Title	On Possible Axiomatic Theory for Kinase Computing : Starting from Axioms (Algebra, Languages and Computation)
Author(s)	Liu, Jian-Qin; Shimohara, Katsunori
Citation	数理解析研究所講究録 (2005), 1437: 29-36
Issue Date	2005-06
URL	http://hdl.handle.net/2433/47491
Right	
Type	Departmental Bulletin Paper
Textversion	publisher

数理解析研究所講究録 1437 巻 2005 年 29-36

On Possible Axiomatic Theory for Kinase Computing: Starting from Axioms

Jian-Qin Liu and Katsunori Shimohara

ATR Network Informatics Laboratories,
2-2-2 Hikaridai, "Keihanna Science City", Kyoto 619-0288, Japan
e-mail: {jqliu, katsu}@atr.jp
URL: http://www.nis.atr.jp/{~jqliu,~katsu}

Abstract: In this summary, our recent research results on formulating axioms for kinase computing, which is regarded as the first step towards a rigorous computation theory, will be reported.

Keywords: Axiomatic theory, kinase computing, computation theory.

Summary

Kinase computing, a new method of molecular computing, was initiated by the authors in 2001 [1]. In kinase computing, computing process is carried out based on the signaling pathways of phosphorylation and dephosphorylation switched by kinases and phosphatases [2]. One of the characteristics of kinase computing is the linear order of control-space complexity when it is applied to 3-SAT problem solving. As we know, in mammalian cells, there exist crosstalks between the pathways of Rho family GTPases and kinases/phosphatases. Thus, constructing a kind of coupling structure considering crosstalked pathways can improve the performance of kinase computing systems in the sense of biochemical engineering. The phosphorylation-dephosphorylation structure of kinase computing proposed in [2] has been extended to a two-layer structure of GTPase pathway and kinase/phosphatase pathway [3] recently for the purpose of reducing control-space complexity to a logarithmic order in the case of applying kinase computing to 3-SAT problem solving. This derivative type of kinase computing is called GTPase-based kinase computing. Cellular molecular switches are the basis of the potential architecture of kinase computing in which the molecular switches in cells are mainly classified into two categories -- phosphorylation/dephosphorylation switches

and GTPase switches [4]. In physical biochemistry, GTP-bound and GDP-bound states of GTPases are determined by GEFs and GAPs, respectively. Phosphorylation and dephosphorylation of signaling proteins are determined by kinases and phosphatases, respectively. During the process of the GTP hydrolysis, the GTP-bound GTPase is hydrolyzed into the GDP-bound state. The above-mentioned signaling mechanism can provide two states of signaling molecules for us to design a molecular computing system by cellular pathways under certain condition of cellular regulation. In order to study a generic kinase computer, research on rigorous computation theory based on cellular pathways under engineered control will be presented here. This theoretical work is regarded as a prerequisite to explore possible generic architecture of a kinase computer, which is expected to be experimentally implemented as a computational nanobiomachine capable of parallel information processing based on the signaling pathways of GTPases as well as kinases and phosphatases.

I. PRELIMINARIES

1.1. Terms

The abbreviations of the terms in molecular cell biology we will use in the later discussion are given as follows:

GEFs -- guanine nucleotide exchange factors;

GAPs -- GTPase-activating proteins;

GTP -- guanosine triphosphate;

GDP -- guanosine diphosphate;

GTPase -- the proteins in cells with two states -- GTP-bound state and GDP-bound state (for more details, Cf., chapter 3 of [4]);

GDI -- guanine nucleotide exchange inhibitors;

Kinase -- a kind of protein in cells that can phosphorylates the signaling molecules in cells.

Phosphatase -- a kind of protein in cell that can detach the phosphate from phosphorylated signaling molecules.

SPKs -- signaling-protein-molecules in cells regulated by kinases/phosphatases and their corresponding pathways. These proteins mainly include serine, threonine and tyrosine [4].

p -- phosphate.

