

形式オントロジーに基づく 遺伝子調節のための数値モデル

秋山卓見¹ 泉直子² 萩原茂樹¹ 米崎直樹¹

¹ 東京工業大学 大学院情報理工学専攻 計算工学専攻

² 十文字学園女子大学 社会情報学部 社会情報学科

近年の分子生物学の進展の中において、遺伝子ノックダウン実験による遺伝子調節ネットワークの解明が進んでいる。本研究では、それらを計算論的モデルを用いて捉えなおし、形式手法を反映したかたちで解析した。すなわち、化学反応の活性/抑制関係を表現した、資源の取り合いを基本概念とする意味論を持つ形式オントロジーの公理を満たす、遺伝子調節量を表す数値間の演算の代数的制約と、そのインスタンスとなる関数を提案した。また、遺伝子調節ネットワークを解析した実験結果に適用し、ダブルノックダウン実験による遺伝子調節値を推定する計算式を提案した。

Numeric Models for Gene Regulation based on Formal Ontology

Takumi AKIYAMA,¹ Naoko IZUMI,² Shigeki HAGIHARA,¹ Naoki YONEZAKI¹

¹ Department of Computer Science, Graduate School of Information Science and Engineering,
Tokyo Institute of Technology

² Department of Social and Information Sciences, Faculty of Social and Information Sciences,
Jumonji University

The gene regulatory network obtained from the gene knockdown experiment is clarified in the progress of the molecular biology. In this paper, we capture those results by using the computational model, and analyze by a formal method, i.e. we proposed a numeric model of gene regulation reflecting our formal ontology that conceptualized promoting/inhibiting relation of the chemical reactions. Moreover, we formally defined requirements for gene regulation and functions which satisfy this requirements. Finally, we estimated the numerical value of the gene regulation by the double-knockdown experiment with this model.

1 はじめに

生物情報に関する静的な属性に基づいた情報検索や、未知の有効な薬物の効果や副作用の発見において、酵素や薬物、遺伝子、タンパク質などの生体物質の相互作用に関する情報を用いることが注目されている。そこでは、様々な異なる物質・反応レベルで共通に同じ概念が使われていることに着目した解析が行われる。たとえば、薬物などの分子、DNAなどの生体物質、さらに大きくは臓器における現象に至るまで、同じ語彙を用いて説明されることが多い。そこで、その共通性の裏に存在していると考えられる共通の概念を統一的に記述する形式オントロジーが提案されている [1]。この形式オントロジーは推論規則の集合として規定されるが、その意味論として

計算論的モデルを持つ。

本稿では、遺伝子調節のための抽象度の高い形式手法の規定する制約の中で考えられる数値モデルを提案し、それを用いて、分子生物学・ゲノミクスにおける成果に対し形式的なモデルに基づく新たな意味を与える。形式手法とは、概念をシンボルで表現し、その間に成立する関係を計算論的な法則として規定し、解析を行うものである。ここでは、生命科学分野への既存のバイオインフォマティクスのシミュレーションベースのアプローチとは異なった、形式手法の適用を行い、実験生物学への有用な情報を与えることを目的とする。

具体的には、化学反応の活性/抑制関係を表現したオントロジーである形式オントロジーを基礎として、

ノックダウン実験に基づく遺伝子調節を表す数値を付与したハイブリッドモデルを提案する。また、それを用いて既存の遺伝子調節データを解析し、新たにダブルノックダウンの数値を予測する計算式を与える。

1.1 関連研究

遺伝子調節などの分子生物学分野への形式手法の応用例としては、項書換えを用いて細胞内シグナル伝達などをモデル化した Pathway logic [3, 4] や、遺伝子発現パターンの実験結果データから遺伝子調節量を推測する研究である [5]、遺伝子調節の際のさまざまな分子の動きを並行システムを用いて表現しようという試みである [6] などが挙げられる。

これらの研究は分子レベル抽象化であったり反応の量にまで言及しており、抽象度が低いこと、また、とくに [5] などは経験に基づく推測であり、数学的厳密さに欠けることなどが本研究との差異であると言える。

2 遺伝子ノックダウンと遺伝子調節

[2] では、カタユウレイボヤの発生段階における調節遺伝子ネットワークの推定が提出されている。具体的には、発生に関わる 76 個の遺伝子について、遺伝子ノックダウン実験から推定されるネットワークを提示している。

