

リカレントニューラルネットワークを用いた 細胞内反応システムにおけるロバストネス瓦解部位予測

北風裕教* 松野浩嗣**
池田信彦*** 宮野悟****

Prediction of Debacle Parts for Robustness in a Cell by using Recurrent Neural Networks

Hironori KITAKAZE*, Hiroshi MATSUNO**,
Nobuhiko IKEDA*** and Satoru MIYANO****

Abstract

Living organisms have sophisticated control mechanism to keep biological system robust against abnormalities from inside/outside of them. However, at the same time, the control mechanism has a critical point at which the stability can be broken easily. This paper proposes a method to find critical points of the control mechanism in a biological pathway described by hybrid functional Petri nets (HFPN). In this method, HFPNs are converted to a recurrent neural networks (RNNs), checking robustness of the biological pathway with the RNN, and finding some crucial points for the robustness. An example to apply this method to an apoptosis pathway is also presented.

Key words : Petri Net, Genomic Object Net, Recurrent Neural Networks, BPTT

1 緒言

生体は巧妙な制御系からなり、外部と内部の異常に対して複数の分子機構が互いに補償することで恒常性を保っている。バイオインフォマティクスの分野においても、多くの研究者が生体モデルを計算機上に作成し、ロバストネスに着目した検討を行い多くの結果が得られるようになった[1]。しかし、生体は何か鍵となる分子やシステム異常により途端に恒常性が崩れるという脆弱性も持ち合わせているが、このようなロバストネスの瓦解部位についてはこれまで殆ど検討がなされていない[2]。

我々の研究グループは、代謝経路、シグナル伝達、遺伝子制御などの生命パスウェイ(経路図)をペトリネット(PN)[3]を基本アーキテクチャとして、計算機上でモデリング・シミュレーションできるツール Genomic Object Net(GON)の開発を行っている[4]。GONを用いて瓦解部位の検討を行なうためには、考慮できる経路の反応を停止(ノックアウト)する必要があり、その都度計算を繰り返す必要があるため膨大な処理時間を要するという問題点がある。

したがって本研究では、ペトリネットで記述された複雑な生命パスウェイを、時間成分を記憶可能なリカレン

トニューラルネットワーク(RNN)[5]に再形成し、自動で連続的にノックアウトを実現することで、ロバストネスの瓦解部位の推定を可能にするシステムについて提案する。またアポトーシス(プログラムされた細胞死)の生命パスウェイにおける瓦解部位について推定実験を行ったので合わせて報告する。

2 HFPN による生命パスウェイの実現

2.1 HPN の基本構成








	discrete elements	continuous elements
place		
transition		
token	integer	real number
arc	 normal	 test
		 inhibitor

図 1: HPN の基本構成要素

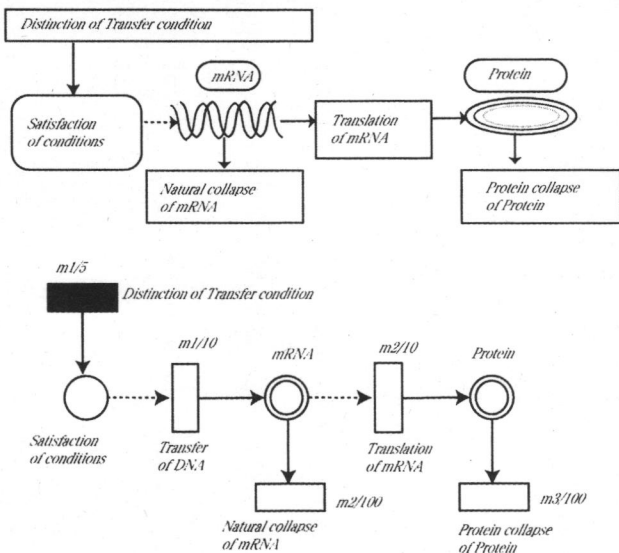


図2: 転写と翻訳

ペトリネット (PN) はグラフ表現の一種であり大別して、プレース、トランジション、アーク、トークンの4つの要素から構成される。最も基本的なPNは、トークンの値に対して離散値のみを許す離散PNであるが、生物事象は物質の濃度変化のような連続値で表現した方が都合の良い場合がある。離散値と連続値の両方を扱うことができるPNにハイブリッドPN (HPN) がある(図1)。このモデルを利用し生命事象を表現する。図2に、DNAの転写からたんぱく質の合成までの生命事象モデルをHPNを用いて表現した例を示す。

