Generating Qualitative Causal Graph using Modeling Constructs of Qualitative Process Theory for Explaining Organic Chemistry Reactions

Alicia Y. C. Tang, Rukaini Abdullah, Sharifuddin M. Zain, and Noorsaadah A. Rahman

Abstract—This paper discusses the causal explanation capability of QRIOM, a tool aimed at supporting learning of organic chemistry reactions. The development of the tool is based on the hybrid use of Qualitative Reasoning (QR) technique and Qualitative Process Theory (QPT) ontology. Our simulation combines symbolic, qualitative description of relations with quantity analysis to generate causal graphs. The pedagogy embedded in the simulator is to both simulate and explain organic reactions. Qualitative reasoning through a causal chain will be presented to explain the overall changes made on the substrate; from initial substrate until the production of final outputs. Several uses of the QPT modeling constructs in supporting behavioral and causal explanation during run-time will also be demonstrated. Explaining organic reactions through causal graph trace can help improve the reasoning ability of learners in that their conceptual understanding of the subject is nurtured.

Keywords—Qualitative reasoning, causal graph, organic reactions, explanation, QPT, modeling constructs.

I. INTRODUCTION

QUALITATIVE Reasoning (QR) formalisms provide the expressive power needed to capture the intuitive and causal notions of many human mental models [1]. One of the goals of QR research has been to understand human-like commonsense reasoning without making use of the preciseness of models that consists of differential algebraic equations, and parameters that are real-valued numbers. QR is opposed to traditional problem solving techniques such as those found in numerical simulators [2]. It has been a challenge to researcher to develop techniques and algorithms that are able to generate explanation about an aspect of the problem being asked. QR approach has the potential to address such needs. Systems developed using qualitative

Y. C. Alicia Tang is with the University of Tenaga Nasional, Selangor, Malaysia (phone: 603-8921-2336; e-mail: aliciat@uniten.edu.my).
R. Abdullah is with the Department of Artificial Intelligence, Malaya University, Kuala Lumpur, Malaysia (email: rukaini@um.edu.my)
S. M. Zain is with the Department of Chemistry, Malaya University, Kuala Lumpur, Malaysia (e-mail: smzain@um.edu.my).
N. A. Rahman is with the Department of Chemistry, Malaya University, Kuala Lumpur, Malaysia (email: noorsaadah@um.edu.my).

reasoning include CyclePad [3], VisiGarp [4], QALSIC [5]-[6], Betty's Brain [7], and complex controllers [8, 9]. Earlier discussion on QR applications mainly placed on physic and engineering. Work on QR simulator in the domain of organic chemistry has not been recorded in available literature. LHASA [10] is one of first computer programs developed for planning of organic syntheses. It is an expert system using database of retro-reactions. There are some associated problems in this approach. First, it is time consuming to prepare the long-range transforms. Second, the program could easily give cumbersome plans for molecules that contained unusual or unforeseen combinations of functional groups. Furthermore, the system cannot generate causal explanation on-the-fly. Other computer systems that have been developed for solving chemistry problems include toxicological systems, structure elucidation, and reaction mechanism analysis. The techniques used in these systems are Self-Organizing Map (SOM), Neural Networks, and Genetic Algorithms [11]. None of the above addresses the issue from the stand point of qualitative reasoning. The work described in this paper is the first use of qualitative reasoning based on Qualitative Process Theory (QPT) [12] in organic reactions modeling, simulation, and explanation.

This work was initiated to find new approaches to develop a learning tool that can help the first year students of a chemistry course at the University of Malaya. The subject is organic chemistry with particular emphasis on organic reactions and mechanisms. The QPT ontology provides the means to embody notions of causality which is important to explain behavior of physical systems. Using QPT, organic reactions can be modeled much like the way a chemist would do in his problem solving, i.e. at intuitive level. Organic reactions involve the study of electrons movement, in which a bond is being made or broken. Much of what is taught in chemistry classes consists of causal theories of chemical phenomena: what happen, what affects it, and what does it affect, etc. The nature of this class of problem is highly qualitative, as presented in [13]. For example, in a given reaction, it is to determine which electrons will start moving in trying to break or form a bond in a molecule. In this work, the functional dependencies among the chemical parameters

(covalent bonds, charges, etc.) of objects (substrates, reagents, nucleophiles, electrophiles, etc.) are represented using the ontological modeling constructs of QPT. This paper will discuss the modus operandi of two software modules of QRIOM (a simulator prototype), namely the reasoning and the explanation engines that help generate explanation and the prediction of products. Explaining the behavioral aspects of a reaction via causal reasoning is an innovation in this work, which is also the focus of this paper.

II. OBJECTIVES

In science education, it is believed that students should understand the qualitative principles that govern the subject, including the causal relationships in processes before they are immersed in complex problem solving. Most of the students learn organic chemistry reactions by memorizing the steps involve in a reaction, and the formulas taught in classes. They faced problem in understanding the principles governing the processes and the cause effect interaction among these processes. As such, some students, particularly weak learners would require additional learning aids such as a software tool to assist them in their learning. Traditional approach for developing science educational software is facing shortcomings, in that the conceptual understanding can only be found in program's document and not in the software itself (e.g. a numerical simulator). The lack of tight coupling between concepts and their embodiment makes it difficult to explain itself [14]. Qualitative reasoning approach addresses this issue to a promising extent. We implemented qualitative representation and qualitative simulation approaches since the problem domain is qualitative in nature. The main educational goals of applying the QR approach to software development are that the approach provides the expressive power needed to capture the intuitive and causal notions of a chemist's mental model. The models people use in reasoning about physical world are called mental models [15]. QRIOM will be used in complementary fashion; as in-class demonstrations and for students to run and interact with the system. Many educators quest for improving instructions using innovative technology. This paper focuses on the programming logic used in generating causal graphs and the interpretation of simulated results.

