Designing organic reaction simulation engine using qualitative reasoning approach

Y.C. ALICIA TANG\textsuperscript{1}, S.M. ZAIN\textsuperscript{2}, N. ABDUL RAHMAN\textsuperscript{2}
\textsuperscript{1}Department of Systems and Networking
University of Tenaga Nasional
KM 7, Jalan Kajang-Puchong, 43009 Selangor
\textsuperscript{2}Department of Chemistry
University of Malaya
MALAYSIA

Abstract: Understanding organic reactions requires good grasp of chemical intuition which is difficult to program using traditional development approach. The paper presents work on a software tool for providing support to learning in the area of organic reaction simulation. The design uses Qualitative Reasoning (QR) and, in particular, Qualitative Process Theory (QPT) and the simulation of basic steps in the chemical reactions such as creating and deleting bonds. QPT ontology is used to represent the chemical theories and chemical intuition of organic reactions. A set of QR algorithms has also been developed to automate the construction of qualitative models, and to perform products prediction. The learning tool can help learners in fostering the reasoning capability and facilitating the mastery of the fundamental principles of organic reactions. We start by presenting the roles of each functional component of the tool. Then, we move on to discuss the qualitative modeling and simulation design for reproducing the chemical behaviors of selected organic reactions. Finally the simulation results are discussed.

Key-Words: Qualitative Reasoning, QPT, Organic Reaction, Simulation, Chemistry, Qualitative Models

1 Introduction
Qualitative Reasoning (QR) is an Artificial Intelligence (AI) technique that attempts to model behavior of dynamic physical systems without involving a bunch of formulas (E.g. chemical equations) and quantitative data in software. Research in QR spans a wide range of topics, from cognitive modeling, task-level reasoning, ontologies, application, to creating new kinds of educational system called articulate software. In [1], an overview of QR research in general is discussed, while an insight view of QR in education can be found in [2]. We have developed a tool abbreviated Qualitative Reasoning in Organic Mechanism (QRIOM) to support the teaching of an organic chemistry course called organic mechanisms at the University of Malaya. A reaction mechanism describes how a reaction takes place by showing what is happening to valence electrons during the making and breaking of bonds. Most of the time, the organic chemists can work out the mechanisms by only using common sense developed from their chemical knowledge. As the chemists have “expertise” which is largely qualitative by nature and therefore best captured and communicated using QR technology.

A large number of the chemistry students learn organic reactions by memorizing the reaction steps and the bunch of formulas taught in classes. Many of them have poor conceptual understanding about the subject, such as not knowing the principles governing the processes and the cause effect interaction among processes. We investigated qualitative representation and qualitative simulation approaches. The ultimate goal is to find a way to nurture learners’ conceptual understanding so that they could solve new or complex problems using their chemical insight.

The simulation environment of our work is different from other QR systems such as CyclePad [4], VisiGarp[5] and QALSIC [6]. The main difference is that the students are not involved in the modeling as part of the learning requirements as in VisiGarp and CyclePad; since our intention is not to train the chemistry students as modelers, rather when the representation and design is implemented, the software can help improve their understanding and the development of reasoning skills. QALSIC is among the earliest applications of QPT in inorganic chemistry for qualitative analysis of a small set of chemical reactions. The QPT processes supported by the software is hand-coded, as such its uses are more...
rigid and limited. QRIOM, however, is able to construct models, and to provide causal explanation. In [7], the two organic reactions “make-bond” and “break-bond” have been identified as reusable components that can be extended to simulate other reaction mechanisms. Sample internal representation for the molecules has also been presented in [8]. This paper will focus on the design of the simulation engine, and extended results cases which are not found in earlier reports.

Section 2 introduces the QPT ontology. Section 3 presents the roles of each software component in the tool. System methodology is given in Section 4. These include system inputs, outputs, and a set of QR algorithms. In Section 5, qualitative simulation algorithms are presented. Section 6 discusses the simulation results. Section 7 concludes the work.

2 QPT as the Knowledge Representation Tool

The work presented in this paper is based on a qualitative reasoning ontology, the Qualitative Process Theory (QPT) [3]. The ontology can be used to model the behaviors of organic reaction at the finest granularity of processes, such as the “make-bond” and “break-bond” processes. QPT provides the means to describe processes in conceptual terms, and embody notions of causality which is important to explain behavior of chemical systems. In QPT, 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: Individuals, Preconditions, Quantity-conditions, Relations and Influences. Relations are statements about functional dependencies among quantities (parameters). An important modeling construct for describing the relationships between quantities is qualitative proportionalities (denoted by \( P+/P- \)), that propagate the effects of processes that express unknown monotonic functions between two chemical parameters. Direct influence (denoted by \( I+/I- \)) supports a process’s direct effect on the object. Note: words typed in italics are QPT modeling constructs. Readers may refer to [3] for a complete reading of the ontology.

3 QRIOM: The Simulation Engine

The main function components of QRIOM are given in Fig. 1. The roles and uses of each module are summarized in Table 1.