生命事象を作成する場合、次の定義に従いHPNでパスウェイを作成する。

- たんぱく質や mRNA などの遺伝子産物の濃度を連続プレースで表す。
- 細胞の形状・状態・条件を離散プレースで表す。
- リン酸化・結合・崩壊などの現象を連続トランジションで表す。その反応速度をトランジションの発火速度で表す。
- シグナルや外部刺激、形態変化のタイミングや条件の判断に要する時間は、離散トランジションで表す。各信号が影響を与えるまでの時間を離散トランジションの遅延時間で表す。
- 反応の方向、影響の種類(状態の移行、活性化、抑制化)をアークで表す。

HPN を用いれば生物事象の模式図に沿って生命パスウェイを実現できる。

2.2 HFPN への拡張

HPN は連続トランジションの発火から消費されるトークン数と出力されるトークン数の比が一定である

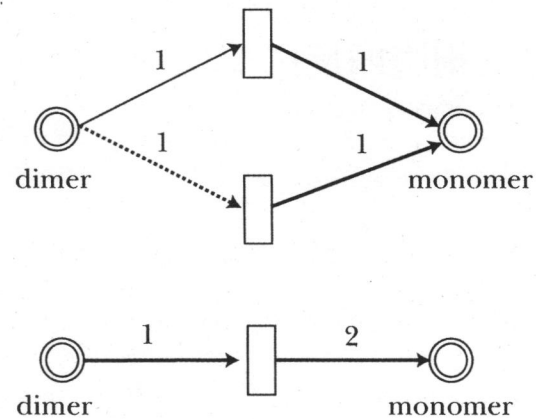


図3: 2量体から1量体への分解(上:HPN,下:HFPN)

ため、自然にモデル化できない生命パスウェイがある。例えば、1分子の2量体が2分子の1量体になる事象や、2分子の1量体が1分子の2量体になる事象である。よって我々はHPNを独自に改良し、トランジション発火速度をアークごとに定義付け可能にしたハイブリッド関数PN (HFPN) [6]を生命パスウェイの表記に利用する。図3に、1分子2量体から1量体への分解の様子をHPNとHFPNによって表現したものを示す。このように1分子2量体が2分子1量体に分解される場合や、逆に2分子の1量体が1分子の2量体に結合する場合において、HFPNはトランジションを1つ記述するのみで生命パスウェイを表現することを可能にする。

2.3 HFPN を生命パスウェイに用いる有効性

HFPN は速度論的な効果を自然に表現可能であり、ゲノムの機能情報や注釈などをモデルに容易に組み込むことができる。ペトリネットの有効性は定量的なモデル化と定性的なモデル化を同時に行える点にある。

3 生命パスウェイの再形成

HFPN で作成した生命パスウェイモデルからリカレントニューラルネットワーク (RNN) へ生命パスウェイの再形成を行い、通時的逆伝播 (BPTT) 法により学習を行った後にロバストネス瓦解部位予測の分析を行う。以下に各々について説明を行う。

3.1 リカレントニューラルネットワーク (RNN)

RNN は、ユニット間の結合が非対称に一般化される相互結合型ネットワークである(図4)。ネットワークを構成するユニットは入力ユニット (I)、出力ユニット (O)、および隠れユニット (H) の3種類のユニットに分類することができる。入力ユニットは全ての出力ユニットと隠れユニットに結合され、出力ユニットと隠れユニットはそれぞれお互いに結合されている。時刻 t におけるユニット i の出力を z_i とする。ここで、入力ユニッ

