@O;sb@O;sb@C)]
6	DiacylPeroxide-FGE		C[(db@O;sb@O;sb@C)]O[(sb@O;sb@C)]O[(sb@O;sb@C)]C[(db@O;sb@O;sb@C)]
7	OrthoCarbonateEstester-FGE		O[(sb@C;sb@C)]C[(sb@O;sb@O;sb@O)]O[(sb@C;sb@C)]O[(sb@C;sb@C)]O[(sb@C;sb@C)]

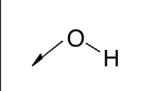
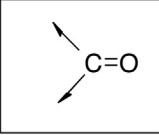
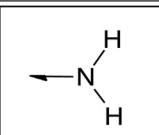
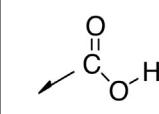
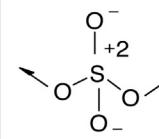
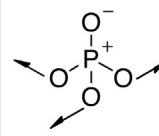
is presented in Fig. 3. The ‘id’ attribute provides a unique alphanumeric value in order to map with the respective concept in the ChemEnt ontology. The ‘title’ and ‘notation’ attributes are used to hold the name and the notation of the entity respectively as already described. The charge on the entity is encoded with ‘charge’ and ‘chargeCount’ attributes. The values of these attributes are “+” or “-” in case of ‘charge’ attribute and numeric values like “0” or “1” or “2” for ‘chargeCount’ attribute. The ‘type’ attribute is used to indicate the type of entity with the values, such as “radical”, “radicallion”, “ion”, “hook”, etc. applicable for some of the entities. Each `<electronEntity>` element is associated with a `<electronLink>` element used for structure description [17] as child element. The complete set of `<electronEntity>` elements is wrapped in a `<electronEntityList>` element and stored as a single XML file. The XML encoding of electron entity is presented in Fig. 3.

4. Encoding of AtomEntity

The concept of AtomEntity is a representation of an atom associated with appropriate electron entities to signify the atom along with the valence electrons. A sp^3 hybridized carbon atom entity can be defined with four “hookPair(00)EE” electron entities. Such an atom entity is a representation of a carbon atom with four covalent bonds in the structure. Similarly, a sp^3 hybridized nitrogen atom entity is defined with four EEs of which three are “hookPair(00)EE” and one is “lonePair(00)EE”. In case of sulphur atom, the sp^3 hybridized entity can be defined with nine EEs comprising two “hookPair(00)EE”, two “lonePair(00)EE” and five “vacantPair(00)EE” to include the vacant d -orbital representations. Likewise, a sp^3d^2 hybridized sulphur atom entity is defined with nine EEs composed of six “hookPair(00)EE”, and three “vacant-

Table 4

Some representatives of AtomGroupEntities.

S.No	Entity Name	Structure Representation	Notation
1	Hydroxyl-AGE		O[(1sh)(sb@H)]
2	Carbonyl-AGE		C[(2sh)(db@O)]
3	Amino-AGE		N[(1sh)(sb@H;sb@H)]
4	Carboxyl-AGE		C[(1sh)(db@O;sb@O)]O[(sb@C;sb@H)]
5	Sulphate-AGE		S[(+2)(sb@O;sb@O;sb@O;sb@O)]O[(1sh)(sb@S)]O[(1sh)(sb@S)]
6	Phosphate-AGE		P[(+1)(sb@O;sb@O;sb@O;sb@O)]O[(1sh)(sb@P)]O[(1sh)(sb@P)]O[(1sh)(sb@P)]

Pair(00)EE" indicating the non-availability of both the lone paired electrons. Further, it is also possible to define the atom entities with double bonded and triple bonded status by grouping the hookPair electron entity appropriately. As the single bond status of EE is denoted with a "hookPair(00)EE", a double bond status is represented by a pair of "hookPair(00)EE" entities and named as "doubleHook". Similarly, a triple bond status of atom entity is defined as a combination of three "hookPair(00)EE" entities in the name of "tripleHook".

The combination of appropriate electron entities for any atom is hypothetical. This approach provides the grouping of electron entities of an atom entity in various possibilities to imply a meaningful hybridization status of the atom. The conceptual grouping of EEs for an atom entity has resulted in defining several atom entity instances for the same atom species. In order to identify various EE combinations for a particular atom, an algorithm is developed for generating different species of the atom entities automatically. It is observed that the algorithm is generating several individual instances for every AtomEntity class. Among the instances generated by the system, only a few of them are already known and the others are not known yet. The algorithm generated 26 carbon atom

entities of which only a few of them represents already known stable species of carbon atom. The remaining unknown entities may be considered as unstable. Similarly, for nitrogen atom, sulphur atom and chlorine atom, the system has generated 15, 39 and 52 number of AE instances, respectively. Considering every element in the Periodic Table as atom entities, a total of 6232 atom entity instances are generated. The representative instances of selected atom entities are presented in Table 2.

In the Table 2, the atom entities are detailed with their entity class name, graphical representation, notation and the meaning of the entity. The entity class is indicated by combining a suffix "AE" to the respective atom class like CarbonAE, NitrogenAE, sulphurAE and so on. The notation system created for AtomEntity is composed of three components viz. atom symbol, a notation string for charge status, and another notation string depicting the complete set of electron entities associated to the atom entity. The charge status of any AE is the resultant charge arising from the sum of charges on the electron entities of the AE. The charge status is represented similar to the method described earlier for electron entities.