遺伝子ノックダウンを調節遺伝子ネットワークの推定に用いるには、ある遺伝子 A をノックダウンした場合の別の遺伝子 B の発現量を計測する。この量を、遺伝子 A をノックダウンしていない通常の状態での遺伝子 B の発現量と比較する。すると、遺伝子 A が発現していることが遺伝子 B の発現 (量) をどう変えているかがわかる。具体的には、

$$\frac{\text{(通常)の遺伝子 B の発現量}}{\text{A をノックダウンしたときの B の発現量}}$$

を計測し、この値がある閾値 $t (> 1)$ より大きければ遺伝子 A は遺伝子 B を activate しており、 $\frac{1}{t}$ より小さければ repress している、としている¹。すなわち、遺伝子 A がない (ノックダウンされている) ときに発現量が増えるのであれば、本来遺伝子 A は遺伝子 B を抑制しており、同様に、遺伝子 A がないときに発現量が減るのであれば本来遺伝子 A は遺伝子 B を活性化していると解釈している。[2] では t として 2.0 を採用している。

¹ 実際には [2] ではこの逆数を用いている

3 薬物相互作用の形式オントロジー

薬物相互作用の形式オントロジーは [1] で提案された、化学反応の活性化・抑制関係を形式化した形式オントロジーである。活性化する・阻害するという 2 つの用語に着目し、動作的意味論を持った推論規則として形式オントロジーを与えている。この意味論は、線型論理を用いてその証明ステップを反応列の反応ステップに対応させた形で定義されている。

本来反応の阻害・活性化を化学的に説明する際には重要とされる反応量、反応速度や反応速度の変化といった概念を排し、線型論理の資源消費性を用いた適切なレベルでの抽象度を持っていることが特徴である。

本稿のこの節より後では、この形式オントロジーを指して単に「形式オントロジー」もしくは「オントロジー」と呼ぶことにする²。

3.1 言語と推論規則

定義 3.1 (シンボル集合) 論理式を記述するためのシンボル集合を以下のように定義する。

\mathcal{O} 物質を表す定数集合

Pre 以下の述語を含む述語集合 $\{Inh, Prom, React\}$

\mathcal{R} 反応名を表す定数集合

定義 3.2 (言語 L) $x, y, z \in \mathcal{O}$ と $R, R' \in \mathcal{R}$ に対して、式を以下のように定義する。これらの式からなる集合を言語 L とする。

$$Inh(R, R'), Prom(R, R'), React(R, x, y, z)$$

$Inh(R, R')$ ($Prom(R, R')$) は R が R' を阻害 (活性化) する、 $React(R, x, y, z)$ は x, y を入力とし z を出力とする反応 R が存在するということを表している。

定義 3.3 (オントロジーの推論規則) 言語 L 上の推論規則を以下のように定義する。

1. $\frac{React(R, x, y, w)}{React(R, y, x, w)}$
2. $\frac{React(R_1, x, y, w_1) \quad React(R_2, y, z, w_2)}{Inh(R_1, R_2)}$
3. $\frac{React(R_1, x, y, w_1) \quad React(R_2, w_1, z, w_2)}{Prom(R_1, R_2)}$

² 「オントロジー」という言葉は、ある特定の領域について用いられる語彙の集合を示すこともあるが、ここでは物や現象の存在を規定する概念の表現方法とそこに内在する法則を定めるという、本来の意味で用いる

4.
$$\frac{Inh(R, R') \quad Inh(R', R'')}{Prom(R, R'')}$$
5.
$$\frac{Inh(R, R') \quad Prom(R', R'')}{Inh(R, R'')}$$
6.
$$\frac{Prom(R, R') \quad Inh(R', R'')}{Inh(R, R'')}$$
7.
$$\frac{Prom(R, R') \quad Prom(R', R'')}{Prom(R, R'')}$$

推論規則 1 は、反応に使用される二つの物質には順序がないことを、推論規則 2 は、物質（ここでは y ）を取り合う 2 つの反応は阻害関係にあることを示している。推論規則 3 は、反応 R_2 が反応 R_1 の生成物を入力とする場合、 R_1 が R_2 を活性化することを示している。推論規則 4 は、 R が R'' を阻害し、 R' が R'' を阻害する関係にあるとき、 R は R'' を活性化することを表している。推論規則 5 以降も同様である。

