Representing S_N1 Reaction Mechanism Using the Qualitative Process Theory

Alicia Tang Y.C.
Department of Computer Science
College of Information Technology
University of Tenaga Nasional
KM 7 Jalan Kajang-Puchong
43009 Kajang, Selangor, Malaysia
aliciat@uniten.edu.my

S.M.F.D. Syed Mustapha[1]
Department of Artificial Intelligence
Faculty of Computer Science and Information Technology
University of Malaya
Lembah Pantai 50603
Kuala Lumpur, Malaysia

Abstract

Qualitative Reasoning (QR), albeit no longer a new research field in Artificial Intelligence (AI), its exploration in chemistry domain remains widely open. The application of Qualitative Process Theory (QPT) in organic reaction mechanisms is another new attempt of QR in chemistry. As far as learning reaction mechanisms is concerned, the ability to understand the underlying principle and the chemical intuition that goes with it are crucial. This paper presents an account for the design of qualitative models for the S_N1 reaction mechanism using QPT. The inclusion of chemical intuition into qualitative models will help to sharpen a student's reasoning ability. Our primary goal is to show that qualitative models constructed for the S_N1 mechanism can support many reaction formulas. The degree of generalization for this application domain is high, which qualifies the problem as a suitable domain for QR modeling and reasoning. To justify this, we included several reactions with different starting materials to demonstrate that QPT models developed for one type of reaction can be reused by a number of reaction formulas. The secondary goal is to use QPT as a tool to provide explanation in a more natural way through local propagation rules which is inherent in QPT. The study of cause effect interaction within these models outperformed passive learning, such as merely memorizing the steps and substrates involve in a reaction. In this work, a reasoning scenario that shows the instantiation of the QPT processes for simulating the S_N1 is also presented. To achieve this, we included the main process flow using QPT approach coupled with the functions and roles played by the reasoning and explanation modules.

1. Introduction

Chemistry problems have been tested as a domain for the QPT, an ontology proposed by Professor Kenneth Forbus and his group (Forbus 1984). GARP (Brederweg 1998), a qualitative reasoning engine implemented in SWI-Prolog, and VisiGARP (Bouwers and Brederweg 2001) that implements a graphical interface to GARP to enable user to inspect qualitative models are among examples of system applying QPT. Earlier works reported about applying QPT in inorganic chemistry are (Pang, Syed and Zain 2001, Syed, Pang and Zain 2002, Syed, Pang and Zain 2005). The application of QPT in the understanding and modeling of organic reaction mechanisms problems charts another milestone for the part taking of QR in chemistry domain. The application of QR in this new domain is appropriate since the nature of the problem is highly qualitative. Many chemistry students stumbled on the understanding of organic reaction mechanisms such the S_N1 (unimolecular nucleophilic substitution) and the S_N2 (bimolecular nucleophilic substitution). Our intention is not to train our chemistry students as modelers, rather when the representation and design is transformed to software could help to uphold student interests in learning reaction mechanism courses, and to improve their understanding and the development of reasoning skills. However, further discussion regarding the aspect of learning of the students is beyond the scope of the paper.

The differences between inorganic chemical reactions and organic reaction mechanisms from the qualitative modeling perspective are summarized in Section 2. Section 3 introduces organic reaction mechanisms. Model abstraction and the identification of general characteristics for the S_N1 are given in Section 4. This section also summarizes the simplification and generalization that can be done to the domain. The individual views needed for S_N1 simulation is listed in Section 5. Section 6 gives the workflow of application of QPT in S_N1. Section 7 presents a reasoning scenario that shows how reasoning is done with the instantiated processes for simulating the S_N1 mechanism. In Sections 8, we examine the cause effect chain via the QPT constructs that could be used to provide behavioral explanation for a given result. Problems and challenges in modeling the reaction mechanism are

---

[1] S.M.F.D. Syed Mustapha is presently attached to the British Telecom (Asian Research Center).

described in Section 9 while Section 10 concludes the work described in this paper.

2. Inorganic Chemical Reaction versus Organic Reaction Mechanism

Qualitative reasoning is concerned with reasoning from the basic principles of a given domain. It is more suited to domains that could meet two basic criteria: first being the problem description is qualitative in nature, and second is the degree of generalization should be high, meaning the model is a logical consequence to a large number of possible values. In the earlier work of QR in inorganic chemistry, it was shown that QPT could model for equilibrium state process (Syed, Pang and Zain 2002, Syed, Pang and Zain 2005), but our further investigation showed that it is limited by a small set of inorganic compound in which the case does not fulfill the second principle. Our further study showed that classifying the experiment types (Synthesis, Decomposition, Combustion, Single-displacement, Double-displacement, etc.) can only help to group reaction types but not to ease characteristic generalization. In many occasion, it is hard to get a pre-defined set of processes that can be applied throughout. Different substances give different behaviors even though they are in the same class or belong to the same reaction type. As a result, a chemical equation can only be constructed (hence simulated) after examining individual atoms and ions that formed those substances. We will give examples for reactions in the double-displacement class to justify the above statement. Double-displacement reactions mostly involving two ionic compounds in an aqueous solution, and usually one of the products is a compound insoluble in water (precipitate), a gas or a slightly ionized compound. Typically the chemical equation is written as $AB + CD \rightarrow AD + CB$, i.e. it involves an exchange of positive and negative groups. Examples in (i) show several acid-base neutralization processes that follow general rules of double-displacement.

(i) Acid-base Neutralization:

$HCl(aq) + KOH(aq) \rightarrow KCl(aq) + H_2O(l)$

$H_2SO_4(aq) + Ca(OH)_2(aq) \rightarrow CaSO_4(aq) + 2H_2O(l)$

$HCl(aq) + NaOH(aq) \rightarrow NaCl(aq) + H_2O(l)$

In most cases, the products can be determined from the knowledge of the ionic charges of the compounds. However, there are still many cases do not follow general principles. Some are given in the following examples (see (ii)). By principle, they should give two and only two products but they give three products and sometimes just one.