1.2 Notations

The notations that we will use in the later discussion are given as follows:

$i = 0, 1, ..., n, n \in N.$

$j = 0, 1, ..., m, m \in N.$

$l = 0, ..., L\text{-}1, L \in N.$

$u = 0, 1, ..., U_l\text{-}1. U_l \in N.$

$v = 0, 1, ..., V_l\text{-}1. V_l \in N.$

$i' = 0, 1, ..., L'\text{-}1, L' \in N.$

G – refers to the GTPase set $\{G_l\} = \{G_0, G_1, ..., G_{L-1}\}$. ($l = 0, 1, ..., L\text{-}1, l \in L$}. $G_l = 1$ for GTP-bound and $G_l = 0$ for GDP-bound.

$Ggef(l)$ – refers to the GEF set, Ggef (l) = $\{G'(l,u)\}$, $G'(l,u)$ refers to the *u-th* GEF in the *l-th* GEF set, $u = 0, 1, ..., U_l\text{-}1. U_l \in N.$ $G'(l,u) = 1$ refers to the activation state of GEF; $G'(l,u) = 0$ refers to the inactivation state of GEF.

Ggap(l) – refers to the GAP set. Ggap (l) = $\{G''(l,v)\}$, $G''(l,v)$ refers to the *v-th* GAP in the *l-th* GAP set, $v = 0, 1, ..., V_l\text{-}1. V_l \in N.$ $G''(l,v) = 1$ refers to the activation state of GAP; $G''(l,v) = 0$ refers to the inactivation state of GAP.

$\Psi_{i'}$ – *i'-th* kinase or *i'-th* phosphatase.

$Y_{i'}$ – refers to the state of phosphorylation/dephosphorylation of the *i'-th* SPK.

$Y_{i'} = 1$ if $Y_{i'}$ is in the state of phosphorylation.

$Y_{i'} = 0$ if $Y_{i'}$ is in the state of dephosphorylation.

II. MAJOR RESULTS

There is no doubt about the significance to use axiomatic representation to formulate a theoretical system for kinase computing. For example, in the 6th Hilbert problem "Mathematical treatment of the axioms of physics", David Hilbert emphasized on the axiomatic work for physics as an important task [5]. This foreseeable argument has been proven to be true in theoretical physics. In the framework of theoretical computer science, axiomatic system has been established many years ago. But, the corresponding axiomatic theory for the purpose of molecular computing based on cellular pathways of GTPases and kinases/phosphatases is still open. The biochemical signaling mechanism, which is the kernel for understanding of the computing process based on cellular pathways, is the starting point for us to formalize the computing process. Here, we propose two axioms. The first axiom dedicates to the description of GTPase molecular switch in the sense of physical biochemistry.

The second axiom also dedicates to the description of GTPase molecular switch in the sense of physical biochemistry but under the condition of engineered pathways in cells.

AXIOM 1

Under the condition that GEFs/GAPs pathways in cells are activated, we have that

$G_l = 1$ if $Ggef(l) \neq \phi$ and $Ggap(l) = \phi$, (1.1)

and

$G_l = 0$ if $Ggef(l) = \phi$ and $Ggap(l) \neq \phi$. (1.2)

In GEFs/GAPs pathways, 1 is used to represent the state of GTP-bound; 0 is used to represent the state of GDP-bound. We assume that the quantity of GEF or GAP is sufficiently enough to catalyze the GTPase activation or GTPase inactivation, respectively. The assumption is also valid for axiom 2.

AXIOM 2

Under the condition that engineered GEFs/GAPs pathways in cells are activated, we have that

$G_l = 1$ if and only if $Ggef(l) \neq \phi$ and $Ggap(l) = \phi$, (2.1)

and

$G_l = 0$ if and only if $Ggef(l) = \phi$ and $Ggap(l) \neq \phi$. (2.2)

As a preliminary status of formalization, a simple form of automata is needed to show how to carry out computation in terms of the GTPase pathway structure in cells. Here, the following assumptions are suggested for the basic formal description of the kinase computing process.

