Design and development of a qualitative simulator for learning organic reactions

Y. C. Alicia Tang, S. M. Zain, and N. A. Rahman

Abstract—Our work features an ontology-supported framework for developing a qualitative simulator for explaining the behaviors of selected sample of organic chemistry reactions. The design of the simulator uses Qualitative Reasoning (QR), and in particular, Qualitative Process Theory (QPT) for constructing qualitative models and the simulation of basic steps in the chemical reactions such as creating and deleting bonds. The qualitative simulator allows learners to access notions of how the behavior of chemical systems evolves in time. Students would benefit from it in terms of improving their reasoning skills and enhancing their understanding in organic processes. The roles of each functional component of the qualitative simulator will first be introduced. Then, we move on to discuss the qualitative modeling and simulation design for reproducing the chemical behaviors of organic reactions. Finally, a discussion on the simulation results and explanation generation capability are presented.

Keywords—Organic reaction, qualitative process theory, qualitative reasoning, simulation.

I. 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 ontologies, cognitive modeling, task-level reasoning, application, to creating new kinds of educational system called articulate software. In [1], an overview of QR research in general is discussed, whereas an insight view of QR in education can be found in [2]. QR, although no longer a new research field in Artificial Intelligence (AI), its exploration in chemistry domain remains widely open. We have developed a tool abbreviated as QRIOM (Qualitative Reasoning in Organic Mechanism) to support the learning and teaching of an undergraduate organic chemistry course (called organic mechanism) 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. Generally, students’ major difficulty in solving organic reaction problems lies in the conceptual understanding of the problem, such as not knowing the principles governing the processes and the cause effect interaction among processes. We investigated qualitative representation of domain knowledge, and qualitative reasoning to predict (and explain) the final products of a reaction. The ultimate goal is, when a learner interacts with the system, his conceptual understanding can be nurtured so that the learner could solve new or complex problems by reading and analyzing the various explanations generated by the software.

The simulation environment of our work is different from other QR systems such as CyclePad [3], VisiGarp[4], and QALSIC [5]. 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 QR in inorganic chemistry for qualitative analysis of a small set of chemical reactions. The qualitative models run in the software are pre-coded. However, QRIOM is able to construct qualitative models, and to provide various forms of explanation on demand. In [6], the “make-bond” and “break-bond” chemical processes have been identified as reusable components in the software, in which the processes can be used for other organic reaction simulation. Sample computer representation for the molecules has been presented in [7]. This paper will focus on the design and implementation of the simulation engine, and extended the simulation and explanation cases which are not found in earlier reports.

The rest of this paper is organized as follows. The second chapter presents the domain suitability and problem formulation. The third chapter introduces the modeling ontology. The roles of each software component in QRIOM are given in chapter four. System methodologies are presented in chapter five. These include system inputs, outputs, and the qualitative modeling and simulation algorithms. Chapter six discusses the simulation results.
Concluding remarks and further work are presented in the last chapter.

II. DOMAIN SUITABILITY AND PROBLEM FORMULATION

The application domain is organic chemistry reaction simulation. We have justified the domain as a suitable field [8]. Learning organic reactions is a challenging task to chemistry students. Most of the students learn organic reactions by memorizing the reaction steps and the bunch of formulas taught in classes. Organic reaction and its mechanism involve the study of electrons movement, in which a bond is being made or broken. 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, and why so. Since explaining organic reactions has the qualitative descriptive nature, therefore, qualitative reasoning is used in the software tool for knowledge articulation, prediction and explanation generation.

A. Organic Reaction and Mechanism

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. A reaction mechanism describes this series of changes. Organic chemists will identify the electron-poor site and electron-rich group when trying to work out a reaction mechanism through their chemical intuition, knowledge and experience developed. Such complication can be modeled by describing the behaviors of reaction mechanism as a series of primitive processes (such as “make-bond” and “break-bond”) to enable lowest level of reasoning.

III. QUALITATIVE PROCESS THEORY (QPT) AS THE KNOWLEDGE CAPTURE TOOL

Reference [9] defines ontology as a knowledge base that describes the concepts and properties of a domain, and their relations (e.g., chemical parameter dependency), providing a common vocabulary in a defined area (e.g., organic chemistry reactions). This work uses a process-based ontology called Qualitative Process Theory (QPT) [10] to model the behaviors of organic reaction at the finest granularity of processes, such as the “make-bond” and “break-bond” organic processes. Ontology has the potential to facilitate the formation of semantic relationships between various portions of useful information to enhance the learning experience in an educational setting [11]. In the same token, QPT plays the role of supporting knowledge acquisition (gather the relevant knowledge) and model construction (creation of relationships among chemical parameters) in the simulation environment.