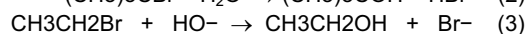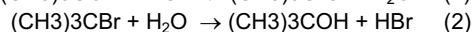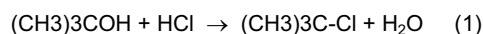## III. PROBLEM DOMAIN AND KNOWLEDGE MODELING LANGUAGE

A. Domain Knowledge Modeling

QPT is a process-based ontology developed by Forbus and his group. In this work, it is used as the language for representing chemical theories and the chemical knowledge possessed by a human chemist. QPT provides a number of modeling constructs (design primitives) for representing the notion of processes needed in expressing the chemical reaction steps. Its modeling ingredients are given as below. A structural description of the model is given by a set of *individual views* and *processes*. The *individual views* describe objects and their general characteristics while the *processes*

support changes in system behavior. A *process* is described by five slots (much like a frame-based structure): *Individuals, Preconditions, Quantity-conditions, Relations, and Direct Influences*. In this formalism, changes are caused by continuous physical processes (such as creating and deleting bonds). These changes propagate through the system via *qualitative proportionalities* which indicate causal relationships between quantities. *Quantity-conditions* contain inequalities involving *quantities* (object's characteristics, e.g. charges, and number of covalent bond), which is crucial in determining the status of a process (being active/inactive). *Relations* slot contains statements about functional dependencies (these are the indirect influences) among *quantities*. An important modeling construct for describing the relationships between *quantities* is the *qualitative proportionalities* (denoted by P+/P-) that propagate the effects of processes that express unknown monotonic functions (increasing/decreasing/unchanged) between two *quantities*. *Quantities* represent the features of entities and agents that change during simulation. A quantity has a magnitude and a derivative, which represent its current value and trend. In this work, the main concepts of the organic reactions involving nucleophiles (electron-rich) and electrophiles (electron-poor) are modeled using the direct and indirect influences modeling construct. The notion of direct effect is used to express the dynamic aspect of a process, represented in the slot called *Direct Influences* (as either I+ or I-). The possibility of encoding declarative knowledge and the explicit representation of direct and indirect influences are characteristics that warrant the ontology to be used in chemistry system modeling for educational purposes.

B. Organic Reactions

An organic reaction is a chemical reaction involving organic compounds, usually between an electrophilic centre and a nucleophilic group. Families of organic compounds are characterized by the presence of distinctive functional groups. Functional groups are the structural units responsible for a given molecule's chemical reactivity. In any chemical reaction, some bonds are broken and new bonds are made. Often, these changes are too complicated to happen in one simple stage. Thus, usually a reaction may involve a series of small changes one after the other. Organic chemists will identify the electron-poor (electrophile) site and electron-rich (nucleophile) group when trying to work out a reaction mechanism through their chemical intuition, knowledge and experience developed. We have tested several reaction formulas on the simulator. Examples of reaction formula tested are (1), (2) and (3). The chemical compounds on the left are the inputs (a substrate and a reagent), and the final products are on the right side of the arrows.

$$(CH3)3COH + HCl \rightarrow (CH3)3C\text{-}Cl + H_2O \quad (1)$$
$$(CH3)3CBr + H_2O \rightarrow (CH3)3COH + HBr \quad (2)$$
$$CH3CH2Br + HO^- \rightarrow CH3CH2OH + Br^- \quad (3)$$

In this paper only (1) will be used to exemplify the behaviour simulation of the organic reactions, and their explanation. For each equation, there are many cognitive steps (see Fig. 1) leading from substrate and reagent (inputs) to final products (outputs). From an in interview conducted, many chemistry students are lacking the reasoning ability to translate reaction formulas into the intermediate steps (and structures) necessary for predicting the outputs. They have problems in seeing the chemical intuition implicit in the above equations which include the main (essential) chemical parameters, a number of relationships among the parameters, and the chemical theories and general principles that govern the behaviours of the concerned reacting species. Noticeably, all these could not be visualized in labs, and could be difficult to visualize in classrooms. QRIOM that uses QR and QPT techniques is able to assist students in articulating these concepts.

To benefit readers from non-chemistry background, (1) is divided into a series of small step. In which, in the first stage, the alcohol oxygen (the "O" from the "OH" group) is protonated. That is, the "O" captures a proton (Step 1). In Step 2, the bond between the tertiary carbon and the alcohol oxygen will break, and this produces a carbocation intermediate. In the last step, the incoming nucleophile (Cl^-) can form a bond to the carbocation to produce a neutral and stable final product (Step 3).
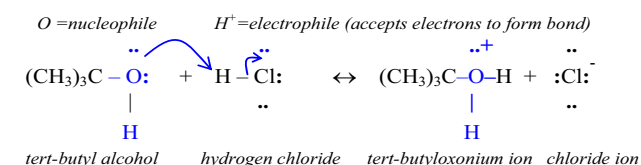
(a) Step 1: This is a "make-bond" process. It is the protonation of *tert*-butyl alcohol to produce an oxonium ion.
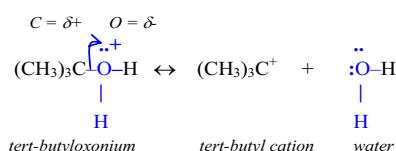
(b) Step 2: This is a "break-bond" process. The process represents the dissociation process of the *tert*-butyloxonium ion to produce a carbocation intermediate product.
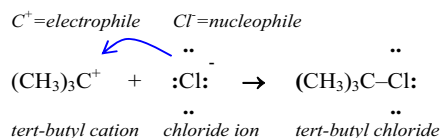
(c) Step 3: This is a "make-bond" process that describes Capturing of *tert*-butyl cation by chloride ion to produce *tert*-butyl chloride.

Fig. 1 Simulation of (1) can be viewed as a series of three reaction steps

Model provides a means to externalize thought. In our approach, the mental model constructed by a human chemist

to solve similar problem is represented as QPT models. For instance, the chemical property and behavior of the "protonation" (Step 1) are modeled as a "make-bond" QPT process (Fig. 2), whereas the chemical property and behavior of the "dissociation" (Step 2) is modeled as a "break-bond" QPT process (Fig. 3). The last reaction step will reuse the process model constructed for "protonation" (since both needing the behavior of a "make-bond" reaction). As such, the qualitative models constructed using our algorithms can support model re-use. Reusing models is made possible by deriving task-oriented model (different organic reactions, for example, the protonation, halogenation, etc.) from generic ones (the "make-bond" and "break-bond" processes). This a desirable feature for building more power tools and for industrial application [16].

All qualitative models are automated by the Model Constructor Module of the QRIOM simulator. Casting expert knowledge into qualitative model is not the focus of this paper. Readers may refer to [13] for knowledge abstraction and QPT modeling algorithms.

A QPT model for a "make-bond" process
Individuals
;there is an electrophile (charged)
1.　　H　　　*;represents hydrogen ion*
; there is a nucleophile (neutral) that has lone pairs of electrons
2.　　O　　　*;represents the alcohol oxygen*
Preconditions
3.　　A_m [no-of-bond(O)] = TWO
4.　　is_reactive(R_3C-OH)
5.　　leaving_group(OH, poor)
Quantity-Conditions
6.　　A_m[lone-pair-electron(O)] >= ONE
7.　　charges(H, positive)
8.　　electrophile(H, charged)
9.　　nucleophile(O, neutral)
10.　　charges(O, neutral)
Relations
11.　　D_s[charges(H)]= -1
12.　　D_s[charges(O)]= 1
13.　　lone-pair-electron(O) P_-^+ no-of-bond(O)
14.　　charges(O) P_+^- lone-pair-electron (O)
15.　　lone-pair-electron(H) P no-of-bond(H)
16.　　charges(H) P_-^+ no-of-bond(H)
Influences
17.　　I_+　　　(no-of-bond(O), A_m[bond-activity])
18.　　I_+　　　(no-of-bond(H), A_m[bond-activity])

Fig. 2 A QPT model representing the chemical behaviors of a "make-bond" process (e.g. protonation)

Supporting students in articulating and reasoning with models can lead to deep understanding of science subjects [17]. It can help sharpen a learner's reasoning ability in the way that the learner has to think hard for why the statements in each slot (of the model) are relevant. A number of learning activities can be derived from articulating the five slots in a QPT process. Some examples are given below. Line numbers are based on the enumeration used in Fig. 2. Note: "P" means "qualitative proportionality", with interpretations: Y

P_+^+ X means "Increasing of X will cause Y to increase", Y

P_-^+ X represents "Y decreases as X increases", and so forth.

