Analysis of Metabolic Pathways by Graph Transformation⋆

Francesc Rosselló[1] and Gabriel Valiente[2]

[1] Department of Mathematics and Computer Science, Research Institute of Health Science (IUNICS), University of the Balearic Islands, E-07122 Palma de Mallorca
[2] Department of Software, Technical University of Catalonia, E-08034 Barcelona

Abstract. Biochemical pathways, such as metabolic, regulatory, and signal transduction pathways, constitute complex networks of functional and physical interactions between molecular species in the cell. They are represented in a natural way as graphs, with molecules as nodes and processes as arcs. In particular, metabolic pathways are represented as directed graphs, with the substrates, products, and enzymes as nodes and the chemical reactions catalyzed by the enzymes as arcs. In this paper, chemical reactions in a metabolic pathway are described by edge relabeling graph transformation rules, as explicit chemical reactions and also as implicit chemical reactions, in which the substrate chemical graph, together with a minimal set of edge relabeling operations, determines uniquely the product chemical graph. Further, the problem of constructing all pathways that can accomplish a given metabolic function of transforming a substrate chemical graph to a product chemical graph using a set of explicit chemical reactions, is stated as the problem of finding an appropriate set of sequences of chemical graph transformations from the substrate to the product, and the design of a graph transformation system for the analysis of metabolic pathways is described, which is based on a database of explicit chemical reactions, a database of metabolic pathways, and a chemical graph transformation system.

1 Introduction

Biochemical pathways, such as metabolic, regulatory, and signal transduction pathways, constitute complex networks of functional and physical interactions between molecular species in the cell. They are represented in a natural way as graphs, with molecules as nodes and processes as arcs. In particular, metabolic pathways are represented as directed graphs, with the substrates, products, and enzymes as nodes and the chemical reactions catalyzed by the enzymes as arcs.

Chemical descriptions in a metabolic pathway can be made at different levels of resolution: a molecular descriptor uniquely identifies a molecule in a chemical database; a molecular formula indicates the number of each type of atom

in a molecule; a constitutional formula or chemical graph also indicates which pairs of these atoms are bonded; and a structural formula also indicates those stereochemical distinctions that are required to uniquely identify a molecule.

In a detailed representation of metabolic pathways, at the level of the constitutional formula or the structural formula, structural change of chemical reactions can be modeled by superimposing the reactant and the product to match up the atoms and bonds that are unchanged in the transformation. A formalism called *imaginary transition structures* was introduced in [1–3] to model chemical reactions, where the chemical graphs representing the reactions' substrate and product are superimposed topologically, and the bonds are then distinguished and classified into three categories: *out-bonds* (bonds appearing only in the substrate molecules), *in-bonds* (bonds appearing only in the product molecules), and *par-bonds* (bonds appearing in both the substrate and the product molecules). Imaginary transition structures can be seen as double-pushout transformation rules [4] over chemical graphs: the left-hand side, context, and right-hand side are chemical graphs with set of labeled nodes corresponding to the atoms in its molecules; the left-hand side graph has edges representing out-bonds, the context graph has edges representing par-bonds, and the right-hand side graph has edges representing in-bonds. This is, essentially, the view of chemical reactions advocated in [5].
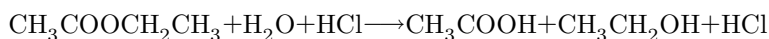
In this paper, chemical reactions in a metabolic pathway are described by edge relabeling graph transformation rules, as explicit chemical reactions and also as implicit chemical reactions, in which the substrate chemical graph, together with a minimal set of edge relabeling operations, determines uniquely the product chemical graph.

Further, the problem of constructing all pathways that can accomplish a given metabolic function of transforming a substrate chemical graph to a product chemical graph using a set of explicit chemical reactions, is stated as the problem of finding an appropriate set of sequences of chemical graph transformations from the substrate to the product, and the design of a graph transformation system for the analysis of metabolic pathways is described, which is based on a database of explicit chemical reactions, a database of metabolic pathways, and a chemical graph transformation system.

2   Modeling Chemical Reactions

A chemical reaction is the change produced by two or more molecules acting upon each other. In a chemical reaction, *substrate* molecules are transformed into *product* molecules, often in the presence of a *catalyst*. For simplicity, we shall assume henceforth that the catalysts of a reaction are part of both its substrate and product.