The third component of the atom entity notation is a string representing all the categories of electron entities associated to the

```

<atomEntity id="AE-N-001" title="NitrogenAE" type="" notation="N(00)(1lp;3sh)" category="" status="">
    <atom id="a1" title="Nitrogen" type="" symbol="N" charge="0" chargeCount="0" hybridization="sp3" geometry="Tetrahedral" x="0" y="0">
        <electronLink id="a1e1" title="s" type="2s" electronStatus="lPair" charge="0" chargeCount="0" linkStatus="" affinity="" bond="" order="" orientation="60" x1="0" y1="0" x2="20" y2="-35"/>
        <electronLink id="a1e2" title="p1" type="2px" electronStatus="uPair" charge="0" chargeCount="0" linkStatus="hook" affinity="" bond="sigma" order="single" orientation="120" x1="0" y1="0" x2="-20" y2="-35"/>
        <electronLink id="a1e3" title="p2" type="2py" electronStatus="uPair" charge="0" chargeCount="0" linkStatus="hook" affinity="" bond="sigma" order="single" orientation="210" x1="0" y1="0" x2="-35" y2="20"/>
        <electronLink id="a1e4" title="p3" type="2pz" electronStatus="uPair" charge="0" chargeCount="0" linkStatus="hook" affinity="" bond="sigma" order="single" orientation="330" x1="0" y1="0" x2="35" y2="20"/>
    </atom>
</atomEntity>

```

Fig. 4. XML encoding of NitrogenAE.

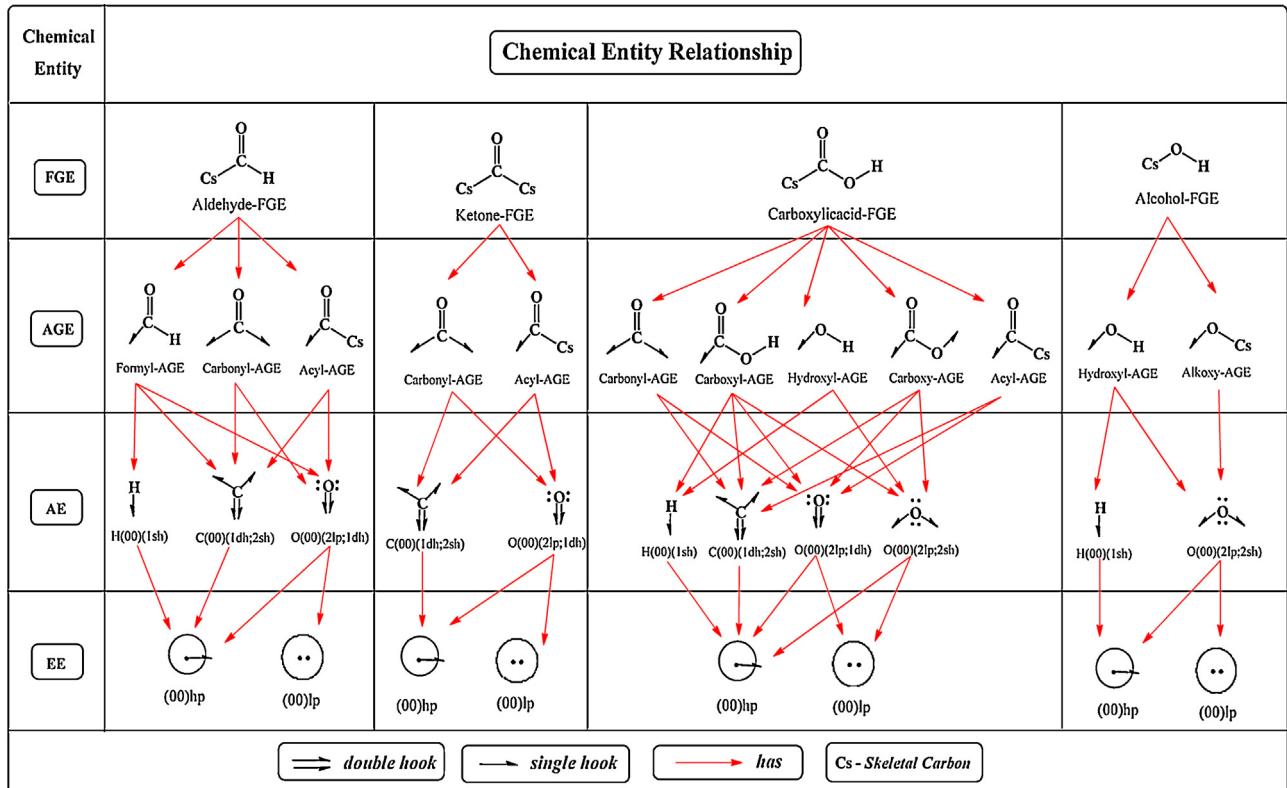


Fig. 5. Relationship of AtomGroupEntity with AtomEntity and include FunctionalGroupEntity.

```

<atomGroupEntity>
    <definition id="AGE-1Gc-005" title="" type="OneGc" gcCount="1" pattern="Point" notation="O[(1sh)(sb@H)]" category="" status="">
        <gcAtom id="" title="Oxygen" type="Gc1" symbol="O" charge="0" chargeCount="0" gcCode="(1sh)(sb@H)"/>
    </definition>
    <atom id="" title="Oxygen" type="Gc1" symbol="O" charge="0" chargeCount="0" gcCode="(1sh)(sb@H)" x="0" y="0">
        <electronLink id="" type="" electronStatus="lPair" charge="0" chargeCount="0" bond="" order="" priority="" gcCode="" orientation="" x1="0" y1="0" x2="0" y2="0"/>
        <electronLink id="" type="" electronStatus="lPair" charge="0" chargeCount="0" bond="" order="" priority="" gcCode="" orientation="" x1="0" y1="0" x2="0" y2="0"/>
        <electronLink id="" type="hook" electronStatus="uPair" charge="0" chargeCount="0" bond="sigma" order="single" priority="" gcCode="" orientation="210" x1="0" y1="0" x2="-35" y2="20"/>
        <electronLink id="" type="bond" electronStatus="bPair" charge="0" chargeCount="0" bond="sigma" order="single" priority="1.1" gcCode="sb@H" orientation="330" x1="0" y1="0" x2="35" y2="20"/>
    </atom>
    <atom id="" title="Hydrogen" type="Ta" symbol="H" charge="0" chargeCount="0" gcCode="(sb@O)" x="35" y="20">
        <electronLink id="" type="bond" electronStatus="bPair" charge="0" chargeCount="0" bond="sigma" order="single" priority="1.8" gcCode="sb@O" orientation="150" x1="35" y1="20" x2="0" y2="0"/>
    </atom>
</atomGroupEntity>