- The proton (H^+) from "HCl" is needed by the first step of reaction formula (1) simulation. Learners would be able to know it by inspecting the *Individuals* slot. Briefly, the slot says "in order to begin the first step, there needs a proton which serves as an electrophile together with a species which is nucleophile". In this case, the nucleophile is the "O" (termed as alcohol oxygen) from the "OH" group which has lone-pair electrons to be donated. Line 1 and Line 2 show exactly the existence of hydrogen ion together with the "OH" functional group from the alcohol which helped to explain why the two substances are required.

- The inequality "lone-pair-electron >= min-electron-pair" speaks for "there is at least one extra lone-pair electrons to be donated to H^+". This indirectly ascertains that a covalent bond could be made between the two species.

- A decrease in the lone-pair electron on "O" can be explained as follows. We know that the immediate cause of the "make-bond" process is that, the number of covalent bond on "O" increases (Line 17). This parameter further influenced its lone-pair electron, and the effect is strictly decreasing through the inverse proportionality relationship (Line 13).

- By using our chemical intuition, when the process ends, the "O" is positively charged (hence unstable) while the "H" will be neutralized. The above can be explained by the following statements in the QPT model: The number of lone-pair electron decreases when more covalent bonds are made on "O" via the inverse qualitative proportionality (Line 13). When the lone-pair-electron of "O" decreases the charges on it will increase (Line 14). The above explains why "O" is positively charged.

QPT model for a "break-bond" process (E.g. Cleavage of C−O bond in ((CH_3)$_3$C-O$^+H_2$))
Individuals
1.	O	;the O from $-O^+H_2$ (oxonium ion) of (CH_3)$_3$C-O$^+H_2$
2.	C	;the C from R_3C- (the alkyl group) of (CH_3)$_3$C-O$^+H_2$

Preconditions
;to show that the two individuals are in the same compound
3.	bond-between(C, O)
4.	electronegativity(O) > electronegativity(C)

Quantity-Conditions
5.	A_m[no-of-bond(O)] > A_m[max-bond-allowed(O)]	; unstable
6.	charges(O, positive)	;oxonium ion is positive charge

Relations
7.	lone-pair-electron(O) P_+^- no-of-bond(O)
8.	charges(O) P_-^+ lone-pair-electron(O)
9.	lone-pair-electron(C) P no-of-bond(C)
10.	charges(C) P_+^- no-of-bond(C)

Influences
11.	I_-	(no-of-bond(O), A_m[bond-activity])
12.	I_-	(no-of-bond(C), A_m[bond-activity])

Fig. 3 General chemical behaviors of a "break-bond" process (e.g. Dissociation) represented using QPT

All the QPT models are viewable by the students. One of the models constructed by QRIOM is shown in Fig. 14.

IV. THE SYSTEM ARCHITECTURE OF QRIOM

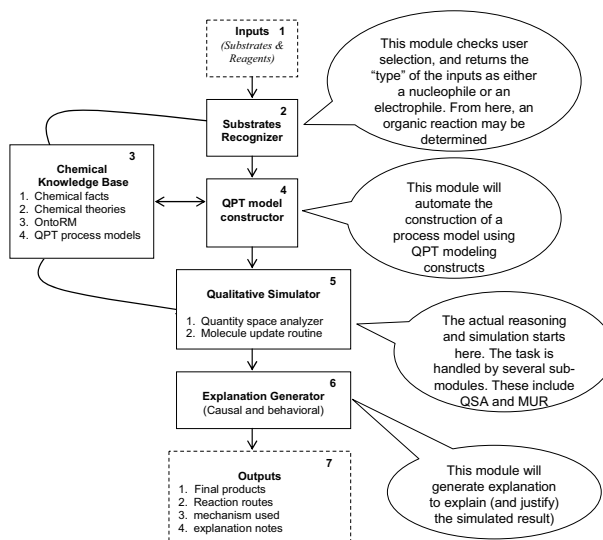QRIOM consists of seven main components, as shown in Fig. 4.

Fig. 4 Software components of QRIOM

A two-tier Knowledge Base (KB) is included to both supplying chemical data and playing the role of knowledge validation during qualitative simulation. In software implementation part, the logic language PROLOG [18] is used to represent all the permanent data or physical quantities. These data are non-changing throughout a reaction. Examples of physical quantities include atomic-number, periodic-number, etc. Chemical theories used in the simulation task are also classified as non-changing data. Examples of permanent data are (in PROLOG syntax):

```
/* --- chemical facts --- */
covalent_bond('C', '4', stable).
covalent_bond('O', '2', stable).
lone_pair('O', '2', stable).
lone_pair('C', '0', stable).

/* --- chemical theories --- */
qprop(make_bond, chargedElec, no_of_bond, charge, plus, minus).
qprop(make_bond, chargedElec, charge, no_of_bond, minus, plus).
qprop(_, _, energy, stability, minus, plus).
qprop(_, _, stability, reactivity, minus, plus).
```

Notice that these relationships do not change in a reaction. The OntoRM (Ontology for Reaction Mechanism) is at the lower tier. OntoRM provides to the learning tool some basic knowledge about the requirements and constraints of organic reaction mechanisms. This is to ensure correct use of the chemical knowledge. Discussion on the chemical KB construction and organization is beyond the scope of this paper.

V. Algorithm Development And Software Implementation

A. Input Scoping

Inputs are assumed to have no noise. Users can only enter alcohols (primary, secondary and tertiary alcohols) and alkyl halides. Other inputs will be rejected with a pop-up window with simple message. This is because in our present work, only the basic facts and theories related to alcohols and alkyl halides are included in the chemical KB. We selected two families of organic substrates as kick start inputs because the number of known organic compounds is more than 10 millions. Nevertheless, more families of organic compound will be included in our second prototype.

