A vertical purple bar on the left side of the slide, featuring a white circular graphic with a yellow dot and a small white circle.

細胞システムのモデリングと パラメータ推定に関する研究

中荃隆(理化学研究所)

仲隆(九州産業大学)

木村周平(鳥取大学)

畠山真里子(理化学研究所)

目次

1. 研究の背景
2. 研究の目的
3. 問題設定
4. 方法論
5. 結果
6. まとめ
7. 謝辞

研究の背景

- 細胞運命の決定は制御されている。
- 細胞内のシグナル伝達系の不具合が癌などの疾病を引き起こすことが知られている。

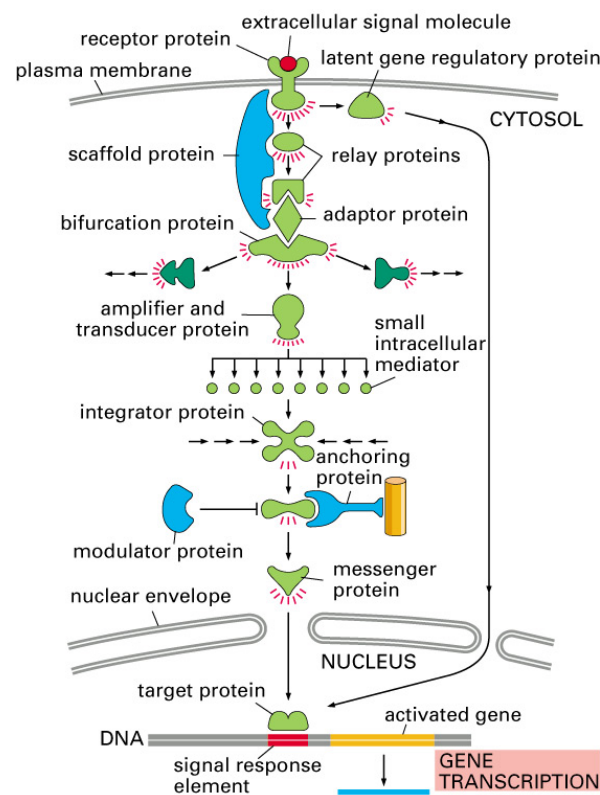
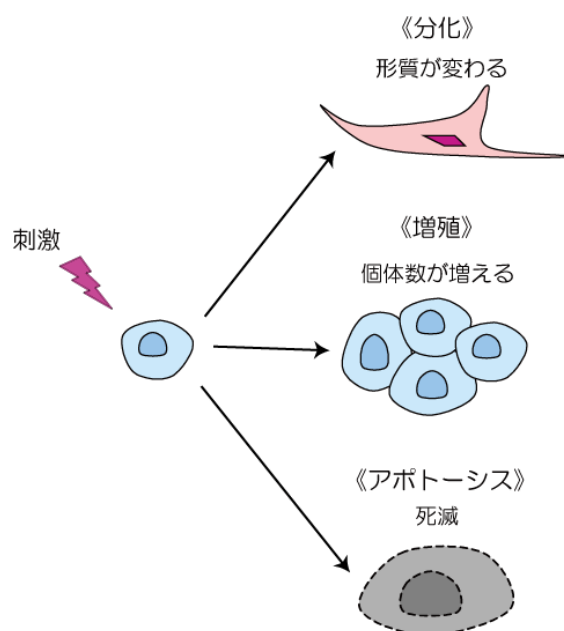


Figure 16-16. Molecular Biology of the Cell, 4th Edition.