```

Fig. 6. XML encoding of Hydroxyl-AGE.

```

<ChemicalEntity>
    <MolecularEntity>
        <FunctionalGroupEntity>
            <OneGcFGEEntity>
                <Instance-FGE-1Gc-Aldehyde-FGE/>
                <Instance-FGE-1Gc-Alcohol-FGE/>.....
                <Instance-FGE-1Gc-AcylBromide-FGE/>.....
            <TwoGcFGEEntity>
                <Instance-FGE-2Gc-Azide-FGE/>
                <Instance-FGE-2Gc-Aldimine-FGE/>
                <Instance-FGE-2Gc-AcylCyanide-FGE/>.....
            <ThreeGcFGEEntity>
                <Instance-FGE-3Gc-Aldoxime-FGE/>
                <Instance-FGE-3Gc-AldHydrazone-FGE/>
                <Instance-FGE-3Gc-Acetal-FGE/>.....
            <FourGcFGEEntity>
                <Instance-FGE-4Gc-DiacylPeroxide-FGE/>
                <Instance-FGE-4Gc-CarboxylicacidAmideAcetal-FGE/>
                <Instance-FGE-4Gc-CarboxylicacidAmidrazone-FGE/>.....
            <FiveGcFGEEntity>
                <Instance-FGE-5Gc-OrthoCarbonateEster-FGE/>
                <Instance-FGE-5Gc-Thiosemicarbazone-FGE/>
                <Instance-FGE-5Gc-Semicarbazone-FGE/>.....
        <AtomGroupEntity>
            <OneGcAGEEntity>
                <Instance-AGE-1Gc-001/>
                <Instance-AGE-1Gc-002/>
                <Instance-AGE-1Gc-003/>.....
            <TwoGcAGEEntity>
                <Instance-AGE-2Gc-001/>
                <Instance-AGE-2Gc-002/>
                <Instance-AGE-2Gc-003/>.....
            <ThreeGcAGEEntity>
                <Instance-AGE-3Gc-001/>
                <Instance-AGE-3Gc-002/>
                <Instance-AGE-3Gc-003/>.....
            <FourGcAGEEntity>
                <Instance-AGE-4Gc-001/>
                <Instance-AGE-4Gc-002/>
                <Instance-AGE-4Gc-003/>.....
            <FiveGcAGEEntity>
                <Instance-AGE-5Gc-001/>
                <Instance-AGE-5Gc-002/>
                <Instance-AGE-5Gc-003/>.....
        <AtomEntity>
            <Group1AEntity>
                <HydrogenAE>
                    <Instance-AE-H-001>
                    <Instance-AE-H-002>
                    <Instance-AE-H-003>
                    <LithiumAE>.....
                    <SodiumAE>.....
                <Group2AEntity>.....
                <Group3AEntity>.....
                <Group4AEntity>.....
                <Group5AEntity>.....
                <Group6AEntity>.....
                <Group7AEntity>.....
                <Group8AEntity>.....
                <Group9AEntity>.....
                <Group10AEntity>.....
                <Group11AEntity>.....
                <Group12AEntity>.....
                <Group13AEntity>.....
                <Group14AEntity>.....
                <Group15AEntity>.....
                <Group16AEntity>.....
                <Group17AEntity>.....
                <Group18AEntity>.....
                <GroupLanthanideAEntity>.....
                <GroupActinideAEntity>.....
            <ElectronEntity>
                <Instance-EE-001/>
                <Instance-EE-002/>
                <Instance-EE-003/>.....

```

Fig. 7. Part of Concept Taxonomy in ChemEnt Ontology.

Entity Table					
Atom Entity	Atom Group Entity	Functional Group Entity	Electron Entity		
Entity Id: EE-001 Notation : (00)lp	Entity Id: EE-002 Notation : (-1)lp	Entity Id: EE-003 Notation : (00)up	Entity Id: EE-004 Notation : (-1)up	Entity Id: EE-005 Notation : (+1)up	Entity Id: EE-006 Notation : (00)hp
(00)lonePairEE	(-1)lonePairEE	(00)unPairEE	(-1)unPairEE	(+1)unPairEE	(00)hookPairEE
Entity Id: EE-007 Notation : (+1)hp	Entity Id: EE-008 Notation : (-1)hp	Entity Id: EE-009 Notation : (+6)hp	Entity Id: EE-010 Notation : (-6)hp	Entity Id: EE-011 Notation : (00)vp	Entity Id: EE-012 Notation : (+1)vp
(+1)hookPairEE	(-1)hookPairEE	(+6)hookPairEE	(-6)hookPairEE	(00)vacantPairEE	(+1)vacantPairEE

Fig. 8. Entity Table showing ElectronEntities.