B. Knowledge Validation

Validation check is handled by OntoRM. The main purpose of OntoRM is to constrain uses of some processes so as to avoid wrong reaction step, or to prevent wrong use of chemical theories during a simulation task. All definitions are represented as Java classes with only attributes, and no processing. Samples of OntoRM are shown in Fig. 5.

```
nucleophile [
        hasName =>> STRING;
        hasNeutral =>> BOOLEAN;
        hasCharge =>> BOOLEAN;
        hasBond =>> NUMBER;
        hasRsDegree =>> NUMBER;
        hasCarbocationStability=>>FUZZY_VALUE;
        hasLonePair =>> NUMBER;
        hasReactivity =>>BOOLEAN;
        hasChargeOperator =>> PLUS_MINUS;
        hasBondOperator =>> ADD_REMOVE; ]
alcohol [
        hasName =>> STRING;
        hasBondType =>>BOND_TYPE;
        hasReactivity =>> BOOLEAN;
        hasDegreeSubstituent =>> DEGREES;
        hasBaseStrength =>> BASE_STRENGTH;
        hasStability =>> FUZZY_VALUE;
        hasLGType =>> LG_STRING;            ]
```

Fig. 5 Definitions of nucleophile and alcohol in the OntoRM chemistry ontology

C. Setting Initial Magnitude

An influence has no effect if the magnitude of the quantity from which it originates is unknown. Therefore this step assigns initial values to quantities. The qualitative simulation algorithm will seek for the initial data (from the chemical KB) in order to perform qualitative arithmetic. Further discussion on this task can be found in Section VII.

D. Workflow of QRIOM

Fig. 6 depicts the workflow of the simulator. In our approach, the construction of qualitative model is automated by a simple pair of substrate and reagent (the inputs). The key is to identify (recognize) the name of the functional group that is attached to the input substrate. When the "input recognizer" identifies the type of the inputs, the nucleophilic and electrophilic centers will be known. This answer can then be used for determining the bonding activity (make/break a

bond). From there, a candidate organic chemistry reaction (process) will be selected and activated. During reasoning, some intermediates are produced (and some are gone, or converted to other molecules) and they are placed in view structure (QPT term). So long as there is still a species left in the structure, a new process will be initiated and repeated the reasoning (but on different process model). When the entire reaction ended, users may ask for an explanation on any aspect of an organic reaction with respect to phenomena that must be learned.
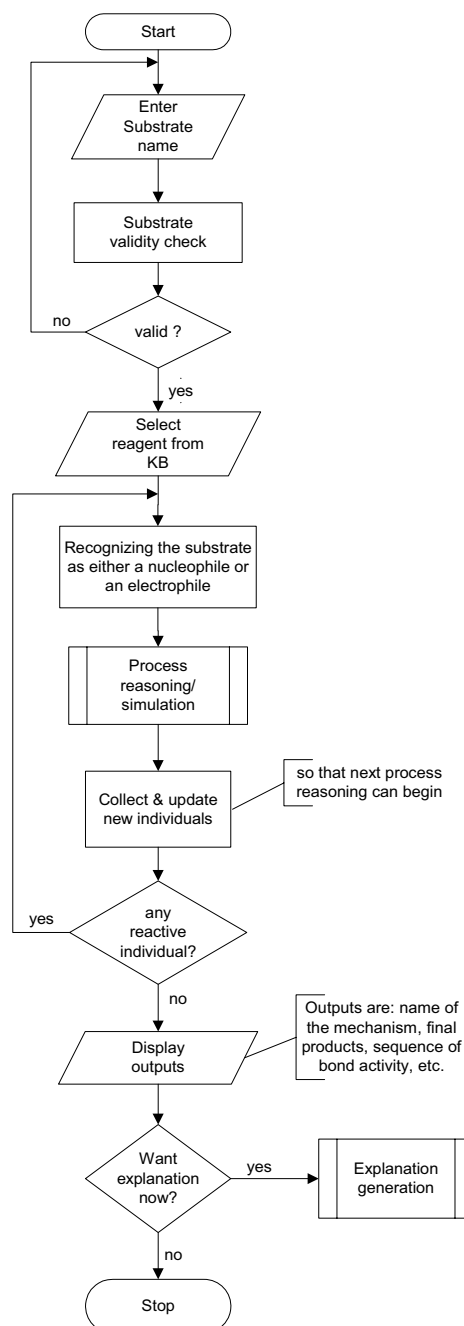

Fig. 6 Workflow of QRIOM

E. Simulation Algorithms

Since QPT is a representation of knowledge and no description about how they are used. This means, it does not describe how the models are learned. Therefore, we have developed a set of QR algorithms to reason the chemical process models encoded in QPT. Fig. 7 outlines the macro steps of the QR algorithm used in reasoning and simulation. Point 3.3 in the algorithm will be elaborated in Fig. 8.

```
QPT based simulation algorithm
Q_Simulation(substrate, reagent, OUTPUT)
1.   Determine a chemical process based on the recognized units
2.   Construct QPT models
3.   Perform QPT process reasoning
     3.1   Initialize multiple data structures
     3.2   Store the directly influenced process's quantities
           3.2.1   Calculate their Ds values and propagate values
                   through the indirect influences
     3.3   Perform quantity space analysis (done by QSA)
     3.4   Store new individuals (intermediates) for next process activation use
     3.5   Update the substrate's molecular structure (done by MUR)
4.   If View_Structure <> EMPTY Then
           Go to step 1
     Else
           Store final products and mechanism used in OUTPUT
     End _If
5.   Return OUTPUT
```

Fig. 7 The QPT based simulation algorithm used in QRIOM

VI. Generating Causal Graph

The basic steps for producing a causal graph during reasoning are shown in Fig. 8.

```
Quantity Space Analyzer (QSA)
QS_Analyzer(quantity_name, initial states of affected quantities, quantity space)
     1.   Store initial states of each quantity in special purposes arrays
     2.   Perform qualitative arithmetic
          2.1   Examine the sign and direct of change (derivative) of the quantity
          2.2   Check the relevant quantity spaces for new values
          2.3   Update qualitative states
          2.4   Keep propagated effects in special purpose data stores
     3.   Stop
```

Fig. 8 Programming logic for causal graph generation

A chain of effect propagation represented as functional dependency among chemical parameters will be constructed during run-time by our QR algorithms (QSA module in specific) that served as embedded intelligence in the tool to produce the causal graphs. A causal graph depicts the set of causal relationships between quantities occurring in the simulation. One such cause-effect relationships are depicted in Fig. 9 (a sketch for illustration purposes) and Fig. 10 (version that generated by the software).

