An Ontology-Based Reasoning Framework for Reaction Mechanisms Simulation

Y.C. Alicia Tang[1], Sharifuddin Mohd. Zain[2], Noorsaadah Abdul Rahman[2], and Rukaini Abdullah[3]

[1] Department of Computer Science, University of Tenaga Nasional
Jalan Kajang-Puchong, 43009 Selangor, Malaysia
[2] Department of Chemistry, University of Malaya, 50603 Kuala Lumpur, Malaysia
[3] Department of Artificial Intelligence, University of Malaya
50603 Kuala Lumpur, Malaysia
aliciat@uniten.edu.my,{smzain,noorsaadah,rukaini}@um.edu.my

Abstract. Many chemistry students have difficulty in understanding an organic chemistry subject called reaction mechanisms. Mastering the subject would require the application of chemical intuition and chemical commonsense adequately. This work discusses a novel framework using Qualitative Reasoning (QR) to provide means for learning reaction mechanisms through simulation. The framework consists of a number of functional components. These include substrate recognizer, qualitative model constructor, prediction engine, molecule update routine, explanation generator, and a knowledge base containing essential chemical facts and chemical theories. Chemical processes are represented as qualitative models using Qualitative Process Theory (QPT) ontology. The construction of these models is automated based on a set of QR algorithms. We have tested the framework on the S_N1 and the S_N2 reaction mechanisms. Representative cases of reaction simulation and causal explanation are also included to demonstrate how these models can serve as a cognitive tool fostering the acquisition of conceptual understanding via qualitative simulation.

Keywords: qualitative reasoning, reaction mechanisms, QPT, ontology.

1 Introduction

Qualitative Reasoning (QR) is a branch of Artificial Intelligence (AI) research that attempts to model behavior of dynamic physical systems without having to include a bunch of formulas and quantitative data in the system. Qualitative representation captures the intuitive and causal aspects of many human mental models. The research spans all aspects of the theory and applications of qualitative reasoning about physical systems. Qualitative Process Theory (QPT)[1] is one of the prominent QR ontology that is widely used to represent the behavior of dynamical systems. CyclePad [2] that teaches analysis and design of thermal cycles is the first smart educational software that employed QPT. QALSIC

Z. Zhang and J. Siekmann (Eds.): KSEM 2007, LNAI 4798, pp. 18–29, 2007.

is among the earliest applications of QPT in inorganic chemistry for qualitative analysis of a limited set of chemical reactions [3]. The chemical processes described in this paper were modeled using QPT ontology.

In the study of science subjects such as chemistry, it is believed that students should deeply understand the qualitative principles that govern the subject, including the physical processes and the causal relationships before they are immersed in complex problem solving. A reaction mechanism describes the sequence of steps that occur during the conversion of reactants to product. Examples of reaction mechanisms are S_N1, S_N2, electrophilic addition, and eliminations. Most of the time, the organic chemists could work out the mechanisms by only using commonsense developed from their chemical intuition and knowledge. A large number of chemistry students had difficulty in understanding reaction mechanisms. They learn the subject by memorizing the steps involved in each reaction. As a result, most students are unable to solve new problems. This finding initiated the work described in this paper. Even though there are many applications of AI techniques in organic chemistry, none has used QPT as the knowledge capture tool. This paper discusses the first use of the QR approach coupled with the QPT ontology to develop a framework that is able to simulate processes such as "make-bond" and "break-bond" in order to reproduce the chemical behaviors of organic reaction mechanism. The paper also introduced OntoRM, which is a set of ontology specifically for use with reaction mechanisms simulation. The framework will later on be transformed into a learning tool called Qualitative Reasoning in Organic Mechanisms (QRIOM).

We have reported in [4] about the modeling decisions and problems faced when trying to cast the expert knowledge into qualitative models. In [5], we justified the problem as a suitable domain by comparing inorganic chemical reactions and organic reaction mechanisms. In the work, we have also grouped all reacting species as either a nucleophile (charged/neutral) or an electrophile (charged/neutral), upon which chemical processes are selected. In [6], we provided guidelines for chemical properties abstraction and a description of how the QPT is used for modeling. In [7], we further ascertained that there are two main reusable processes, namely "make-bond" and "break-bond", for the entire reaction mechanisms, specifically on S_N1 and S_N2. An algorithm for automating the "make-bond" and the "break-bond" processes in QPT terms has also been discussed in the paper, where the initiation of the entire process is from a simple substrate.

