Cell Signaling Networks Ontology

Takako Takai-Igarashi [1] and Riichiro Mizoguchi [2]
[1] Computer Science, Graduate School of Information Science and Technology, University of Tokyo
[2] The Institute of Scientific and Industrial Research, Osaka University

Keywords: ontology, cell signaling pathway, biological function

Summary
Although databases for cell signaling pathways include numbers of reaction data of the pathways, the reaction data cannot be used yet to deduce biological functions from them. For the deduction, we need systematic and consistent interpretation of biological functions of reactions in cell signaling pathways in the context of "information transmission". To address this issue, we have developed a functional ontology for cell signaling pathways, Cell Signaling Network Ontology (CSN-Ontology), which provides framework for the functional interpretation presenting some important concepts as *information, selectivity*, *movability*, and *signaling rules* including *passage of time*.

1. Introduction

Databases for cell signaling pathways represent functions of chemical changes in the pathways. The representation is absolutely true, because any reaction in organisms is nothing else but a chemical reaction. However, functions of cell signaling pathways, which are consequences of chains of the chemical reactions, do not belong to chemical but biological knowledge domain. The discrepancy of knowledge domain between functions of reactions and pathways is a particular feature of cell signaling pathways. That can be compared with metabolic pathways in which functions both of reactions and of pathways belong to chemical knowledge domain. This conceptual discrepancy prevents us from deducing pathway functions from information on reactions in cell signaling pathways. We need an ontology that bridges the gap between the two of knowledge domains. For instance, we need an ontology that provides us with an answer to a question like how "phosphorylation of TGF-beta receptor 1 by TGF-beta receptor 2" relates to its pathway function of "cell growth arrest". Ontology provides a clear structured framework for systematic, consistent, and shareable description of target phenomena. For building such ontologies, we need, in the first place, explication of implicit domain knowledge so as to obtain a framework that every phenomenon in a domain bases on. In the second place, we

have to demonstrate how to reconstruct the phenomena in computers based on the framework [1]. This kind of ontologies differs in purpose of development from those mainly for control vocabulary represented by Gene Ontology [2].

We developed an ontology, named CSN-ontology (Cell Signaling Networks Ontology), which provides a framework for systematization of functional knowledge in cell signaling pathways. The framework enables us to capture the target functions consistently in the context of "information transmission" which is the basis of biological functions in cell signaling pathways. If all the functions are consistency grounded on the basis of "information transmission", the functions will be systematically decomposed from pathways to individual reactions. Actually CSN-ontology tells us how to functionally relate "phosphorylation" with "cell growth arrest", bridging the gap between biological and chemical knowledge domains. In addition, the ontology abstracts some important concepts of cell signaling mechanisms which databases to date have failed to represent explicitly. They are selectivity, amplification, divergence, restriction, movability, biochemical rule, and passage of time.

2. Methods

CSN-Ontology is based on *Device Ontology* and *Functional Concept Ontology* developed by Kitamura and Mizoguchi [3] aiming at systematization of functional knowledge for design. The role of *Device Ontology* is to provide a consistent viewpoint for capturing the target functions. That of *Functional Concept Ontology* is to provide well-defined concepts for description of knowledge, and then to give a basis for systematization of knowledge [3]. Correspondence of concepts in the ontologies to those in CSN-ontology is indicated in Table 1. We are preparing a separate manuscript that reports the details of applying *Device Ontology* and *Functional Concept Ontology* to biological knowledge domain. We refer to [4, 5] for knowledge of TGF-beta pathway.

In this article we use term "information transmission" instead of term "signal transduction", although we consider that both of them have the same meanings in knowledge domain of cell signaling. We prefer the former term, because in the future we would like to apply our ontology to another kinds of "information transmission" in organisms such as activation of information encoded in genome DNA, series of chemical structural changes occurred in metabolic pathways, mutual growth regulations in neuronal networks, and communications between cells and organs in endocrine system. We define "information" as what is transmitted continuously along a series of chemical reactions in organisms which causes certain biological functions as its consequences.