Causality can be used to manifest order (arrange) upon the world. For example, when given 'X causes Y', we believe that if we want to obtain Y we would create X. As such, when we observe Y we will think that X might be the reason for it. We will demonstrate how the ontological design primitives of QPT can provide casual explanation about some aspects of chemical system behavior. In this work, Qualitative proportionality (the Ps) helps propagate effects of change caused by process quantity. For example, given two proportionalities (qp1 and qp2):

$$\text{lone-pair-electron (O)} \quad P_-^+ \quad \text{no-of-bond(O)} \qquad \dots qp1$$

$$\text{charges(O)} \quad P_+^- \quad \text{lone-pair-electron (O)} \qquad \dots qp2$$

A hypothetical question that can be asked (or derived) from the above two statements could be: "How would the above qualitative proportionalities explain the "O" atom becoming positively charged?" A causal explanation that could be generated is: The number of lone-pair electron will decrease when more covalent bonds are made on the "O" atom via the inverse proportionality defined in qp1. In qp2, when the lone-pair electron on "O" decreases the charge on it will increase. The above functional dependencies can explain why "O" is positively charged; simply it donated electrons to form a covalent bond. We perceive this type of explanation to be more natural and dynamic (run-time generation).

Another advantage of representing chemical behaviour using the qualitative proportionality of QPT is that the state of the chemical system can be tracked over time. This tracking helps explain the underlying cause-effect chain which is implicit in the qualitative models. The causal graph produced by QRIOM can be used to capture such dependency at run-time, and then be used to generate explanation on-the-fly. Figure 9 is a causal diagram that depicts the overall chemical change of the substrate during simulation of reaction formula (1). Overall, the diagram is produced by the QSA module that reasons the qualitative models following the programming logic presented in Fig. 8.

For cross checking purposes, label "a" in step 1 of Fig. 9 represents line 17 of the "make-bond" process depicted in Fig. 2, where "charges(O) increased" means that the charge on "O" atom will increase, and so forth. Similarly, from the same graph, label "b" is derived from line 13 and label "c" represents line 14 in the same process model. The symbols "d" and "e" denote the direct and indirect influences of the QPT modeling constructs.

A. Interpreting Causal Graph

Fig. 9 interpretation is given next. An organic reaction is triggered by an electrophile and a nucleophile. Step 1 (a "make-bond" process) was activated by the <H⁺, O> pair. In this case, the nucleophile is the "O" from the OH group which has extra lone-pair electrons to be donated to the proton (H⁺, electrophile). The direct effect (immediate cause of the process) is that a covalent bond will be made between "O" and "H". These effects are propagated to other quantities, as follows. The charge on "H" is decreased (i.e., from positive to neutral), while the lone-pair on "O" is decreased (donated to the electrophile). Decreasing in its lone-pair will cause the charge on it increasing. The charge of oxygen atom is now turning from neutral to positive. Assigning quantity values to each reacting species is coordinated by the QSA module (under "perform qualitative arithmetic" of Fig. 8). All values that are assigned to each parameter are retrieved from the quantity spaces maintained in the chemical KB. Table 1 gives the three main quantities used in this problem. The first reaction step produces R_3CO+H_2 (an intermediate).

Step 1: make-bond (E.g. Protonation)

H+ (Hydrogen ion, a charged electrophile)   R3C-**O**H(Alcohol oxygen, a nucleophile)

no-of-bond(H) **increased**   no-of-bond(O) **increased**

charges(H) **decreased**   lone-pair-electron(O) **decreased**

                                                       charges(O) **increased**

The make-bond process
produced the R3C-O+H2

Step 2: break-bond (E.g. Dissociation)

R3C(the C is an electrophile)   **O+H2** (the electronegativity of O is higher)

no-of-bond(C) **decreased**   no-of-bond(O) **decreased**

charges(C) **increased**   lone-pair-electron(O) **increased**

                                                       charges(O) **decreased**

The break-bond process produced
the carbocation intermediate

Step 3: make-bond (E.g. Capturing of the carbocation by a halide ion)

Cl- (the chloride ion serves as a nucleophile)   **R3C+** (the carbocation serves as an electrophile)

no-of-bond(Cl) **increased**   no-of-bond(C) **increased**

lone-pair-electron(O) **decreased**   charges(C) **decreased**

charges(Cl) **increased**

The make-bond process
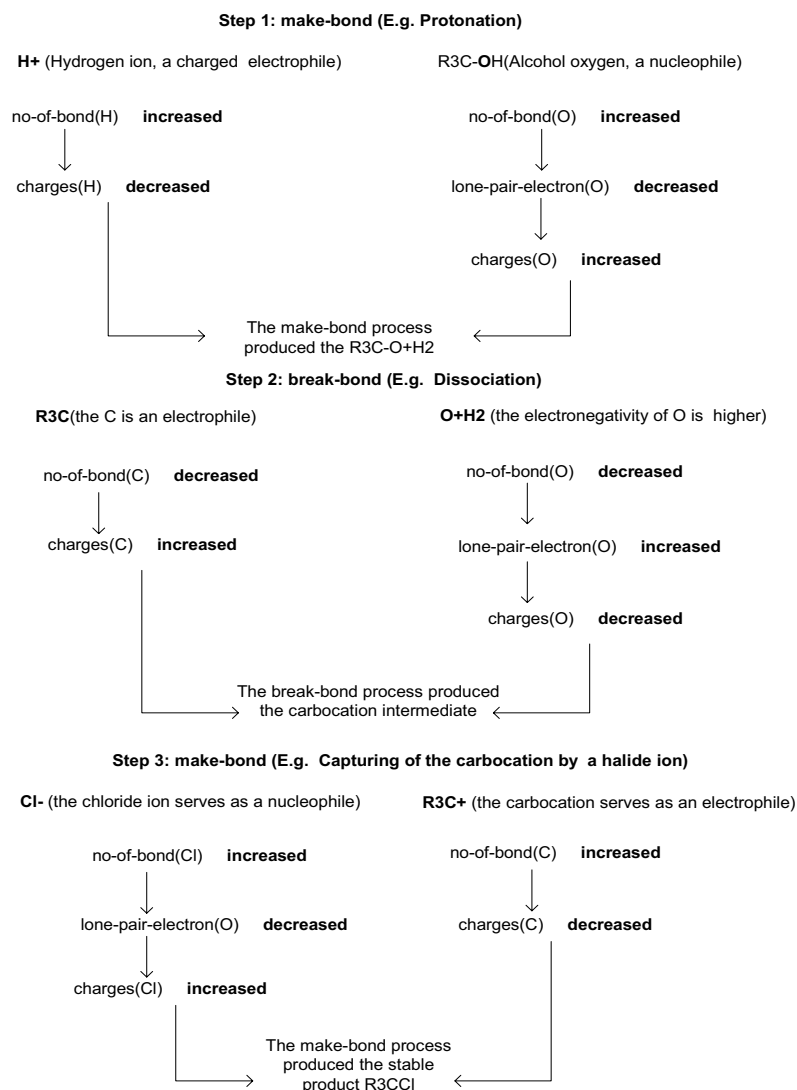produced the stable
product R3CCl