Entity Table					
Atom Entity	Atom Group Entity	Functional Group Entity	Electron Entity		

Fig. 9. Entity Table showing AtomEntities.

atom. There are four categories of electron entities namely, lonePair, unPair, hookPair and vacantPair. The strings to denote these electron entities are "lp", "up", "hp" and "vp", respectively. There are three possibilities for hookPair combinations namely singleHook, doubleHook and tripleHook. The notations "sh", "dh" and "th" are used to denote them, respectively. Accordingly, an atom entity defined with a singleHook (sh) indicates that the atom is bonded with some other atom through a single bond in a chemical structure. Similarly, a doubleHook (dh) defined with an atom entity indicates the atom represented by the entity is bonded to another atom through a double bond in a chemical structure. So, the tripleHook (th) is a representation of a triple bond. The notation string is created by prefixing the count of each type of electron entities. For example, three lonePair EE can be denoted as '3lp'; two singleHook as '2sh'; one doubleHook as '1dh'; and so on. These notation bits are grouped in a common order separated by ";" like 'lp'; 'up'; 'vp'; 'th';

'dh'; 'sh' with the respective counts as prefixes. The complete notation of an AtomEntity is fixed by combining the three components as *atomSymbol* (*notation for charge status*) (*notation for electron EE*). For example, the notation of a carbon atom entity, C(00)(1dh;2sh) means a neutral carbon atom is associated with one doubleHook and two singleHook electron entities. The notation of a chlorine atom entity, Cl(00)(3lp;5vp;1sh) indicates that the entity charge is neutral and is associated with three lonePair, five vacantPair and one singleHook electron entities.

Encoding of atom entity in XML is performed by the ChemEd system automatically as per the direction of an algorithm. The XML construct is similar to the structure description frame reported [17]. Accordingly, an atom entity is encoded with an *<atom>* element and appropriate number of *<electronLink>* child elements. Then the *<atom>* element is placed inside an *<atomEntity>* element with necessary semantics. The XML description of a NitrogenAE is shown

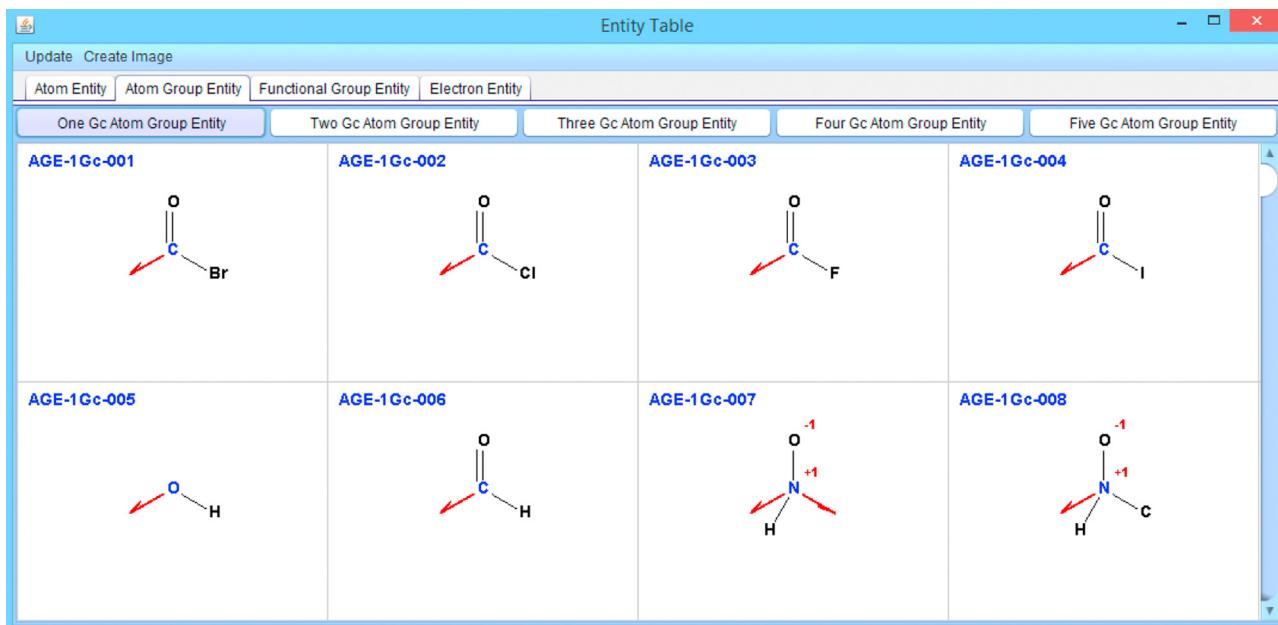


Fig. 10. Entity Table showing AtomGroupEntities.

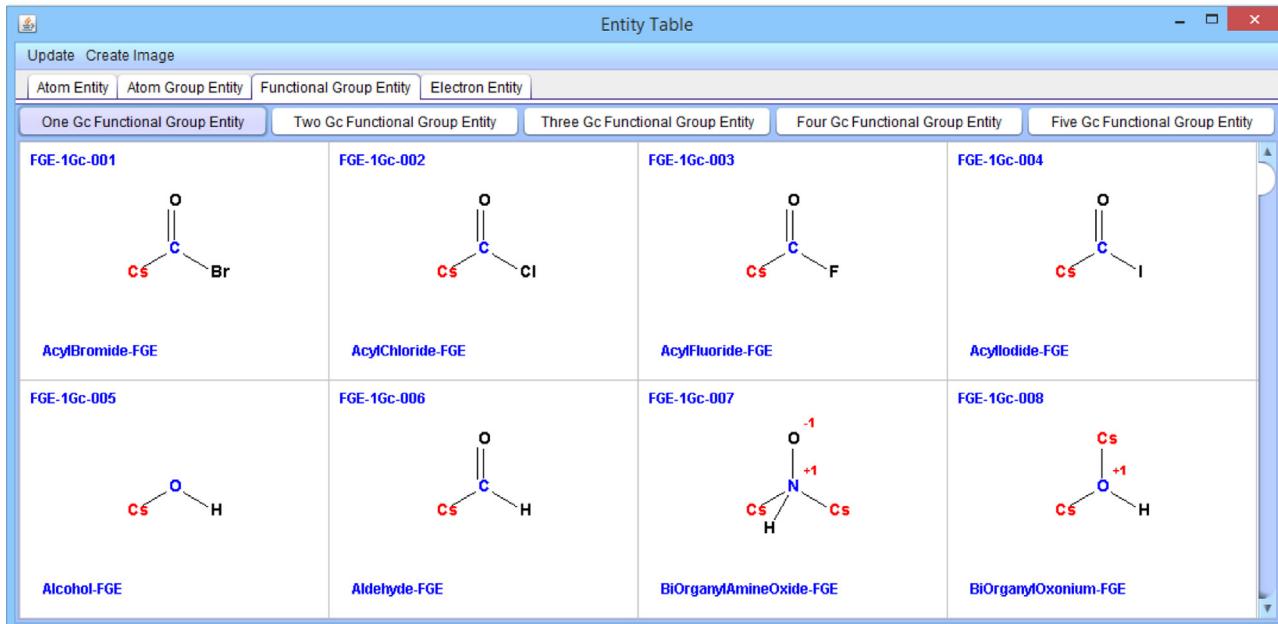


Fig. 11. Entity Table showing FunctionalGroupEntities.