(ii) Double-displacement reactions that produced three products:

$NH_4Cl + NaOH \rightarrow NaCl + NH_{3(g)} + H_2O$

$H_2SO_4 + Na_2CO_3 \rightarrow Na_2SO_4 + H_2O + CO_2(g)$

$K_2CO_3(aq) + 2HNO_3(aq) \rightarrow 2KNO_3(aq) + H_2O(l) + CO_2(g)$

We take another example: $CaO + CO_2$ is supposed to give $CaO_2 + CO$ as products, since $CaO + CO_2$ has the format of $AB + CD$ but the reaction $CaO + CO_2 \rightarrow CaO_2 + CO$ will never occur. The product is CaO_3 instead. This is because even it has the format of the left-hand side of a double-displacement reaction but it is actually a combination reaction because both compounds contain oxygen gas. Again, it is the substance that decides what process in a reaction. A reasoner that based solely on ground principle will not be able to predict this outcome. In inorganic chemistry, when an individual view is identified as 'ion', one still needs to look at what is the specific ion (e.g. Cl^-, Mg^{2+}). As we can see, the prediction of inorganic reactions requires performing analysis down to very elementary level, at the price of having massive pre-coded knowledge-base. The system will eventually become a database approach, to store as many specific cases as possible, without which the system is unable to return reliable answers.

On the other hand, organic reaction mechanism involves the study of electrons movement, in which a bond is being made or broken. The nature of the problem is qualitative. 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? The set of general behavior and properties are easier to identify when study the organic reaction mechanism (explained in Section 4). In addition, automatic construction of individual views is also possible through recognizing the reacting species as either a nucleophile or an electrophile (see Section 5).

3. Organic Reaction Mechanisms

The study of organic reaction mechanisms normally includes the examination of the molecular structure and the chemical bonding of the starting material or substrate. A reaction mechanism describes a step-by-step sequence of reactions by which overall chemical change occurs, going from the starting material or substrate to the final product and these chemical changes occur by specific routes. Even though the number of known organic compounds is more than 10 millions but they belong to a relatively few structural types and that are even fewer reaction types than structural types (Robert 1977, Audrey & Philippa 2000, William 2000). 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 order to characterize and reason with instances of functional units, qualitative representations are tested and found to be appropriate. In which the individual views can be limited to either an electron rich or an electron poor species. As such, there is no need to study the unique properties of each reactant.

3.1 Organic Synthesis Planning System

LHASA (LHASA homepage, http://lhasa.harvard.edu/) is one of the first computer programs developed for planning of organic syntheses. The LHASA program suite is a sophisticated expert system that has been under development at Harvard since 1969. It is an expert system using database of retro-reactions (transforms). During a synthesis work, time is spent on finding the retrosynthetic route based on some searching algorithms. There are some associated problems with 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. In addition, since the modules (e.g. retrosynthetic and toxicology) were not dynamically updated when new reactions were added the modules slowly slipped out of date as new reactions were discovered. The development of the system is based on traditionally knowledge-based approach. LHASA technology relies on chemistry knowledge bases written in a proprietary language called CHMTRN.

3.2 Chemist Way of Modeling Organic Reaction Mechanism

Reaction mechanisms are normally used to explain why certain compounds were made. If one does not understand how X is changed to Y (the mechanisms used), the creation of Y cannot be explained. An organic chemist usually will look into the reaction mechanism to help explain the outcome of a reaction. When the organic chemists want to create a novel compound, they would first draw the reactant structure and then drawing the structure of the compound they want to create. With their chemistry knowledge and chemical intuition, they then work out possible mechanisms from reactant to product. In this scenario, the chemists are doing organic synthesis by following the mechanisms they proposed. Most of the time, the organic chemists could work out the mechanisms by only using common sense developed from their chemical intuition and knowledge. This is a very suitable field for QPT as common sense reasoning and qualitative description are necessitated. We will discuss in the following section the modeling of a type of nucleophilic substitution reaction called the S_N1.

4. Representing S_N1 Characteristics Using QPT

In representing S_N1 using the QPT ontology, the first task was the identification of chemical properties and parameters (*quantities*) of the reaction mechanism. It is important to note that there are many properties that a substrate (e.g. an alcohol) can hold thus performing model abstraction is essential. In order to conduct the best fit mapping between the general behavior of the S_N1 and the QPT, general behavior for the substrates must be first sought. We will explain S_N1 characteristics and their mapping to QPT constructs in the following subsection. Refer to (Forbus 1984) for the full account of QPT.

4.1 S_N1 Chemical Characteristics and Behaviors Abstraction

In this section, we discuss one type of reaction which involves an *OH* (Hydroxyl) functional group transformation. The mechanism used was S_N1. The 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 a proton (the H^+). This is to make the O^+H_2 as a good leaving group, in order to break the bond between the *C* and the O^+H_2. Once broken, a carbocation will be produced. In the second stage of the mechanism, the incoming nucleophile (X^-) can bond to the carbocation to form a neutral and stable final product. Our design is based on the following general reaction formula:

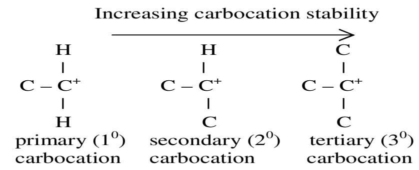
R-OH	+	HX	→	R-X	+	H_2O
Alcohol		Hydrogen halide		Alkyl halide		Water

'X' is any Halogen (group VII of periodic table), e.g. Cl, Br, etc., while 'R' can take the forms CH_3, CH_3CH_2, CH_3CH_3CH, or $(CH_3)_3C$.

We will use the following specific equation for the model construction activity described in the next few sections.

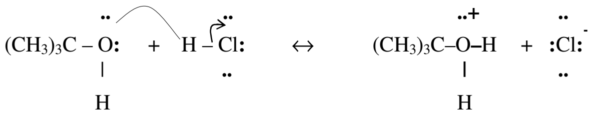
$(CH_3)_3C\text{-}OH + HCl \rightarrow (CH_3)_3C\text{-}Cl + H_2O$ Equation (1)

The S_N1 necessitates an intermediate product called carbocation. Below gives the stability of various structures of carbocations under S_N1 reaction.



