$Sub(\mathbb{Z})$ IMPLEMENTATION OF BIOLOGICAL ACTIVITIES CONDUCTED BY PROTEINS

NAOTO MORIKAWA

1. Introduction

We write $Sub(\mathbb{Z})$ for the collection of all finite subsets of the set $\mathbb{Z}$ of all integers. Assigning prime numbers to proteins and their complexes, we represent biological activities conducted by proteins as actions of integers on a subset of the set $\mathbb{N}$ of all natural numbers. In particular, a biological process is described as an orbit of the action of an element of $Sub(\mathbb{Z})$ on a subset of $\mathbb{N}$. (See [1] for $Sub(\mathbb{Z}^N)$ implementations of protein-protein interactions for $N \in \mathbb{N}$.)

2. Action of $\mathbb{Z}$ on $\mathbb{N}$

In this paper, we shall consider addition $m = a + n$ of two integers a and n $(n, m > 0)$. Let $n = \prod_{1 \le i \le N} p_i^{l_i}$ and $m = \prod_{1 \le i \le M} q_i^{k_i}$ be prime factorizations of n and m. Then, addition $+$ induces an action $*$ of a on finite sets of primes:

$$a * \{\overbrace{p_1, \dots, p_1}^{l_1 \text{ times}}, \ \dots \ , \overbrace{p_N, \dots, p_N}^{l_N \text{ times}}\} := \{\overbrace{q_1, \dots, q_1}^{k_1 \text{ times}}, \ \dots \ , \overbrace{q_M, \dots, q_M}^{k_M \text{ times}}\}.$$

For example, $1+18 = 19$ defines an action of 1 on $\{2, 3, 3\}$: $1*\{2, 3, 3\} = \{19\}$. That is, one protein 2 and two proteins 3s form a protein complex 19 by the "interaction" of 1.

In general, we define action $*$ of $\mathbb{Z}$ on $P := \{p_1^{l_1}, \dots, p_k^{l_k} \mid 0 \le l_i \in \mathbb{Z}\} \cup \{0\} \subset \mathbb{N}$ for a finite set $\{p_1, \dots, p_k\}$ of primes as follows:

$$* : Z \times P \to P, \quad a * s := \begin{cases} a + s & \text{if } a + s \in P, \\ 0 & \text{else.} \end{cases}$$

$A \in Sub(\mathbb{Z})$ is called *definable* at $s \in P$ if $a * s \ne 0$ and $a' * s \ne 0$ imply $a = a'$ for $a, a' \in A$. The action of $A \in Sub(\mathbb{Z})$ on $s \in P$ is defined by

$$A * s := \begin{cases} \sum_{a \in A} a * s & \text{if } A \text{ is definable at } s, \\ \text{undefined} & \text{else,} \end{cases}$$

where $\sum_{a \in A} a * s$ denotes the sum of natural numbers $a * s$.

3. Arithmetic implemantation of processes

Let $P = \{p^l q^m r^n \mid 0 \le l, m, n \in \mathbb{Z}\}$ and $A \in Sub(\mathbb{Z})$. Here we shall consider an orbit of the action of A on P: $s_{n+1} := A * s_n (n \ge 0)$. And we describe two simple processes as an orbit of the action.

Date: Octorber 20, 2004.

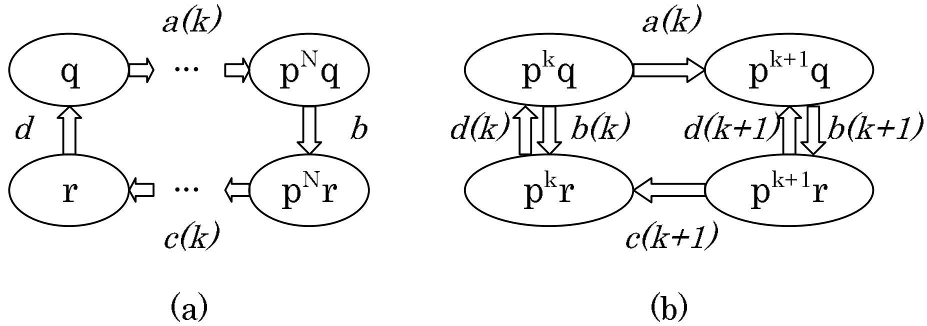

FIGURE 1. Transition diagrams. (a): A deterministic process. (b): A non-deterministic process.