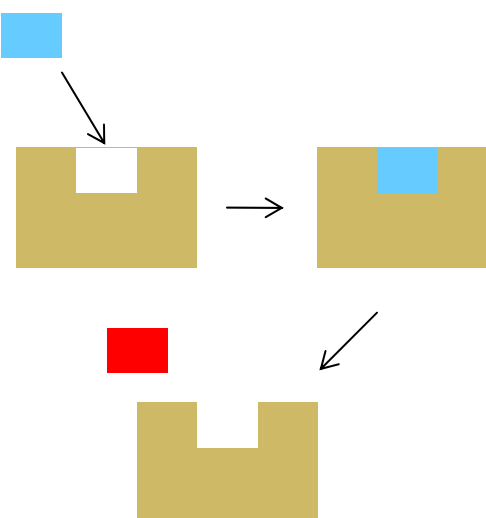
タンパク質のネットワーク

研究の背景

- シグナル伝達系の振る舞いを特徴付ける因子の1つとしてタンパク質が注目されている。

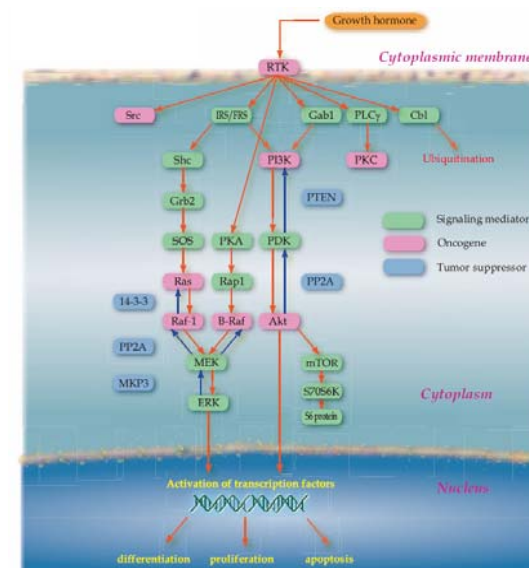


立体構造



タンパク質間の相互作用

個々の性質や相互作用



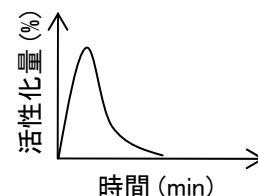
活性化の連鎖反応の
シグナルフローと時空間動態

集団としての振る舞い

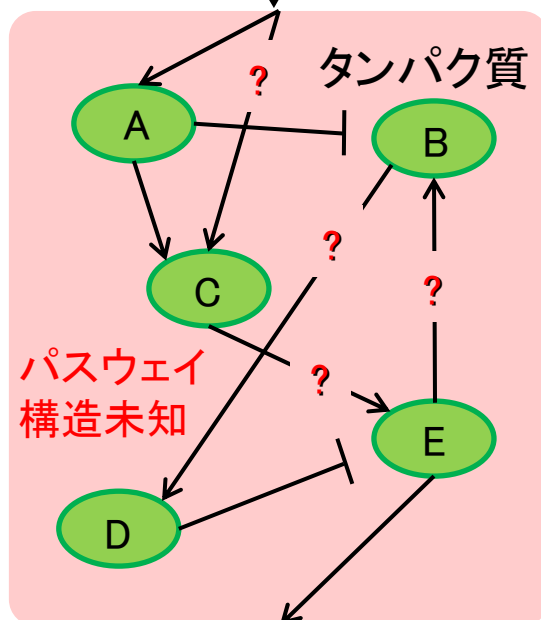
- シグナル伝達系を制御システムと捉え、そのメカニズムを解明しようとするシステムバオロジーが注目されている。

研究の目的

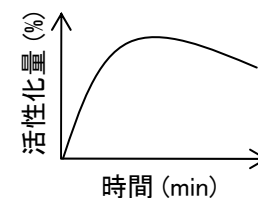
上流タンパク質の活性化



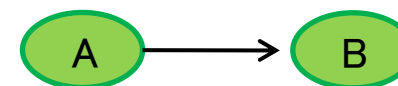
シグナル伝達系



下流タンパク質の活性化



正の制御



負の制御



シグナル伝達系のパスウェイ構造を
生化学実験データに基づき、
情報学的な手法で、
予測する手法を提案する。

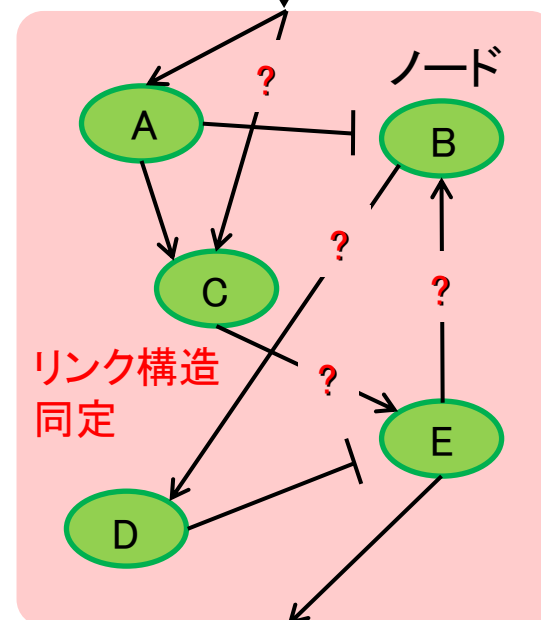
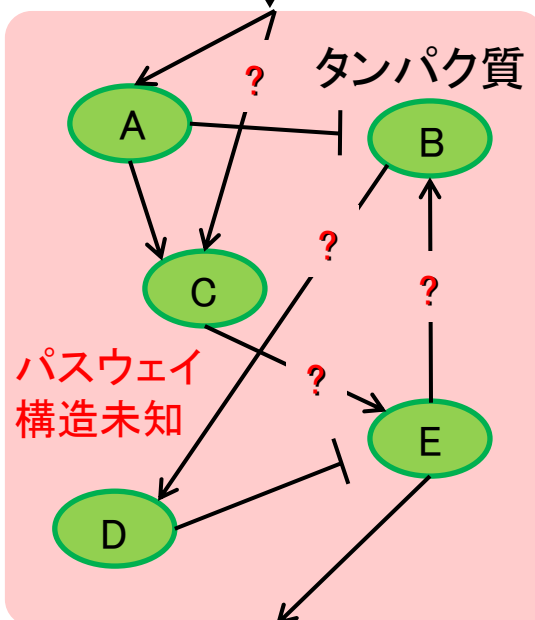
研究の目的