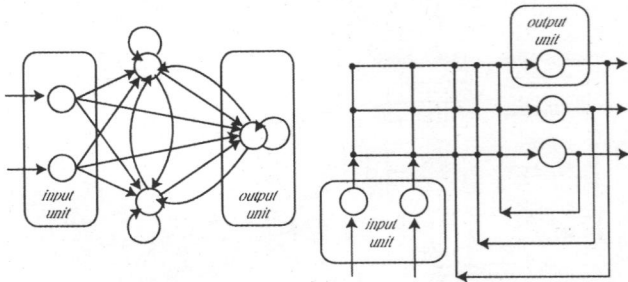


図 4: RNN 基本モデル

トの出力を $x_i(t)$, 隠れユニットと出力ユニットの出力を $y_i(t)$ とすれば, $z_i(t)$ は,

$$z_i(t) = \begin{cases} x_i(t), & i \in I \text{ のとき} \\ y_i(t), & i \in H \cup O \text{ のとき} \end{cases} \quad (1)$$

のように書き表すことができる。ここで, I, H, O はそれぞれ入力ユニット, 隠れユニット, 出力ユニットの添字の集合を表している。

リカレントニューラルネットワークは学習していない場合には, 出力ユニットの出力が外部出力として使われるということ以外には, 隠れユニットと出力ユニットの相違はないので, 離散時間 $t+1$ における隠れユニットと出力ユニットの関係は, 次のように表すことができる。

$$y_i(t+1) = f_i(s_i(t+1)), \quad i \in H \cup O \quad (2)$$

$$\begin{aligned} s_i(t+1) &= \sum_{j \in I} w_{ij} x_j(t) + \sum_{j \in H \cup O} w_{ij} y_j(t) \\ &= \sum_{j \in I \cup H \cup O} w_{ij} z_j(t) \end{aligned} \quad (3)$$

ここで, w_{ij} はユニット j からユニット i への結合荷重, $s_i(t+1)$ はユニット i の内部状態を表し, 入出力関数 $f_i(\cdot)$ は微分可能な単調増加関数と仮定されている。ここでは, $f_i(\cdot)$ の具体的な関数系として, ロジスティック関数であるシグモイド関数

$$f_i(s_i(t)) = \frac{1}{1 + e^{-s_i(t)}} \quad (4)$$

を用いる。これを用いることで, 微分を次のように求めることが可能である。

$$f'_i(s_i(t)) = f_i(s_i(t))(1 - f_i(s_i(t))) \quad (5)$$

式 (2), 式 (3) から分かるように時刻 t の入力データは時刻 $t+1$ まで出力に影響を及ぼすことなない。

3.2 通時的逆伝播 (BPTT) 法 [学習方法]

RNN の学習は, 一般的な学習として, 再帰的構造を時間展開して取り扱う手法を用いる。フィードバック結合のある非階層的なネットワークを離散時刻に対応する

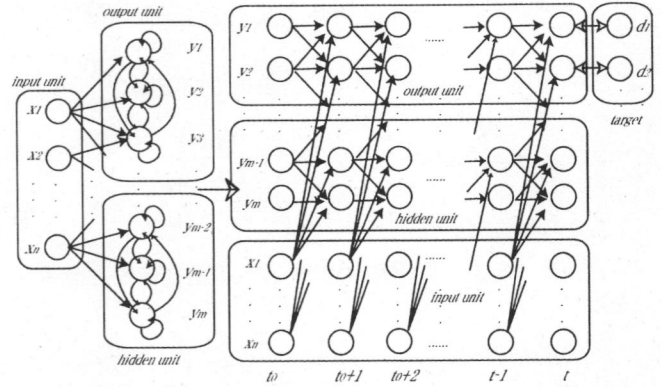


図 5: 通時的逆伝播 (BPTT) 法

階層型ネットワークとみなしてバックプロパゲーション法を適用する通時的逆伝播 (BPTT) 法 [5] を用いる。

BPTT 法はネットワークのダイナミクスを時刻 t_0 から t まで計算しながら, 過去の入出力と結合荷重の値を保持しておき, その後時刻 t から t_0 まで誤差の計算を行う。したがって, BPTT は, 離散時間モデルの場合には, 図 5 で示されるように, RNN をフィードフォワードの階層型ネットワークとみなして, 誤差逆伝播法を適用することと等価である。