Example 1. Consider, for example, the acidic hydrolysis of ethyl acetate, which is described by the following equation:

$$CH_3COOCH_2CH_3 + H_2O + HCl \longrightarrow CH_3COOH + CH_3CH_2OH + HCl$$

The substrate of the reaction, ethyl acetate ($CH_3COOCH_2CH_3$) and water (H_2O), is transformed into acetic acid (CH_3COOH) and ethanol (CH_3CH_2OH) in the presence of a catalyst, hydrochloric acid (HCl).
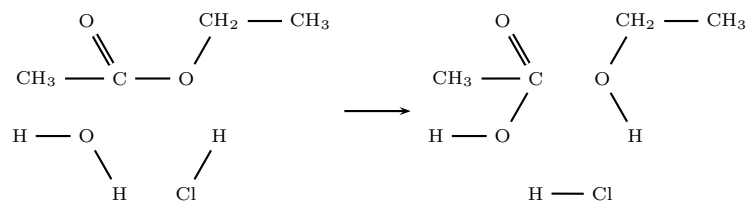
It is usual to describe molecules as graphs, with nodes representing the atoms, each one of them labeled by the name of the corresponding element, and edges representing the bonds, with a positive weight describing the type of bond (1 for a single bond, 2 for a double bond, etc.). A set of molecules is consequently described by the disjoint union of the graphs representing them. For reasons that will be clear below, we shall allow the existence in these graphs of one or more edges labeled 0: they should be seen as non-existent. Let us call these graphs representing sets of molecules *chemical graphs*.

Definition 1. *A **chemical graph** is a weighted graph (V, E, μ), where (V, E) is an undirected graph (without multiple edges or self-loops) all whose nodes are labeled by means of chemical elements, and $\mu : E \to \mathbb{N}$ is a weight function. The **valence** of a node in a chemical graph is the total weight of the edges incident to it.*

Chemical reactions consist of breaking, forming and changing bonds in sets of molecules. Therefore, a chemical reaction can be represented by the transformation of the chemical graph representing the reaction's substrate into the chemical graph representing the product. This transformation will satisfy a set of specific conditions. First, the number and type of the atoms in the substrate and the product must be the same, and therefore the transformation must induce the identity on the set of labeled nodes. Besides, and for simplicity, we shall restrict ourselves in this paper to chemical reactions where each individual atom has the same valence in the substrate and in the product: from the point of view of graphs, this corresponds to ensure that the total weight of edges incident to each node remains constant after the transformation. In a more general setting we could simply impose the conservation of the total number of valence electrons, which would correspond to the conservation of the sum of all edges' weights.

A systematic study of organic chemical reactions was made in [1–3], where a formalism called *imaginary transition structures* was introduced to model chemical reactions. The imaginary transition structure of a given reaction is a structural formula in which, using our language, the unweighted chemical graphs representing the reactions' substrate and product are superimposed topologically, and the bonds are then distinguished and classified into three categories: *out-bonds* (bonds appearing only in the substrate molecules), *in-bonds* (bonds appearing only in the product molecules), and *par-bonds* (bonds appearing in both the substrate and the product molecules).

Example 2. The acidic hydrolysis of ethyl acetate from Example 1 can also be depicted as a transformation between the chemical graph to the left (representing the substrate) and the chemical graph to the right (representing the product) in the following diagram:

O   CH$_2$ —— CH$_3$ O   CH$_2$ —— CH$_3$

CH$_3$ —— C —— O   CH$_3$ —— C   O

H —— O   H   H —— O   H

H   Cl   H —— Cl

The corresponding imaginary transition structure is shown in Fig. 1, where out-bonds are denoted by solid lines crossed by a bar, par-bonds are denoted by plain solid lines, and in-bonds are denoted by solid lines crossed by a small circle; for simplicity, we have assigned single nodes to groups of atoms like CH$_3$ and CH$_2$ that are not broken in the reaction.