上流タンパク質の活性化

入力シグナル

シグナル伝達系

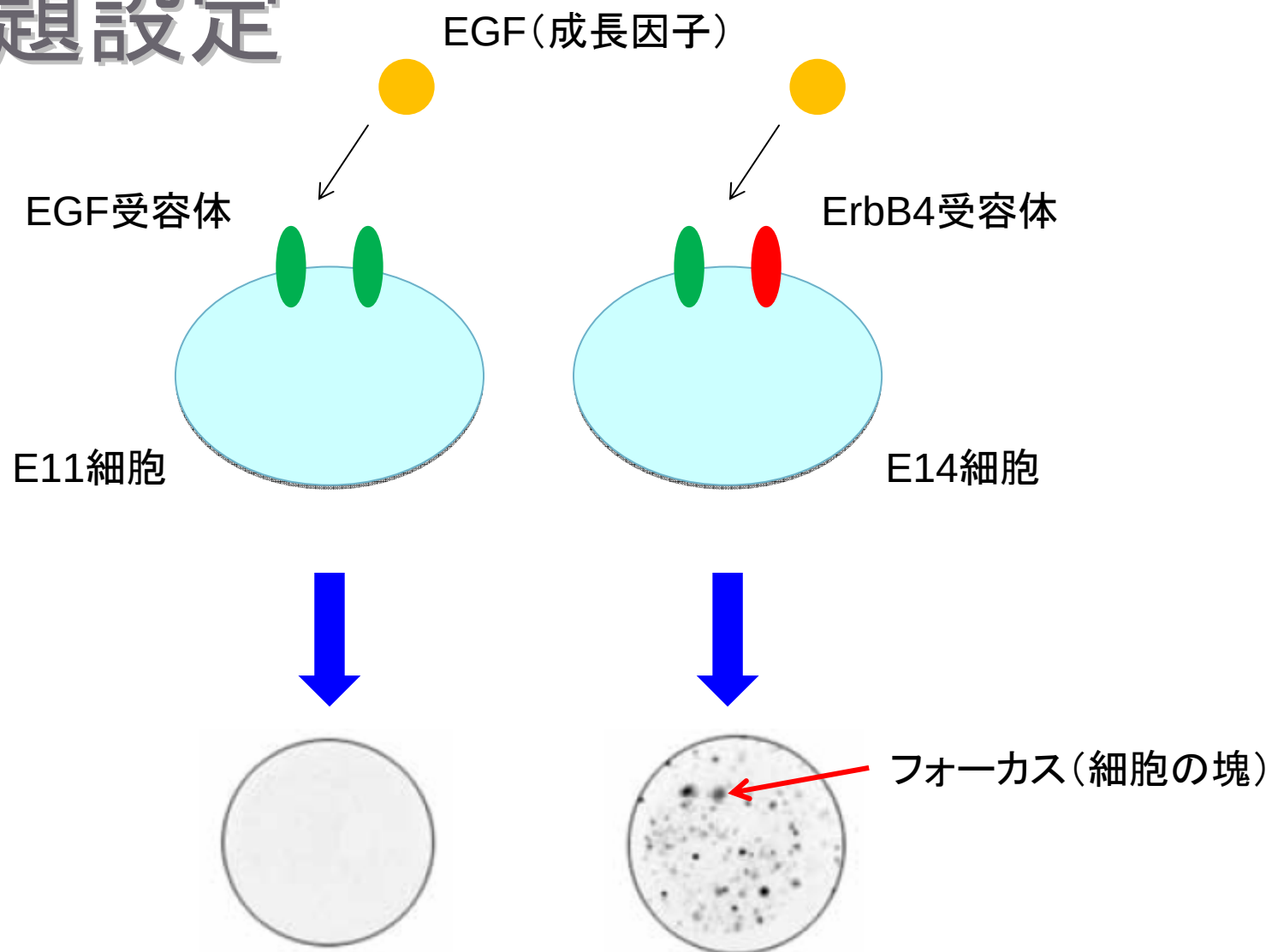
動的ネットワーク



下流タンパク質の活性化

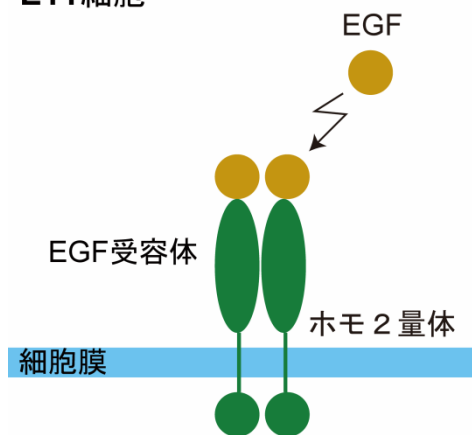
観測出力シグナル

問題設定

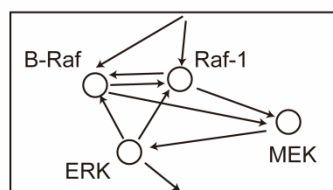


問題設定

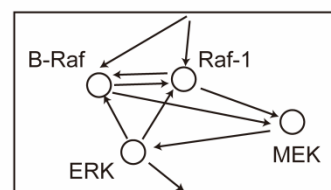
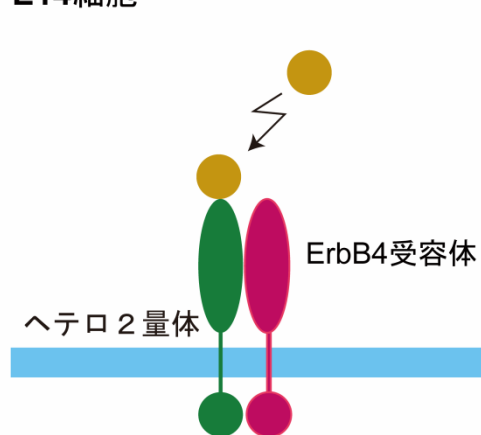
E11細胞



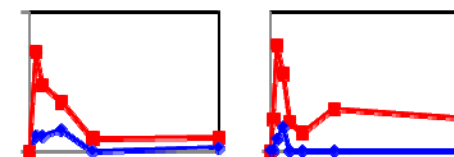