in the Fig. 4 as an example. The NitrogenAE is described with a sp^3 hybridized tetrahedral nitrogen atom associated with one lonePair <electronLink> and three hookPair “electronLink” elements as evidenced from the ‘electronStatus’ and ‘linkStatus’ attributes shown in Fig. 4.

5. Encoding of FunctionalGroupEntity and AtomGroupEntity

All the functional group definitions described earlier [18] are considered as individual functional group entity instances in this study. Accordingly, the notations of functional groups become the notations of the functional group entities. Some representative

functional group entities are shown in Table 3. An atom group entity is perceived as a combination of one or more atom entities representing a characteristic chemical group like chloro group, formyl group, acyl group, carbonyl group, hydroxyl group, etc. The combinations of atom entities can be effected through the hookPairs suitably. A carbonyl group can be created with a carbon atom entity and an oxygen atom entity with notations C(00)(1dh;2sh) and O(00)(2lp;1dh), respectively. In this process, the doubleHooks of both the atom entities are used for the conversion of a double bond between the atom entities. The two singleHook EE are still left with the carbon atom entity. Similarly, a hydroxyl group is a combination of a hydrogen atom entity and an oxygen atom entity with the notations H(00)(1sh) and O(00)(2lp;2sh), respec-

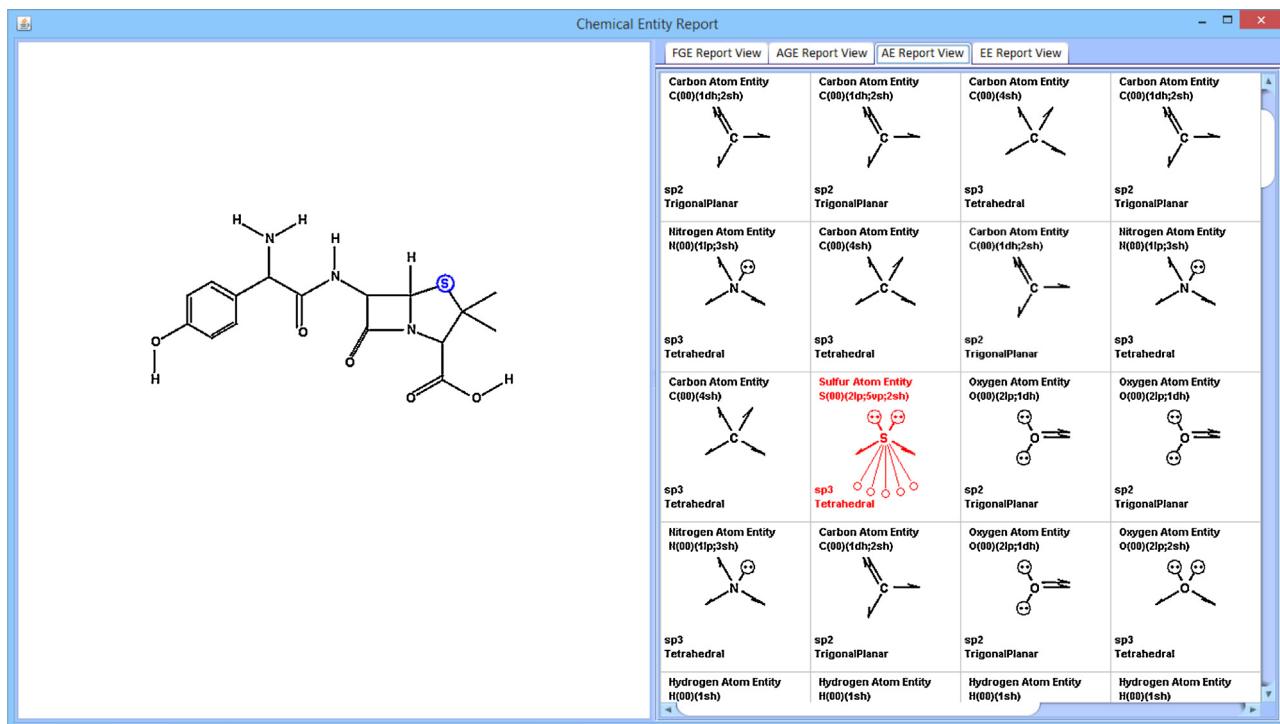


Fig. 12. Chemical Entity Detection Report on ChemEd editor for Amoxicillin molecule.

tively. This atom group entity is formed by a single bond between the atom using the singleHook of Hydrogen atom entity and one of the singleHook of oxygen atom entity leaving the other one for further combination.

It is to be noted that the atom group entities are described with at least one singleHook entity. When the single hook in atom group entity is replaced with a bond pair entity connecting a carbon atom in a structure, the resultant entity is a functional group entity. The carbon atom to which the AGE is connected in the structure is termed as "skeletal carbon" (Cs) [18]. This implies that the atom group entities can be automatically defined from the definitions of functional group definitions. An algorithm is developed to create the atom group entity definitions from the definitions developed for functional group definitions [18]. A total of 204 number of atom group entities are generated by the system. The relationship among the chemical entities starting from electron entities to functional group entities is depicted in Fig. 5.