In Sect. 2, we provide the methodology of our work. These include data sets, algorithms and functional components used by the framework. Section 3 discusses the simulation algorithm together with a process reasoning scenario and an explanation generation example. Section 4 concludes the work.

2 Methodology

The following data, modeling constructs, algorithms, and components are needed in order to simulate and reproduce the behavior of reaction mechanisms.

2.1 QPT as the Knowledge Capture Tool

Among the well-known QR ontology are component-centered [8], constraint-based [9] and process-centered [1]. QPT is a process-centered ontology. The theory provides the necessary means for representing qualitative knowledge, and the notion of processes needed in expressing chemical reaction steps (E.g. "protonation" and "dissociation" processes). In QPT, a description of the model is given by a set of *individual views* and *processes*. The *individual views* (E.g. a nucleophile) describe objects and their general characteristics while the *processes* (E.g. "make-bond") support changes in system behavior. A process is described by five slots: *Individuals*, *Preconditions*, *Quantity-conditions*, *Relations* and *Influences*. The *Quantity-conditions* slot contains inequalities involving *quantities*. A *quantity* is used to represent an object's characteristic, which is crucial in determining the status of a process (active/inactive). The statements in the Relations slot defined functional dependencies among *quantities*. Other important design constructs are the *qualitative proportionalities* and the *quantity spaces* (see Sect. 2.2). Note: words typed in italics are QPT modeling constructs. Further discussion about the ontology is beyond the scope of this paper. Readers may refer to Forbus [1] for a complete description of the ontology.

2.2 Qualitative Proportionality, Direct Influence, and Quantity Space

This section discusses the main modeling constructs used in the framework. *Qualitative proportionalities* are responsible for propagating the effects of processes execution. For example: lone-pair-electron (O) P^{+}_{-} no-of-bond (O) means "an increase in covalent bond on the 'O' atom will cause a decrease in the number of lone-pair electrons on it". In QPT, the dynamic aspects are expressed by the notion of direct influence, represented in the slot called *Influences* as either I+ or I-. For example, I+(no-of-bond(O),Am[bond-activity]) indicates that in a chemical process, the direct effect on 'O' is the extra covalent bond it would gain. This effect will be propagated to other statements in the *Relations* slot. Examples of *quantities* are number of covalent bond, lone-pair electrons and nucleophilic reactivity. A quantity consists of two parts: amount (magnitude) and derivative (sign). *Quantity space* is a collection of numbers which form a partial order. Processes start and stop when orderings between the *quantities* change. Table 1 gives the three main *quantities* used in the problem. The main task of qualitative simulation based on this formalism is to keep track of the changing states in each quantity for each individual view used in a reaction, in order to explain why and how a particular process happens or ends.

2.3 Data Sets

Reaction formulas tested by the framework include the production of alkyl halide from tertiary alcohol and the production of alcohol from tertiary alkyl halide.

Table 1. Quantities and quantity spaces used in the framework for modeling the behaviors of nucleophiles and electrophiles

Quantity	Quantity Space	Remarks
Charges	[negative, neutral, positive]	At any time the charge on any atom can either be negative, neutral or positive.
no-of-bond	[one, two, three, four]	We consider only the important atoms for nucleophilic substitution reaction. For example, the 'four' goes to carbon; the 'one' is for hydrogen when they are in the most stable state.
lone-pair-electrons	[zero, one, two, three, four]	The maximum value 'four' is for halide ions. The minimum 'zero' goes to hydrogen ion.

Examples of the reaction formulas used in modeling and simulation are (1) and (2).

$$(CH_3)_3COH + HCl \rightarrow (CH_3)_3CCl + H_2O. \quad (1)$$

$$(CH_3)_3CBr + H_2O \rightarrow (CH_3)_3COH + HBr. \quad (2)$$

Both the equations can be explained by the S_N1 mechanism. In general, S_N1 is a two-stage mechanism. In the first stage, the alcohol oxygen (the 'O' from the 'OH' group) is protonated. Meaning, the 'O' captures the electrophile ('H^+'). This is to make the 'O^+H2' as a good leaving group, in order to break the bond between the 'C' and the 'O^+H2'. Once broken, a carbocation will be produced. In the second stage, the incoming nucleophile ('X^-') can bond to the carbocation to form a neutral and stable final product. 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. Thus, (1) can be subdivided into a series of small step, as shown in Fig. 1. The main ideas of the reactions in Fig. 1 will be modeled as QPT processes. These processes are automatically generated by the Qualitative Model Constructor (module 2, Fig. 4). In which, the chemical properties of each organic reaction are represented as chemical theories using the qualitative proportionality construct. To avoid being too technical in terms of chemistry contents, we provided in this paper only the results (qualitative models in QPT syntax), rather than the entire modeling activity. A qualitative simulation scenario for the first reaction step in (1) is demonstrated in Sect. 3.