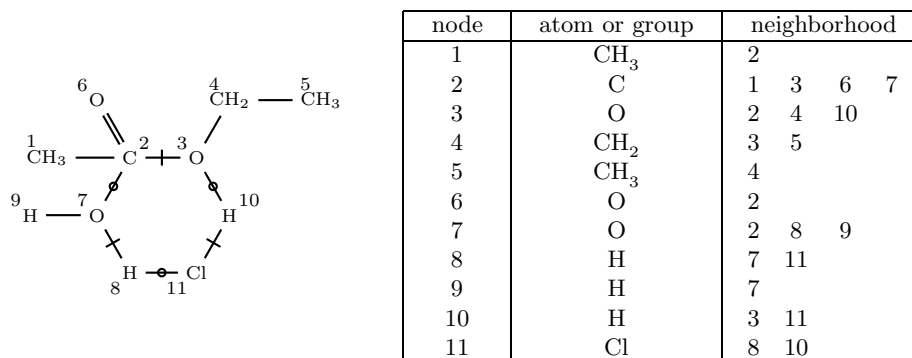| node	atom or group	neighborhood			
1	CH$_3$	2			
2	C	1	3	6	7
3	O	2	4	10	
4	CH$_2$	3	5		
5	CH$_3$	4			
6	O	2			
7	O	2	8	9	
8	H	7	11		
9	H	7			
10	H	3	11		
11	Cl	8	10		

Fig. 1. The imaginary transition structure of acidic hydrolysis of ethyl acetate

An imaginary transition structure can be seen as a double-pushout transformation rule [4] over chemical graphs. The left-hand side, context, and right-hand side are chemical graphs with set of labeled nodes corresponding to the atoms in its molecules; the left-hand side graph has edges representing out-bonds, the context graph has edges representing par-bonds, and the right-hand side graph has edges representing in-bonds. This is, essentially, the view of chemical reactions advocated in [5].

An imaginary transition structure can also be seen as a set of edge relabeling operations applied to a chemical graph. The substrate of the chemical reaction is represented by a chemical graph with set of labeled nodes corresponding to the atoms in its molecules and edges representing bonds that exist in the substrate or in the product of the reaction. The edges' weights are then assigned according to the definition of chemical graph given above: an edge is labeled 0 if the corresponding bond does not exist in the substrate (and then, by construction,

it *must exist* in the product), and if the bond exists in the substrate, then it is weighted according to its type. The chemical reaction is then simply described by relabeling of the edges in this graph: a bond existing in some substrate molecule that breaks in the product molecules is relabeled by 0; a new bond appearing in some product molecule is assigned the label corresponding to its type; a bond that exists both in a substrate molecule and in a product molecule but they are of a different type, is relabeled according to its new type; and, finally, labels of bonds that are not modified at all by the chemical reaction are not modified either.

This description of a chemical reaction, based on Fujita's imaginary transition structures, motivates the introduction of the notion of an *explicit chemical reaction*.

Definition 2. *An **explicit chemical reaction** is a structure (V, E, σ, π), where (V, E, σ) and (V, E, π) are chemical graphs, called the **substrate** and the **product** chemical graphs respectively, satisfying the following conditions:*

(i) There is no $e \in E$ such that $\sigma(e) = \pi(e) = 0$.
(ii) For every $v \in V$, if $e_1, \ldots, e_k$ are the edges incident to it, then

$$\sigma(e_1) + \cdots + \sigma(e_k) = \pi(e_1) + \cdots + \pi(e_k) \geqslant 1.$$

Every imaginary transition structure, and hence every chemical reaction, can be represented by means of an explicit chemical reaction (V, E, σ, π) with (V, E, σ) and (V, E, π) being the chemical graphs describing its substrate and product.

Example 3. Consider again the acidic hydrolysis of ethyl acetate from Examples 1 and 2. The diagram in Fig. 2 represents the explicit chemical reaction describing it: the left-hand side graph depicts the graph (V, E), and the weight functions σ and π are given in the right-hand side table.

Application of an explicit chemical reaction to a given chemical graph, consists of relabeling the substrate by the product within the given chemical graph.

Definition 3. *A chemical graph (V, E, μ) is a **subgraph** of a chemical graph (V', E', μ') if $V \subseteq V'$, $E \subseteq E'$ and for all edges $e \in E$, $\mu(e) \leqslant \mu'(e)$. An explicit chemical reaction (V, E, σ, π) **can be applied** to a chemical graph (V', E', μ) if (V, E, σ) is a subgraph of (V', E', μ). In such a case, the **application** of (V, E, σ, π) to (V', E', μ) results in a chemical graph (V', E', μ'), where $\mu'(e) = \mu(e)$ for all edges $e \in E' \setminus E$ and $\mu'(e) = \pi(e)$ for all edges $e \in E$.*