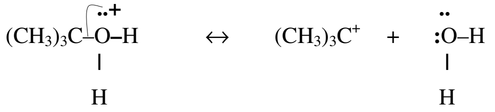
In order visualize the changes made to the individuals, the thought processes underlying the Equation (1) can be explicitly stated in a series of steps as follows.

Step 1: Protonation of *tert-Butyl alcohol* to give an *oxonium ion.* This is a make bond process.



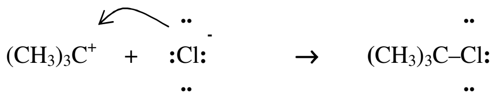
tert-butyl alcohol *hydrogen chloride* *tert-butyloxonium ion* *chloride ion*

Step 2: Dissociation of *tert-butyloxonium* ion to give a *carbocation*. This is a break bond process.



tert-butyloxonium *tert-butyl cation* *water*

Step 3: Capturing of *tert-butyl cation* by *chloride ion*. This is also a make bond process.



tert-butyl cation *chloride ion* *tert-butyl chloride*

The above three steps will be modeled as QPT processes. Essential chemical properties and characteristics of each process are given below:

- *Process I: Protonation.* There is a proton (H^+, in our case) serving as an electrophile (electron seeking species), and a non-bonded electron pair on O. The OH (not a good leaving group) is also needed. This process produces $–O^+H_2$ which is unstable so that the next process can begin.
- *Process II: Dissociation.* This describes the bond cleavage in $(CH_3)_3C\text{-}O^+H_2$. This process happens when the O in $(CH_3)_3C\text{-}O^+H_2$ is unstable since there are three covalent bonds (valency for oxygen is two), and there should not be a positive charge on O. At the end of this step, it will produce H_2O (a stable, neutral molecule) and R^+ (the carbocation intermediate).
- *Process III: Capturing of carbocation by halide anion.* This process finishes the S_N1 by producing $R\text{-}X$. This step entails the formation of a covalent bond between the X^- (nucleophile) and the R^+ (electrophile). The process is called upon since the reacting species are still unstable (they are charged species).

A 'mechanism' describes how a reaction takes place by showing what is happening to valence electrons during the making and breaking of bonds. Based on this definition, we have identified three main *quantities*, namely, charges, non-bonded electron pairs and number of covalent bonds. We have included the above chemical knowledge and intuition into our QPT models (Figures 4 – 6).

4.2 Comparing Two Reaction Equations for Properties Generalization

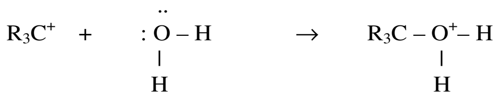
Our earlier statement about 'generalization is easier in organic reaction mechanism' is justified here. When we examine the behavior of another equation (see Equation 2), it is found that even though the substrates used in both reactions are different but the processes designed for Equation (1) can be reused by Equation (2), and other reaction formulas that use S_N1. These are briefly explained below, with supporting facts and data given in Table 1 and Table 2. The preparation of $(CH_3)_3COH$ (an alcohol) can be explained using the S_N1, as shown below:

$(CH_3)_3CBr + H_2O \rightarrow (CH_3)_3COH + HBr$ Equation (2)
Alkyl halide Alcohol

The sequence of steps that occurred during the conversion from alkyl halide to alcohol is depicted as under:

1st step: Dissociation

2nd step: Reaction with water



3rd step: Fast acid-base reaction

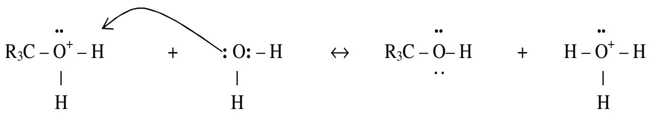


Table 1: The 'protonation' process of S_N1 (Equation 1) for the production of alkyl halide. The process takes place between a nucleophile and an electrophile.

Nucleophile (the 'O' of OH)	**Before**	**After**	**Remarks**	**Electrophile** (H^+)	**Before**	**After**	**Remarks**
Charge	Neutral	Positive	Still unstable	**Charge**	Positive	Neutral	Stable
No. of covalent bond	2	3	Still unstable	**No. of covalent bond**	0	1	n.a
Non-bonded electron pair	2	1	Not maximum pair yet	**Non-bonded electron pair**	0	0	No change

Table 2: The 'reacts with water' process of S_N1 (Equation 2) for the formation of alcohol.

Nucleophile (the 'O' of OH_2)	**Before**	**After**	**Remarks**	**Electrophile** (C^+)	**Before**	**After**	**Remarks**
Charge	Neutral	Positive	Unstable	**Charge**	Positive	Neutral	Stable
No. of covalent bond	2	3	More than maximum bond allowed	**No. of covalent bond**	3	4	'4' will put C in stable state
Non-bonded electron pair	2	1	Not the maximum yet	**Non-bonded electron pair**	0	0	No change

When the numerical data in the above tables are analyzed, the following functional dependency and effect propagation can be applied throughout. In QPT, they are represented as *qualitative proportionalities* (a) – (f).

non-bonded-electron-pair(O) P_{-}^{+} no-of-bond(O) …(a)

charges(O) P_{+}^{-} non-bonded-electron-pair(O) …(b)

non-bonded-electron-pair(H^+) P no-of-bond(H^+) …(c)

charges(H^+) P_{-}^{+} no-of-bond(H^+) …(d)

non-bonded-electron-pair(C^+) P no-of-bond(C^+) …(e)

charges(C^+) P_{-}^{+} no-of-bond(C^+) …(f)

We have extended the qualitative proportionality ('P') of the QPT to include extra symbols that gives rise to a set of four as shown below. Interpretation is more straightforward with these four relationships. Let $P_{-}^{+} \cong \alpha Q_{-}^{+}$.