Table 1. Correspondence of concepts in Device Ontology and Functional Concept Ontology to those in CSN-ontology

Device Ontology Functional Concept Ontology	CSN-ontology
object	information (type of chemical transformation, selectivity, and movability)
device	protein, compound, ion
conduit (= medium)	molecular interaction (recognition)
function	functions of reaction, complex, stage (cellular function), and pathway
achievement way	rule (in cell signaling mechanism)

3. Results

3-1 Explication of "information" and "information transmission"

Although reactions in cell signaling pathways are nothing but chemical reactions, they are also considered as "information transmissions". Although this is a consensus in biology, it is neither explicated what "information" is nor how chemical reaction and information transmission differ. What most biologists agree is that "something like information is transmitted continuously along a pathway" and that "chemical reactions are observed along the pathway".

Chemical reactions in general and in cell signaling pathways differ in sizes of their reactants. Reactants in cell signaling pathways are mostly large molecules called proteins having much more extended "recognition" surfaces than general chemical compounds. For instance, in a phosphorylation in cell signaling pathways, a reactant is recognized not only at its phosphorylation site but also at a separate site specialized for recognition. In a protein-protein binding reaction, we can separate effect of recognition from effect of chemical transformation. The former corresponds to recognition between the two protein surfaces, while the latter corresponds to conformational changes of the proteins induced by the binding.

Separating "recognition" from "chemical reaction", we found that a cell signaling pathway is a process of alternate "recognitions" and "chemical transformations" (Figure 1). On the separation, we define "recognition" as any kind of interaction between reactants, while we define "chemical transformation" as any physicochemical force that changes a physicochemical state of an individual reactant. Then we refer to two of the most important

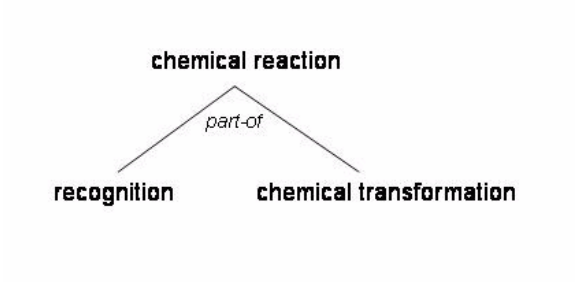

Figure 1 (Ontological analysis shows us that any chemical reaction in cell signaling pathways consists of "recognition" and "chemical transformation".)

features of cell signaling pathways in such a way that a "property" of chemical transformation is changed along the pathways and that the necessary recognition is changed progressively.

For instance, a change from "binding" to "phosphorylation" can be found in an allosteric enzyme. The change is one of the essential roles of cell signaling mechanisms. The essence is not "to change chemical structures of molecules" but "to change properties of chemical transformations". The property includes chemical effects, substrate specificity, turnover, and cofactors of the transformations. This essence exactly constitutes what biologists call “information transmission”, and we can interpret "properties of chemical transformations" as an element of "information (something transmitted)". Since changes of the properties occur in molecules, we regard the molecules as having roles to change the properties, namely to transmit information. Similarly we also find "something transmitted" in metabolic pathways. In metabolic pathways "change of chemical structure" is transmitted, while "change of properties of chemical transformation" is transmitted in signaling pathways. "Change of chemical structure" is an internal motion of a process of chemical transformation, which implies "what is changed in the process". On the other hand, "change of properties of chemical transformation" is an external motion of a process of chemical transformation, which implies "what the process itself is changed to". Although we observe nothing but chemical reactions in both cell signaling and metabolic pathways, we find distinction between them in conceptualization of the reactions in the context of what is transmitted in the different kinds of pathways.