Now, in general, it will not be necessary to provide the full substrate and product edge weight functions σ and π of an explicit chemical reaction, to describe the corresponding edge labeling transformation. Indeed, it is enough to specify the substrate chemical graph and a minimal set of edge relabeling operations which, when applied to a graph taking into account conditions (i) and (ii) in Definition 2, determine uniquely the product chemical graph. Since the
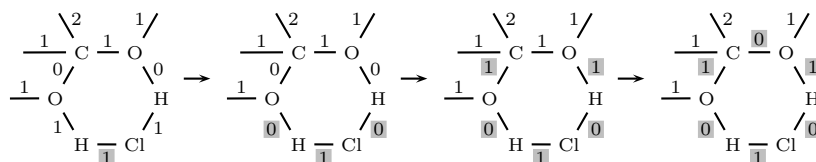
bond node—node	substrate weight	product weight
1—2	1	1
2—3	1	0
2—6	2	2
2—7	0	1
3—4	1	1
3—10	0	1
4—5	1	1
7—8	1	0
7—9	1	1
8—11	0	1
10—11	1	0

Fig. 2. The explicit chemical reaction of acidic hydrolysis of ethyl acetate

undirected graph underlying the substrate chemical graph is finite, such a minimal set of edge relabeling operations will always exist, although it need not be unique.

Example 4. Consider one more time the acidic hydrolysis of ethyl acetate. As it can be seen in the table given in Fig. 2, this chemical reaction corresponds to six edge relabeling operations (see also Example 2). Now, it can be easily checked that each one of these edge relabeling operations, together with the conservation of the atoms' valences and the structure of the underlying undirected graph, entails the other five ones and hence it describes completely the chemical reaction.

For instance, assume we relabel from 0 to 1 the edge 8—11, joining a hydrogen in the water molecule and the chlorine in the hydrochloric acid. To preserve these atoms' valences, edges 7—8 and 10—11 must be relabeled to 0. And then to preserve the oxygen and hydrogen atoms' valences involved in these two last edges, the only possibility is to relabel edges 2—7 and 3—10 to 1. And finally, in order to preserve the valences of the carbon and oxygen atoms involved in these two edges, it can be checked that the only possibility is to relabel edge 2—3 to 0: any other relabeling modifies the valence of other atoms bound to the carbon or the hydrogen.

This leads us to the definition of an *implicit chemical reaction.*

Definition 4. *An **implicit chemical reaction** is a finite set* $\mathbf{R}$ *of edge relabeling operations of the form*

$$r = (A_{r,1}, A_{r,2}, w_{r,s}, w_{r,p}),$$

with $A_{r,1}$ *and* $A_{r,2}$ *atomic elements and* $w_{r,s}, w_{r,p} \in \mathbb{N}$ *such that* $w_{r,s} \neq w_{r,p}$.

*Such an implicit chemical reaction **can be applied** to a chemical graph* (V, E, σ) *when the following conditions are satisfied:*

(a) *For every* $r \in \mathbf{R}$ *there is one edge* e_r *in* E *whose nodes have labels* $A_{r,1}$ *and* $A_{r,2}$, *whose weight is* $w_{r,s}$, *and such that* $w_{r,p}$ *is less than or equal to the valence of both these nodes.*

(b) *There is one, and only one, chemical graph* (V, E, τ), *with the same nodes and edges as* (V, E, σ), *such that*
- *all its nodes have the same valence as in* (V, E, σ);
- $\tau(e_r) = w_{r,p}$ *for every* $r \in \mathbf{R}$;
- *if* $\sigma(e) = 0$, *then* $\tau(e) \neq 0$.

The result of this application is, then, the chemical graph (V, E, τ), *and this application is said to **represent** the explicit chemical reaction* (V, E, σ, τ).

Notice that an implicit chemical reaction need not admit an application to a given chemical graph, even if it satisfies condition (a). And also that a given implicit chemical reaction may admit several applications to a given chemical graph, possibly giving different results. It could be interesting to study application and uniqueness conditions for implicit chemical reactions. But we shall not consider them here, as we are dealing with already existing chemical reactions, which we represent by means of an application to a suitably defined chemical graph of an implicit chemical reaction that has been determined beforehand.