A. QPT Modeling Constructs

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 process supports changes in system behavior. A QPT process is described by five slots (see Fig. 1): Individuals (contains a list of objects upon which the process is applicable), Preconditions (it contains statements referring to external conditions), Quantity-conditions (it contains inequalities involving object’s characteristics, which is essential in determining the status of a process active or inactive), Relations and Direct Influences. Relations are statements about functional dependencies among quantities (or 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 [10] for further description of the ontology.

![Fig. 1 The five slots of a QPT process](image-url)

IV. QRIOM: THE SIMULATION ENGINE

The main software components of QRIOM are given in Fig. 2. The roles of each software module are summarized in Table 1.

![Fig. 2 Software components of QRIOM for modeling and simulating organic reactions based on QPT](image-url)
A. 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 producing the reaction mechanism behaviors. Examples of tier that defines the reaction mechanisms. OntoRM describes the basic chemical facts and chemical theories.

Module 2 (Substrate Recognizer)
- To initialize a number of tables (e.g., 2D arrays) to hold the running results of various chemical parameters during simulation.

Module 4 (Model Constructor)
- To compose a QPT process specification (the qualitative model) based on the identity of user input.
- It will generate a model as depicted in Fig. 9.

Module 5 (Reasoning Engine)
- This module does the actual simulation.
- The main reasoning functions are handled by the Quantity Space Tracker (QST) and the Molecule Update Routine (MUR).

Module 6 (Molecule Update Routine)
- This module keeps track of the structural change (pattern) of the substrate, from one organic reaction to another.
- It will display reaction route as shown in Fig. 14.

Module 7 (Explanation generator)
- To retrieve various data structures (produced by the prediction engine) in order to generate explanation on-the-fly.
- Identifying chemical properties for organic reactions. This is for model composition use (Module 3).
- Classifying the possible reacting species and types (Module 2).
- Developing the automated model construction algorithms (Module 4).
- Developing the reasoning steps for predicting and simulating the chemical behaviors of selected organic chemistry reactions (Module 5).
- Generating explanation based on QPT modeling constructs (Module 7). Various forms of explanation are produced by this module. Explanation interfaces are given in Fig. 8 – Fig. 9, and Fig 12 – Fig. 14.

A. Inputs, Outputs, and Reaction Types

Chemical scientists deal with a variety of structures and transformation which can usually be decomposed into clearly identifiable entities. We decomposed the organic compounds (also called substrates) into the “Rs” chain (e.g., “CH$_3$CH$_2$CH$_3$”), and the attachments (e.g., the functional group “OH”). In our approach, substrate validity check is performed before a simulation is started. As such, we focus our representation on the nucleophiles (i.e., an electron-rich species) to be substituted. As for the outputs, the simulator will return the following results: (1) final products, (2) intermediates produced at each step, (3) sequence of processes used to re-produce the behavior of the proposed reaction mechanism, (4) overall structural change of the substrate (see Fig. 13), and (5) explanation or justification for a question being asked (refer to Fig. 8 – Fig. 9, Fig. 12 – Fig. 14). Sample results for (1), (2), and (3) can be found in Fig. 7.

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 reaction formula. It is the production of alkyl halide from a tertiary alcohol. To benefit readers from non-chemistry background, (1) is subdivided into a series of small step, as shown in Fig. 3. In which, in the first stage, the alcohol oxygen (the “O” from the “OH” group) is protonated. That is, the “O” captures a proton (refer to Step 1). In Step 2, the link between the tertiary carbon and the alcohol oxygen will break, and this produces a carbocation intermediate. In the last stage, the incoming nucleophile (in this case, it is “Cl”−) can bond to the carbocation to form a neutral and stable final product (refer to Step 3). The three steps will be modeled as three QPT processes. Note: A = tert-Butyl alcohol, B = Hydrogen chloride, C= tert-Butyl chloride, D = Water molecule. Labels A and B are the inputs while C and D represent the final products.

\[
(CH_3)_3C-OH + HCl \rightarrow (CH_3)_3C-Cl + H_2O \tag{1}
\]

V. METHODS

In this work, a chemical process’s functional characteristics (the “what”) are represented using QPT and its reasoning/processing description (the “how”) is controlled by a set of QR algorithms. The system methodology is divided into a number of tasks (Modules numbering are based on Fig. 2):
Using the model construction steps, a QPT model for the “make-bond” process representing the chemical behavior of Step 1 can be constructed (see Fig. 5). 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)₃C-OH) protonated by H⁺)

**Individuals**
1. H⁺ represents hydrogen ion
2. O⁻ represents the alcohol oxygen
3. Aₖ([non-bonded-electron-pair(O)]) = TWO
4.  is_reactive(R-C(OH))
5.  leaving_group(OH, poor)

**Quantitative-Conditions**
6.  Aₖ([non-bonded-electron-pair(O)]) >= ONE
7.  charges(H, positive)
8.  electrophile(H, charged)
9.  nucleophile(O, neutral)
10. charges(O, neutral)

**Relations**
11. D_{chargers(H)} = -1
12. D_{chargers(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), Aₖ([bond-activity]))
18. I, (no-of-bond(H), Aₖ([bond-activity]))

Fig. 5 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.