The encoding structure of atom group entity is packed inside a <atomGroupEntity> element composed of one <definition> element and one or more <atom> elements with appropriate number of <electronLink> elements. The <definition> element is described with the attributes id, title, type, gcCount, pattern, notation, etc. and is associated with one or more <gcAtom> elements [18] along with the necessary attributes. For example, the XML description of a hydroxyl atom group entity is shown in the Fig. 6. The hydroxyl atom group entity is described with a definition part with attributes 'id' to hold unique alpha numeric identification value to be mapped with the ChemEnt ontology. The 'type' attribute is used to indicate the group centre type and the 'gcCount' captures the number of group centre (Gc) atoms [18]. The 'pattern' attribute stores the group centre pattern already explained in FGE description [18]. The attribute 'notation' captures the unique notation of the AGE. The unique atom group entity notation is also generated similar to that developed for functional groups [18]. In case of atom group entity, the notation changes only in the indication of a singleHook in place of skeletal carbon of the functional group notation. Some of the representative atom group entities along with

the graphical representations and their notations are presented in Table 4.

6. Entity Table for ChemEd structure editor

Since there are only twelve electron entities, no further classification is required for the electron entities. The atom entities, atom group entities and functional group entities are classified further into appropriate classes. The individual instances of respective entity classes are organized in the ChemEnt ontology. A part of the concept taxonomy is shown in Fig. 7. A GUI named as Entity Table comprising graphical representations of all the chemical entities is developed and programmed as a functionality of ChemEd structure editor. This is a facility to view the graphical representation of chemical entities on screen and to be used for reactivity modelling. The Entity Table developed for each category is presented in the Figs. 8–11.

7. MolecularEntity

The concept of MolecularEntity is included in the chemical entity organization in the top level in order to consider any molecular structure described in XML through ChemEd. Accordingly, it is possible to detect the functional group entities, atom group entities, atom entities and their electron entities in the structural description of any molecule. As the detection of functional groups is already demonstrated [18], new algorithms are developed to detect the atom group entities inside any functional group. Subsequently, the detection of atom entities inside the atom group entities followed by the electron entities inside atom entities is achieved. For example, the fundamental chemical entities detected in the structure of Amoxicillin [27] molecule are shown in Fig. 12. The report window has the provision to view the detected FGE, AGE, AE and EE instances on separate panels. The system is developed in such a way to locate any of the detected entities on the structure by a selection on the entity table.

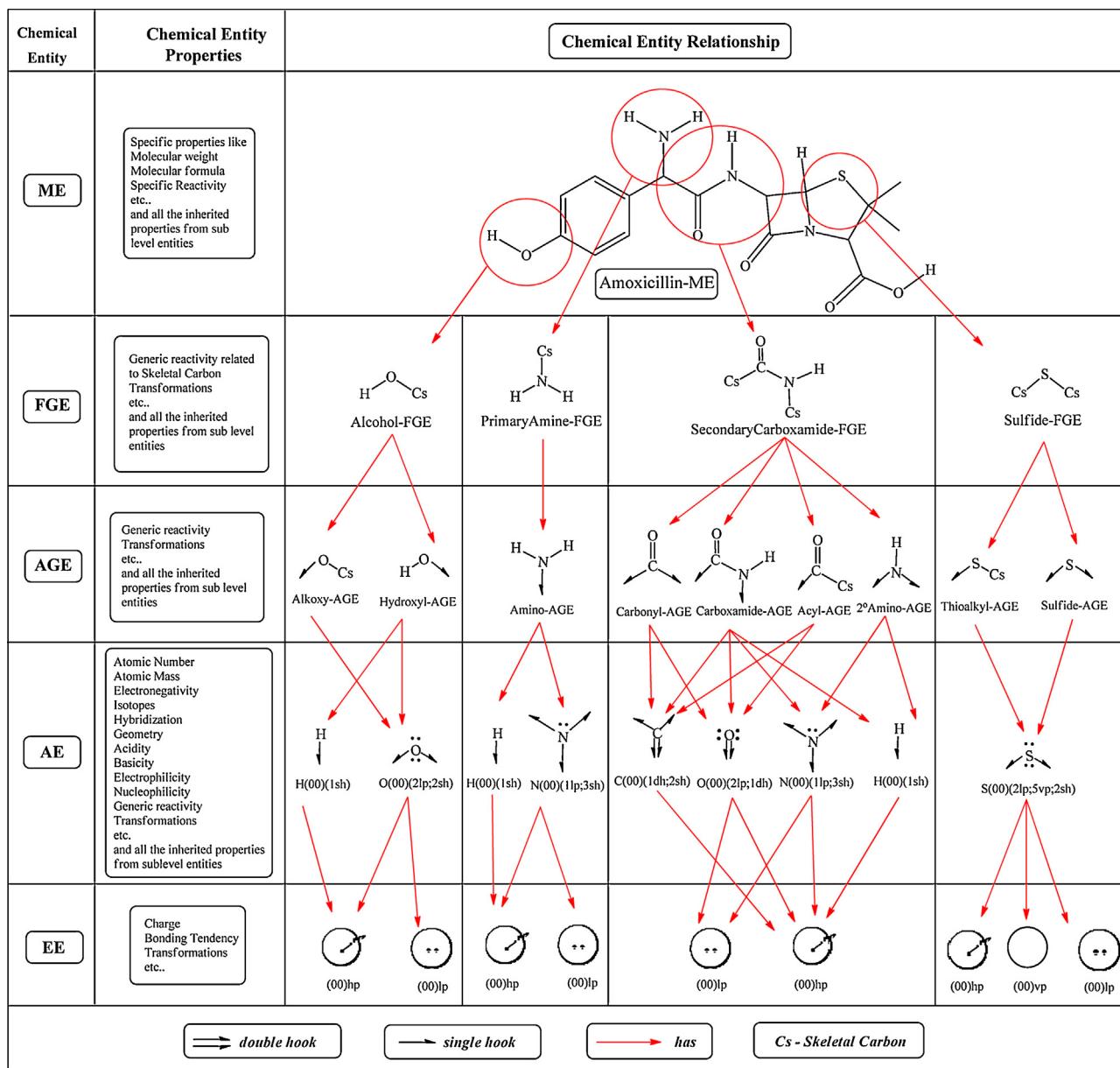


Fig. 13. Property flow in fundamental Chemical Entities.