Then one might wonder about missing the internal motion of chemical transformation from cell signaling pathways. The internal motion is not missed but conceptualized differently as "change of recognition", since the chemical transformations actually change complementarity of surface structures of reactants that determines recognition. Recognition fills two roles of "selectivity" and "movability" in the context of information transmission, two of which are additional constitutive elements of "information". "Selectivity" is an attribute of "recognition" and represents how many molecules of how

many sorts take a reaction. "Selectivity" is an interpretation of such biochemical behaviors of recognition as "how long the recognition is retained" and "how strict the recognition is". Selectivity subsumes biologically important concepts of "divergence", "amplification", and "restriction" (figure 2). These concepts represent quantitative changes of information in reactions. "Divergence" and "amplification" mean increase of information with low and high specificity respectively, while "restriction" means no increase of information. Conceptual specialization of "selectivity" gives the three concepts. If many molecules of many sorts take a reaction, we regard the reaction as "divergence". We regard the reaction as "amplification", if many molecules of few sorts take the reaction. A reaction taken few molecules of few sorts is regarded as "restriction". The other constitutive element of information, "movability", is an attribute of a molecule and represents an area where a molecule can exist. Because cells are compartmented into sub-cellular structures, movability consists with either an area within one of the sub-cellular structures or an area between two of the sub-cellular structures. Movability is also determined by "recognition", because a molecule determines its existing area by "recognizing" specific proteins fasten at a certain area in cells. Since each area permits only certain reactions to occur, movability should change reaction by reaction along a pathway to transmit information to certain destination areas.

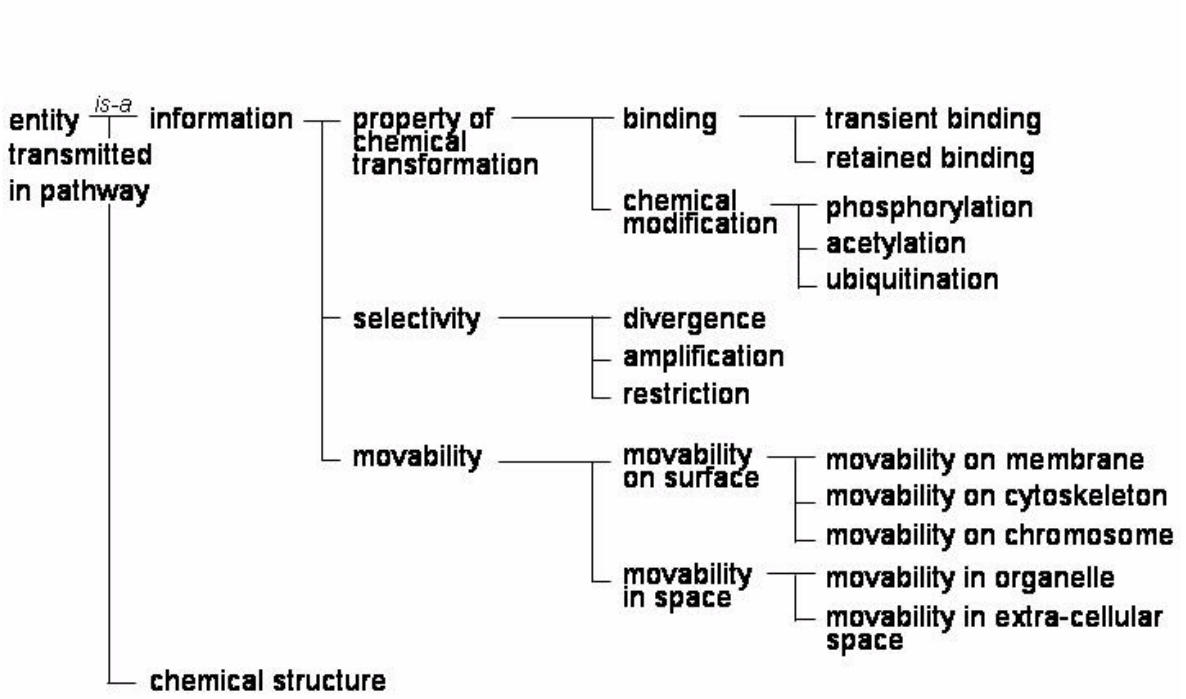

Figure 2 (Is-a hierarchy of concepts represent "transmitted entity" in cell signaling and metabolic pathways.)