Y P_{+}^{+} X	to denote 'Y increases as X increases'
Y P_{-}^{+} X	to denote 'Y decreases as X increases'
Y P_{-}^{-} X	to denote 'Y decreases as X decreases',
Y P_{+}^{-} X	to denote 'Y increases as X decreases'

The meanings of (a) to (f) are explained below:

- In both cases, an increase in *no-of-bond* of the nucleophile will cause a decrease in its *non-bonded-electron-pair* (see (a)). This will in turn increase the charges of the affecting species either from neutral to positive or from negative to neutral (refer (b)).
- Notice that the charge of electrophile is neutral in either of the processes (shown in equations (d) and (f)). For example, regardless of whether it is C^+ or H^+, so long as it is an electrophile then it will demonstrate same chemical behavior and properties (equations (c) – (f)).

5. Individual Views

Individual views are used to model the behavior of *individuals* (objects), and to provide explanation about their general characteristics. Individual views constructed for Equation (1) are the following:

- Individual-View Proton (e.g. H^+)
  (An electrophile used by step 1)
- Individual-View Alcohol (e.g. $(CH_3)_3C\text{-}OH$)
  (A nucleophile used by step 1)
- Individual-View Oxonium ion (e.g. $(CH_3)_3C\text{-}O^+H_2$)
  ('C' is δ+, 'O' is a δ- & they are used by step 2)
- Individual-View Halide-Ion (e.g. Cl^-)
  (A nucleophile used by step 3)
- Individual-View Carbocation (e.g. $(CH_3)_3C^+$)
  (An electrophile used by step 3)

A nucleophile (electron-rich species) can donate a lone pair of electrons to an electrophile (electron-poor species) An electrophile will accept electrons in order to fill up their valence shell. The symbol δ+ refers to a partial positive charge species while δ- symbolizes partial negative charge species that has a tendency to pull electrons towards it (in order to break a bond). With the above chemical properties, two templates for individual views have been designed, as shown in Figure 1 and Figure 2 respectively.

Individual-View 'Electrophile'

Individuals
p ;a piece-of-stuff

Preconditions
electrophile(p)

Relations
non-bonded-electron-pair(p) P no-of-bond(p)

(a)

Individual-View 'Charged-electrophile'
Individuals
p ;a piece-of-stuff, e.g. H^+
Quantity-Conditions
charges(p, positive)
Relations
charges(p) P_{-}^{+} no-of-bond (p)

(b)

Individual-View 'Neutral-electrophile'
Individuals
p ;a piece-of-stuff, e.g. the C next to the O
Quantity-Conditions
charges(p, neutral)
Relations
charges(p) P_{+}^{-} no-of-bond (p)

(c)

Figure 1 – An 'electrophile' view described using QPT

Individual-View 'Nucleophile'
Individuals
p ;a piece-of-stuff

Preconditions
nucleophile (p)

Relations
D_s[charges(p)]= 1
non-bonded-electron-pair(p) P_{-}^{+} no-of-bond(p)
charges(p) P_{+}^{-} non-bonded-electron-pair(p)

(a)

Individual-View 'Charged-nucleophile'
Individuals
p ;a piece-of-stuff e.g. Cl^-
Quantity-Conditions
charges(p, negative)
A_m[non-bonded-electron-pair(p)] ≥ ONE

(b)

Individual-View 'Neutral-nucleophile'
Individuals
p ;a piece-of-stuff e.g. the O in OH_2
Quantity-Conditions
A_m[non-bonded-electron-pair(p)] ≥ ONE
charges(p, neutral)

(c)

Figure 2 – A 'nucleophile' view described using QPT

The constructed views cater for both the charged and neutral nucleophile/electrophile. In some reactions, the reacting species involved could be the *C* (neutral electrophile) rather than the C^+ and the H^+ (charged electrophile) as used in the above example. Also, a nucleophile can be charged (e.g. Cl^-) beside the alcohol oxygen (the *O*, which is a neutral nucleophile) we demonstrated above. The above specifications can serve as generic views for any functional group.

6. Workflow of Application of QPT in S_N1

The overall flow is depicted in Figure 3 (a). There are two main components in the framework, namely the reasoning and the explanation modules. These two components are further illustrated in Figure 3 (b) and Figure 3 (c).