3.1. **Deterministic processes.** The process specified by the transition diagram of Fig.1(a) corresponds to the orbit S defined by $s_{n+1} = A * s_n$ $(n > 0)$,

$$s_0 = q \text{ and } A = \{a(k), b, c(k+1), d \mid k = 0, 1, \ldots, N-1\},$$

where

$$a(k) := p^{k+1}q - p^k q,\ b := p^N r - p^N q,\ c(k) := p^{k-1}r - p^k r, \text{ and } d := q - r.$$

For example, $s_N = p^N q$, $s_{N+1} = p^N r$, $s_{2N+1} = r$, and $s_{2N+2} = s_0$. If $N = 2$, the process is implemented as an orbit of the action of

$$A = \{34, 102, -36, -78, -26, 4\} \text{ on } P = \{3^l 17^m 13^n \mid 0 \le l, m, n \in \mathbb{Z}\}.$$

If $N = 2$ and $(p, q, r) = (5, 7, 11)$, then A is not definable at $pr \in S$ because $c(1) * pr \neq 0$ and $d * pr \neq 0$.

3.2. **Non-deterministic processes.** For $a_1, a_2 \in \mathbb{Z}$, we introduce the notion of "non-deterministic or" a_1 or a_2 to denote an action $(a_1 \text{ or } a_2) * n$ which behaves either like $a_1 * n$ or $a_2 * n$, where the selection between them is made arbitrarily, without the knowledge or control of the external environment.

The process specified by Fig.1(b) corresponds to the orbit S defined by $s_{n+1} = A * s_n$ $(n > 0)$,

$$s_0 = q \text{ and } A = \{\alpha(k), \beta(k+1), b(N), d(0) \mid k = 0, 1, \ldots, N-1\},$$

where

$$\alpha(k) := a(k) \text{ or } b(k), \quad a(k) := p^{k+1}q - p^k q, \quad b(k) := p^k r - p^k q,$$
$$\beta(k) := c(k) \text{ or } d(k), \quad c(k) := p^{k-1}q - p^k q, \quad d(k) := p^k q - p^k r.$$

For example, $s_1 = \alpha(0) * q$, $s_2 = \alpha(1) * s_1$ if $\alpha(0) = a(0)$, and $s_2 = d(0) * s_1$ if $\alpha(0) = b(0)$. If $N = 2$, the process is implemented as an orbit of the action of

$$A = \{7 \text{ or } 16,\ 14 \text{ or } 32,\ 64,\ -64 \text{ or } -46,\ -32 \text{ or } -23,\ -16\}$$
$$\text{on} \quad P = \{2^l 7^m 23^n \mid 0 \le l, m, n \in \mathbb{Z}\}.$$

If $N = 2$ and $(p, q, r) = (3, 17, 13)$, then A is not definable at $r \in S$ because $b(0) * r \neq 0$, $b(1) * r \neq 0$, and $d(0) * r \neq 0$.

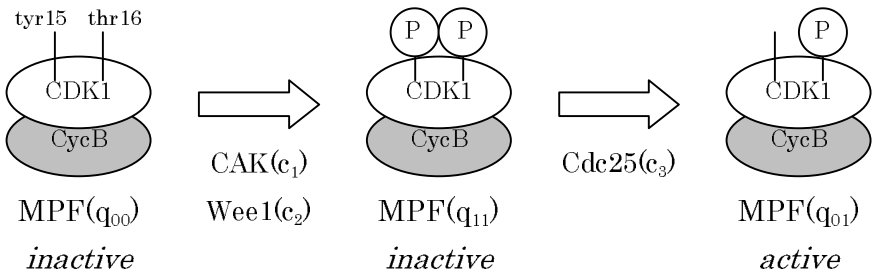

FIGURE 2. Modification (Phosphorylation) of MPF.

4. EXAMPLE OF BIOMOLECULAR ACTIVITIES: CELL DIVISION

The cell division in eukaryotes consists of four phases: G1 $\to$ S $\to$ G2 $\to$ M. Between successive cell divisions, the cell stays in a gap phase G1. The cell division starts with the replication of the DNA in the synthesis phase S. And the cell divides into two in the mitotic phase M after a second gap phase G2. The following is the simplified description of the progression from G2 to M which is controlled by a protein complex called MPF (maturation promoting factor).

For $N \in \mathbb{N}$, let

$$P = \{{p_1}^{l_1}{p_2}^{l_2}{q_{00}}^{l_3}{q_{01}}^{l_4}{q_{10}}^{l_5}{q_{11}}^{l_6}{c_1}^{l_7}{c_{20}}^{l_9}{c_{21}}^{l_9}{c_3}^{l_{10}}{c_4}^{l_{11}} \mid 0 \le l_1, \ldots, l_{11} \in \mathbb{Z}\}.$$