We indicate in figure 2 three constitutive elements of "information"; "properties of chemical transformations", "selectivity", and "movability", and their subsumed concepts. Next we explain how the information transmissions contribute to biological functions of pathways. We show how "phosphorylation" contributes to "cell growth arrest".

3-2 Functional decomposition of biological functions

We explicate that the reactions transmit "properties of chemical transformation", "selectivity", and "movability" in cell signaling pathways. More kinds of elements should be transmitted in cell signaling pathways than in metabolic pathways, in which only "chemical structure" is simply transmitted. Accordingly reactions should be multi-functional in cell signaling pathways. In fact, many single reactions cannot transmit "information" independently. Many single reactions should construct complexes to obtain multi-functionality required for information transmission. "Ligand-receptor complex in TGF-beta pathway" is an example for a multi-functional unit of information transmission. This complex changes "binding" of TGF-beta into "phosphorylation" of TGF-beta receptor 1, changes selectivity of TGF-beta receptor 1, and changes movability from "extra-cellular space" to "cytosolic surface of plasma membrane". Similarly individual reactions construct or modify any complexes and contribute to information transmission taken by the complexes. Most of complexes are dynamically constructed as well as broken down along cell signaling pathways. The dynamical construction is special to cell signaling pathways, compared with that most of complexes are constructed beforehand and unchanged in metabolic pathways. The dynamical construction enables flexibility and adaptability of cell signaling pathways. The complex construction is strictly regulated, because individual reactions that construct the complexes are strictly regulated with their selectivity and movability.

A function of a complex can be decomposed and assigned to individual component reactions. For example, there is a reaction as "phosphorylation of TGF-beta receptor 1 by TGF-beta receptor 2" occurred in a "ligand-receptor complex" in "TGF-beta pathway". A function assigned to the reaction is "to fix the (allosteric) regulation at interior of the cell", because the reaction is a sub-process of information transmission taken by the complex with a biochemical mechanism called "allosteric regulation" (Figure 3).

To answer a question like how to functionally relate "phosphorylation" with "cell growth arrest", we may decompose the biological function ("cell growth arrest") down to reactions by way of complexes. Figure 3 shows a result of the functional decomposition. The figure shows how a biological function on "pathway level" is decomposed into "stage level", "complex level", and "reaction level" in sequence. Since the decomposition into

