A hypergraph and perspective approach to organic synthesis design

TERUAKI ITO*1 and SHUICHI FUKUDA2

1Department of Mechanical Engineering, The University of Tokushima, 2-1, Minami-Josanjima, Tokushima 770, Japan
2Department of Production, Information and Systems Engineering, Tokyo Metropolitan Institute of Technology, 6-6, Asahigaoka, Hino, Tokyo 191, Japan

Received January 1996 and accepted July 1996

Various computer programs for organic synthesis design greatly assist chemists in the task of synthesizing chemical compounds. Despite the value of such programs, however, their common and collective failing is that they disrupt the chemists' natural flow of thought, by interrupting their concentration on chemistry and repeatedly forcing them to attend to computing. This problem originates in the lack of a chemist's view among the various programs in the process of design. A new idea called hypergraph, with the concept of perspective in design, is proposed as a solution to the problem. The idea is designed to represent a dynamic perspective view over a graph-structured object, such as a chemical structure, and to support a perspective-oriented shift of view during organic synthesis design. An example of precursor generation for organic reaction design is presented to illustrate the desired capabilities of hypergraph during designing. The goal of the research is to create a modelling environment in which the processing of creative ideas is enhanced: for example, in organic synthesis design.

Keywords: Hypergraph, dynamic perspective, perspective-oriented shift of view, idea processing, organic synthesis, precursor generation

1. Introduction

With today's GUI drawing utilities, even a chemist who is a computer novice can easily create graphic formulae of chemical structures. Such graphic representations constitute a common gateway to access various computer programs that greatly assist the chemist in the task of synthesizing chemical compounds. Despite the value of such programs, however, their common disadvantage for creative thinking is that they disrupt the chemist's natural flow of thought (McCracken and Akscyn, 1984) by interrupting their concentration on chemistry and repeatedly forcing them to attend to computing. This kind of interruption is observed even if chemists are working on a single program such as organic synthesis design, because even such a program is also a combination of programs designed for various purposes.

This problem originates in the lack of the chemist's view among the various programs in the process of design. We propose a new idea called hypergraph (Ito and Fukuda, 1994) as a solution to this problem. The idea is designed to represent a dynamic perspective view and a perspective-oriented shift of view. The paper describes the concept with some examples in organic synthesis design using the hypergraph to demonstrate its capability. A system dealing with an object based on the hypergraph will provide an environment in which the computer follows the chemist's natural train of thought without interruption of the chemist's view.

2. Organic synthesis design

2.1. Organic synthesis design and computer-assisted systems

The purpose of organic synthesis design (Fieser and Fieser, 1957) is said to be either the creation of new chemical substances, such as pharmaceutical chemicals and agricultural materials, or the development of better synthesis route proposals for existing materials. Generally, synthesis planning of chemical substances requires huge expenditures, and is very labour intensive. Moreover, it is very difficult to make good, detailed plans at the beginning of organic synthesis design. The
development of new tools that assist chemists to find appropriate synthesis routes greatly attracts those who are engaged in either academic research or industrial work.

Some chemists have developed computer-assisted systems (Pensak and Corey, 1977; Wipke et al., 1978; Nakayama, 1991; Lindsay, 1993), which can organize chemists' knowledge including huge volumes of chemical reactions, and can generate synthesis trees by AI technologies. They have been proposed to provide computational support for the selection of starting materials appropriate for a given target molecule, to generate and evaluate synthetic routes from the selected materials to the target molecule, and to search for known synthetic reactions or methodologies analogous to each reaction step in the generated synthetic sequence.

When chemists use one of these systems, they are very likely to encounter the problems mentioned above, of having their natural flow of thought for the design constantly interrupted. But most of the research has been more related to such topics as knowledge extraction, knowledge representation, and so on. Less attention is paid to user interfaces on the part of program integration to avoid interruption and enhance creative thinking.

2.2. Chemical structures and chemical graphs

To describe the mechanism of how to deal with chemical compounds in these systems, an internal description of chemical compounds is briefly presented below, because hypergraph is developed from the general form of internal descriptions based on chemical graph expression.

Chemical structures can be described by several notations, such as empirical formulae, component formulae, molecular formulae, structural formulae and 3D models. The structural formula is the most popular expression, thanks to its ability to store information for the molecule and give basic knowledge about structural information, which can be extracted if necessary.

Molecular design systems (Hendrickson et al., 1981; Gasteiger, 1982) employ a mathematical expression of chemical structural formulae called a chemical graph (Smith, 1968; Lynch et al., 1971), which is based on graph theory. A chemical graph is expressed by nodes and edges. Each node of a chemical graph is given a number and a name. A node is described by indicating the types of atom, its connectivity to other nodes, the number of unsaturations (i.e. double bonds) from the node and, the number of hydrogen atoms attached to it. The expression is translated into an adjacency matrix in the computer system, and can be processed quite easily for structural manipulation with only a basic knowledge of chemistry. An example of a chemical graph expression is shown in Fig. 1.

2.3. Precursor generation for reaction design

The primary purpose of organic synthesis design is to design an organic reaction to generate a precursor candidate for a given target structure. Because a whole synthesis route design or synthesis route planning is a combination of a series of consecutive organic reactions, and even though chemists undertake various tasks even for a single organic synthesis design, procedures in the design can be classified into five phases. One cycle composed of these five phases corresponds to a single reaction design, or a single precursor generation if it is conducted in a reverse way to generate a precursor for a given target structure.

The precursor generation for organic synthesis design is presented based on this cycle composed of five phases.

Phase 1 of a reaction design is started by creating a target structure. The structure may be created by chemists using drawing utilities, or may just be assigned based on the result of previous design procedures.

Phase 2 is feature recognition, which studies a chemical structure, recognizes chemically meaningful

![Fig. 1. An example of chemical graph expression.](image)