3.1 Two-tier knowledge base

The knowledge base has two-tier architecture. Upper tier stores the basic chemical facts and chemical theories. OntoRM (Ontology for Reaction Mechanism) is at the lower tier that defines the reaction mechanisms. OntoRM describes the
knowledge, requirements and constraints needed in producing the reaction mechanism behaviors. Examples of reaction mechanism are unimolecular nucleophilic substitution ($S_N 1$) and bimolecular nucleophilic substitution ($S_N 2$). As an example, the OntoRM can be used to check if a primary alcohol can undergo $S_N 1$.

4 Methods
A chemical process’s functional characteristics (what) are represented using QPT and its reasoning/processing description (how) is coordinated and controlled by a set of QR algorithms. The system methodology is divided into a number of main tasks:

- Identifying chemical properties for organic reactions for model composition use
- Classifying the possible reaction species and types (Module 2)
- Developing the automated model construction logic (Module 4)
- Developing the reasoning steps for predicting and simulating the chemical behaviors of selected organic reactions (Module 5)
- Designing the means for generating explanation (Module 7)

4.1 Inputs, outputs and reaction type
Chemical scientists deal with a variety of structures and transformation which can usually be decomposed into clearly identifiable entities. We decomposed the organic compounds (substrates) into the “Rs” chain (R=CH$_3$) and the attachment (functional groups). The approach performs substrate validity check before a simulation can start, as such the “Rs” will normally not be needed in the nucleophilic substitution reactions. Therefore, we focus our representation on the nucleophiles (electron-rich species) to be substituted. As for the outputs, the tool will return the following results: (1) products formed (2) intermediates produced at each step (3) processes simulated (to re-produce the behavior of the proposed reaction mechanism) (4) overall structural change of the substrate, and (5) explanation or justification for a question being asked. 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. Usually, a reaction may involve a series of small changes one after the other. A reaction mechanism describes this series of changes.

We have selected $S_N 1$ as the test case. It is the substitution of one nucleophile by another. Equation (1) is used to exemplify the behavior simulation of the reactions. It is the production of alkyl halide from a tertiary alcohol. To benefit readers from non-chemistry background, Equation (1) is subdivided into a series of small step, as shown in Fig. 2. In which, in the first stage, the alcohol oxygen (the ‘O’ from the ‘OH’ group) is protonated. Meaning, the ‘O’ captures a proton (Step 1). In Step 2, the link between the tertiary carbon and the alcohol oxygen will break, and this produces a carbocation. In the last stage, the incoming nucleophile (Cl$^-$, in this case) can bond to the carbocation to form a neutral and stable final product (Step 3). The three steps will be modeled as three QPT models.

\[(\text{CH}_3)_2\text{C-CH}_2\text{OH} + \text{HCl} \rightarrow (\text{CH}_3)_2\text{C-CH}_2\text{Cl} + \text{H}_2\text{O}\] (1)

a) Step 1: Protonation of tert-Butyl alcohol by a proton (H$^+$). This is a “make-bond” process.

\[\text{CH}_3\text{CH}(_2)\text{CH}_3\text{OH} + \text{H}^+ \rightarrow \text{CH}_3\text{CH}(_2)\text{CH}_3\text{O}^+\] (2)

(b) Step 2: Dissociation of oxonium ion ($\text{H}_2\text{O}_\text{+}$). This is a “break-bond” process.

\[\text{CH}_3\text{CH}(_2)\text{CH}_3\text{O}^+ + \text{Cl}^- \rightarrow \text{CH}_3\text{CH}(_2)\text{CH}_3\text{Cl} + \text{H}_2\text{O}\] (3)

(c) Step 3: Capturing of tert-butyl cation by chloride ion. This is a “make-bond” process.

Fig. 2 The production of alkyl halide can be explained by a series of three reaction steps.

4.2 Qualitative model construction algorithms
Fig. 3 outlines the steps for constructing a QPT process. It is the micro steps of point 3 in Fig 5.
QPT VIEWS AND PROCESSES MODELING ALGORITHM
Qualitative_Modeling(substrate, reagent, QPT_MODEL)
1. Examine user inputs (substrate and reagent)
2. Decompose inputs into functional units
3. Recognize functional units
4. Compose the Individual Views where the structure of the input is used to recognize chemical intuition
5. Retrieve general properties of the groups
6. Check Individual Views in the VS
7. Assign units to either nucleophile or electrophile group
8. Retrieve general properties for the selected process
9. Compose the individuals for a QPT process
10. Stop

Fig. 3 Model construction logic

Using the model construction steps, a QPT model for the “make-bond” process can be constructed (Fig. 4). Note: P means ‘qualitative proportionality’, with interpretations: Y P X represents ‘Y increases as X increases’, Y P - X represents ‘Y decreases as X increases’, and so forth. To avoid being too technical in chemistry contents, we included only the minimal and essential properties in our illustrations.