ネットワーク



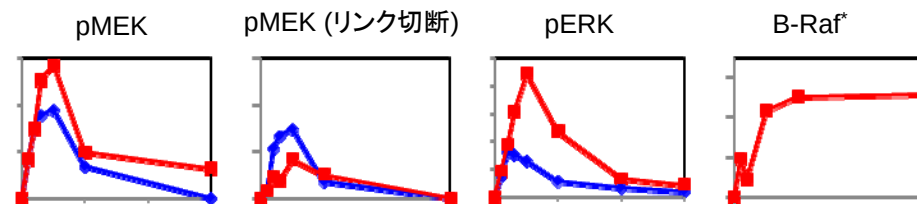
E14細胞



2種類の入力刺激の生成



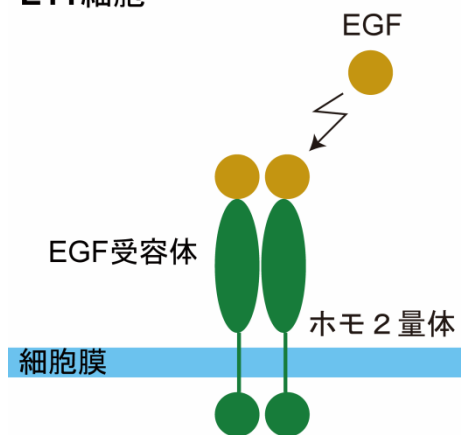
予測対象のネットワーク



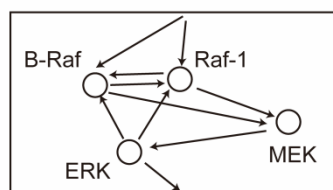
中間及び出力ノードの
シグナルの観測
(~30分)

問題設定

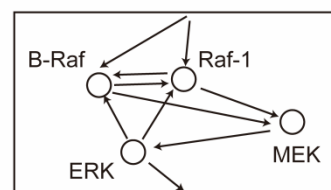
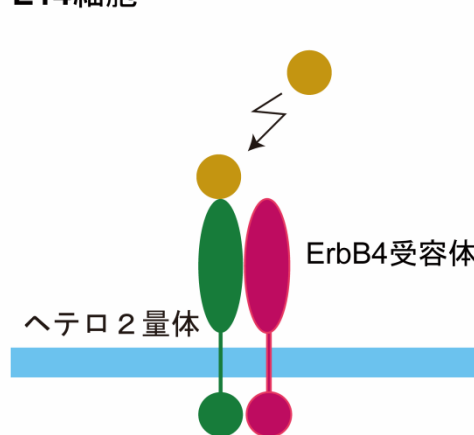
E11細胞