C. Simulation Engine Design

The simulation algorithm is given in Fig. 6. Detailed explanation is given in Chapter VI.
relevant or negligible. After a simulation is performed, learners may view the entire reaction route. This function is handled by button $D$, when it is clicked, Fig. 13 will be displayed. Users can also examine how and why things happen by calling up the explanation generator (button $E$). This button will lead the learner to various forms of explanation one at a time; upon user selection.

![Fig. 7 The main interface of the reasoning and simulation engine](image)

![Fig. 8 A model inspection page that shows a chemical process represented using QPT constructs](image)

Provision is also made for learners who needed further explanation, especially on the QPT ontology. Each slot of a QPT process is explained as shown in display area $A$ of Fig. 9. 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. 9 Parameter functional dependency can be checked in a more interactive way](image)

### A. Simulation Scenario

The simulation workflow of the combined use of QR and QPT approach is depicted in Fig. 10.

![Fig. 10 Workflow of the QPT based reasoning used in QRIOM](image)
This section explains how qualitative reasoning is performed on a “make-bond” process (Fig. 5). The “make-bond” process is the first reaction step for predicting the final product of “(CH3)3COH + HCl”. Using the simulation algorithm in Fig. 6, prediction begins with the Influences slot, where the number of covalent bond on “O” will increase (Line 17). 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 on “O” decreases, the charges on it will 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 describes knowledge which is common to the organic chemists, and QPT is able to capture this type of chemical intuition using only its qualitative proportionality modeling construct.

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. 11), while Fig. 12 shows the contents of an atom property table during chemical processes reasoning. The information in Fig. 12 can help a learner to examine in greater details the step-by-step chemical changes acted on each atom. Examples of atoms are the functional groups and the incoming nucleophile to be substituted.

When a reasoning task is performed, the runtime results are kept in various arrays (tables). These data will be used to generate more explanation about the underlying concept of an organic reaction. We will discuss one good use of such data for generating the reaction route of the initial substrate (input), as below.

### B. The Molecule Update Routine (MUR)

During reasoning (from a process to another one), the MUR will be called upon to handle the structural change of the substrate’s functional group. The reaction route taken by the substrate from the start state until the entire reaction ended is depicted in Fig. 13. In the figure, the substrate’s molecular structure displayed was translated by the MUR (making use of the results generated by the QST). As an example, when the charge on “C” atom is positive (see Fig. 12, under the “After reaction step 2” heading), then there is a positive (+) sign printed next to the “C” atom (see Fig. 13, under the “After reaction step 2” heading). Functional dependency of chemical parameters can also be examined by selecting the “Inspect Causal Graph” function of the simulator (refer to Fig. 14). With these multiple tabulated results, learners are able to make appropriate mental connections, and as such one’s reasoning ability can be enhanced, especially in improving their understanding of the organic processes and the cause-effect chain (of chemical parameters) that is implicit in the QPT models.
I. CONCLUSION AND FUTURE WORKS

The combined use of QR and QPT is able to automatically construct qualitative models for chemical processes and to reproduce the behavior of organic reactions through qualitative reasoning. The approach also supports causal and behavioral explanation generation. QPT ontology is good for capturing the intuitive and causal aspects of human mental models. This is also a new test domain for QPT. The presented computational approach can serve as alternative learning technology in developing educational software for subjects that require application of domain knowledge at intuitive level.

The qualitative models constructed using our algorithms can support model re-use. Reusing models is made possible by deriving task-oriented model, i.e., the different organic reactions (e.g., protonation, and halogenation) from generic ones (the “make-bond” and “break-bond” processes). This is also a desirable feature for building more power tools and for industrial application, as described in [1]. We hope that the results of this study may facilitate a widespread use of qualitative model development technique to other sub-fields of organic chemistry.
The work will be continued from a number of aspects. These include generating 3D animated output (currently, outputs are presented in plain 2D format); the development of a protocol converter to handle protocol between the reasoning shell and the 3D output. A problem ontology that handles user queries much like the one presented in [12] is also the direction of our future work. The main purpose of having the problem ontology is to deal more specifically and accurately with questions that may be asked by the learners. QRIOM can also be improved by adding the pedagogical elements (such as the different learning styles) in the “technogogical” three-dimensional (technology, content, and the pedagogy) learning environment as proposed by Idrus [13].

REFERENCES