このネットワークにおいて, 図 5 のように時刻 t_0 における入力 $x_i(t_0)$ と時刻 t の教師信号 $d_i(t)$ を与えたとき, 時刻 t における出力ユニット $i \in O$ の出力 $y_i(t)$ が教師信号 $d_i(t)$ となるように学習を進める。

ネットワーク全体の 2 乗誤差関数を考える。

$$E(t) = \frac{1}{2} \sum_{i \in O} (e_i(t))^2 \quad (6)$$

教師信号 $d_i(t)$ と出力ユニット $i \in O$ の出力 $y_i(t)$ との誤差 $e_i(t)$ は次のように表される。

$$e_i(t) = \begin{cases} d_i(t) - y_i(t), & i \in O \text{ のとき} \\ 0, & \text{その他のとき} \end{cases} \quad (7)$$

誤差関数 $E(t)$ を極小化するように学習則を導出する。各時刻 τ ごとの結合荷重 $w_{ij}(\tau)$ が, 時刻 τ には依存しない定数 $w_{ij}(\tau) = w_{ij}$ であると仮定する。誤差関数 $E(t)$ を結合荷重で偏微分すれば次のように計算される。

$$\begin{aligned} \frac{\partial E(t)}{\partial w_{ij}} &= \sum_{\tau=t_0}^{t-1} \frac{\partial E(t)}{\partial w_{ij}(\tau)} \frac{\partial w_{ij}(\tau)}{\partial w_{ij}} = \sum_{\tau=t_0}^{t-1} \frac{\partial E(t)}{\partial w_{ij}(\tau)} \\ &= \sum_{\tau=t_0}^{t-1} \frac{\partial E(t)}{\partial s_i(\tau+1)} \frac{\partial s_i(\tau+1)}{\partial w_{ij}(\tau)} \\ &= \sum_{\tau=t_0}^{t-1} z_j(\tau) \frac{\partial E(t)}{\partial s_i(\tau+1)} \end{aligned} \quad (8)$$

これより, $\partial E(t)/\partial s_i(\tau+1)$ が求まれば, $\partial E(t)/\partial w_{ij}$ が得られる。 $\partial E(t)/\partial s_i(\tau)$ は, 式 (2) と式 (3) を用い

て求まるため、 $\tau = t$ と $t_0 + 1 \leq \tau \leq t - 1$ に分けて考える。

$\tau = t$ (出力ユニットへの結合荷重に対する場合)

$$\begin{aligned} \frac{\partial E(t)}{\partial s_i(\tau)} &= \frac{\partial E(t)}{\partial e_i(t)} \frac{\partial e_i(t)}{\partial y_i(t)} \frac{\partial y_i(t)}{\partial s_i(\tau)} \\ &= f'_i(s_i(t))(y_i(t) - d_i(t)) \end{aligned} \quad (9)$$

$t_0 + 1 \leq \tau \leq t - 1$ (隠れユニットへの結合荷重に対する場合)

$$\begin{aligned} \frac{\partial E(t)}{\partial s_i(\tau)} &= \sum_{l \in HUO} \frac{\partial E(t)}{\partial s_l(\tau + 1)} \frac{\partial s_l(\tau + 1)}{\partial y_i(\tau)} \frac{\partial y_i(\tau)}{\partial s_i(\tau)} \\ &= f'_i(s_i(\tau)) \sum_{l \in HUO} w_{li} \frac{\partial E(t)}{\partial s_l(\tau + 1)} \end{aligned} \quad (10)$$

このように $\partial E(t)/\partial s_i(\tau + 1)$ は式 (9), 式 (10) により再帰的に計算可能である。ここで,

$$\delta_i(\tau) = \frac{\partial E(t)}{\partial s_i(\tau)}, \quad t_0 + 1 \leq \tau \leq t - 1 \quad (11)$$