2.4 Knowledge Validation

The chemical knowledge used by QRIOM has two-tier architecture (Fig. 2). We are exploring the development of OntoRM, working on top of the QPT. It is to

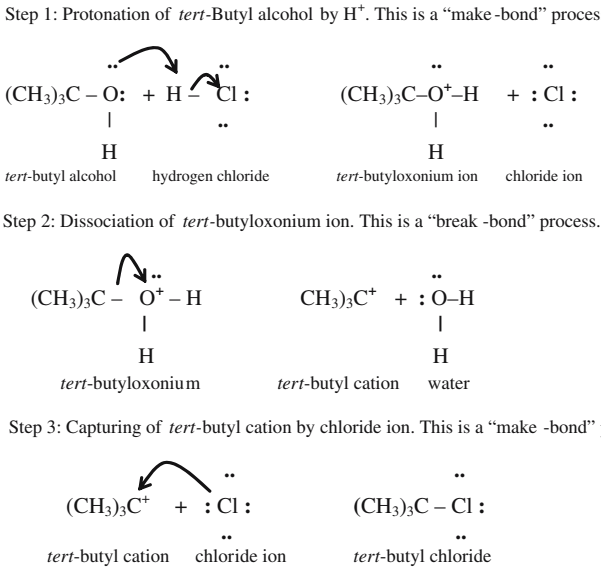

Fig. 1. The production of alkyl halide can be explained by a series of three reaction steps

facilitate knowledge validation during simulation such as to constrain the use of the chemical knowledge base. As an example, the OntoRM can be used to check if a primary alcohol can undergo a S_N1 reaction. Representative examples of OntoRM are given in Fig. 3. Further discussion about OntoRM is not the focus of this paper.

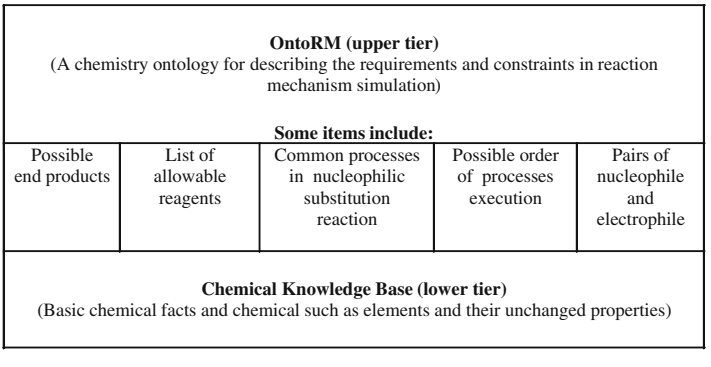

Fig. 2. Knowledge-base used by QRIOM

```
ReactionMechanism
Sn1 [
      hasAlias =>> STRING;
      hasReactants =>> FuncUnit;
      hasReactantNames =>> STRING;
      hasProduct =>> PROD_STRING;
      hasProductNames =>> STRING;
      hasDegreeSubstituent =>> NUMBER;
      hasReactivity =>> BOOLEAN;
      hasRateDetermineStep =>> WHAT_STEP_STRING;
      hasProcessOrder =>> PROCESS_ORDER_STR;
      hasViewsPairConstraint =>> SPECIES_TYPE;
      hasSpecialCause = >> SOLVENT_TYPE;   ]

ElectrophileView[
      hasName =>> STRING;
      hasNeutral =>> Electrophile;
      hasCharge =>> Electrophile;
      hasBond =>> NUMBER;
      hasRsDegree =>> NUMBER;
      hasCarbocationStability=>>FUZZY_VALUE;
      hasLonePair =>> NUMBER;
      hasReactivity =>>BOOLEAN;
      hasChargeOperator =>> PLUS_MINUS;
      hasBondOperator =>> ADD_REMOVE;      ]
```