ネットワーク

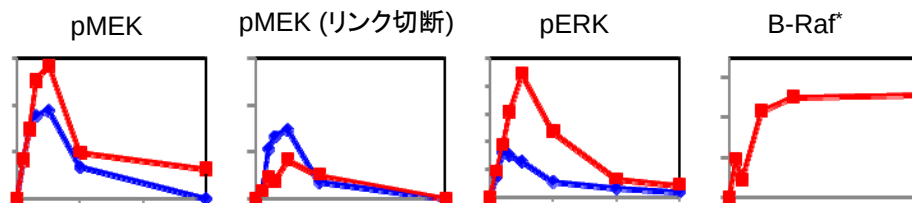


E14細胞



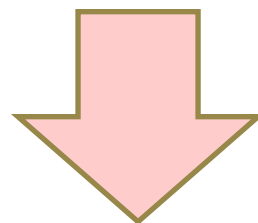
パスウェイ構造を予測

Raf-1, B-Raf, MEK, ERK
タンパクをノードとして持つ
ネットワーク



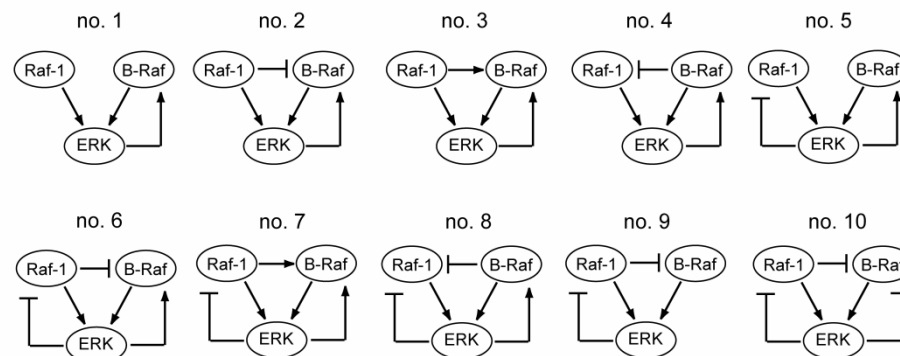
方法論

- 事前情報に基づくパスウェイ構造の候補の絞り込み
 - 1) MEK→ERKノードのリンクは既知(先行研究)
 - 2) ERK→Raf-1 or B-Rafへ正のフィードバック(実験結果)

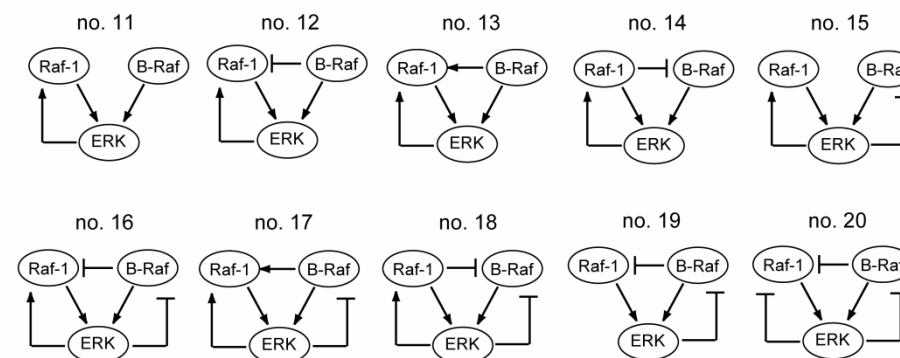


以上の条件を満たすパスウェイ構造は29種類

Positive feedback to B-Raf



Positive feedback to Raf-1



Positive feedback to B-Raf and Raf-1

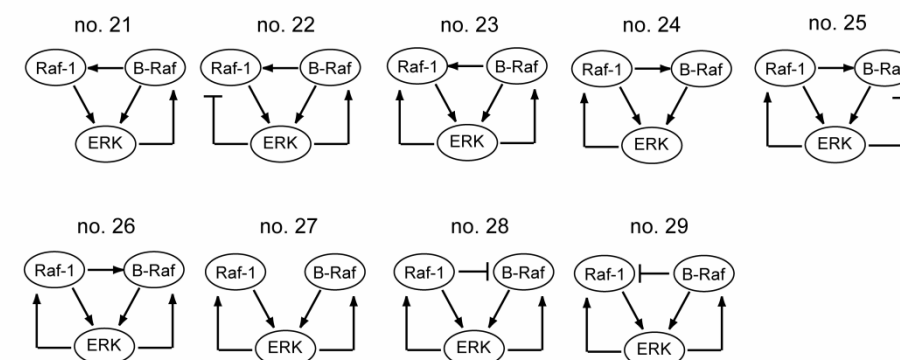
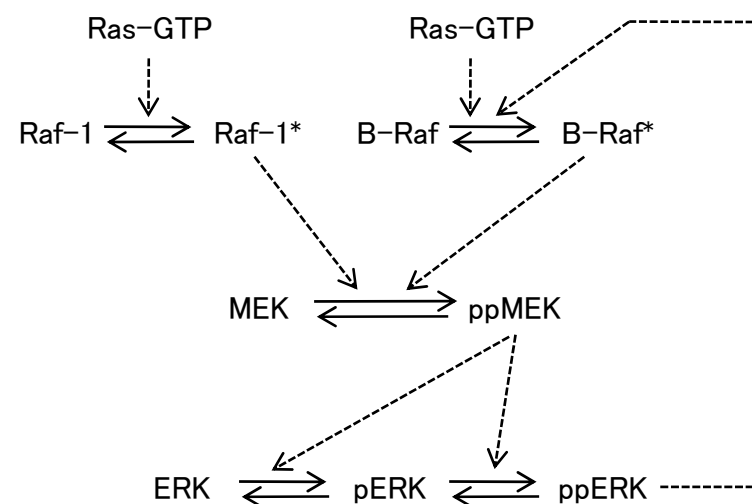
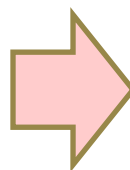
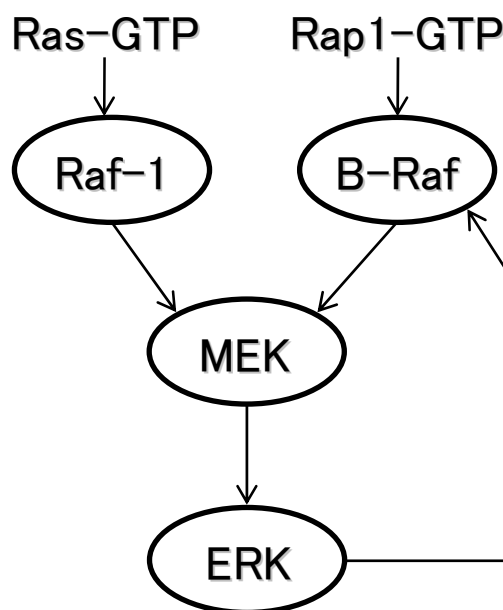
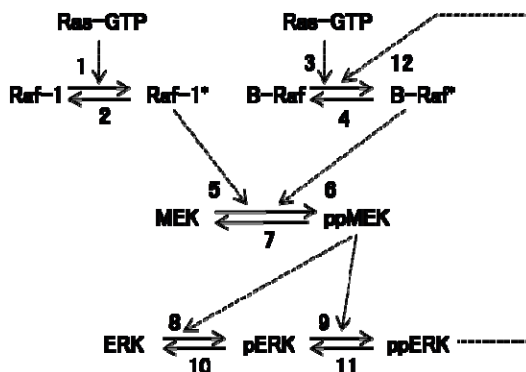