3 Analyzing Metabolic Pathways

Metabolism can be regarded as a network of chemical reactions catalyzed by enzymes and connected via their substrates and products, and a metabolic pathway can be regarded as a coordinated sequence of chemical reactions [6]. The definition of a metabolic pathway is not exact, and most pathways constitute indeed highly intertwined cyclic networks. In a cell, a pathway's substrates are usually the products of another pathway, and there are junctions where pathways meet or cross [7]. For the purposes of this paper, we shall adopt the following definition.

Definition 5. *A **metabolic pathway** is a connected directed graph* (C, R), *where* C *is a set of chemical graphs and* $R \subset C \times C$ *is a set of explicit chemical reactions. The **substrate** of* (C, R) *is the set of chemical graphs* $S \subset C$ *such that for all* (V, E, π) *in* S, *there is no explicit chemical reaction of the form* (V, E, σ, π) *in* R. *The **product** of* (C, R) *is the set of chemical graphs* $P \subset C$ *such that for all* $(V, E, \sigma) \in P$, *there is no explicit chemical reaction of the form* (V, E, σ, π) *in* R.

The analysis of metabolic pathways is motivated by the rapidly increasing quantity of available information on metabolic pathways for different organisms. One of the most comprehensive sources of metabolic pathway data is [8]. There are also several databases on metabolic pathways, such as BRENDA [9], EcoCyc [10], KEGG [11], and WIT [12]. These databases contain hundreds of metabolic pathways and thousands of chemical reactions, and even the metabolic pathway for a small organism constitutes a large network. For instance, the proposed metabolic pathway for the bacterium *E. coli* consists of 436 compounds (substrates, products, and intermediate compounds) linked by 720 reactions [13].

One aspect of metabolic pathway analysis is *flux analysis*: the decomposition of a metabolic pathway into a complete set of nondecomposable steady state flux distributions. Two similar approaches to flux analysis are known which are based on the set of *elementary flux modes* [14] and on the set of *extreme pathways* [15]. In large metabolic networks, however, these approaches are hampered by the combinatorial explosion of possible routes: the maximal number of elementary flux modes in a metabolic pathway is exponential in the number of reactions, substrates, and products [16].

Another, complementary aspect of metabolic pathway analysis is *pathway synthesis*: the construction of all pathways that can accomplish a given metabolic function, which is: the transformation of a given set of substrates to a given set of products. Pathway synthesis belongs in pathway analysis, because it allows biologists and biochemists to contrast those metabolic pathways which exist in the cell for different organisms, against feasible metabolic pathways obtained by synthesis.

In pathway synthesis, much like in retrosynthetic analysis in organic chemistry [17, 18], the target chemical graph is subjected to a disconnection process, which corresponds to the reverse of a chemical reaction. As a result, the target chemical graph is transformed to a sequence of simpler chemical graphs in a stepwise manner, along a path that ultimately leads to simple chemical graphs. For a complex target chemical graph, some intermediate chemical graphs may undergo further retrosynthetic analysis. Thus, the repetition of this process eventually will result in a hierarchical synthesis tree for the target chemical graph.

In order to synthesize meaningful metabolic pathways, axioms on reaction pathways have been established in [19, 20], based on [21]. A first set of axioms, the *feasible reaction pathway* axioms, establish that (1) every product is totally produced by reactions represented in the pathway; (2) every substrate is totally consumed by reactions represented in the pathway; (3) intermediate compounds are entirely produced by previous reactions and completely consumed by subsequent reactions; (4) each reaction represented in the pathway is defined a priori; (5) the network representing the pathway is acyclic; and (6) at least one reaction represented in the pathway affects the activation of a substrate.

Among these axioms for feasible reaction pathways, (4) follows from Definition 5. The remaining ones are enforced by the following definition.

Definition 6. *A metabolic pathway (C, R) is said to be a **feasible reaction pathway** if it satisfies the following axioms:*