Fig. 9 Schematic view of the chemical parameters dependencies during simulation of (1)
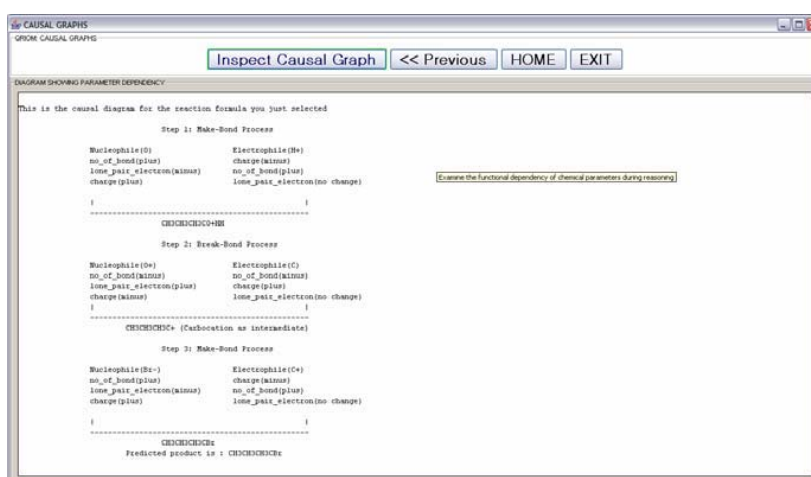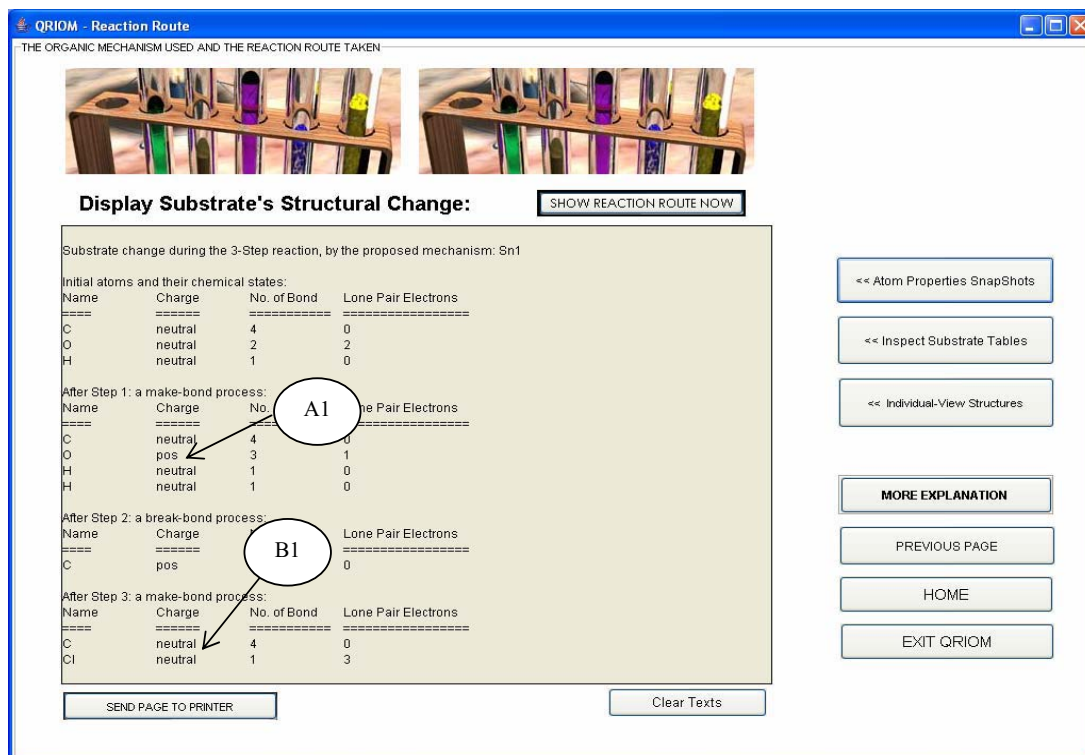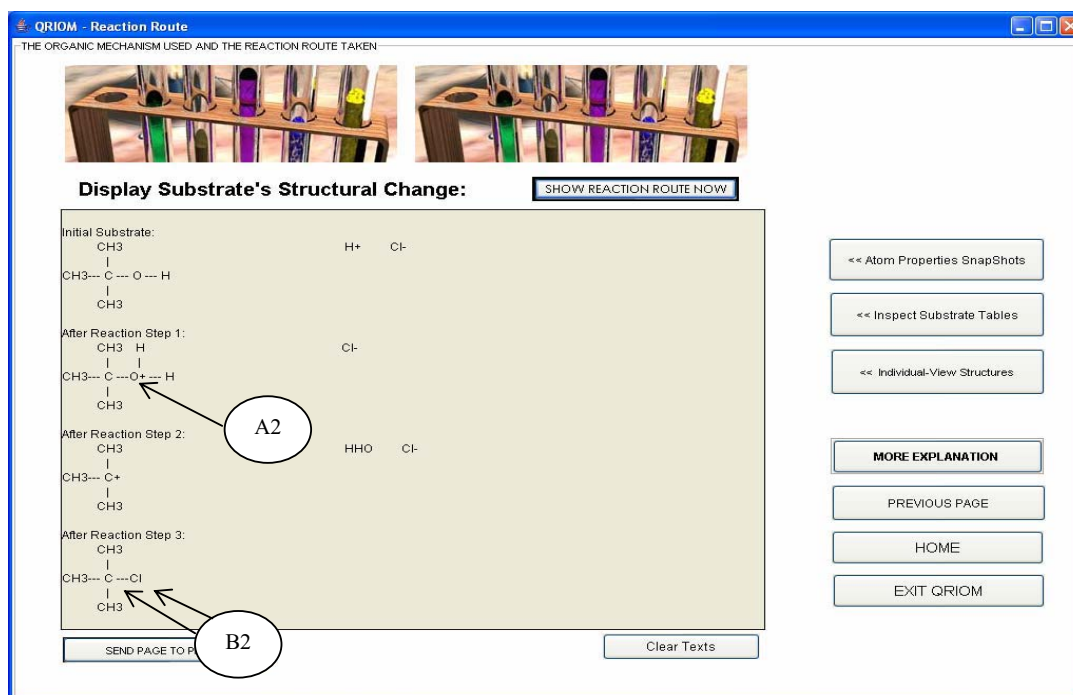

Fig. 10 Causal graph generated by QRIOM at the end of a simulation

(a) A snapshot of the chemical states possessed by each reacting unit during simulation

(b) Step-by-step change of the molecular structure of an organic substrate

Fig. 11 More simulation results to justify explanation about chemical behavioral change

From the chemical KB, this intermediate has "C" and "O+" as the reacting species for a "break-bond" process to be activated.