図 2. 29 種類のネットワーク構造

- 常微分方程式 (ODE) モデルの構築
29種類のパスウェイ構造それぞれに対応するODEモデルを作成

例) モデルNo.1





状態変数	9個
リアクション	12個
パラメータ	24個
未知パラメータ	28個

$$\frac{d[\text{Raf-1}]}{dt} = -\frac{V_1 \cdot [\text{Ras-GTP}] \cdot [\text{Raf-1}]}{K_{m_1} + [\text{Raf-1}]} + \frac{V_2 \cdot [\text{Raf-1}^*]}{K_{m_2} + [\text{Raf-1}^*]}$$

$$\frac{d[\text{Raf-1}^*]}{dt} = \frac{V_1 \cdot [\text{Ras-GTP}] \cdot [\text{Raf-1}]}{K_{m_1} + [\text{Raf-1}]} - \frac{V_2 \cdot [\text{Raf-1}^*]}{K_{m_2} + [\text{Raf-1}^*]}$$

$$\frac{d[\text{B-Raf}]}{dt} = -\frac{V_3 \cdot [\text{Ras-GTP}] \cdot [\text{B-Raf}]}{K_{m_3} + [\text{B-Raf}]} + \frac{V_4 \cdot [\text{B-Raf}^*]}{K_{m_4} + [\text{B-Raf}^*]} - \frac{V_{12} \cdot [\text{ppERK}] \cdot [\text{B-Raf}]}{K_{m_{12}} + [\text{B-Raf}]}$$

$$\frac{d[\text{B-Raf}^*]}{dt} = \frac{V_3 \cdot [\text{Ras-GTP}] \cdot [\text{B-Raf}]}{K_{m_3} + [\text{B-Raf}]} - \frac{V_4 \cdot [\text{B-Raf}^*]}{K_{m_4} + [\text{B-Raf}^*]} + \frac{V_{12} \cdot [\text{ppERK}] \cdot [\text{B-Raf}]}{K_{m_{12}} + [\text{B-Raf}]}$$

$$\frac{d[\text{MEK}]}{dt} = -\frac{V_5 \cdot [\text{Raf-1}^*] \cdot [\text{MEK}]}{K_{m_5} + [\text{MEK}]} - \frac{V_6 \cdot [\text{B-Raf}^*] \cdot [\text{MEK}]}{K_{m_6} + [\text{MEK}]} + \frac{V_7 \cdot [\text{ppMEK}]}{K_{m_7} + [\text{ppMEK}]}$$

$$\frac{d[\text{ppMEK}]}{dt} = \frac{V_5 \cdot [\text{Raf-1}^*] \cdot [\text{MEK}]}{K_{m_5} + [\text{MEK}]} + \frac{V_6 \cdot [\text{B-Raf}^*] \cdot [\text{MEK}]}{K_{m_6} + [\text{MEK}]} - \frac{V_7 \cdot [\text{ppMEK}]}{K_{m_7} + [\text{ppMEK}]}$$

$$\frac{d[\text{ERK}]}{dt} = -\frac{V_8 \cdot [\text{ppMEK}] \cdot [\text{ERK}]}{K_{m_8} + [\text{ERK}]} + \frac{V_{10} \cdot [\text{pERK}]}{K_{m_{10}} + [\text{pERK}]}$$