Fig. 3. Definitions of S_N1 mechanism and electrophile in OntoRM

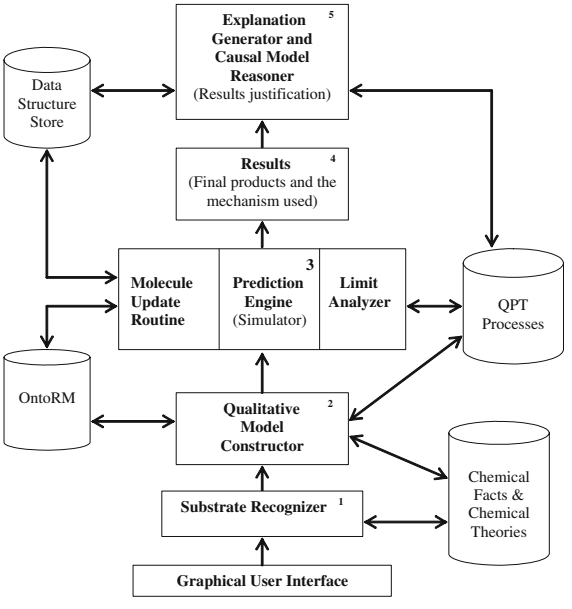

Fig. 4. Functional components of the framework

2.5 Functional Components

The QR framework is presented in Fig. 4. The framework consists of a set of reusable components such as the substrate recognizer, qualitative models, OntoRM, prediction engine, and causal explanation generator. The roles of the prediction engine and the explanation generator are discussed in Sect. 3.

The modules in Fig. 4 serve as embedded intelligence to the simulator. When used, it is expected to generate the following outputs: Final products; intermediates produced at each step; the sequence of use of the chemical processes; the name of the mechanism used; the structural change of the substrate; and the parameters change of each nucleophile and electrophile.

3 Qualitative Simulation

The QR algorithm for reaction mechanism simulation is outlined in Fig. 5.

```
QUALITATIVE SIMULATION ALGORITHM
Simulation(substrate, reagent, OUTPUT)
1. Recognize substrate
2. Construct individual views
3. Determine candidate processes
4. Construct QPT process
5. Perform processes reasoning
      Store process's quantity from the direct influence slot
      Perform limit analysis
      Check qualitative proportionalities in Relation-slot
      Store propagated effects based on quantities dependency
      Update atom table and atom property table
      Store new individuals in view structure
      Update view structure array
6. If process_stopping_condition = true Then
      Check if any reactive units in the view structure
      If reactive units <> EMPTY Then
           Go to step 3
      Else
           Suggest the mechanism used in the simulation
           Show the overall reaction route
           Display final products
        End_If
   End_If
7. Generate explanations
```

Fig. 5. QR algorithm based on QPT for reaction mechanism simulation

3.1 Top Level Design of the Simulation Algorithm

The overall simulation can be summarized as follows: Given a formula in the form "A (substrate) + B (reagent)", *individual views* will be constructed based on their chemical properties. These *views* will be stored in Instance Structure

(IS). Next, it is the checking of what processes can be used. When a process is in active state, reasoning will begin (details are given in Sect. 3.2). Briefly, the reasoning engine will keep track of the values of the affected *quantities*, starting from the first process until the entire reaction ends. A process will stop when the statements in its *quantity-condition* slot are invalid. If there are still reactive units (E.g. charged species or species that have not completed their valences), the reasoning process will be repeated. The entire reaction will end when there is no more *views* in the IS. When a reaction ends, outputs are displayed, together with all steps/processes involved in producing the outputs. If a user needs an explanation for the results or has a question regarding the behavior of a *quantity*, then the explanation module will be run.

3.2 Process Reasoning

A qualitative model for the "protonation" process is shown in Fig. 6. All complexities in constructing a qualitative model are hidden from the users, since the QPT model construction process is automated.