8. Property assertion and inheritance of chemical entities

The main objective of developing the fundamental chemical entities is to demonstrate the possibility of asserting suitable properties to the entities at different levels and to inherit them appropriately in molecular structures. It is also possible to bring property restrictions or modifications when one entity is transforming into another. If a lonePair entity is used for the formation of a coordinate covalent bond, the asserted properties also changes accordingly. During this change the notation changes from "(00) lonePairEE" to "(+1) hookPairEE" and so the properties too. The property inheritance is also possible from an atom entity to an atom group entity and subsequently to functional group entity. Finally, the native and inherited properties of the chemical entities become the properties of a molecular entity. Accordingly, the asserted property of an electron entity can be detected on EE, AGE, FGE and ultimately on ME. Similarly, the properties of AE can be

carried into AGE and subsequently into FGE. Ultimately, the properties of all sub-level entities, such as FGE, AGE, AE and EE can be detected as the properties of molecular entity.

The approach of packing the semantics of entities at different levels enables the property flow from the bottom level entity, EE to the top level entity ME. Further, the inheritance of properties through hierarchical entity taxonomy brings the possibility to process the properties of entities in appropriate levels with respect to some context. It is a useful approach to arrive at some meaningful descriptors to detail the organic reactivity. For example, property of a lone pair to form an *onium ion*, once defined for a suitable ElectronEntity, can be detected on ME without assigning the property to the molecular entity level separately. It is also possible to modify the property at different entity levels based on the nature of other entities in each level. The mechanism of property flow along the fundamental chemical entities is shown in Fig. 13.

9. Conclusion

The definitions of fundamental chemical structural entities encoded in this study may be considered as building blocks possessing the generic properties organized in a common platform. The system allows any new properties to be associated further consensually by the community through ontology. The property changes during the transformation of one entity into another provide the automatic inference of property transformation. This could be used to build knowledge based chemical information systems. Unlike the reaction database search retrieval, the proposed approach is useful to develop reactivity prediction systems which can provide precise information. The approach is extendable to describe the supramolecules which are formed by the association of two or more chemical species [28] through atom entity relations. It may be associated in virtual screening technique [29] and for the evolving proposals in academia [30]. The ontology supported encoding and detection system in XML media is suitable for the development of intelligent open source applications for the evolving Semantic Web [31–33].

Acknowledgment

The authors P.S. and D.V. acknowledge the financial support by the Department of Science & Technology (DST), New Delhi, India (Project No. SR/S1/OC-86/2009).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jmgm.2015.06.001>