The "break-bond" process (refer to Step 2 in Fig. 9) describes the cleavage of the carbon-oxygen bond in *tert*-butyloxonium ion $((CH_3)_3C-O^+H_2)$ which is unstable since the "O" is charged. The immediate cause of this process is the bond between "O" and "C" will cleave (bond will break). State changes that propagated via functional dependencies among quantities are: The acceptance of two electrons from the dissociation activity will neutralize the "O" in "O^+H_2". On the other hand, donation of electrons (lone-pair decreases, as indicated in the graph) will cause the charge on "C" becomes positively charged (charge increases). This propagation produces a tertiary carbocation (R_3C+) for next process activation use.

Atom "C" in the carbocation is now unstable and it is reactive. Since only the carbocation and the chlorine ion (Cl⁻) remain in the view structure, another "make-bond" process ("capturing of carbocation by anion") can be initiated. This is the third reaction step in the entire simulation. The start of this process can be explained by the incomplete octet of the carbocation and the chloride ion. The following describes Step 3: When a covalent bond is made between "Cl" and "C+", the chlorine's lone-pair electrons will decrease. This effect is further propagated to changing its charge (from negative to neutral state), while the carbon's charge is decreased (from positive to neutral). The entire process ended here because both "Cl" and "C" are in neutral state (i.e. their valences having been completed).

In QPT reasoning, a process is deemed to stop when there is no more reacting species left in the view structure. From the graph, the final products are alkyl chloride (R_3CCl) which is very stable, and a side product (water molecule). The sequence of process activations are protonation, dissociation, followed by capturing of carbocation by chloride ion. These three steps (make-bond, break-bond, and make-bond again) can be used to explain the overall chemical change occurred. The presentation of chemical effect propagation as a cause-effect diagram would help appreciates the general chemical principles underlying the chemical phenomenon, and hence useful in improving one's conceptual understanding on the subject.

The causal graph presented in Fig. 10 is used for two purposes. First, it is for user inspection (to detect chemical changes of reacting species), and second, it is used for generating the overall results tabulated in Fig. 11. When Fig. 11(a) is coupled with Fig. 11(b), the prediction of the final product is made more apparent. For example, when the charge on "O" is positive (A1, Fig. 11 (a)), then there is a positive sign next to the "O" atom (A2, Fig. 11 (b)) displayed on the current substrate structure. Also, when "C" and "Cl" are both in neutral state (B1, Fig. 11(a)), the reaction is claimed ended as shown in B2, Fig. 11(b). Besides, Fig. 11 (b) depicts the step-by-step conversion of the substrate to form the final product.

TABLE I
MAIN QUANTITIES AND QUANTITY SPACES FOR MODELING ORGANIC REACTION

Quantity	Quantity Space	Remarks
charges	[negative, neutral, positive]	At any one time, the charge on any atom can be negative, neutral or positive.
lone-pair electron	[zero, one, two, three, four]	The 'four' comes from Halide ion, and "zero" from Hydrogen ion.
no-of-bond	[zero, one, two, three, four]	We consider important atoms for nucleophilic substitution reaction only. The 'four' comes from Carbon; "zero" from Hydrogen when in their most stable state.

VI. QRIOM AND SIMULATION RESULTS DISCUSSION

A. QRIOM Walkthrough

The QRIOM problem solving model (Fig. 12) consists of several parts that made up of any number of pages. The user could repeat any function as many times as desired or go to a page that is unclear. In the present work, the tool is designed in such a way that the user is taken through the modeling, simulation and explanation pages step-by-step. The tool also provides explanation via special functions to emphasize certain chemical theories and general concepts.
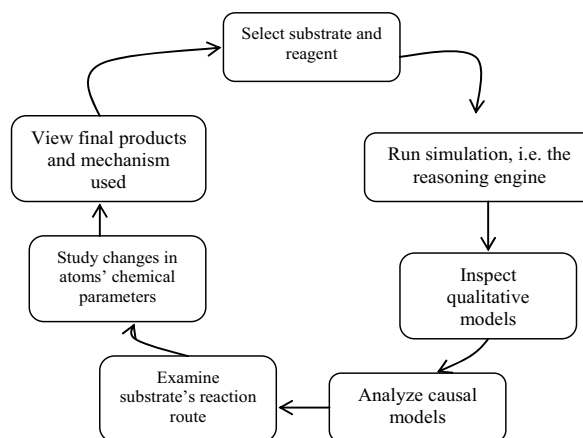

Fig. 12 The problem solving model of QRIOM

B. Simulation Results

All results and explanations are occurred dynamically. The computational complexities (modeling, reasoning, and simulation) described in earlier sections are hidden from the users. The system is rather interactive (users would be able to select the inputs and parameters) and generative (knowledge articulation is supported). Fig. 13 shows the main interface of QRIOM. After a simulation is performed, the entire reaction route may be viewed; this function is handled by button *A*. Button *B* is used to compose qualitative models (see Fig. 14

for a sample QPT model). Clicking on button *D* will lead to more explanation and learning pages. All these features are designed to increase a student's enthusiasm, as they would tend to find out more explanation and compare the obtained information to what they have in mind (or thought to be appropriated). With these multiple tabulated results, learners are able to make appropriate mental connections. From an interview conducted, one major obstacle most learners faced was the use of the QPT ontological terms in some of the screenshots.
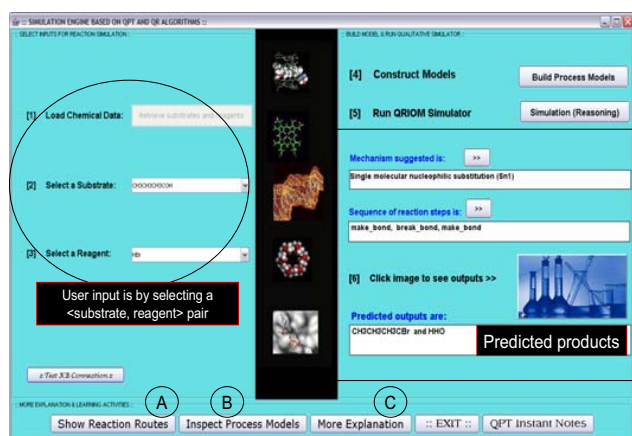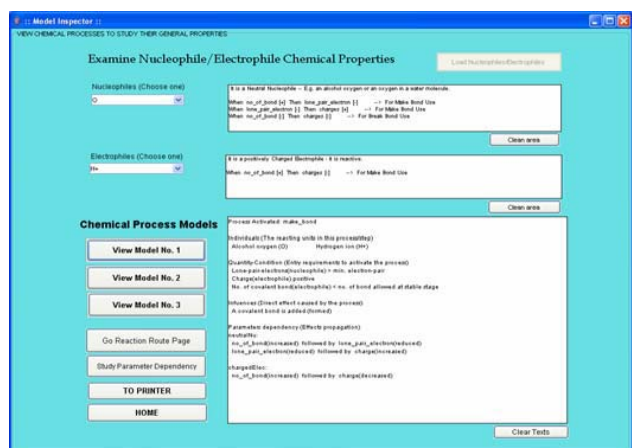