Process "Protonation" (e.g. $((CH_3)_3C\text{-}OH)$ is protonated by H^+)

Individuals
;there is an electrophile (charged)
1. H *; hydrogen ion*
; there is a nucleophile (neutral) that has lone pairs electron
2. O *; 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[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_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])

Fig. 6. A chemical process modeled using QPT ontology

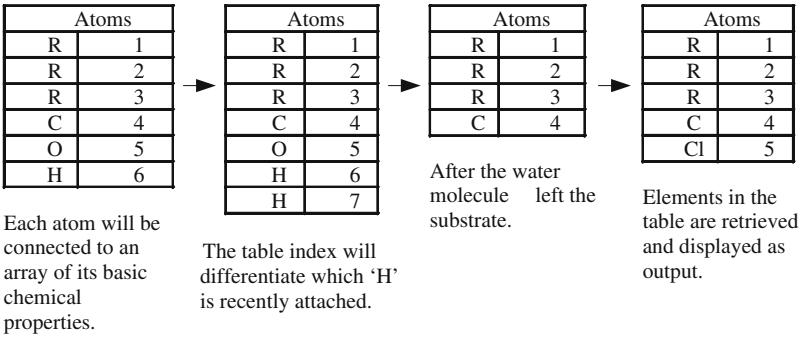

Fig. 7. A snapshot of the contents of an atom array during simulation. $R = CH_3$.

"Protonation" is the first reaction step of the S_N1 mechanism for predicting the final product for $(CH_3)_3COH + HCl$. Prediction begins with the *Influences* slot where the number of covalent bond on the '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 of 'O' decreases, the charges on 'O' will also increase (Line 14). This will make the 'O' become positively charged and having an extra covalent bond (hence it is unstable). When 'O' is protonated, the 'H' is no longer positively charged (Line 16), thus violating the statement in the *quantity-conditions* slot. All values assigned

Table 2. Atom property array keeps track of the step-by-step changing of values in various quantities

	Charges	No. of covalent bond	Lone -pair electron
C	Neutral	4	0
O	Neutral	2	2
H	Neutral	1	0
		(+)	
	Charges	No. of covalent bond	Lone -pair electron
H	Positive	0	0
		(=)	
	Charges	No. of covalent bond	Lone -pair electron
C	Neutral	4	0
O	Positi ve	3	1
H	Neutral	1	0
H	Neutral	1	0

to each individual view are taken from the *quantity spaces* by the limit analyzer, in that it keeps track of the current values of each *quantity* and their direction of change. The running result of the limit analyzer, acted on the atom array will be used to produce the molecule structure for the final product. An atom array is a table that stores the elements of a substrate during reasoning in order to produce the structure of the final product. Figure 7 depicts the contents of an atom array during processes simulation for reaction formula in (1). When performing limit analysis, the atom property array will also be updated. This is when the processes move from one to another until the entire reaction ended. A snapshot of the atom property array during process reasoning is given in Table 2. The above is achieved through constant updating of the *qualitative proportionality* statements (Lines 13-16) using values in the *quantity spaces*. In QRIOM, qualitative models can be inspected by learners at any stage of the learning process. This helps to sharpen a learner's logical and critical thinking in the way that the learner has to think hard for why the statements in each slot are relevant or negligible.

3.3 Explanation Generation

The ability to generate causal explanation has been one of the promises of the QR approach. Causality is normally used to impose order. For example, when given 'P causes Q', we believe that if we want to obtain Q we would bring about P. As such, when we observe Q we will think that P might be the reason for it. We will demonstrate how the modeling constructs of QPT can provide this nature of explanation. Figure 8 shows a partial causal diagram derived from Fig. 6.

Two qualitative proportionalities (abstracted from the left branch of Fig. 8) are presented (3) and (4) to manifest the explanation generation ability of the approach.

$$lone\,pair\,electron(O)\;\; P^{+}_{-}\;\; no\,of\,bond(O). \tag{3}$$

$$charges(O)\;\; P^{-}_{+}\;\; lone\,pair\,electron(O). \tag{4}$$

Based on the above parameters dependency statements, a set of hypothetical Q&A and answers can be devised, as follows:

- Question 1: How would you explain a decrease in the lone-pair-electron on the 'O' atom?
- Answer 1: The immediate cause of the process is the number of covalent bond on 'O' will increase. This quantity will influence the lone-pair-electron on 'O', and the influence is strictly decreasing through the inverse proportionality relationship. Thus, a decrease in the lone-pair-electron on 'O' is observed.
- Question 2: How would the above qualitative proportionalities explain the 'O' atom become positively charged?
- Answer 2: The number of lone-pair-electron will decrease when more covalent bonds are made on 'O' via the inverse proportionality defined in (3). In (4), when the lone-pair-electron on 'O' decreases, the charge on it will increase.