References

- [1] J. March, Advanced Organic Chemistry: Reactions, Mechanisms, and Structure, third ed., Wiley, New York, 1985.
- [2] J. Clayden, N. Greeves, S. Warren, P. Wothers, Organic Chemistry, New York, Oxford, 2001.
- [3] L.G. Wade Jr., Organic Chemistry, fifth ed., Pearson Education, Delhi, 2004.
- [4] M.G. Moloney, Structure and Reactivity in Organic Chemistry, first ed., Wiley-Blackwell, 2008.
- [5] K. Vollhardt, C. Peter, N.E. Schore, Organic Chemistry: Structure and Function, W.H. Freeman and Company, New York, 2007.
- [6] WIKIPEDIA, Hydroxyl Group 2015; <<http://en.wikipedia.org/wiki/Hydroxyl>>.
- [7] WIKIPEDIA, Functional Group, <http://en.wikipedia.org/wiki/Functional_group>.
- [8] IUPAC, Compendium of Chemical Terminology, 2nd ed. (the Gold Book). Compiled by A.D. McNaught, A. Wilkinson. Blackwell Scientific Publications, Oxford (1997). XML on-line corrected version: <http://goldbook.iupac.org> (2006) created by M. Nic, J. Jirat, B. Kosata; updates compiled by A. Jenkins. ISBN 0-9678550-9-8. <http://dx.doi.org/10.1351/goldbook>
- [9] L.L. Thomas, Review of Organic Functional Groups: Introduction to Medicinal Organic Chemistry, fourth ed., Lippincott Williams & Wilkins, Baltimore, MD, Philadelphia, PA, 2003.
- [10] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Isbel-Tarver, A. Kasarskis, S. Lewis, J.C. Mateo, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene ontology: tool for the unification of biology. The gene ontology consortium, Nat. Genet. 25 (2000) 25–29.
- [11] C. Brooksbank, G. Cameron, J. Thornton, The European bioinformatics institutes data resources: towards systems biology, Nucleic Acids Res. 33 (2005) D46–D53.
- [12] J. Hastings, D. Magka, C. Batchelor, L. Duan, R. Stevens, M. Ennis, C. Steinbeck, Structure-based classification and ontology in chemistry, J. Cheminf. 4 (2012) 8.
- [13] J. Hastings, L. Chepelev, E. Willighagen, N. Adams, C. Steinbeck, M. Dumontier, The chemical information ontology: provenance and disambiguation for chemical data on the biological semantic web, PLoS One 6 (10) (2011) e25513, <http://dx.doi.org/10.1371/journal.pone.0025513>
- [14] P. Sankar, G. Aghila, Design and development of chemical ontologies for reaction representation, J. Chem. Inf. Model. 46 (2006) 2355–2368.
- [15] P. Sankar, G. Aghila, Ontology aided modeling of organic reaction mechanisms with flexible and fragment based XML markup procedures, J. Chem. Inf. Model. 47 (2007) 1747–1762.
- [16] M. Kotera, A.G. McDonald, S. Boyce, K.F. Tipton, Functional group and substructure searching as a tool in metabolomics, PLoS One 2 (2008) e1537.
- [17] P. Sankar, A. Krief, G. Aghila, Model tool to describe chemical structures in XML format utilizing structural fragments and chemical ontology, J. Chem. Inf. Model. 50 (2012) 755–770.
- [18] P. Sankar, A. Krief, D. Vijayasarathi, A conceptual basis to encode and detect organic functional groups in XML, J. Mol. Graphics Model. 43 (2013) 1–10.
- [19] D. Magka, M. Krötzsch, I. Horrocks, A rule-based ontological framework for the classification of molecules, J. Biomed. Semant. 15 (5) (2014) 17, <http://dx.doi.org/10.1186/2041-1480-5-17>
- [20] C. Bobach, T. Böhme, U. Laube, A. Püschel, L. Weber, Automated compound classification using a chemical ontology, J. Cheminf. 29 (4) (2012) 40, <http://dx.doi.org/10.1186/1758-2946-4-40>
- [21] W3C, OWL Working Group, <<http://www.w3.org/2007/OWL/wiki/OWLWorkingGroup>>.
- [22] (a) WIKIPEDIA, ChEBI Home Page, <<http://www.ebi.ac.uk/chebi/>>;
 (b) K. Degtyarenko, M. de Matos, J. Ennis, M. Hastings, A. Zbinden, M. Alcántara, M.M. Guedj Ashburner, ChEBI: a database and ontology for chemical entities of biological interest, Nucleic Acids Res. 36 (2008) D344–D350, <http://dx.doi.org/10.1093/nar/gkm791> (Database issue);
 (c) R. de Matos, A. Alcantara, M. Dekker, J. Ennis, K. Hastings, I. Haug, S. Turner, C. Steinbeck, Chemical entities of biological interest: an update, Nucleic Acids Res. 38 (2010) D249–D254, <http://dx.doi.org/10.1093/nar/gkp886> (Database issue);
 (d) N. Hill, M. Adams, C. Bada, T.Z. Batchelor, H. Berardini, H.J. Dietze, M. Drabkin, R.F. Ennis, M.A. Foulger, J. Harris, N.S. Hastings, C.J. de Matos, G. Mungall, P. Owen, C. Roncaglia, S. Turner, J. Lomax, Dovetailing biology and chemistry: integrating the Gene Ontology with the ChEBI chemical ontology, BMC Genomics 29 (2013) 14:513, <http://dx.doi.org/10.1186/1471-2164-14-513>
- [23] (a) Data Base, Glycomics Ontology, <<http://bioportal.bioontology.org/ontologies/GLYC0>>;
 (b) W.S. York, A. Sheth, K. Kochut, J.A. Miller, C. Thomas, K. Gomadam, X. Yi, M. Nagarajan, Semantic Integration of Glycomics Data and Information, Human Disease Glycomics/Proteome Initiative 1st Workshop, Functional Glycomics in Disease, Osaka, Japan, 2004.
- [24] (a) WIKIPEDIA, ChEMBL, <<http://en.wikipedia.org/wiki/ChEMBL>>;
 (b) L.J. Gaulton, A.P. Bellis, J. Bento, M. Chambers, A. Davies, Y. Hersey, S. Light, D. McGlinchey, B. Al-Lazikani, J.P. Overington, ChEMBL: a large-scale bioactivity database for drug discovery, Nucleic Acids Res. 40 (2012), <http://dx.doi.org/10.1093/nar/gkr777> (Database issue).
- [25] L.L. Chepelev, M. Dumontier, Chemical entity semantic specification: knowledge representation for efficient semantic cheminformatics and facile data integration, J. Cheminf. 3 (2011) 20.
- [26] W3C, Extensible Markup Language (XML), <<http://www.w3.org/XML>>.
- [27] ChemsSpider, Amoxicillin Structure, CSID. 31006, <<http://www.chemspider.com/Chemical-Structure.31006.html>>
- [28] Jean-Marie Lehn, Supramolecular Chemistry-Scope and Perspectives Molecules – Supermolecules – Molecular Devices. Nobel Lecture, Institut Le Bel, Université Louis Pasteur, 4, rue Blaise Pascal, 67000 Strasbourg and Collège de France, 11 Place Marcelin Berthelot, 75,005 Paris, December 8 (1987).
- [29] P. Badrinarayanan, G.N. Sastry, Virtual high throughput screening in new lead identification, Comb. Chem. High Throughput Screening 14 (2015) 840–860.
- [30] (a) A. Krief, J. Castillo-Colau, C. Juilliard, S. Peraldi, Building the EnCorE dictionary collaboratively: strategy and practice, Int. J. Knowl. Learn. 4 (2/3) (2008);
 (b) A. Krief, P. Sankar, The hook-concept and its application for functional groups and groups name-to-structure disambiguation, (Oral Presentation), in: 44th World Chemistry Congress, IUPAC, Istanbul 11–16 August, 2013.
- [31] W3C SEMANTIC WEB, <<http://www.w3.org/standards/semanticweb/>>.
- [32] L.L. Chepelev, M. Dumontier, Semantic web integration of cheminformatics resources with the SADI framework, J. Cheminf. 3 (2011) 16, <http://dx.doi.org/10.1186/1758-2946-3-16>
- [33] L.L. Chepelev, J. Hastings, M. Ennis, C. Steinbeck, M. Dumontier, Self-organizing ontology of biochemically relevant small molecules, BMC Bioinf. 6 (2012) 3, <http://dx.doi.org/10.1186/1471-2105-13-3>.