The progression corresponds to the orbit S defined by $s_{n+1} = A * s_n$ $(n > 0)$,

$$s_0 = {p_1}^N {p_2}^N \text{ and } A = A_F \cup A_M \cup A_A \cup A_D,$$

where

$$
\begin{aligned}
&A_F := \{a_F(s,l,d) \mid s + l = N, d = 0, 1\},\\
&A_M := \{a_M(l,m,n,k) \mid l + m + n + k = N\},\\
&A_A := \{a_{A1}(m,k), a_{A2}(m,k) \mid m + k = N\},\\
&A_D := \{a_D(s,m,d) \mid s + m = N, d = 0, 1\},\\
&a_F(s,l,d) := a_1(s,l)c_1{c_{20}}^{1-d}{c_{21}}^d c_3 c_4,\\
&a_M(l,m,n,k) := \big(a_{21}(l,m){q_{10}}^n{q_{11}}^k c_{20} \text{ or } a_{22}(n,k){q_{00}}^l{q_{01}}^m c_{20}\\
&\qquad\qquad \text{or } a_{23}(l,n){q_{01}}^m{q_{11}}^k c_1 \text{ or } a_{24}(m,k){q_{00}}^l{q_{10}}^n c_1\big)\, c_3 c_4,\\
&a_{A1}(m,k) := a_{31}{q_{01}}^m{q_{11}}^k c_1 c_4,\\
&a_{A2}(m,k) := a_{32}(m,k)c_1 c_{21} c_4,\\
&a_D(s,m,d) := a_4(s,m)c_1{c_{20}}^{1-d}{c_{21}}^d c_3.
\end{aligned}
$$

$a_1(s,l), a_{21}(l,m), \ldots, a_4(s,m)$ are given below with their biological meanings.

Step 1. Formation of a complex: ${p_1}^N{p_2}^N \mapsto {q_{00}}^N$. N pairs of proteins cyclin-dependent kinase CDK1 (p_1) and cyclin CycB (p_2) form N protein complexes MPF (q_{00}). And $a_1(s,l)$ denotes this formation:

$$a_1(s,l) + {p_1}^s{p_2}^s{q_{00}}^l := {p_1}^{s-1}{p_2}^{s-1}{q_{00}}^{l+1}.$$

MPF (q_{00}) is inactive when it is first made.

Step 2. Modification (Phosphorylation): ${q_{00}}^N c_1 c_{20} \mapsto {q_{11}}^N c_1 c_{20}$. MPF ($q_{00}$) is modified at two sites (thy15 and thr16) by adding a small molecule called phosphate group if corresponding catalysis proteins exist. Protein CAK (c_1) catalyzes the addition of a phospharate group to thr16 and protein Wee1 (c_{20}) catalyzes the modification at thr15 as follows:

$$\begin{aligned} a_{21}(l,m) + {q_{00}}^l {q_{01}}^m c_1 &:= {q_{00}}^{l-1} {q_{01}}^{m+1} c_1, \\ a_{22}(n,k) + {q_{10}}^n {q_{11}}^k c_1 &:= {q_{10}}^{n-1} {q_{11}}^{k+1} c_1, \\ a_{23}(l,n) + {q_{00}}^l {q_{10}}^n c_{20} &:= {q_{00}}^{l-1} {q_{10}}^{n+1} c_{20}, \\ a_{24}(m,k) + {q_{01}}^m {q_{11}}^k c_{20} &:= {q_{01}}^{m-1} {q_{11}}^{k+1} c_{20}, \end{aligned}$$

where we denote MPF modified at thy15 by q_{10}, MPF modified at thr16 by q_{01}, and MPF modified at both thy15 and thr16 by q_{11}. MPF triggers mitosis machinery if it is modified only at thr16. Thus, MPF (q_{11}) is still inactive because of the modification at thy15 by Wee1 (c_{20}). See Fig.2.

Step 3. Activation (Dephosphorylation): ${q_{11}}^N c_{20} c_3 \mapsto {q_{01}}^N c_{21} c_3$ ($\mapsto$ *mitosis*). Protein Cdc25 (c_3) modifies Wee1 (c_{20}) to inhibit it and activate MPF (q_{11}) by removing the phosphate at tyr15:

$$\begin{aligned} a_{31} + c_{20} c_3 &:= c_{21} c_3, \\ a_{32}(m,k) + {q_{01}}^m {q_{11}}^k c_3 &:= {q_{01}}^{m+1} {q_{11}}^{k-1} c_3. \end{aligned}$$

MPF (q_{01}) is now active and mitosis starts.

Step 4. Disassembling: (*mitosis* $\mapsto$) ${q_{01}}^N c_4 \mapsto {p_1}^N c_4$. Anaphase promoting complex APC (c_4) disassembles MPF (q_{01}) and destructs CycBs (p_2):

$$a_4(s,m) + {p_1}^s {q_{01}}^m c_4 := {p_1}^{s+1} {q_{01}}^{m-1} c_4.$$

And the cell exits from mitosis.

(Combinations of values of $\{p_1, p_2, q_{00}, q_{01}, q_{10}, q_{11}, c_1, c_{20}, c_{21}, c_3, c_4\}$ which make A definable at all $s \in S$ has not computed.)

References

1. N.Morikawa, *Towards Sub($\mathbb{Z}^N$) implementation of protein-protein interactions*, 2004 (manuscript).

E-mail address: nmorika@f3.dion.ne.jp