Assumption 1

Provided that there exist the feedback from GTPase to GEFs/GAPs and a set of signaling molecules $\{Q(l)\}$ $(l = 0, 1, ..., L-1)$ that are intermediate components between GTPase and GEF/GAP. The mapping from $\{G_l\}$ to $\{Q(l)\}$ is many-to-one, and the mapping from $\{Q(l)\}$ to $\{Ggef(l)\}$ and $\{Ggap(l)\}$ are one-to-many.

Then, we can formalize this feedback mechanism.

Let A be an alphabeta. Gw is the set of all the states $\{w(2^L)\}$ encoded by $\{Ggef(l) \cup Ggap(l)\}$ based on axiom 1 and axiom 2. s_0 is the start state in Gw. Gwq is the set of the final states that are acceptable when halt happens. $Gwq \subseteq Gw$.

We can conceptually design a deterministic finite state automaton [6]

$$W = (A, Gw, s_0, Gwq, \delta) \qquad (3.1)$$

where δ is defined as following function:

$$\delta: Gw \times A \dashrightarrow Gw \qquad (3.2)$$

The input to this automaton is defined as a single molecule in $\{Q(l)\}$ and denoted as q in A. The current state of the automaton is denoted as $Gw(t)$ and next state of the automaton is denoted as $Gw(t+1)$. According to the automaton structure, any possible implementation of the state transition from the current state of GEF/GAP to the next state of GEF/GAP will be controlled by the input in A. Here, we assume that a single molecule $Q(l)$ exist in the above process. This fits the definition of automaton in the basic form. It is possible to be extended into a multiple automata system under certain condition (we denote this condition as Θ to be studied in future work). In addition to the assumption on a single molecule from $\{Q(l)\}$, other possible situations include manual operation of the state transition based on detection of GTPase and automated state transition by control of engineered pathways. In theory, multiple automata for the above-mentioned process can be formalized equivalently to the corresponding single automaton based on the abstract representation given above.

By automaton W, the configuration $C_0, C_1, ..., C_n$ can be defined as

$$(q, Q_w)$$

where Q_w refers to the sequence that is constructed by the molecules from $\{Q(l)\}$ and is to be activated in succeeding steps. They can be generated by controlling the related pathways. Thus, certain sequences for computation can be formulated by this sequence

$$\{C_i\} \ (i = 0, 1, ..., n).$$

Here, the configuration process is only presented by formalization. We conclude this speculation as follows:

Assumption 2

The configuration $C_0, C_1, ..., C_n$ can give rise to a computing process of deterministic finite state automaton under the condition of assumption 1, axiom 1 and axiom 2.

Now, we discuss the above-mentioned process for computing in the definition of automaton. The feasibility of this computing process is based on the "controllability" of biochemical signaling mechanism. The range of the biochemical concentration is (0,1). If we can prove that any arbitrary state in the range of (0,1) can be obtained in theory, the

corresponding feasibility of the computing process we present above can be inferred consequently.

THEOREM 1

Based on the axiom1, axiom 2, assumption 1, and assumption 2, the automaton defined above is feasible in the sense of controllability on engineered pathways.

Proof:

The quantitative relation between $Q(l)$ and $Ggef \cup Ggap$ is represented in the matrix $B_{L\times L}$. Because the total number of the activated GTPase is L, the activated GEF in Ggef and the activated GAP in Ggap satisfy the condition that L1+L2 = L where L1 is the number of the activated GEF in Ggef and L2 is the number of activated GAP in Ggap.

The feasibility of the integrated pathways can be inferred from the study of pathway of signaling molecule $\{Q(l)\}$ whose input is the $\{Q(l)\}^{(t)}$ and output is GEF/GAP in $\{Ggef \cup Ggap\}^{(t+1)}$, and pathway of GEF/GAP whose input is GEF/GAP in $\{Ggef \cup Ggap\}^{(t+1)}$ and output is $\{G(l)\}^{(t+1)}$.

Now, let's consider an equation for a quantitative description of the GEF/GAP pathways that control the GTP-bound/GDP-bound states of GTPases:

$$d/dt\ (X) = A(t)\ X + B(t)\ U(t), \qquad (4.1)$$

where

X – refers to the vector represented by activated GEFs/GAPs whose total number is L,

$A(t)$ – refers to the matrix of the cross-talks in the above-mentioned integrated pathways among the GEFs/GAPs, whose size is $L \times L$,

$U(t)$ – a vector constructed from the set of $\{Q(l)\}$,

$B(t)$ – a matrix constructed by the cross-talks of the integrated pathway between $\{Q(l)\}$ and GEFs/GAPs where the interaction is physically available among the engineered-pathways. Its rank is $L \times L$.