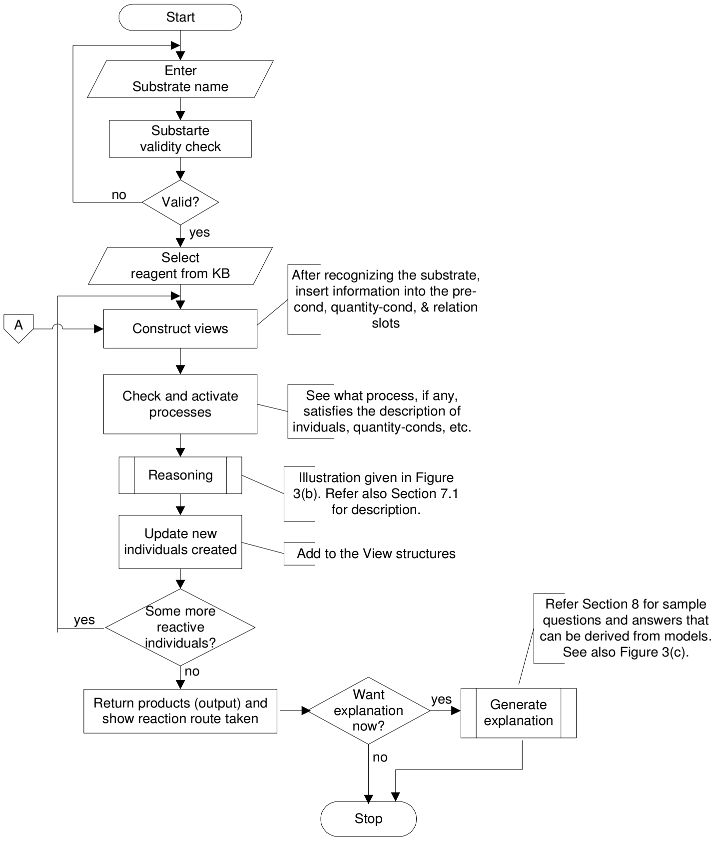


Figure 3(a) – Workflow for S_N1 simulation using QPT

The architecture is 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 Individual Structure (IS). Next, it is the checking of what processes can be used. A candidate process is one that satisfies quantity-conditions and pre-conditions. Individuals needed by the process must also available in the IS. When a process is in active state, reasoning will begin (reasoning details are given in Section 7.1, showing how a process is activated and switched to the next, and so on). Briefly, the reasoning engine will keep track of the values of states changing of the quantities being affected, starting from the first process until the entire reaction ends. A process will stop when the statements in its quantity-condition slot are invalid. When a process is applied, some individuals may cease (become a new individual). Therefore, updating of the IS is necessary. If there is a stable product produced, it will be stored separately for future retrieval. If there are still reactive units (charged species or species that still have not completed their valences), the reasoning process will be repeated, acting on other process instances. The entire reaction will end when there is no more views instances. When a reaction ends, outputs are displayed, together with all the steps/processes taken to produce the outputs. This task is rather straightforward since all the processes that have been applied are recorded with their sequence of use, and the changes made to each individual (described in term of functional dependency among quantities) are also recorded. 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 executed. An account of explanation generation is given in Section 8.

Using this architecture, for each reaction, three main outputs can be derived. These include: (1) The products (2) The steps/processes taken (3) Behavioral explanation generation. We will explain in Sections 7 and 8 how each of these deliverables can be obtained.

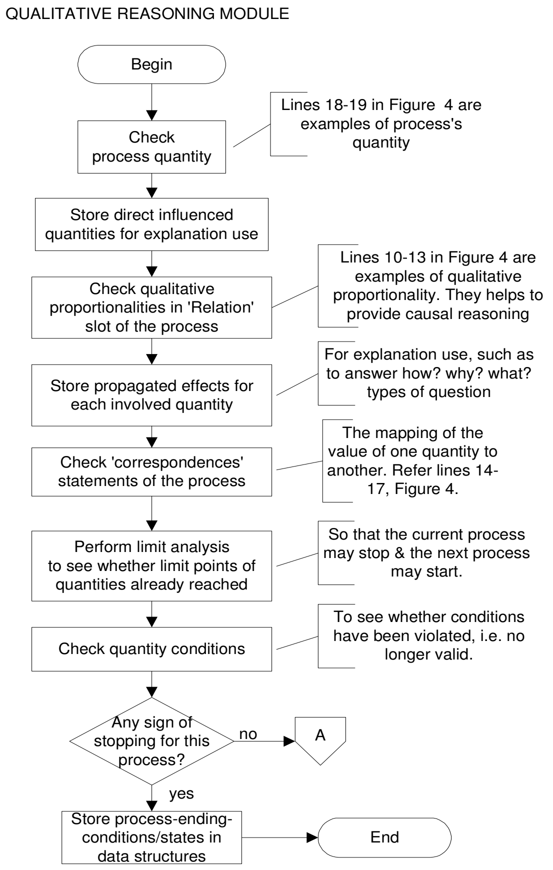


Figure 3(b) – Main steps in the qualitative reasoning module

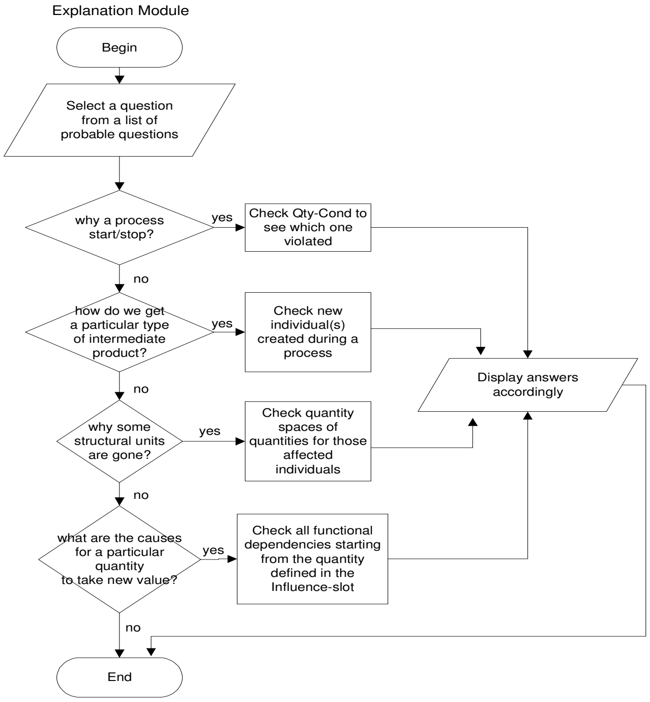


Figure 3(c) – A sample of questions showing how explanations can be generated

7. Qualitative Reasoning and Simulation

This section shows the simulation results, in particular, how reasoning is done with the instantiated QPT processes for simulating the S_N1 mechanism. Note: all works are based on Equation (1), and the three QPT processes (Figure 4 – Figure 6).

7.1 Simulation and Results

Recall that when an alcohol is used as the substrate and a hydrogen halide as the reagent to produce alkyl halide, S_N1 mechanism is necessitated. We will show how the qualitative models can reproduce S_N1. Initially, there are 3 species (candidate *individual instances*): a proton, the chlorine ion, and the alcohol. Simulation begins by calling up the *Protonation* process (Figure 4) since the *individuals* it needs are available in the view structure. The statements in *quantity-conditions* also satisfied (Lines 6 – 7), which speak for "there needs a proton and alcohol oxygen with at least one pair of non-bonded electron to be donated to the proton in order to make a bond". This is the first step of simulating the S_N1 for the reaction formula $(CH_3)_3C–OH + HCl \rightarrow (CH_3)_3C–Cl + H_2O$.