- *For each product chemical graph of the form (V, E, μ) in P, there is a set of explicit chemical reactions $R' \subseteq R$ of the form (V', E', σ', π') such that (V, E, μ) is a subgraph of the union of the product chemical graphs (V', E', π') in R'.*
- *For each substrate chemical graph of the form (V, E, μ) in S, there is a set of explicit chemical reactions $R' \subseteq R$ of the form (V', E', σ', π') such that (V, E, μ) is a subgraph of the union of the substrate chemical graphs (V', E', σ') in R'.*
- *For each chemical graph of the form (V, E, μ) in $C \setminus (S \cup P)$, there is a set of explicit chemical reactions $R' \subseteq R$ of the form (V', E', σ', π') such that (V, E, μ) is a subgraph of the union of the product chemical graphs (V', E', π') in R', and there is a set of explicit chemical reactions $R'' \subseteq R$ of the form $(V'', E'', \sigma'', \pi'')$ such that (V, E, μ) is a subgraph of the union of the substrate chemical graphs (V'', E'', σ'') in R''.*
- *(C, R) is acyclic.*
- *There is an explicit chemical reaction of the form (V, E, σ, π) in R, with (V, E, σ) a substrate chemical graph in $S \subset C$.*

A second set of axioms, the *combinatorially feasible reaction pathway* axioms, establish that (1) every product is represented in the network; (2) every substrate is represented in the network; (3) each reaction represented in the network is defined a priori; (4) every compound represented in the network has at least one path leading to a product of the network; (5) every compound represented in the network is a substrate or a product for at least one reaction represented in the network; (6) a substrate of any reaction represented in the network is a substrate of the network if it is not a product of any reaction represented in the network; and (7) each reaction represented in the network is either forward or backward, but not both.

Among these axioms for combinatorially feasible reaction pathways, (3) and (4) follow from Definition 5. The remaining ones are enforced by the following definition.

Definition 7. *Let S' and P' be fixed chemical graphs. A metabolic pathway (C, R) is said to be a **combinatorially feasible reaction pathway** with respect to S' and P' if it satisfies the following axioms:*

- *$P' \subseteq C$.*
- *$S' \subseteq C$.*
- *For each chemical graph of the form (V, E, μ) in C, there is an explicit chemical reaction of the form (V', E', σ', π') in R such that either (V, E, μ) is a subgraph of (V', E', σ'), or (V, E, μ) is a subgraph of (V', E', π').*
- *For each explicit chemical reaction of the form (V, E, σ, π) in R, (V, E, σ) is in the substrate S of (C, R) if there is no explicit chemical reaction of the form (V', E', σ', π') in R such that (V, E, σ) is a subgraph of (V', E', π').*
- *There are no two explicit chemical reactions of the form (V, E, σ, π) and (V, E, π, σ) in R.*

Now, in the analysis of metabolic pathways, the problem of constructing all pathways that can accomplish a given metabolic function can be stated as follows.

Problem 1 (Synthesis of metabolic pathways). Given a substrate chemical graph S', a product chemical graph P', and a set R' of explicit chemical reactions, find one or all feasible or combinatorially feasible metabolic pathways (C, R) with substrate $S = S'$, product $P = P'$, and set of explicit chemical reactions $R \subseteq R'$.

Since explicit chemical reactions are edge relabeling graph transformation rules, the metabolic pathway synthesis problem can be solved by graph transformation, as follows. Given a substrate chemical graph S', a product chemical graph P', and a set R' of explicit chemical reactions, a metabolic pathways (C, R) with substrate $S = S'$, product $P = P'$, and set of explicit chemical reactions $R \subseteq R'$ is given by a set of sequences of chemical graph transformations with substrate S and product P.

A graph transformation system for the analysis of metabolic pathways is being developed at the Technical University of Catalonia. The system is based on the following three main components:

- Database of explicit chemical reactions.
- Database of metabolic pathways.
- Chemical graph transformation system.

The efficient implementation of the chemical graph transformation system relies on the CANON method for labeling a molecular structure with canonical labels [22, 23], in which a molecular structure is treated as a graph with nodes (atoms) and edges (bonds), and each atom is given a unique numerical label on the basis of the topology of the molecular structure.

4   Conclusion

Chemical reactions in a metabolic pathway are described in this paper by edge relabeling graph transformation rules, both as explicit chemical reactions and as implicit chemical reactions, in which the substrate chemical graph, together with a minimal set of edge relabeling operations, determines uniquely the product chemical graph. On the basis of explicit chemical reactions, the problem of constructing all pathways that can accomplish a given metabolic function of transforming a substrate chemical graph to a product chemical graph using a set of explicit chemical reactions, is stated as the problem of finding an appropriate set of sequences of chemical graph transformations from the substrate to the product. The design of a graph transformation system for the analysis of metabolic pathways, based on a database of explicit chemical reactions, a database of metabolic pathways, and a chemical graph transformation system, is also described.