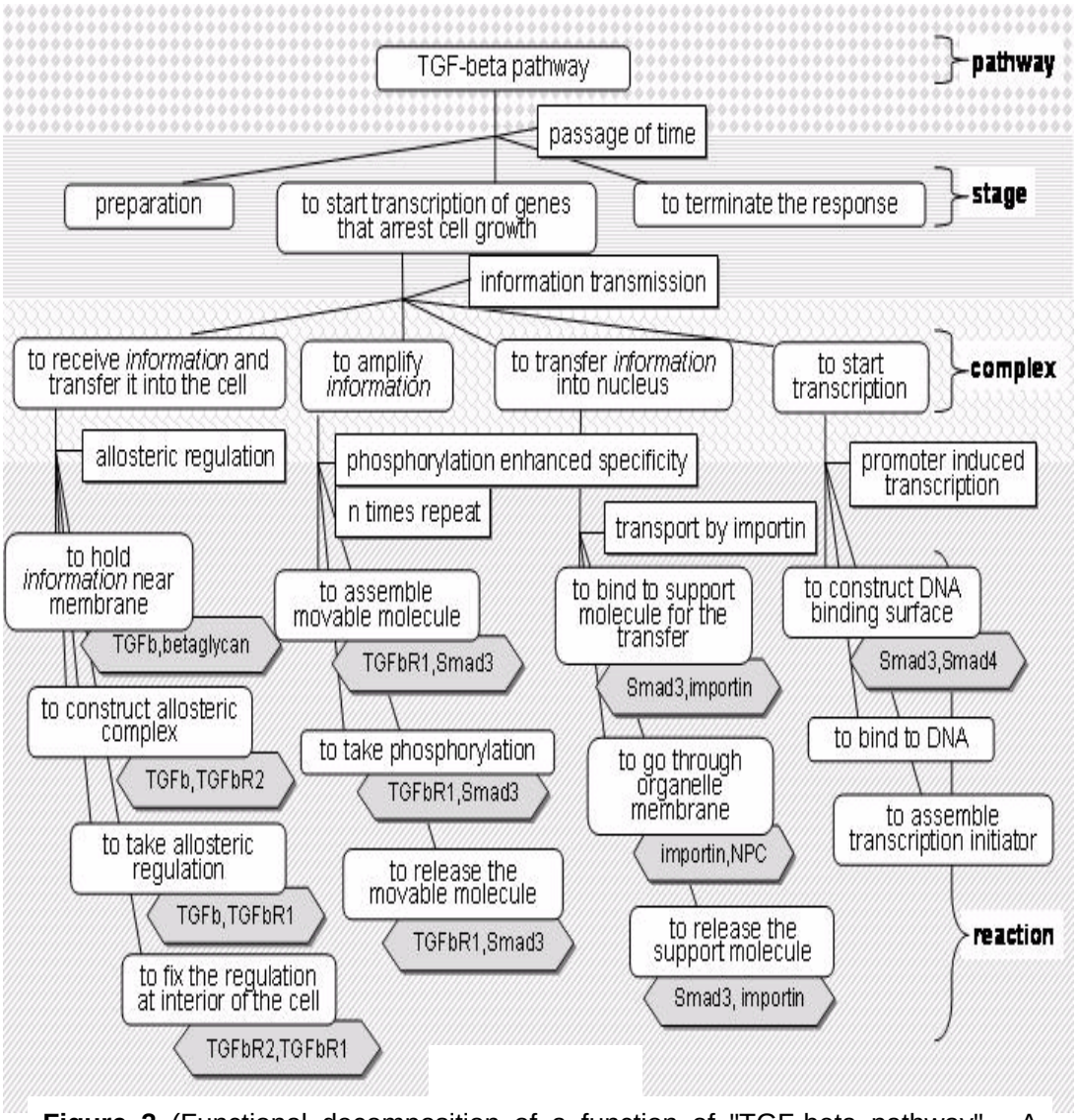

Figure 3 (Functional decomposition of a function of "TGF-beta pathway". A concept in a rounded rectangle box represents "biological function" ("base-function" in the Device Ontology). A concepts in a rectangle box represents "rule in cell signaling mechanism" ("achievement way" in the Functional Concept Ontology). Molecules written in a hexagonal box represent reactants that take part in individual reactions. Abbreviations used in figure 3; TGFb (TGF-beta), TGFbR1 (TGF-beta receptor 1), TGFbR2 (TGF-beta receptor 2), NPC (nuclear core complex).)

"pathway level" is general in any cell signaling pathways, we start our explanation from "stage level" that represents so called "cellular function". The cellular function of "TGF-beta pathway" is "to start transcription of genes that arrest cell growth". The cellular function is decomposed into "to receive information and transfer it into the cell" in "complex

level". Then the complex function is further decomposed into "reaction level" like as "to fix the (allosteric) regulation at interior of the cell", that is a function of phosphorylation in the complex.

Functions can be decomposed, because the functions obey any "rules" in cell signaling mechanisms that enable component reactions achieve the functions. The decomposition tree in figure 3 shows that "information transmission" is a common rule in "complex level". The rules in "reaction level" represent "biochemical mechanisms of complex activities". Generally any pathway function is decomposed into "preparation", "primary response", and "secondary response" in its first decomposition level, that we call "stage level" (Figure 3). The cellular function of TGF-beta pathway (e.g. "to start transcription of genes that arrest cell growth") is a subsumption of "primary response". The decomposition in "stage level" is ruled by "passage of time", that represents regulation by "time". Abstraction of such rules is significant for more precise and consistent representation of cell signaling mechanisms.