Here, controllability of the computing process refers to the ability to make the states of m Rho family GTPases fall into ranges corresponding to all of the related combinatorial forms.

Let $A(t) = A$ and $B(t) = B$. The existence of controllability of related pathways is dependent on the rank of

$$B{:}AB{:}\ldots{:}A^{L-1}B. \qquad (4.2)$$

The rank of $B{:}AB{:}\ldots{:}A^{L-1}B$ is L because the related pathway units encoded for GTPases whose number is L always generate the maximum number of $L \times L$ cross-talked pathways. Therefore, the signaling mechanism for the computing process presented based on axiom 1 and 2 and assumption 1 and 2, which is constructed by the signaling pathways between GEF/GAP and $\{Q(l)\}$ pathways, is feasible in the sense of quantitative measurement for the related crosstalking mechanism, according to the necessary and sufficient condition of controllability. The controllability refers to that the related system (i.e., for equation (4.1)) is controllable if and only if the rank of $B{:}AB{:}\ldots{:}A^{L-1}B$ is L. (For more details on related controllability theory for control systems, Cf. [7]. Some of the notations used for controllability in this summary are different from those used in [7]). Therefore, the conclusion is obtained.

QUESTION 1

Is there any code designed rigorously by GTPases in mathematics?

Acknowledgments

This research was conducted as part of "Research on Human Communication" with funding from the National Institute of Information and Communications Technology, Japan. The authors sincerely thank Prof. Masami Ito, Prof. Yuji Kobayashi, Prof. Kozo Kaibuchi, Dr. Shinya Kuroda, Dr. Mutsuki Amano, Prof. Reiji Nakajima, Prof. Minoru Kanehisa, Prof. Tatsuya Akutsu, and Prof. Katsumasa Watanabe for their advice. Also appreciated is Min-Jie Wei for reading this manuscript and giving comments.

REFERENCES

[1] Jian-Qin Liu and Katsunori Shimohara, Kinase computing in Rho-GTPases: analysis and simulation, IPSJ SIG Notes, vol.2001, No.91, Sept.17, 2001, pp.29-32 (in Japanese).

[2] Jian-Qin Liu and Katsunori Shimohara, A biomolecular computing method based on Rho family GTPases, IEEE Transactions on Nanobioscience, 2 (2), June 2003, pp. 58 -62.

[3] Jian-Qin Liu and Katsunori Shimohara, A new algorithm of biomolecular computation based on Rho family GTPases with a logarithmic order of complexity of GEFs/GAPs, submitted, Dec. 2004.

[4] Bruce Alberts, et al., Molecular Biology of the Cell, 4th Edition, Garland Science, 2002.

[5] David Hilbert, Mathematical problems, Lecture delivered before the International Congress of Mathematicians at Paris in 1900,
http://babbage.clarku.edu/~djoyce/hilbert/problems.html

[6] Harry R. Lewis and Christos H. Papadimitriou, Elements of the theory of computation, Prentice-Hall, Inc., 1981.

[7] Katsuhiko Ogata, Discrete-time control systems, Second Edition, Prentice-Hall International, Inc., 1995.

APPENDIX:

AXIOM 3

Under the condition that kinase/phosphatase pathways in cells are activated, we have that

$Y_{i'} = 1$ if $\Psi_{i'} = i'$-*th* kinase

and

$Y_{i'} = 0$ if $\Psi_{i'} = i'$-*th* phosphatase.

AXIOM 4

Under the condition that engineered kinase/phosphatase pathways in cells are activated, we have that

$Y_{i'} = 1$ if and only if $\Psi_{i'} = i'$-*th* kinase

and

$Y_{i'} = 0$ if and only if $\Psi_{i'} = i'$-*th* phosphatase.