References

1. Fujita, S.: Description of organic reactions based on imaginary transition structures. Part 1–5. Journal of Chemical Information and Computer Sciences **26** (1986) 205–242
2. Fujita, S.: Description of organic reactions based on imaginary transition structures. Part 6–9. Journal of Chemical Information and Computer Sciences **27** (1987) 99–120
3. Fujita, S.: Computer-Oriented Representation of Organic Reactions. Yoshioka Shoten, Kyoto (2001)
4. Corradini, A., Montanari, U., Rossi, F., Ehrig, H., Heckel, R., Löwe, M.: Algebraic approaches to graph transformation. Part I: Basic concepts and double pushout approach. In Rozenberg, G., ed.: Handbook of Graph Grammars and Computing by Graph Transformation, Volume 1: Foundations. World Scientific (1997) 163–246
5. Benkö, G., Flamm, C., Stadler, P.F.: A graph-based top model of chemistry. Journal of Chemical Information and Computer Sciences **43** (2003) 1085–1093
6. Deville, Y., Gilbert, D., van Helden, J., Wodak, S.J.: An overview of data models for the analysis of biochemical pathways. Briefings in Bioinformatics **4** (2003) 246–259
7. Karp, P.D., Mavrovouniotis, M.L.: Representing, analyzing, and synthesizing biochemical pathways. IEEE Expert **9** (1994) 11–21
8. Michal, G., ed.: Biological Pathways: An Atlas of Biochemistry and Molecular Biology. John Wiley & Sons, New York (1999)
9. Schomburg, I., Chang, A., Schomburg, D.: BRENDA, enzyme data and metabolic information. Nucleic Acids Research **30** (2002) 47–49
10. Karp, P.D., Riley, M., Saier, M., Paulsen, I.T., Collado-Vides, J., Paley, S.M., Pellegrini-Toole, A., Bonavides, C., Gama-Castro, S.: The EcoCyc database. Nucleic Acids Research **30** (2002) 56–58
11. Kanehisa, M., Goto, S.: KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Research **28** (2000) 27–30
12. Overbeek, R., Larsen, N., Pusch, G.D., D'Souza, M., Selkov, E., Kyrpides, N., Fonstein, M., Maltsev, N., Selkov, E.: WIT: Integrated system for high-throughput genome sequence analysis and metabolic reconstruction. Nucleic Acids Research **28** (2000) 123–125
13. Edwards, J.S., Palsson, B.O.: The escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc. Natural Academy of Science, USA **97** (2000) 5528–5533
14. Schuster, S., Fell, D.A., Dandekar, T.A.: A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nature Biotechnology **18** (2000) 326–332
15. Schilling, C.H., Letscher, D., Palsson, B.O.: Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. Journal of Theoretical Biology **203** (2000) 229–248
16. Klamt, S., Stelling, J.: Combinatorial complexity of pathway analysis in metabolic networks. Molecular Biology Reports **29** (2002) 233–236
17. Corey, E.J., Cheng, X.M.: The Logic of Chemical Synthesis. John Wiley & Sons, New York (1995)
18. Serratosa, F., Xicart, J.: Organic Chemistry in Action: The Design of Organic Synthesis. 2nd edn. Number 51 in Studies in Organic Chemistry. Elsevier, Amsterdam (1996)

19. Seo, H., Lee, D.Y., Park, S., Fan, L.T., Shafie, S., Bertók, B., Friedler, F.: Graph-theoretical identification of pathways for biochemical reactions. Biotechnology Letters **23** (2001) 1551–1557
20. Fan, L.T., Bertók, B., Friedler, F.: A graph-theoretic method to identify candidate mechanisms for deriving the rate law of a catalytic reaction. Computers and Chemistry **26** (2002) 265–292
21. Friedler, F., Tarján, K., Huang, Y.W., Fan, L.T.: Graph-theoretic approach to process synthesis: Axioms and theorems. Chemical Engineering Science **47** (1992) 1973–1988
22. Weininger, D.: SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. Journal of Chemical Information and Computer Sciences **28** (1988) 31–36
23. Weininger, D., Weininger, A., Weininger, J.L.: SMILES. 2. Algorithm for generation of unique SMILES notation. Journal of Chemical Information and Computer Sciences **29** (1989) 97–101