4. Discussion

Two kinds of models have been used to represent cell signaling pathways. One model is called "binary relationship" and applied in CSNDB[6], TRANSPATH[7], and BIND[8]. Theses databases regard pathways as chains of relationships between molecules such as bindings and chemical modifications. For instance they regard phosphorylation as a relationship between an enzyme and a substrate that causes a kind of chemical modification. We consider that this model emphasizes effect of "recognition" in information transmission. The other model is applied in aMAZE[9] and PATIKA[10]. There databases regard pathways as chains of processes in which physicochemical states of individual molecules change. For instance they regard phosphorylation as a process of state-change from unphosphorylated to phosphorylated state occurred in a substrate catalyzed by an enzyme. We consider that the latter model emphasizes effect of "chemical transformation". Obviously it is controversial which model is appropriate to represent cell signaling pathways. CSN-ontology offers how to relate the two models and argues that both the effects (recognition and chemical transformation) are required for "information transmission" in cell signaling pathways (Figure 1, 2). Recently tremendous numbers of interaction data among proteins are produced from proteomic researches. CSN-Ontology is also useful for computational analyses of the data, because the ontology provides framework for assignment of biological functions to individual protein-protein interactions.

In our ontological analysis to explicate "information transmission" in cell signaling pathways, we could abstract concepts that databases to date have failed to represent

explicitly. They are selectivity, amplification, divergence, restriction, movability, biochemical rule, and passage of time. We are under a development of a database for cell signaling pathways based on CSN-ontology including these newly abstracted concepts.

This study analyzed only one pathway (TGF-beta pathway). We will pursue our development of CSN-ontology to cover all the domain knowledge of cell signaling mechanisms.

5. Acknowledgement

We are grateful to Dr. Edgar Wingender for useful comments on this manuscript.

6. References

[1] Mizoguchi, R., and Kitamura, Y. (2000) Foundation of Knowledge Systematization: Role of Ontological Engineering, *Industrial Knowledge Management - A Micro Level Approach*. Roy, R. (ed.), Springer-Verlag, London, pp.17-36.

[2] Consortium, T.G.O. (2001) Creating the gene ontology resource: design and implementation. *Genome Res*. **11**, 1425-1433.

[3] Kitamura, Y., Mizoguchi, R. (2003) Ontology-based description of functional design knowledge and its use in a functional way server. *Expert Systems with Application*. **24**, 153-166.

[4] Shi Y, Massague J. (2003) Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell*. **113**, 685-700.

[5] Heldin CH, Miyazono K, ten Dijke P. (1997) TGF-beta signaling from cell membrane to nucleus through SMAD proteins. *Nature*. **390**, 465-471.

[6] Takai-Igarashi, T. and Kaminuma, T. (1999) A pathway finding system for the cell signaling networks database. *In Silico Biol*. **1**,129-146.

[7] Schacherer, F., Choi, C., Gotze, U., Krull, M., Pistor, S. and Wingender, E. (2001) The TRANSPATH signal transduction database: a knowledge base on signal transduction networks. *Bioinformatics*. **17**,1053-1057.

[8] Gary D. Bader, Doron Betel and Christopher W. V. Hogue (2003) BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Research*. **31**, 248-250.

[9] van Helden, J., Naim, A., Lemer, C., Mancuso, R., Eldridge, M. and Wodak, S.J. (2001) From molecular activities and processes to biological function. *Brief Bioinform*. **2**, 81-93.

[10] E. Demir, O. Babur, U. Dogrusoz, A. Gursoy, G. Nisanci, R. Cetin-Atalay, and M. Ozturk, (2002) "PATIKA: An Integrated Visual Environment for Collaborative Construction and Analysis of Cellular Pathways". *Bioinformatics*. **18**, 996-1003.