線型論理による意味論は、この推論規則が健全で完全であることを確かめるために用いられる [1]。

4 遺伝子調節のための数値モデル

閾値で切られた 2 値モデルとしての調節遺伝子ネットワークは形式オントロジーへ非常に直感的に適用することができる。しかしその反面、2 値化によってノックダウン実験の情報量が大幅に減少してしまい、そのために生じる不整合を解決することができない。たとえば、activate と repress が相殺して表面上関係が見えなくなっている状態などを表現できない。ここでは、そのような 2 値モデルの限界を乗り越えるための新しい数値モデルを提案し、その妥当性とその応用の有用性を明らかにするためのデータを提示する。

4.1 Pathway ベースモデルの代数的制約

上記の目的のためにまず、[2] のノックダウン実験で示された数値への意味付けのために、pathway を基にした汎用的な遺伝子間の抑制/活性の数値関係の間に汎用的に定義されるふたつの演算子、およびその演算子に求められる制約を定義する。

定義 4.1 (演算子 \oplus) ある遺伝子 A の発現が、ある pathway を通じて遺伝子 B の発現量を a 倍にし、また遺伝子 A の発現が別の pathway を通じて遺伝子 B の発現量を b 倍にすると、遺伝子 A の発現が、こ

のふたつの pathway を通じて遺伝子 B の発現量を $a \oplus b$ 倍にする。(図 1)

定義 4.2 (演算子 \otimes) ある遺伝子 A の発現が遺伝子 B の発現量を a 倍にするとする。また、遺伝子 B の発現が遺伝子 C の発現量を b 倍にするとする。このとき、遺伝子 A の発現が、遺伝子 B を介して遺伝子 C の発現量を $a \otimes b$ 倍にする。(図 2)

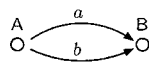


図 1 $a \oplus b$

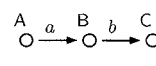


図 2 $a \otimes b$

括弧を省略して書く場合に備えて、演算子の優先度を $\otimes < \oplus$ としておく。

これら 2 つの演算子は汎用的に定義されているが、これら 2 つに成り立たなければならない非常に小さな制約を考えることができる。以下に述べる制約は、定義 4.1, 4.2 が、pathway および遺伝子調節を元に定義されていることから考えて、この定義による 2 つの演算子に対する最小の制約となっていると思われる。

定義 4.3 (\otimes, \oplus に対する制約) 演算子 \oplus, \otimes には、以下の性質が成り立つものとする。

- \oplus に対する可換則 - $a \oplus b = b \oplus a$
- \oplus に対する結合則 - $(a \oplus b) \oplus c = a \oplus (b \oplus c)$
- \otimes に対する結合則 - $(a \otimes b) \otimes c = a \otimes (b \otimes c)$
- \oplus, \otimes に対する分配則 - $a \otimes (b \oplus c) = a \otimes b \oplus a \otimes c$,
 $(b \oplus c) \otimes a = b \otimes a \oplus c \otimes a$

まず、 \oplus に対する可換則については、定義 4.1 上の文脈において \oplus の左右に現れる a と b のふたつの値についてそれらを交換してもまったく同じように定義を読むことができることから導かれる。

\oplus に対する結合則については、その意味に根拠を見出すことができる。式 $(a \oplus b) \oplus b$ は、ある遺伝子 A の発現がある遺伝子 B を、3 つの pathway を通じてそれぞれ a, b, c 倍にしている状況に定義される。このとき、pathway を足しあわせる操作には順序があってはならない。なぜならば、3 つの pathway が

この並行に並んでいるというこの pathway のならばにおいて a, b, c の3者は対等で、結合順の違いは存在しないからである。

⊗ の結合則については、 $a \otimes b$ を考えたとき、その a という値自身がすでになんらかの遺伝子を経由した値である、つまり、 $a = x \otimes y$ であったということがありうる。このとき、pathway が前から後ろにつながる (⊗ が左結合である) と仮定した場合、 $x \otimes y \otimes z \otimes w$ について、その演算結果は $((x \otimes y) \otimes z) \otimes w$ となる。一方、 $a \otimes b = (x \otimes y) \otimes (z \otimes w)$ であるから、この両者が等しくなるためには ⊗ が結合的でなければならないことがわかる。