$$\frac{d[\text{pERK}]}{dt} = \frac{V_8 \cdot [\text{ppMEK}] \cdot [\text{ERK}]}{K_{m_8} + [\text{ERK}]} - \frac{V_{10} \cdot [\text{pERK}]}{K_{m_{10}} + [\text{pERK}]} - \frac{V_9 \cdot [\text{ppMEK}] \cdot [\text{pERK}]}{K_{m_9} + [\text{pERK}]} + \frac{V_{11} \cdot [\text{ppERK}]}{K_{m_{11}} + [\text{ppERK}]}$$

$$\frac{d[\text{ppERK}]}{dt} = \frac{V_9 \cdot [\text{ppMEK}] \cdot [\text{pERK}]}{K_{m_9} + [\text{pERK}]} - \frac{V_{11} \cdot [\text{ppERK}]}{K_{m_{11}} + [\text{ppERK}]}$$

$$[\text{Raf-1}](0) = R_0$$

$$[\text{Raf-1}^*](0) = 0$$

$$[\text{B-Raf}](0) = B_0$$

$$[\text{B-Raf}^*](0) = 0$$

$$[\text{MEK}](0) = M_0$$

$$[\text{ppMEK}](0) = 0$$

$$[\text{ERK}](0) = E_0$$

$$[\text{pERK}](0) = 0$$

$$[\text{ppERK}](0) = 0$$

● パスウェイ構造予測アルゴリズム

「パスウェイ構造の推定問題」を「ODEのパラメータ推定問題」に帰着させる。

モデル*i* をピックアップ

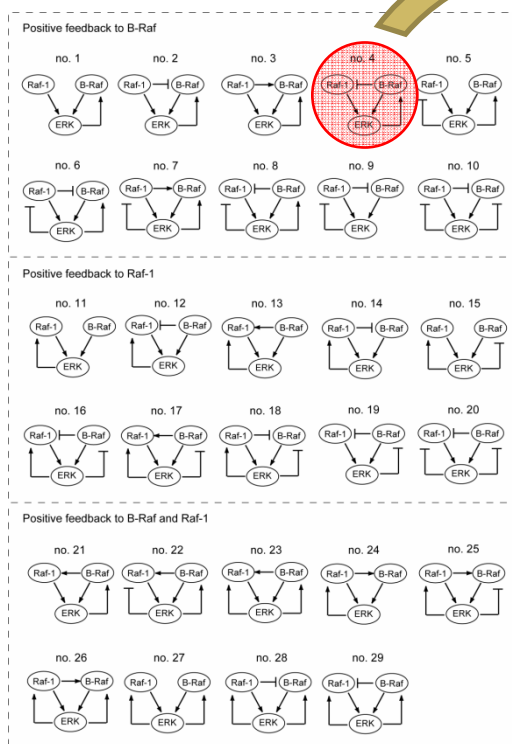


図 2. 29 種類のネットワーク構造

パラメータ推定

シミュレーション

実験データと比較

判定条件

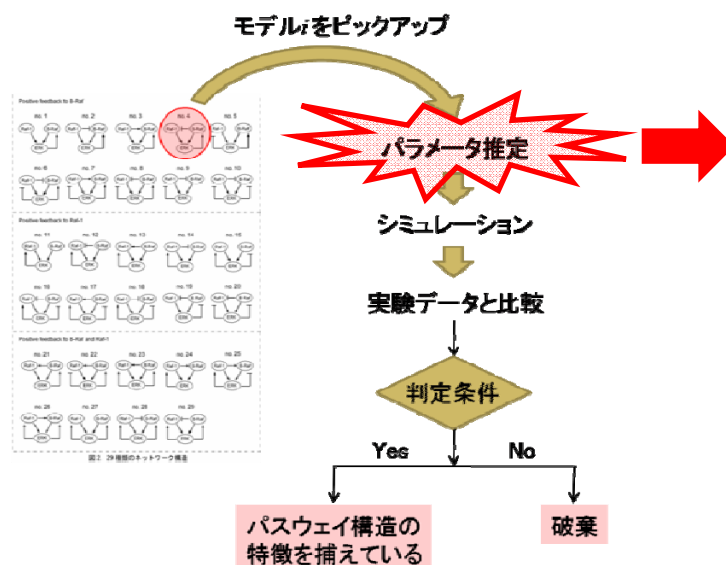
Yes

No

パスウェイ構造の
特徴を捕えている

破棄

- ・推定誤差
- ・パターン(傾向)
- ・リンク切断時の挙動



問題点

- 非線形ODEモデルのパラメータ推定は多くの計算時間を必要とする。
- 推定対象のモデルは全29種類
- 各モデルに対して、推定の際に初期探索点を変えて何度かTryする必要がある(今回は3試行)。