This process will directly influence the charges of the proton and the number of covalent bond on the *O*, defined as two direct influence statements using the I+/I- notation of the QPT (Lines 18 & 19). These effects will propagate to other dependent quantities. For example, the number of non-bonded electrons will decrease when more covalent bonds are made on the *O* via the inverse qualitative proportionality defined in Line 10. When the non-bonded electron pair of *O* decreases, the charges on *O* will increase (Line 11). This will make the *O* a positively charged species and having three covalent bonds (unstable). When the *O* is protonated, the *H* is no longer positively charged (explained in Line 12 & Line 16), thus violating the statement in the *quantity-conditions* slot. The new quantity created by this process is the *tert*-butyloxonium ion (Line 17). All values assigned to each individual are taken from the quantity spaces (Table 3) by the limit analyzer that keeps track of the current values of each quantity and the direction of change.

At this point, the oxonium ion *individuals* required by the *Dissociation* process (Figure 5) are available, the simulation thus continues by switching to the second step of the mechanism. Interestingly, the two species come from the same compound. To emphasize this we include *bond-between(C, O)* in the *Preconditions* slot (Line 3) which reads as "there exists a bond between the carbon and the oxygen atoms." This process 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 and having three covalent bond (valency for oxygen is two). Changes that propagated via functional dependencies among quantities are: The acceptance of two electrons from the dissociation activity will neutralize the *O* of O^+H_2 (Line 9 & Line 10), hence the formation of water molecule (refer *correspondence* statement, Line 16). At this point, both the conditions in *quantity-condition*s slot are no longer valid. On the other hand, donation of electrons will cause the charges on *C* becoming positive (in this case, it is a tertiary carbocation). This in turn caused a decrease in the number of covalent bond formed on it (Line 8). Atom *C* (of the carbocation) is now unstable and is reactive.

Since the carbocation $((CH_3)_3C^+)$ and the chlorine ion (Cl^-) are now left in the solution, and the conditions required to run the process 'Capturing of carbocation by anion' (Figure 6) are all met, this qualifies it as the third process in the reaction. The start of this process can be explained by the incomplete octet of carbocation and chloride ion and it stops due to the production of most stable species where both ions completed their valence. Lines 8–9 describe the following scenario: The increase of charge on Cl^- at the *direct influence* slot (Line 15) propagates its effect to bring about the reduction of the number of non-bonded electrons and further affecting the number of covalent bond on it. In this case, the number of covalent bond is being increased. As for the other reacting species, the formation-activity necessitates an increase in the number of covalent bond for the *C*. When this is coupled with the *correspondence* statement (Line 11), the atom *C* will regain its maximum bonds. The process ends as *Cl* and *C* are both in neutral state (their valences having been completed).

At this point the view structure is empty and the entire reaction deemed to stop. The final products are alkyl chloride $((CH_3)_3C–Cl)$ and water molecule (H_2O) which are very stable. While the sequence of process activation are *Protonation*, *Dissociation*, followed by *Capturing of carbocation by chloride ion*. These three steps (reaction route) can be used to explain the overall chemical change occurred, and to describe the S_N1 that done to Equation (1).

Process: Protonation (($(CH_3)_3C$-OH) protonated by H^+)
Individuals
1. H ;*hydrogen ion (the H^+ from hydrogen halide)*
2. O ;*alcohol oxygen that has extra pair of non-bonded electrons*

Preconditions
3. A_m [no-of-bond(O)] = TWO
4. is_reactive($(CH_3)_3C$-OH) ;*KB will tell what are not reactive*
5. leaving_group(OH, poor) ;*check KB for good/poor leaving group*

Quantity-Conditions
6. A_m[non-bonded-electron-pair(O)] $\geq$ A_m[min-electron-pair(O)]
7. charges(H, positive)

Relations
8. D_s[charges(H)]= -1
9. D_s[charges(O)]= 1
10. non-bonded-electron-pair(O) P_-^+ no-of-bond(O)
11. charges(O) P_+^- non-bonded-electron-pair(O)
12. no-of-bond(H) P_+^- charges(H)
13. formation($(CH_3)_3C$-O^+H_2) P_+^+ bond-activity(O)
14. correspondence((non-bonded-electron-pair(O), ONE),
    (no-of-bond(O), THREE))
15. correspondence((charges(O), positive),
    (non-bonded-electron-pair(O), ONE))
16. correspondence((no-of-bond(H), ONE),
    (charges(H), neutral))
17. correspondence((formation($(CH_3)_3C$-O^+H_2), ON),
    (bond-activity, make-a-bond))

Influences
18. I_+ (no-of-bond(O), A_m[protonation-activity])
19. I_- (charges(H), A_m[protonation-activity])
20. I_+ (bond-activity(O), A_m[protonation-activity])

Figure 4 – Protonation process described using QPT

Process: Dissociation
(Bond between the C and the O of ($(CH_3)_3C$-O^+H_2) is broken)
Individuals
1. O *;the O from* $-O^+H_2$ *(oxonium ion) of* $(CH_3)_3C$-O^+H_2
2. C *;the C from the* $(CH_3)_3C-$
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)]
6. charges(O, positive) *;oxonium ion is positive charge*
Relations
7. dissociate-activity P_{+}^{+} (A[no-of-bond(O)] – A[max-bond-allowed(O)])
8. no-of-bond(C) P_{-}^{+} charges($(CH_3)_3C$)
9. non-bonded-electron-pair(O) P_{+}^{-} no-of-bond(O)
10. charges(O) P_{-}^{+} non-bonded-electron-pair(O)
11. formation(H_2O) P_{+}^{-} bond-activity(C, O)
12. formation($(CH_3)_3C^+$) P_{+}^{-} bond-activity(C, O)

13. correspondence((no-of-bond(C), THREE),
(charges($(CH_3)_3C$), positive))
14. correspondence((charges(O), neutral),
(non-bonded-electron-pair(O), TWO))
;let $(CH_3)_3C^+$*, a carbocation be a quantity*
15. correspondence((formation($(CH_3)_3C^+$), ON),
(bond-activity(C, O), break-a-bond))
16. correspondence((formation(H_2O), ON),
(bond-activity(C, O), break-a-bond))
Influences
17. I_- (no-of-bond(O), A_m[dissociate-activity])
18. I_+ (charges(C), A_m[dissociate-activity])
19. I_- (bond-activity(C, O), A_m[dissociate-activity])

Figure 5 – Dissociation process described using QP theory

Process: Capturing of carbocation by anion
(Bond formation between Cl^- and the tertiary carbocation)
Individuals
1. C *;the carbocation* $(CH_3)_3C^+$
2. Cl *;chloride ion*
Quantity-Conditions
; yet to complete their octets
3. A_m[no-of-bond(C)] < A_m[max-bond-allowed(C)]
4. A_m[non-bonded-electron-pair(Cl)] > A_m[max-non-bonded-pair-allowed(Cl)]
5. charges(Cl, negative)
6. charges($(CH_3)_3C^+$, positive)
Relations
7. charges(C) P_{-}^{+} A_m[max-bond-allowed(C)] - A_m[no-of-bond(C)]
8. non-bonded-electron-pair(Cl) P_{-}^{+} charges(Cl)
9. no-of-bond(Cl) P_{+}^{-} non-bonded-electron-pair(Cl)
10. formation($(CH_3)_3C$-Cl) P_{+}^{+} bond-activity(C, Cl)
11. correspondence((charges(C), neutral),
(no-of-bond(C), FOUR))
12. correspondence((non-bonded-electron-pair(Cl), THREE),
(charges(Cl), neutral))
13. correspondence((formation($(CH_3)_3C$-Cl), ON),
(bond-activity, make-a-bond))
Influences
14. I_+ (no-of-bond(C), A_m[formation-activity])
15. I_+ (charges(Cl), A_m[formation-activity])
16. I_+ (bond-activity(C, Cl), A_m[formation-activity])

Figure 6 – Formation of alkyl halide modeled as a QPT process

Table 3: Quantities and associated quantity spaces for S_N1.

Quantity	**Quantity Space**	**Remarks**
charges	[negative, neutral, positive]	At any given time the charge of any atom is either –ve, neutral or +ve.
non-bonded-electron-pair	[ZERO, ONE, TWO, THREE, FOUR]	We only consider important atoms for S_N1. The 'FOUR' comes from halide ion, while 'ZERO' from the hydrogen ion.
no-of-bond	[ONE, TWO, THREE, FOUR]	The 'FOUR' comes from Carbon. 'ONE' for the Hydrogen.
bond-activity	[break-a-bond , none, make-a-bond]	During a process, a bond is either being made or broken. 'none' used as starting point.
nucleophile-reactivity	[charged, neutral]	Charged species is more reactive than a neutral counterpart.
electro-negativity	[LOW, HIGH]	Index for comparing two species in the same compound for their electro-negativity levels.

8. Roles of Direct and Indirect Influences in Providing Explanation

When S_N1 is described in qualitative terms, the 'causality' notion is inherent in the model. This helps to provide behavioral explanation. For example, given two qualitative proportionalities (*qp1* and *qp2*):

non-bonded-electron-pair(O) P_{-}^{+} no-of-bond(O) … *qp1*
charges(O) P_{+}^{-} non-bonded-electron-pair(O) … *qp2*

Explanations that can be derived from the above are: The number of non-bonded electron will decrease when more covalent bonds are made on the *O* atom (via the inverse qualitative proportionality defined in *qp1*). In *qp2*, when the *non-bonded-electron-pair* for the *O* decreases the *charges on it* will increase. These altogether can explain why the *O* becomes positive charge; simply it donated electrons to form a covalent bond.

8.1 Sample Questions and Answers

A set of hypothetical questions is given below. Answers derived from qualitative models are typed in italics. We provided one Q&A example for each QPT process.

- Question 1: "How could we explain the decrease in the non-bonded electron pairs of *O* in the *Protonation* process?" A possible conclusion would be "*We know that the immediate cause of the process is the number of covalent bond of O increases. This quantity in turn influenced the non-bonded electron pairs of O, and the influence is strictly decreasing through the inverse proportionality relationship*".
- Question 2: "What are the influencing factors that could affect the reduction of the number of covalent bond of C in the *Dissociation* process?" A conclusion

of this form is expected: "*Assuming the process of dissociation does happen then one of the covalent bond on the C will break, and this is caused by the following factors. First, the dissociation activity will directly influence the charges on C and this is strictly increasing. Next, the increase in the charges will decrease the number of covalent bond the C has*".

- Question 3: "In the third process, what is the main cause for the charges on the *C* (of carbocation) to reach neutral state?" A possible conclusion would be "*The formation activity will add a covalent bond to C, and this effect is propagated to the charges on C via the strictly decreasing notation. So, the main cause is the parameter called no-of-bond*".

We perceive this type of explanation to be more natural, exact (run-time generation), and the significant part is that these answers can be derived from the qualitative model.

9. Challenging Aspects in Modeling Reaction Mechanisms

The group faced two challenges during early part of the work. First, knowledge abstraction is difficult because many chemical commonsense are required. Humans tend to make a lot of assumption in their reasoning, and the chemical intuition required to suggest reaction mechanisms is largely depending on the commonsense knowledge one has. Second, the setting of inequality for the quantity-condition is challenging. Unlike other physical systems, the modeling of reaction mechanisms is not a straight forward task, that it is difficult to write (differential) equations to establish relationships among variables. For example one can easily set up equation $F = m .a$ to mean "*net-force* P^{+}_{+} *mass*" and "*acc* P^{+}_{-} *mass*". Or, in ecology context, "*growth-rate* P^{+}_{+} *recruitment*" and "*growth-rate* P^{+}_{-} *mortality*" to represent the expression *growth-rate* = *recruitment* – *mortality* (Salles, Robert and Pain 1996, Salles and Brederweg 1997). In the description of a heat-flow process, we could also easily identify that there should be a difference between the source and the sink temperatures (*source-temp* > *sink-temp* or *source-temp* – *sink-temp* > *zero*). However, this type of relationship is not clear in our problem. Nevertheless, we have identified and used the most fundamental aspect to trigger the series of steps in a reaction, namely the reacting species should be in their unstable states such as incomplete octets (valences have not been completed). Apart from the aforesaid challenges, there also found some useful tips during the modeling activity, especially in the mapping of the chemical properties to QPT primitives. Some are given below:

- In inorganic chemistry, a reaction takes place by dissolving the reactants to produce ions, and these ions have the same chemical properties such as "an increase in concentration will result in the increase of product formation, etc." but when organic compound is used as substrate, it is the *structure* of the compound that determines what reaction mechanism to apply in the synthesis route. Since we must study the structure of a compound and there could have many structural units in a given substrate, this suggests that more than one view is required, one for each units, (and not one view per substrate). Figure 7 depicts this idea. Even though all the substrates are alcohol but they are having different degrees of carbons and thus exhibiting varying reactivity under S_N1. This is what we meant by 'looking at structures' is needed. For example, in the 3^0 case, one view should be designed for the *OH* portion, and one more view for the $(CH_3)_3C$–. This is because it is the functional group that responsible for a reaction. Illustration follows:

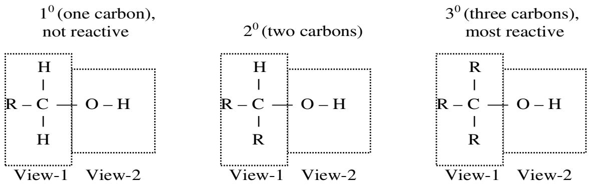


Figure 7 – Alcohol reactivity under S_N1

- A QPT *view* can contain more than one individual, one for each functional unit of the substrate. For example, the *tert*-butyloxonium ion ($(CH_3)_3C\text{-}O^+H_2$) is a view in our model with two individuals ($(CH_3)_3C$–, the alkyl part of the substrate, and the oxonium ion, $–O^+H_2$).
- Modeling of processes for S_N1 requires the inclusion of the 'equilibrium' phenomena. In our context, equilibrium is achieved when all of the reacting species reached the so called 'complete valence' state. This also serves as the stopping condition. On the other hand, if valency is incomplete, reasoning on the next process will proceed.

During the development of the qualitative models, we really have to think hard for which part of the expert knowledge to be included and how they could be related. This must be made very clear before they are cast into QPT modeling constructs. However, once the embodiment of these chemical commonsense in a QPT specification is done, it is proven robust and can be used to explain the underlying concepts very well.

10. Conclusions and Future Works

We have demonstrated that the QPT models are able to reproduce S_N1 reactions, and to generate explanations. We are still at the stage of modeling, and our ultimate goal is to develop a full system that could explain predicted result in a more natural way after running a simulation based on QPT models. As part of the ongoing work, we will include parameters such as 'bonding' and 'stability' in our models. For example, the carbocations can be classified as primary (1^o), secondary (2^o), and tertiary (3^o), according to the number of carbons that are directly attached to the positively charged carbon. Carbocations are stabilized by substituents that release or donate electron density to positively charged carbon. Alkyl groups (the 'R's) release electron density better than hydrogen substituents, so the more alkyl groups attached to the positively charged carbon, the more stable the carbocation. Currently, the models only support tertiary carbocations. In order to assess the generalization degree of the QPT models and templates we developed thus far, we will study other reaction mechanisms such as the electrophilic addition reaction.

References

Atkins, R. C., and Carey, F. A. 1997. *Organic Chemistry: A Brief Course;* McGraw-Hill Companies, Inc.

Miller, A., and Solomon, P. H. 2000. *Writing Reaction Mechanisms in Organic Chemistry*. Advanced Organic Chemistry Series, Harcourt, Academic Press.

Groutas, W. C. 2000. *Organic Reaction Mechanisms – Selected Problems and Solutions*; John Wiley & Sons, Inc.

William, C.G. 2000. Organic Reaction Mechanisms – Selected Problems and Solutions. John Wiley & Sons, Inc.

Forbus, K. 1984. Qualitative Process Theory. Artificial Intelligence 24 (1984) 85 – 125.

Pang, J.S.; Syed Mustapha, S.M.F.D.; and Zain, S.M. 2001. Preliminary studies on embedding qualitative reasoning into qualitative analysis and simulation in the laboratory. The Pacific Asian Conference on Intelligent Systems 2001 (PAIS 2001), Seoul, Korea, pp 230 – 236.

Syed Mustapha, S.M.F.D.; Pang, J.S.; and Zain, S.M. 2002. Application of Qualitative Process Theory to Qualitative Simulation and Analysis of Inorganic Chemical Reaction. In Proceedings of the Sixteenth International Workshop of Qualitative Reasoning, Barcelona, Spain.

Syed Mustapha, S.M.F.D.; Pang, J.S.; and Zain, S.M. 2005. 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.

LHASA group at Harvard University homepage, http://lhasa.harvard.edu/

Bredeweg, B. 1992. Expertise in qualitative prediction of behavior. PhD thesis, University of Amsterdam, Amsterdam, The Netherlands.

Bouwers, A., and Bredeweg, B. 2001. VisiGarp: Graphical representation of qualitative simulation models. In Artificial Intelligence in Education: AI-ED in the wired and wireless future, IOS-Press/Ohmsha, Osaka, Japan, pp: 294 – 305.

Salles, P.; Robert, I.; and Pain, H. 1996. Qualitative Models in Ecology and Their Use in Intelligent Tutoring Systems. In Proceedings of the Tenth International Workshop of Qualitative Reasoning, 21 – 24 May, California, USA.

Salles, P., and Bredeweg, B. 1997. Building Qualitative Models in Ecology. In Proceedings of the Eleventh International Workshop on Qualitative Reasoning, 3 – 6 June, Tuscany, Italy.