最後の性質、⊕, ⊗ に対する分配則については、遺伝子調節システムが遺伝子をノードとするネットワークをなしていることを考えると、以下で述べるように導かれる。この性質を導くためのもっとも簡単な例として、図3を見てみることにする。この図に示した遺伝子調節ネットワークを演算子を使って書けば、A は C の発現量を $(a \oplus b) \otimes c$ 倍にすることになる。しかし、このネットワークを pathway に分解する (図4は図3)。これを見ると、A は C の発現量を $(a \otimes c) \oplus (b \otimes c)$ 倍にすることになる。ネットワークを pathway ベースで見る考え方にもとづけば、前者と後者は等しい値をもたなければならない。したがって、⊕, ⊗ に対する分配則が理解されるだろう。

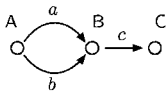


図3 簡単な例

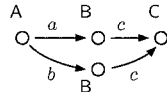


図4 パスに分解

最後に、⊕, ⊗ に関する簡便のための省略記号を2つ定義しておく。それぞれの演算子に対する性質 (可換則・結合則) にもとづく定義になっていることに注意されたい。

定義 4.4 (⊕ の省略記号) ある集合 $\{a_1, a_2, \dots, a_n\}$ に対して、

$$\bigoplus \{a_1, a_2, \dots, a_n\} = a_1 \oplus a_2 \oplus \dots \oplus a_n$$

また、上の定義を用いて以下のように、有限のイン

デックス集合 I と列 a_i に対して、

$$\bigoplus_{i \in I} a_i = \bigoplus \{a_i \mid i \in I\}$$

さらに、これらを無限集合に拡張したものととして、無限集合 $\{a_1, a_2, \dots\}$ に対して

$$\bigoplus \{a_1, a_2, \dots\} = \lim_{n \rightarrow \infty} \bigoplus \{a_1, a_2, \dots, a_n\}$$

また、無限のインデックス集合 I と無限列 a_i に対しても同様に定義する。

定義 4.5 (⊗ の省略記号) 有限列 $\langle a_1, a_2, \dots, a_n \rangle$ に対して、

$$\bigotimes \langle a_1, a_2, \dots, a_n \rangle = a_1 \otimes a_2 \otimes \dots \otimes a_n$$

4.2 代数的制約を満たす関数

定義 4.3 による制約の実現となる関数の構築のために、この代数、つまりこれら2つの演算子を、 \cdot : 正実数 \times 正実数 \rightarrow 正実数の型をもつ関数として具体的に定義しなければならない。この型は、定義 4.1, 4.2 から導かれるもので、「 a 倍にする」という概念から必然的に導かれる。

定義 4.3 の性質を満たす、上述の型をもつ関数の組は多くある。たとえば、演算子 \oplus, \otimes の見た目どおり、 \oplus に $+$ を、 \otimes に \times を割り当てると、定義 4.3 による制約を満たすことは明らかである。しかし、この関数モデルは遺伝子調節システムに対するモデルとして適当でない。たとえば、 $0.8 \oplus 0.8$ を考える。これは、ある遺伝子 A が遺伝子 B の発現量のある pathway によって 0.8 倍に、別の pathway によって 0.8 倍にしていることを意味する。ここで $\oplus := +$ では、この2つの pathway による和として、A は B を 1.6 倍にすることになる。発現量を減らす (0.8 倍にする) 2つの pathway が並行していると、発現量を増やしてしまう (1.6 倍にする) というのは遺伝子調節システム概念に合わない。したがって $\oplus := +$ 、 $\otimes := \times$ に割り当てる関数モデルは、ここでは採用され得ない。

そこで、ノックダウン実験のしくみをモデル化し、そこから関数を導くことを考える。そのため、ノックダウン実験によって得られた、遺伝子調節量がどのような意味をもっているのかを考えていく。

図5に示したように、ノックダウン実験では、A をノックダウンしたときの B の発現量とノックダウ

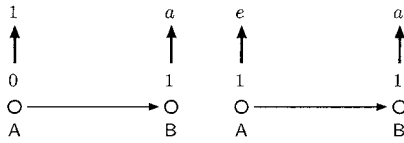


図5 ノックダウン実験の
意味

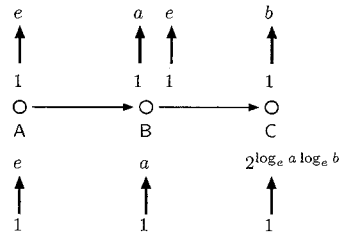


図6 より実際的な
クダウン実験の意味

図7 演算子 \otimes に対する関数

ンしないときのBの発現量との比をとったものをその値としている。すなわち、Aの発現量が0から1に増えたとき、Bの発現量が1からいくつに増えるかがその値である。しかし、生物学的実験において発現量が0になることは実際にはまずありえない。