とおけば,

$$\delta_i(t) = f'_i(s_i(t))(y_i(t) - d_i(t)) \quad (12)$$

$$\delta_i(\tau) = f'_i(s_i(\tau)) \sum_{l \in HUO} w_{li} \delta(\tau + 1) \quad (13)$$

$$(t_0 + 1 \leq \tau \leq t - 1)$$

となる。この $\delta_i(t)$ と $\delta_i(\tau)(t_0 + 1 \leq \tau \leq t - 1)$ は、バックプロパゲーション法 (BP 法) における δ 項に対応している。したがって、結合荷重の更新は式 (14) に従う。

$$\Delta w_{ij}(t) = -\eta \sum_{\tau=t_0}^{t-1} \delta_i(\tau + 1) z_j(\tau) \quad (14)$$

ここで、 η は学習率を表す。

4 再形成による分析

4.1 再形成定義

HFPN によりモデル化した生命パスウェイを RNN による BPTT で学習できる形態に再形成を行う。我々の提案する形成方法では、図 6 左上の HFPN は、図 6 右上の RNN の形で表される。HFPN のプレースから得られたトークンの時系列データ (濃度変化) を全てのプレースから抽出する。抽出したプレースから得られる時系列データは、RNN の入力ユニットとして定義する。さらにプレースとトランジションを RNN の隠れユニットとして定義する。但し、隠れユニットと出力ユニットとは同等である。したがって、RNN のユニット数は、HFPN のプレース数とトランジション数によって増減する。また、プレースからトランジションまたはトランジションからプレースへ結合するアークを、RNN の出力ユニット間 (隠れユニット) の結合ネットワークとする。

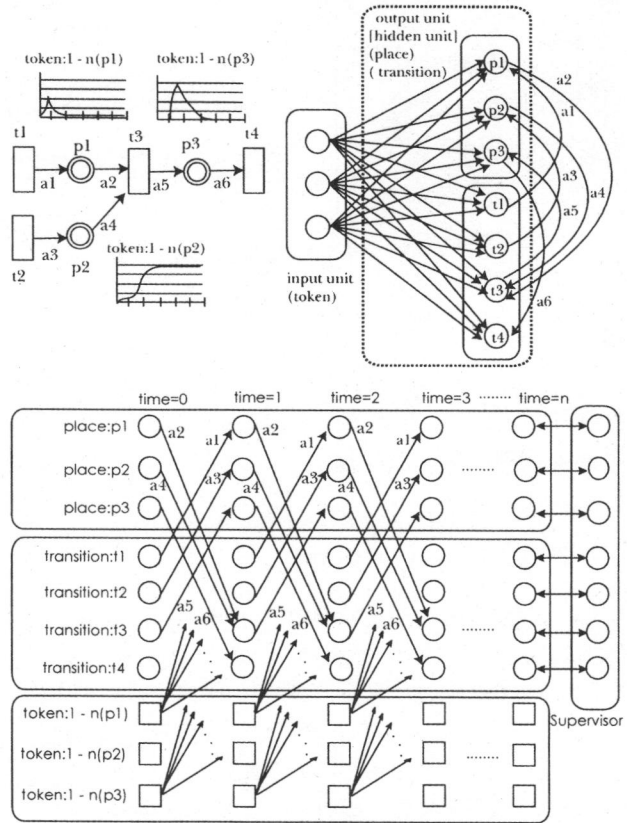


図 6: RNN への形成と BPTT への展開

BPTT で学習するために、RNN を時間展開したものを図 6 下に示す。出力ユニット間の連結は時刻が変化した場合においても基本となる RNN の連結には変化はないため、同様の連結が続く。階層数は、HFPN で得られる時系列データ数として定義する。HFPN でサンプリング間隔が小さい場合は、RNN の階層数が増加する為、リサンプリングを行う。結合荷重の初期設定は乱数とし、教師信号は、定数として定義する。生命パスウェイが複雑化すればするほど、連結されるアークの数が増加するため、安定状態からの誤差量の変動を大きく観測することが可能となり、後述する瓦解部位分析の分析制度を向上させることが可能となる。

4.2 分析方法 (瓦解部位推定)