The inspection of cause effect chain can help a learner to sharpen his or her reasoning ability, in that the learner is able to pick up the underlying concept better than merely memorizing the reaction steps and formulas.

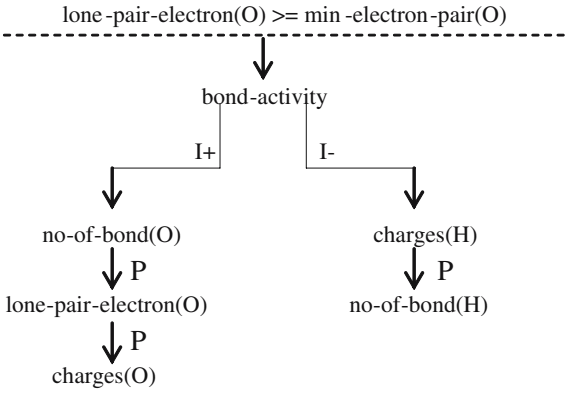

Fig. 8. The inequality above the dotted line is the entry condition to the process. Effects propagation is modeled using the 'Is' and the 'Ps' of the QPT design constructs.

4 Conclusion

We proposed a QR framework using the QPT ontology to systematically gather, reuse, and explain the knowledge about reaction mechanisms. The ontology provides the means to describe processes in conceptual terms, and embody notions of causality which is important to explain the behavior of chemical systems. The framework can provide a learning environment that assists learners in understanding the 'How', 'Why', 'Why-not', and 'What' aspects of the general principles of organic reactions. Our approach enables prediction (for the outputs) to be made, as well as causal explanation generation about theories of chemical phenomenon. The system can be expected to provide explanation to the following questions: (1) What are the chemical processes used? (2) What was their sequence of use? (3) What happened to each functional group (E.g. a particular nucleophile) in a reaction? With this promise, students can better understand the underlying chemical concepts. Their critical thinking can be improved especially from inspecting the cause-effect chain that explains an aspect of system behavior using only the ontological primitives. We envisage that the tool will improve a student's intuitive learning, in that it can lead to deeper and systematic understanding of chemical processes.

References

1. Forbus, K.: Qualitative Process Theory. Artificial Intelligence 24, 85–168 (1984)
2. Forbus, K., Whalley, P., Everett, J., Ureel, L., Brokowski, M., Baher, J., Kuehne, S.: CyclePad: An Articulate Virtual Laboratory for Engineering Thermodynamics, Artificial Intelligence Journal 114, 297–347 (1999)
3. Syed Mustapha, S.M.F.D., Pang, J.S., Zain, S.M.: 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)

4. Alicia Tang, Y.C., Zain, S.M., Abdul Rahman, N., Abdullah, R.: Modelling of Organic Reaction Mechanism Using Qualitative Process Theory: A Modeller's Experience. In: ATC. Conference on Intelligent System and Robotics, Advanced Technology Congress, Malaysia (2005)
5. Alicia Tang, Y.C., Syed Mustapha, S.M.F.D.: Representing SN1 Reaction Mechanism using the Qualitative Process Theory. In: Bailey-Kellogg, C., Kuipers, B. (eds.) International Workshop on Qualitative Reasoning 2006, Hanover, USA, pp. 137–147 (2006)
6. Alicia Tang, Y.C., Zain, S.M., Abdul Rahman, N., Abdullah, R.: Knowledge Representation and Simulation of Nucleophilic Substitution Reactions using Qualitative Reasoning Approach. In: IEEE TENCON, Hong Kong (2006) ISBN: 1-4244-0549-1
7. Alicia Tang, Y.C., Syed Mustapha, S.M.F.D., Abdullah, R., Zain, S.M., Abdul Rahman, N.: Towards Automating QPT Model Construction for Reaction Mechanisms Simulation. In: Price, C., Snooke, N. (eds.) International Workshop on Qualitative Reasoning 2007, Aberystwyth, United Kingdom (2007) (poster)
8. de Kleer, J., Brown, J.S.: Qualitative Physics Based on Confluences. Artificial Intelligence Journal 24, 7–83 (1984)
9. Kuipers, B.J.: Qualitative Simulation. Artificial Intelligence Journal 29, 289–338 (1986)