そこで、実際にノックダウンしたときの発現量は0ではなく $\frac{1}{e}$ (e は $e \geq 1$ の任意の実数であり、ここでは自然対数の底に限ったものではない) になると考えることができる。その状況を図にしたものが図6である。

ここで、ノックダウン実験の値の意味を新たにとらえ直してみることにする。図6によれば、Aの発現量を1から e にしたとき(つまり、 e 倍にしたとき)、Bの発現量が1から a になる(つまり、 a 倍になる)ことがノックダウン実験のもつ値の意味である。このとき、もしAの発現量を1から e^2 にしたらどうなるだろうか。このとき、Bを2回 a 倍にすると考えればいいので、Bの発現量は a^2 倍となる。またさらに、Aの発現量を1から $x (> 0)$ にすると、同様に考えて、Bの発現量は $a^{\log_e x}$ 倍となる。

さらに、これを調節の連鎖としてつなぎ合わせることができる。図6の状況で、AがBの発現量を a 倍にしたとする。すると、上述の議論にしたがって、Cの発現量は $b^{\log_e a}$ 倍になる。ここで、 $b^{\log_e a} = e^{\log_e a \log_e b}$ であるので(図7)、演算子 \otimes に対して以下の割り当てができる。

定義 4.6 (演算子 \otimes への関数の割り当て)

$$a \otimes b := e^{\log_e a \log_e b}$$

実際は、体内に存在する多くの生体物質の介在(その中にはそのほかの遺伝子も含まれる)によって遺伝子AとBの間にさまざまな pathway が存在する。ノックダウン実験で得られる値はその総体を表していると考えられる。これを考慮にいれたときにも上述の議論が成立するノックダウン実験では、pathway

の集合がブラックボックスとして扱われた結果が得られるため、そのまま図7で書かれているような一つの矢印で表される関係として取り扱ってよいからである。また、例えば5節で述べるような、AとBの間にある pathway 全てに対するものでないような小さな関係に対しても、やはりその小さな pathway 集合の総体とみなせるため、同じ議論が展開できる。つまり、ノックダウン前後の発現量の比を捉えるという立場ならば、常にここで述べた議論が成り立つ。

次に、演算子 \oplus について考える。ふたつの pathway があつたとき、その影響力の“和”はどうなるかを考えればよい(図1)。遺伝子Aの発現が遺伝子Bの発現量を、ある pathway で a 倍に、別の pathway で b 倍にすると、このふたつの pathway で遺伝子Aの発現が遺伝子Bの発現量を何倍にするだろうか。

$a \oplus b$	$a < 1$	$1 \leq a$
$b < 1$	$a \oplus b < a, c < b$	$b < a \oplus b \leq a$
$1 \leq b$	$a < a \oplus b \leq b$	$a \leq a \oplus b, b \leq a \oplus b$

表1 \oplus に対する性質

そのために、表1のような性質を考える。この性質は、repress (activate) する pathway 同士の和はより強力な repress (activate) に、repress する pathway と activate する pathway の和はその中間になる、という性質である。この性質を満たす最も基本的な関数は

定義 4.7 (演算子 \oplus への関数の割り当て)

$$a \oplus b := a \times b$$

であろう。

さて、定義 4.7 によって \oplus の、定義 4.6 によって \otimes の割り当てをした関数を考えたときに、それは前節の代数的制約の実現となっている、すなわち、定義 4.3 による性質を満たしている。ここでは証明は省略する。

4.2.1 数値モデルとオントロジーの公理

ここでは、上で定義した関数が、3 節で述べた形式オントロジー体系の公理 (定義 3.3) を満たしていることを示す。これは、形式オントロジーが調節遺伝子とノックダウン実験という特定の実験問題に対して妥当な体系となっていることの根拠となることを示している。

まず、形式オントロジーを数値モデルで見たときの定義 3.3 による推論規則の対応を示さなければならない。遺伝子調節ネットワークでは、ある閾値 t に対し、遺伝子 A の発現量が遺伝子 B の発現量を a 倍にするときに、 $a \geq t$ であれば $Inh(A, B)$ とし、 $a \leq \frac{1}{t}$ であれば $Prom(A, B)$ とする。ここで考える遺伝子調節システムでは、先に述べたように遺伝子がオントロジーでの反応に相当しているので、オントロジーでの物質に相当するものは存在しない。したがって、定義 3.3 の規則 1 から 3 について議論することはできない。規則 4 から 7 では、 R, R', R'' の 3 つの反応間に関係が記述されているが、これは遺伝子調節システムではちょうど図 2 の状況に相当する。そのため、数値モデルでの演算子 \otimes が定義 3.3 の規則 4 から 7 を満たすことを示せばよい。

$a \otimes b$	$a \leq \frac{1}{t}$	$t \leq a$
$b \leq \frac{1}{t}$	$t \leq a \otimes b$	$a \otimes b \leq \frac{1}{t}$
$t \leq b$	$a \otimes b \leq \frac{1}{t}$	$t \leq a \otimes b$

表 2 \otimes に対するオントロジーの公理の解釈

このとき、定義 3.3 による推論規則は、数値モデル上では表 2 に表される性質に対応することになる。わかりやすく述べれば、 Inh と $Prom$ 、つまり repress と activate の連鎖はちょうど正負の掛け算のように反転した結果を生み出す。つまり、たとえば repress \otimes repress=activate となる。ここで、 \otimes に割り当てられた関数モデルでの関数 (定義 4.6) の対数の底に閾値 t を選んだときに、表 2 の性質を満たすことが示せる。証明は簡単であるので省略するが、 $t > 1$ であることに注意されたい。

定理 4.1 定義 4.6, 4.7 による関数は、オントロジーの公理を満たす。

以上の議論によって、遺伝子調節システムを pathway ベースで解釈した代数モデルに対して、そこからノックダウン実験の性質を考えて導いた数値モデルのインスタンスとなる具体的な関数が存在し、それが 3 節の形式オントロジーの公理体系を完全に満たしていることを示すことができた。これにより、オントロジー体系が遺伝子調節システムに対して妥当な体系を成しており、遺伝子調節システムに形式オントロジーを適用することの正しさとなる根拠を得ることができた。

5 ダブルノックダウンによる遺伝子調節量

ここでは数値モデルのアプリケーションとして、ダブルノックダウンによる遺伝子調節量の推定を提案する。ダブルノックダウン実験によって遺伝子の発現量を比較測定する試みはいまのところないようである。ここでは、ダブルノックダウン実験を仮定して発現量を比較することを考えてみる。

具体的には、ある遺伝子 (例えば C) をノックダウンしておき、その上である遺伝子 (例えば A) をノックダウンしたときとしないときで、また別の遺伝子 (例えば B) の発現量がどれだけ変化するかを比でとらえることを考えることにする。このとき C はノックダウンしたままである。

このような値を考えることにはたとえば以下のようなメリットがある。図 8 のような状況を考えてみる。図の左では、A から B, B から C, A から C のそれぞれに activate エッジが伸びている。ここで、A から C に伸びる activate エッジは、A が B を activate し、B が C を activate した結果によって生まれたものなのだろうか？それとも、A が B に C を経由することなくいわば直接の影響をあたえているのだろうか？図 8 の右を見ればわかるように、C をノックダウンした状態で A の B に対する影響力を測ればこれは明らかになる。

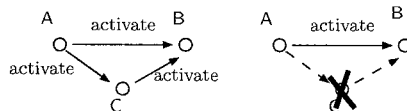


図 8 ダブルノックダウン

以下では、このようなダブルノックダウン時の遺

伝子同士の発現調節の定義とその計算手法を述べる。

5.1 定義と計算法

定義 5.1 (準原子的関係) ノックダウン遺伝子の集合 G_d と試験遺伝子の集合 G_e としたとき、 $A \in G_d$ と $B \in G_e$ に対して a_{AB} は、 A が B を準原子的に何倍にするかを表す。準原始的関係とはここでは、間に (ノックダウン実験に現れる) 他の遺伝子をはさまない遺伝子調節量をさす。

定義 5.2 (シングルノックダウン) ノックダウン遺伝子の集合を G_d 、試験遺伝子の集合を G_e としたとき、遺伝子 $g \in G_d \cap G_e$ 、 $d \in G_d$ と $e \in G_e$ ($d, e \neq g$) に対して k_{de} は、 d の発現が e の発現量を何倍にするかを表す。

定義 5.3 (ダブルノックダウン) ノックダウン遺伝子の集合を G_d 、試験遺伝子の集合を G_e としたとき、遺伝子 $g \in G_d \cap G_e$ 、 $d \in G_d$ と $e \in G_e$ ($d, e \neq g$) に対して w_{de}^g は、 g をノックダウンした際に、 d の発現が e の発現量を何倍にするかを表す。

定義 5.4 (ノックダウンによる調節値) ノックダウン遺伝子の集合 G_d と試験遺伝子の集合 G_e と遺伝子 $d \in G_d$ 、 $e \in G_e$ および遺伝子 $g \in G_d \cap G_e$ に対して定義される k_{de} を、遺伝子間の準原子的関係 a を用いて、以下のように定義する。

$$k_{de} = \bigoplus_{n \in \mathbb{N}, c_i \in G_d \cap G_e} \otimes \langle a_{dc_{i_1}}, a_{c_{i_1}c_{i_2}}, \dots, a_{c_{i_n}e} \rangle$$

ここで、 $\mathbb{N} = \{0, 1, 2, \dots\}$ である。また、 c_i は重複を許す。

この定義の意図するところは、準原子的関係の全ての相互作用の和が k になるということである。準原子的関係の相互作用は、 $G_d \cup G_e$ をノードとし、 $\{(d, e) \mid d \in G_e, e \in G_e\}$ をエッジとするグラフとなる。このグラフに対して、4.1 節で述べたような pathway への展開を行えば、上式が導かれる。

c_i に対して重複を許しているのは、遺伝子が自分自身に対してフィードバックをかけている (つまり、準原始的に自己抑制/活性化がある) 場合に、重複を許さない経路ではそれを考慮できないためである。

c_i の重複を許して任意の長さの経路を考えるため、(大きな) \oplus の引数が無限集合となっており、収束性を考慮しなければならない。しかし、この定義によ

て K が収束した値として存在するとした定義のため、この級数の収束は仮定してよい。

定義 5.4 を見たときに、準原子的関係のみからノックダウン実験の値が表現されることに疑問を感じるかもしれない。つまり、ノックダウン実験の値は生体内のあらゆる物質・反応が関わっている総体的な値であるのに対し、準原子的関係には特定の (数少ない) 遺伝子のみしか現れていないようにみえる。しかし、準原子的関係は「真に原子的な関係」とは異なる。準原子的関係はその定義から、間にノックダウン実験に現れる (数少ない) 他の遺伝子をはさまない関係の総体である。したがって、ここに登場しないいわゆる隠れパラメータについても、その影響を含んだ値となっている。そのため、準原子的関係のみからノックダウン実験の値が表現される。

また、今度は逆に隠れパラメータによって多くの activate/repress 経路を重複して加えているように見えるかもしれない。しかし、定義 5.4 中の準原子的関係は、すべて始点と終点が指定されており、それぞれが異なっている。そのため、すべて異なる経路の和を表現しているため、経路の重複は起きていない。

さらに、ダブルノックダウン関係に対して遺伝子間の準原子的関係を用いた定義を与える。

定義 5.5 (ダブルノックダウンによる調節値) ノックダウン遺伝子の集合 G_d と試験遺伝子の集合 G_e と遺伝子 $d \in G_d$ 、 $e \in G_e$ および遺伝子 $g \in G_d \cap G_e$ (ただし $g \neq d, e$) に対して定義される w_{de}^g を、遺伝子間の準原子的関係 a を用いて以下のように定義する。

$$w_{de}^g = \bigoplus_{n \in \mathbb{N}, c_i \in (G_d \cap G_e) \setminus \{g\}} \otimes \langle a_{dc_{i_1}}, a_{c_{i_1}c_{i_2}}, \dots, a_{c_{i_n}e} \rangle$$

この定義は、定義 5.4 にならったものである。定義 5.4 との違いは、 c を $G_d \cap G_e$ からとっていたのが、 $(G_d \cap G_e) \setminus \{g\}$ からとるようになったところである。つまり、ノックダウンされた遺伝子 g を遺伝子調節ネットワークの経路から排除して pathway の集まりを計算する。

ここで、定義 5.5、5.4 の元で、以下の定理が導かれる。

定理 5.1 \otimes が連続である、つまり

$$\lim_{x \rightarrow x_0, y \rightarrow y_0} (x \otimes y) = \lim_{x \rightarrow x_0} x \otimes \lim_{y \rightarrow y_0} y$$

であるとき、ノックダウン遺伝子の集合 G_d と試験遺伝子の集合 G_e と遺伝子 $d \in G_d, e \in G_e$ および遺伝子 $g \in G_d \cap G_e$ に対して、

$$w_{de}^g = k_{de} \ominus k_{dg} \otimes k_{ge}$$

この定理の直観は、 w_{de}^g は、 k_{de} から、 d から e に至るパスのうち g を経由するものすべての総和を引いたものである、ということである。 \otimes に対する連続性は、無限和の分配則を成立させるために仮定されているが、関数に対しては連続性が成り立っているのに加えて、遺伝子調節に関する数値演算子が連続的であるというのは常に妥当な仮定であるといつてよい。詳細は省略するが、この定理は \oplus, \otimes に対する代数的制約と \otimes の連続性から証明される。

この定理 5.1 を用いて、任意の $g \in G_d \cap G_e, d \in G_d, e \in G_e$ ($g \neq d, e$) に対して、 k から w_{de}^g を計算することができる。これは、前述したように、この計算によって得られた w^g を、実際のダブルノックダウン実験によって追試することが可能であるため、生物学分野から見ても非常に意味深い結果になる。

そのため、本節の冒頭で述べたようにある遺伝子 A の B への影響が C を経由したものなのかどうかを確認する手段となりうる点において意義がある。実際の実験による手法よりもコストが低いため、実験の目星を付けるための予測手段として用いることができる。ダブルノックダウン実験は通常のノックダウン実験よりもパラメータが一つ多くなるため、実験コストが冪乗で増ため、本手法のような予測手段が求められる。

5.2 実験と考察

定理 5.1 による計算実験を、[2] の二つのノックダウン実験 (早期胚と後期胚) について行った。しかし先に述べたように、本手法の目的の一つは計算結果をもとに生物学的実験手法によって追試・検証を行うことであるし、また紙幅の都合からここには結果は示さない。

6 まとめ

本稿では、遺伝子調節量の間定義される 2 つの演算子とその満たすべき制約を定義し、ノックダウン実験から導かれた演算子のインスタンスとなる関数が、その代数的制約に加えて形式オントロジーの公理を満たすことを示した。加えて、この数値モデルのアプリケーションとして、ダブルノックダウン

に基づく遺伝子調節の相互作用値という概念を数値モデルにもとづいて定義・導入し、それを計算する式を示した。この計算値は、今後生物学領域の研究・実験に活用されることが期待される。

本研究の、今後に残された研究課題として以下が挙げられる。

後方の影響を減じる関数を加えたモデル 反応速度的観点からは、遠くの反応の影響力は弱まることが予想される。この概念を反映したモデルの構築が数値計算精度の向上などに役立つと思われる。

マルチノックダウンに関して ダブルノックダウンからの拡張として、マルチノックダウンによる遺伝子調節量を考慮することができそうである。ノックダウン実験のコストはノックダウン遺伝子が増えるほど増すので、マルチノックダウンを扱う意義は大きい。

参考文献

- [1] 米崎直樹, 泉直子, 秋山卓見. 「薬物相互作用の形式オントロジー」日本ソフトウェア科学会第 22 回大会 (2005 年度) 論文集
- [2] Kaoru S. Imai, Michael Levine, Nori Satoh, Yutaka Satou. Regulatory Blueprint for a Chordate Embryo. *Science* 26 May 2006: Vol. 312. no. 5777, pp. 1183-1187
- [3] Steven Eker, Merrill Knapp, Keith Laderoute, Patrick Lincoln, Jose Meseguer, Kemal Sonmez. Pathway Logic: Symbolic Analysis of Biological Signaling. *Proceedings of the Pacific Symposium on Biocomputing*. January, 2002. pp. 400-412.
- [4] Steven Eker, Merrill Knapp, Keith Laderoute, Patrick Lincoln, and Carolyn Talcott. Pathway Logic: Executable Models of Biological Networks. *Fourth International Workshop on Rewriting Logic and Its Applications (WRLA'2002)*. Elsevier. 2002.
- [5] Saeed Tavazoie, Jason D. Hughes, Michael J. Campbell, Raymond J. Cho, George M. Church. Systematic determination of genetic network architecture. *Nature Genetics* 22, pp. 281-285 (1999)
- [6] Céline Kuttler, Joachim Niehren. Gene Regulation in the Pi Calculus: Simulating Cooperativity at the Lambda Switch. *BioConcur workshop proceeding*, 2004.