RNN の学習後にノックアウトを連続して行い、瓦解部位の推定分析を行う。単純ではあるが重要性の高い仮想的な生命パスウェイにおいて、ノックアウトを行い RNN の結合状態を変化させたモデルを図 7 に示す。これは HFPN のプレース p_8 からトランジション T_{16} への連結をノックアウトした場合に相当する。したがって、HFPN のアーク a_{23} が削除された状態であり、RNN では $time = 1$ から $time = n$ までの全ての t_2 から p_2 への連結 a_{23} が削除される。換言すると、「HFPN におけるノックアウトは再形成した RNN では再帰的構造に時

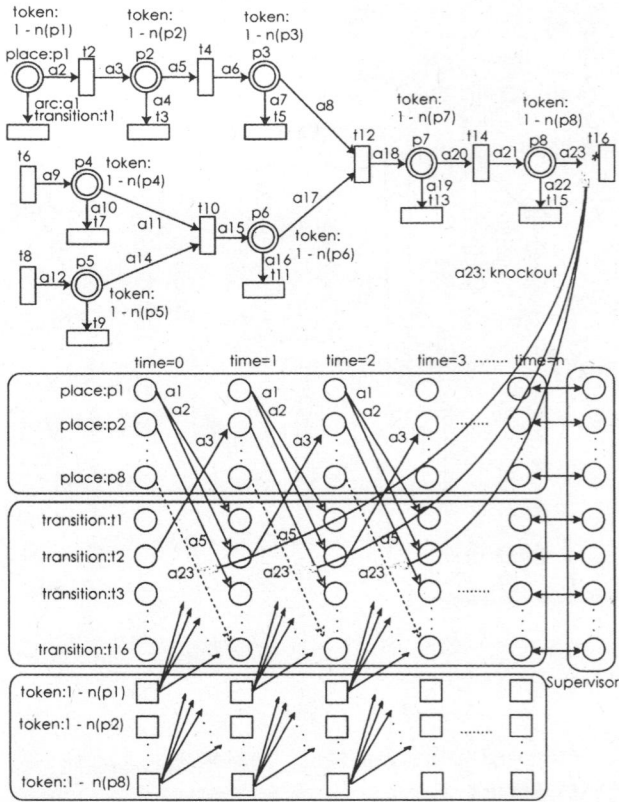


図 7: ノックアウト時における連結

間展開した各時刻における連結ネットワークの削除に相当する。展開した時間分だけユニット間の連結が同時に削除される。」といえる。このユニット間の連結をカットした状態で入力信号を与えた場合と、通常の状態を入力信号を与えた場合では、出力信号と教師信号の間には誤差量の大きさが変化して出力される。この誤差量の変化が大きいのほど、細胞内システムにおいてロバストネス瓦解部位の可能性が高いと言うことができる。

瓦解部位の推定を全ての連結部で確認するために、 a_1 から a_{23} まで順々にノックアウトを行う。

また、ノックアウトに関係するアークに接しているプレース中のトークンはゼロと考えられる。したがって、RNN においても入力ユニットからの入力信号を全てゼロとする。これにより各ユニット間に与える影響が大きくなるため、最終層からの出力信号と教師信号との誤差が大きくなり、よりはっきりと瓦解部位を推定可能とすることができる。

5 ロバストネス瓦解部位推定実験

Fas によるアポトーシスのシグナル伝達機構 [7] におけるロバストネス瓦解部位推定実験を行い、システムの有効性の検討を行う。システムの有効性を検討するための瓦解部位推定実験にこの生命パスウェイを用いた理由は、他の生命モデルに比べてフィードバックループなどの構造が少なく経路が複雑でないことにある。