QPT Process for “make-bond” (e.g. ((CH₃)₃COH) protonated by H⁺)

<table>
<thead>
<tr>
<th>Individuals</th>
<th>Preconditions</th>
<th>Quantity-Conditions</th>
<th>Relations</th>
<th>Influences</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. H⁺</td>
<td>Aₙ[no-of-bond(O)] ~ TWO</td>
<td>charges(H⁺, positive)</td>
<td>lone-pair-electron(O) P⁺ no-of-bond(H)</td>
<td>Iₙ(no-of-bond(O)) Aₙ[bond-activity]</td>
</tr>
<tr>
<td>2. O⁻</td>
<td>is_reactive(R, C-OH)</td>
<td>electrophile(H⁻, charged)</td>
<td>charges(O⁻, neutral)</td>
<td>Iₙ(no-of-bond(H⁻), Aₙ[bond-activity])</td>
</tr>
<tr>
<td>3. Aₙ[nuclearophile(neutral)] = ONE</td>
<td>nucleophile(O⁻, neutral)</td>
<td>lone-pair-electron(H⁻) P⁻ no-of-bond(O)</td>
<td>lone-pair-electron(O⁻) P⁻ no-of-bond(O)</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 4 A “make-bond” process described in QPT terms. It is read as “If Individuals and Quantity-conditions are true then Influences and Relations are executed”. In this case, the statements in Influences and Relations slots are qualitatively reasoned.

**4.3 Simulation engine design**
The main simulation steps are given in Fig. 5. The reasoning technique is based on QPT ontology.

**5 Simulation Scenario**
This section explains how qualitative reasoning is performed on a “make-bond” process (Fig. 4). The process is the first reaction step for predicting the final product of “[(CH₃)₃COH + HCl]”. Using the simulation algorithm in Fig. 5, prediction begins with the Influences slot where the number of covalent bond on ‘O’ will increase (Line 16). Such effect will propagate to other dependent quantities. For example, the number of lone-pair electrons will decrease when more covalent bonds are made on the ‘O’ via the inverse qualitative proportionality (Line 13). When the lone-pair electron of ‘O’ decreases, the charges on ‘O’ will also increase (Line 14). This will make the ‘O’ a positively charged species and having an extra covalent bond (hence it is unstable). When the ‘O’ is protonated, the ‘H’ is no longer positively charged (Line 16), thus violating the statement in the quantity-conditions slot. The above scenario is rather “common sense” to the organic chemists; and QPT is able to capture this type of chemical intuition. Quantity Space Tracker (QST) is a sub module of the reasoning engine that keeps track of the current values of each quantity and their direction of change. The QST is also responsible for maintaining a number data structures such as the substrate table that stores the constituent elements of a substrate during reasoning in order to produce the final product and its structure (Fig. 6). Figure 7 shows the contents of one of the atom property arrays during chemical processes reasoning.
6 Results Discussion

The main graphical interface of the qualitative simulator is shown in Fig. 9. Button A is used to construct QPT processes. When a learner is ready to run a simulation, button B is clicked. In QRIOM, qualitative models can be inspected at any stage of the learning process. When button C is clicked, the constructed process models will be displayed (see Fig. 10). After a simulation is performed, learners may view the entire reaction route. This function is handled by button D. This mechanism helps manifest the “knowledge articulation” learning pedagogy, as the learner has to think hard for why the statements in each slot (of the model) are relevant or negligible. Users can also examine how and why things happen by calling up the explanation generator (button E).

5.1 The Molecule Update Routine (MUR)

When reasoning is switched from a process to another one, the MUR will be called upon to handle the structural change of the substrate’s functional group (the unit that is involved in the reaction). The reaction route taken by the substrate from the start state until the entire reaction ended is shown in Fig. 8.

Fig. 6 Schematic view of the substrate during simulation

Fig. 7 Step-by-step change of the substrate’s chemical properties. The change is governed by the qualitative proportionality statements in the QPT model. Notice that all species in the last table are in neutral states.

Fig. 8 Reaction route showing the molecular patterns of the alcohol substrate from the first process until the formation of the alkyl halide.
Provision is also made for learners who needed further explanation, especially on the QPT ontology. For this, each slot of a QPT process is explained as shown in display area A of Fig. 11. In B, users may choose a bond activity and select a specific pair of parameters to examine their dependency. By doing so, users are able to investigate how the different set of processes may affect the chemical parameters.

Fig. 11 Parameter functional dependency can be checked in a more interactive way

7 Conclusion
Qualitative reasoning allows learners to access notions of how the behavior of systems evolves in time, while QPT ontology is good for capturing the intuitive and causal aspects of human mental models. The qualitative models in QRIOM captured and communicated knowledge that is common to chemistry people (via the QPT constructs). This lends itself as a promising representation and simulation technique to explain phenomenon in science subjects such as chemistry. The new computational approach can serve as alternative learning technology in developing educational software for subjects that require application of domain knowledge at intuitive level.

References: