

Satoh, Hiroko

Associate Professor, Material and Life Informatics, Principles of Informatics Research  
Division

### **Chemoinformatics - Towards Making a Guide to The Chemical Reactions' Complex World**

Chemical reactions are fundamental phenomena in nature. Much variety compounds have been generated by natural resources and synthesized artificially. Prediction of the reaction products from reactants and reaction condition is one of my research themes. This is one of the difficult problems in chemistry.

#### **Attempts to chart a course through the “universe of chemical reactions”**

Solving the chemical reaction prediction problem entails finding a solution from a huge number of possibilities of reactions that occur as a result of complicated interactions between several factors concerning structural and electronic properties of reactants, reagents, catalysts, and solvents, and conditions such as temperature, density, pressure, and reaction time. To find a solution, the space with broad diversity to be searched ought to be reduced. Hence, chemists must seek ways to reach the solution within an acceptable time by reducing the space in a rational way.

Chemoinformatics is a new discipline that has a possibility to reduce the size of the search space by using chemical information as well as informatics technologies. Chemoinformatics is an interdisciplinary field that aims at solving chemical problems by applying informatics methods. We have been attempting to chart a course through the “universe of chemical reaction” by using the chemoinformatics approaches.

My idea originated from considering chemical reactions as numerical data as well as learning from the cognitive and reasoning processes of chemists. We have a conceptual

scheme for it, whereby (1) chemical reactions are represented by numerical data on factors that control reactions, (2) reactions are classified based on the similarities of the factors, and (3) models are constructed for each group to predict reactions using the factors that distinguish the differences from the others. If the factors controlling reactions are considered as multi-dimensional data, chemical reactions that result from their interactions can be considered as non-linear relationships between these data in a multi-dimensional space. The purpose of classification is to know how the data distribute in the space as well as to determine their degree of contribution for each reaction group.

### **Reaction classification to view the universe of chemical reactions**

The task of classification based on simple physicochemical parameters was the first phase of this research. Simple physicochemical parameters were used as descriptors for the classification. The similarity of reactions, each of which was represented with a set of parameters, was investigated by using a self-organizing neural network called a Kohonen neural network. A notable point of this study is that it demonstrated that primitive classification is possible even with simple physicochemical parameters, such as  $\sigma$ -,  $\pi$ -charges, electronegativity, polarizability, and  $pK_a$  values, which were calculated in an empirical way based on topological relationships of atoms and bonds. This was an attempt to view the universe of chemical reactions in a physicochemically rational way.

Chemical reactions are phenomena of complex systems. To have success in predicting them, a rational transdisciplinary approach will be necessary. Chemoinformatics is a discipline that can act as guide to the complex world of chemical reactions.

(Interviewed and summarized by Tomoaki Yoshito)