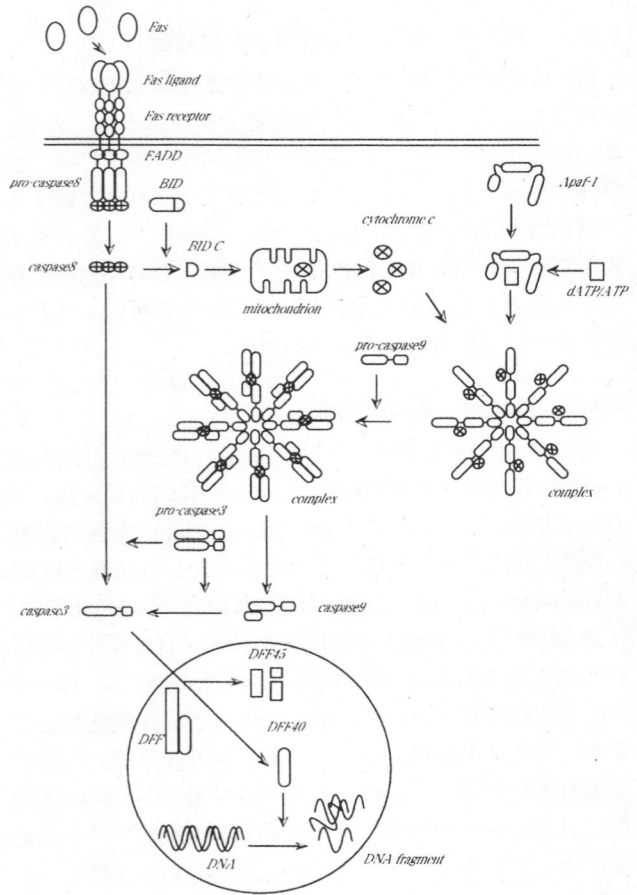


図 8: アポトーシスの経路図

5.1 Fas によるアポトーシスのシグナル伝達機構

アポトーシスとは細胞自滅や細胞自殺、プログラム細胞死、能動的細胞死と呼ばれるもので、アポトーシスによる細胞死は、生体にとって有害な細胞や不要となった細胞を速やかに除去する役割を担っている [8]。近年では、癌、劇症肝炎、自己免疫疾患、エイズ、糖尿病、アルツハイマー病やプリオン病を含む神経性疾患など数多くの病気がこのアポトーシスと関連性を持つことが報告されている。

アポトーシスは細胞の種類によっても、アポトーシスを誘導する刺激の種類によっても経路が異なり、また幾つもの経路が平行して利用されると考えられている。ここでは比較的解明されつつある Fas からアポトーシスの指標である DNA 断片化までの経路についてのロバストネス瓦解部位についての解明を行う。

図 8 に Fas によるアポトーシスのシグナル伝達機構 (Fas の多量体化から DNA の断片化) の経路図を示す。図より、経路が大きく二つの経路に分岐し、分岐した経路がまた一つに結合されていることが分かる。

5.1.1 Fas 受容体と Fas リガンドの結合

Fas は腫瘍壊死因子 (TNF)/神経成長因子 (NGF) 受容体ファミリーに属する 型膜貫通タンパク質であり、細胞内にアポトーシスのシグナルを伝える特徴を持つ受容体である。Fas リガンドは、TNF ファミリーに属する 型膜貫通タンパク質であり、Fas 受容体の on/off を調節している。受容体の活性化に三量体以上の複合体の形成が必要となる。通常三量体として存在する Fas リガンドによって受容体の三量体化を導く。

5.1.2 caspase8 の活性化

受容体の三量体が起こると death-inducing signaling complex(DISC) と呼ばれるタンパク質複合体が Fas リガンド依存的に Fas の death domain(DD) を含む細胞内領域に結合する。DISC には2つのタンパク質 FADD と caspase8(pro-caspase8) が含まれている。FADD の N 末端側には、death effector domain (DED) と呼ばれる細胞死実行に必要な領域が、C 末端側には Fas の DD を含む細胞内領域への結合に必須な DD 構造が存在する。また、前駆体 caspase8 の N 末端側には、FADD の DED への結合に必須である二つの DED が、C 末端側には caspase 様構造を持っていて、これが、Fas の刺激によって切断され活性型 caspase8 に変換される。前駆体 caspase8 は特定の酵素によって切断されるわけではなく、前駆体でも弱いプロテアーゼ活性が存在する。そして、Fas が多量体化されたことで、それに結合する DISC も多量体化を引き起こし、前駆体 caspase8 が細胞の中で部分的に濃縮される。酵素は、活性が低い場合も基質が多いと触媒できるという性質があるため、前駆体 caspase8 同士が互いに活性化しあうことで、前駆体を切断でき、活性型 caspase8 ができる。