Fig. 13 The qualitative simulator interface

Fig. 14 Model inspection helps articulate the chemical theories that are implicit in the model

VII. CONCLUSION

Qualitative reasoning is concerned with both prediction and explanation, whereas qualitative modeling focuses on causal and conceptual knowledge. This is exactly the kind of knowledge that students are trying to master in science subjects. Our approach satisfied the following: (1) Representation of domain knowledge in qualitative terms. (2) Utilization of commonsense knowledge in problem solving (chemical theories via QPT constructs). (3) Dynamic causal graph generation for explaining theories of chemical

phenomena. We hope that the results of this work may facilitate a widespread use of the new approach in building educational software for subjects that require application of domain knowledge at intuitive level. Currently, the simulator generates output in text based environment and the molecule rendering is in 2D format. We plan to incorporate a multimedia interface to produce animated 3D structures for molecules during simulation. Site testing will also be carried out for more effective implementation.
.

REFERENCES

[1] K. Forbus, and D. Gentner, "Qualitative mental models: Simulations or memories?" Proc. of the 11th International workshop on qualitative reasoning, Cortona, Italy, 1997.
[2] I. Iwasaki, "Real-world applications of qualitative reasoning", IEEE Expert, 1997, pp. 16–21.
[3] K. Forbus, P. Whalley, J. Everett, L. Ureel, M. Brokowski, J. Baher, and S. Kuehne, "CyclePad: an articulate virtual laboratory for engineering thermodynamics," Artificial Intelligence Journal, Vol. 114 , 1999, pp. 297–347.
[4] A. Bouwer, and B. Bredeweg, "VisiGarp: graphical representation of qualitative simulation models," in: J.D. Moore, G. Luckhardt Redfield, J.L. Johnson (Eds.), Artificial Intelligence in Education: AI-ED in the Wired and Wireless Future, Amsterdam, The Netherlands, 2001, pp. 294-305.
[5] S.M.F.D. Syed Mustapha, J.S. Pang, and S.M. Zain, "Application of Qualitative Process Theory to Qualitative Simulation and Analysis of Inorganic Chemical Reaction", Proc. of the 16th International Workshop of Qualitative Reasoning, Barcelona, Spain, 2002.
[6] S.M.F.D. Syed Mustapha, J.S. Pang, and S.M. Zain, "QALSIC: towards building an articulate educational software using qualitative process theory approach in inorganic chemistry for high school level," International Journal of Artificial Intelligence in Education, 15 (3): 229–257, 2005.
[7] G. Biswas, D. Schwartz, J. Bransford, and The Teachable Agents Group at Vanderbilt, "Technology Support for Complex Problem Solving: From SAD Environment to AI", in K. Forbus, and P. Feltovich, (Eds.), Smart Machines in Education, Menlo Park Calif.: AAAI Press, pp. 71–97, 2001.
[8] I. Bratko, and D. Suc, "Qualitative explanation of controllers", Proc. of the 16th International Workshop on Qualitative Reasoning, June 10-12, Barcelona, Spain, 2002, pp. 1–2.
[9] I. Bratko, and D. Suc, "Learning Qualitative Models", AI Magazine 24(4): 107–119, 2003.
[10] LHASA research group home page. Available: http://derek.harvard.edu/.
[11] D. P. Dolata, "Artificial intelligence in Chemistry", Ohio University, Athens, OH, USA, 1993.
[12] K. Forbus, "Qualitative process theory," Artificial Intelligence, Vol. 24, 1984, pp. 85–168.
[13] Y.C. Alicia Tang, and S.M.F.D. Syed Mustapha, "Representing SN1 reaction mechanism using the qualitative process theory," in: C. Bailey-Kellogg, B. Kuipers, (Eds.), Proc. of the 20th International Workshop on Qualitative Reasoning, Hanover, USA, 2006, pp. 137-147.
[14] K. Forbus, "Articulate software for science and engineering education", in K. Forbus, P. Feltovich, and A. Canas (Eds), Smart machines in education: the coming revolution in educational technology, AAAI Press, 2001.
[15] D. Gentner, and A. Stevens, "Mental models", IEA Associates, 1983.
[16] B. Bredeweg, and K. Forbus, "Qualitative modeling in education," AI Magazine, 24(4): 35-46, 2003.
[17] D.E. Penner, "Cognition, Computers and Synthetic Science: Building knowledge and meaning through modeling", Review of research in education, 25: 1–36, 2001.
[18] Amzi! Prolog Home Page. Available: http://www.amzi.com

Alicia Tang obtained her Master of Science from The University of Manchester, United Kingdom. She is a doctoral student in the Faculty of Computer Science and Information Technology, University of Malaya. She is also a lecturer at the University of Tenaga Nasional. Her research fields include Qualitative Reasoning, Prediction Engine Design for Organic Chemistry Reactions, and Computer-based Learning. She is a member of APSCE, IEEE, and IEEE Computer Society.

Rukaini Abdullah obtained her Ph.D. from University of Leeds, United Kingdom. She is attached to the Faculty of Computer Science and IT, University of Malaya. Her research interests are Natural Language Processing, AI in Education, Information Retrieval, and Computer Based Learning. Dr. Rukaini is currently the Head of Artificial Intelligence Department in the Faculty.

Sharifuddin Mohd. Zain received his Ph.D. in 1995 from Imperial College, London, United Kingdom. He is an Associate Professor in the Department of Chemistry at University of Malaya, Malaysia. His research interests include Computational Chemistry, Application of Computers in Chemical Research and Education, Artificial Intelligence in Chemistry, Molecular Spectroscopy, and Environmental Modeling. Dr. Sharifuddin is leading several projects supported by the Ministry of Science, Technology and Innovation (MOSTI), Malaysia. He is also a Regional Editor for the Journal of Global Environmental Engineering.

Noorsaadah Abdul Rahman received her Ph.D. in 1990 from Cambridge University, United Kingdom. She is a Professor at the Department of Chemistry, University of Malaya, Malaysia. Her areas of specialization are Organic Synthesis and Synthetic Methods, Drug Design and Modeling. Professor Noorsaadah also heads projects on drug design and development under the Biopharmacy and the Structural Biology Programs funded by the National Biotechnology Directorate, Ministry of Science, Technology and Innovation. In addition, she serves on the International Advisory Panel to GETRIB – Bioinformatics, Hyderabad, India.