5.1.3 ミトコンドリア機能消失によるシトクロム c の放出

活性型 caspase8 により、Bcl-2 ファミリーである BID が切断され、切断された C 末端が細胞質からミトコンドリアに運ばれることによって、ミトコンドリアの機能の低下が起こる。通常ミトコンドリアの内膜部位に局在する、13kDa の水溶性電子伝達タンパク質であるシトクロム c が、ミトコンドリアの機能低下により細胞質へ漏出される。

5.1.4 caspase9 の活性化

130kDa 細胞質モノマーである Apaf-1 に dATP/ATP が結合し、dADP/ADP へ加水分解され、そこへ露出したシトクロム c が Apaf-1 に結合する。すると、Apaf-1 とシトクロム c は多量体化を起こす。そして、前駆体 caspase9 が Apaf-1 と 1 対 1 の割合で結合するので、前駆体 caspase9 が細胞の中で部分的に濃縮される。よっ

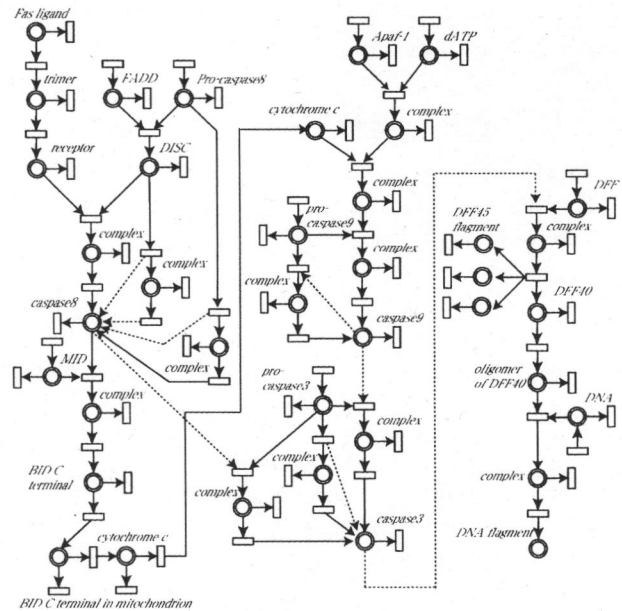


図 9: アポトーシスの経路図

て、caspase8 の場合と同様に、前駆体 caspase9 同士が互いに活性化しあうことで、前駆体を切断でき、活性型 caspase9 になる。

5.1.5 caspase3 の活性化

caspase9 は、前駆体 caspase3 に作用して前駆体を切断し、活性型 caspase3 をつくる。caspase3 の場合は、caspase8 や caspase9 の場合とは異なり、前駆体酵素が部分的に濃縮されて活性化されるのではなく、前駆体 caspase3 に対して前駆体 caspase9 が酵素として働き、活性型 caspase3 となる。

5.1.6 DNA の断片化

caspase3 が DNA fragmentation factor(DFF) に作用する。DFF は、40kDa(DFF40) と 45kDa(DFF45) の異なった二量体からなる。caspase3 により DFF45 は、3 つの小さな断片に分解される。分解された DFF45 の断片が DFF40 から離れると、DFF40 の酵素である DNase ドメインが活性化される。DFF40 は多量化し、DFF40 オリゴマーが DNase として働き、DNA を断片化する。

5.2 生命パスウェイによる実験

図 9 にアポトーシスのパスウェイを HFPN で表現したモデルを示す。プレース、トランジション、アークは各々 39, 78, 126 の要素からなり、それに合わせて、RNN の入力ユニット、隠れユニット (出力ユニット) の構成を行った。濃度変化の時系列データをダウンサンプリングし 1002 とする。

5.3 学習による誤差の推移

出力層の誤差の推移の様子を、図 10 に示す。この結