現実的な時間で全ての計算が終わるか？

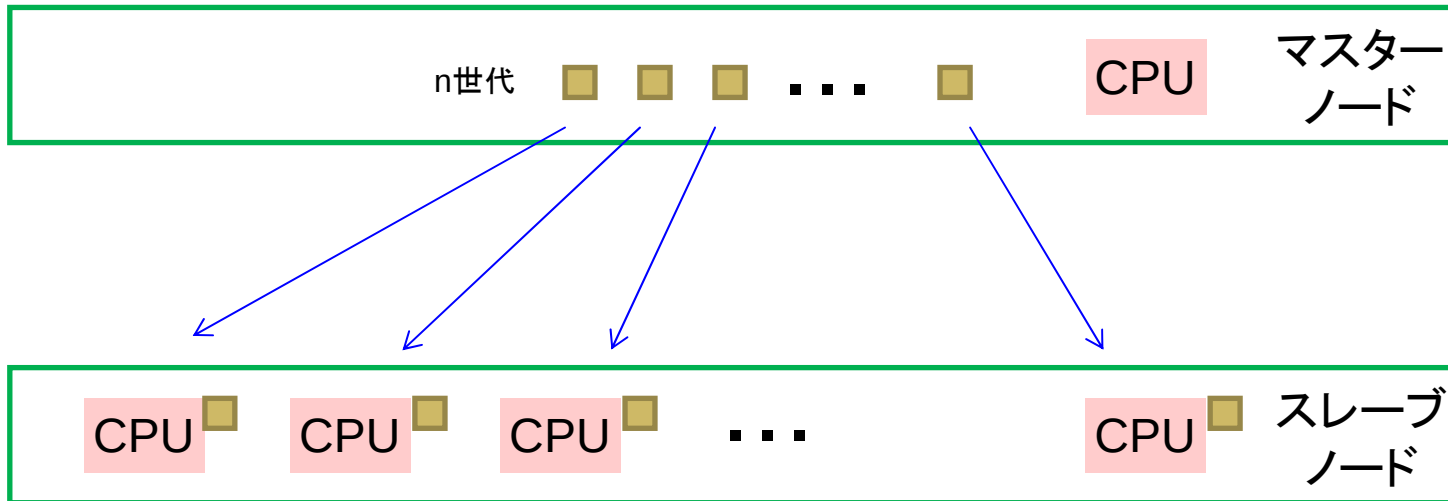


解決方法

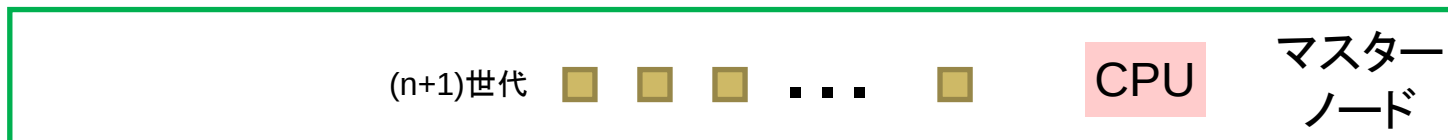
並列計算機を用いて計算時間の短縮を図る。

● パラメータ推定 (遺伝的アルゴリズム^[4])

(未知パラメータ数 × 3) 個分のパラメータセットを生成



各ノードはローカルサーチを行う



結果を用いて、(n+1)世代目のパラメータセットを生成

●パラメータ推定(誤差の計算)

$$ERR = \sum_{\substack{\text{Cell Type} \\ \in \{E1, E1/4\}}} \sum_{\substack{\text{Protein} \\ \in P}} \sum_{\substack{\text{Data Point} \\ \in D}} (\text{Experimental data} - \text{Simulated value})^2$$

ここで、集合 P 、 D は以下の表で与えられる。

No.	タンパク質 (P)	細胞	リンク切断	観測時刻(分)								実験データ(教師データ)							
1	ppERK	E1	無	0	1	2	3	5	10	20	30	0.000	7.388	15.831	15.091	12.761	5.534	3.230	1.870
		E1/4	無	0	1	2	3	5	10	20	30	0.000	9.035	18.341	30.326	44.061	23.542	6.344	4.313
2	ppMEK	E1	無	0	1	2	3	5	10	30		0.000	0.406	0.765	0.886	0.943	0.340	0.000	
		E1/4	無	0	1	2	3	5	10	30		0.000	0.414	0.744	1.262	1.415	0.486	0.319	
3	B-Raf*	E1/4	無	0	1	2	5	10	30			0.000	0.189	0.087	0.434	0.502	0.514		

結果

● 計算環境

【パラメータ推定】

- ・PCクラスタ: RSCC (160CPU)
- ・開発: C言語 (MPI)

【シミュレーション & 条件判定】

- ・Linuxマシン (AMD Opteron 2.2GHz)
- ・開発: MATLAB (The MathWorks, Inc.)

結果

●計算時間

一連の処理フローの中で、パラメータ推定部分の計算時間が支配的である。

1つのODEモデルのパラメータ推定に要した平均時間:

8時間(16CPU使用)

【比較】40時間(1CPU使用※AMD Opteron 2.2GHz)

全計算時間:

8時間 × 29モデル × 3試行 ÷ 10並列ジョブ
= 69.6時間(2.9日)

【比較】40時間 × 29モデル × 3試行 = 3480時間(145日)

160CPUを利用することで1/50倍の時間短縮

結果

抜粋
(No.1~7)

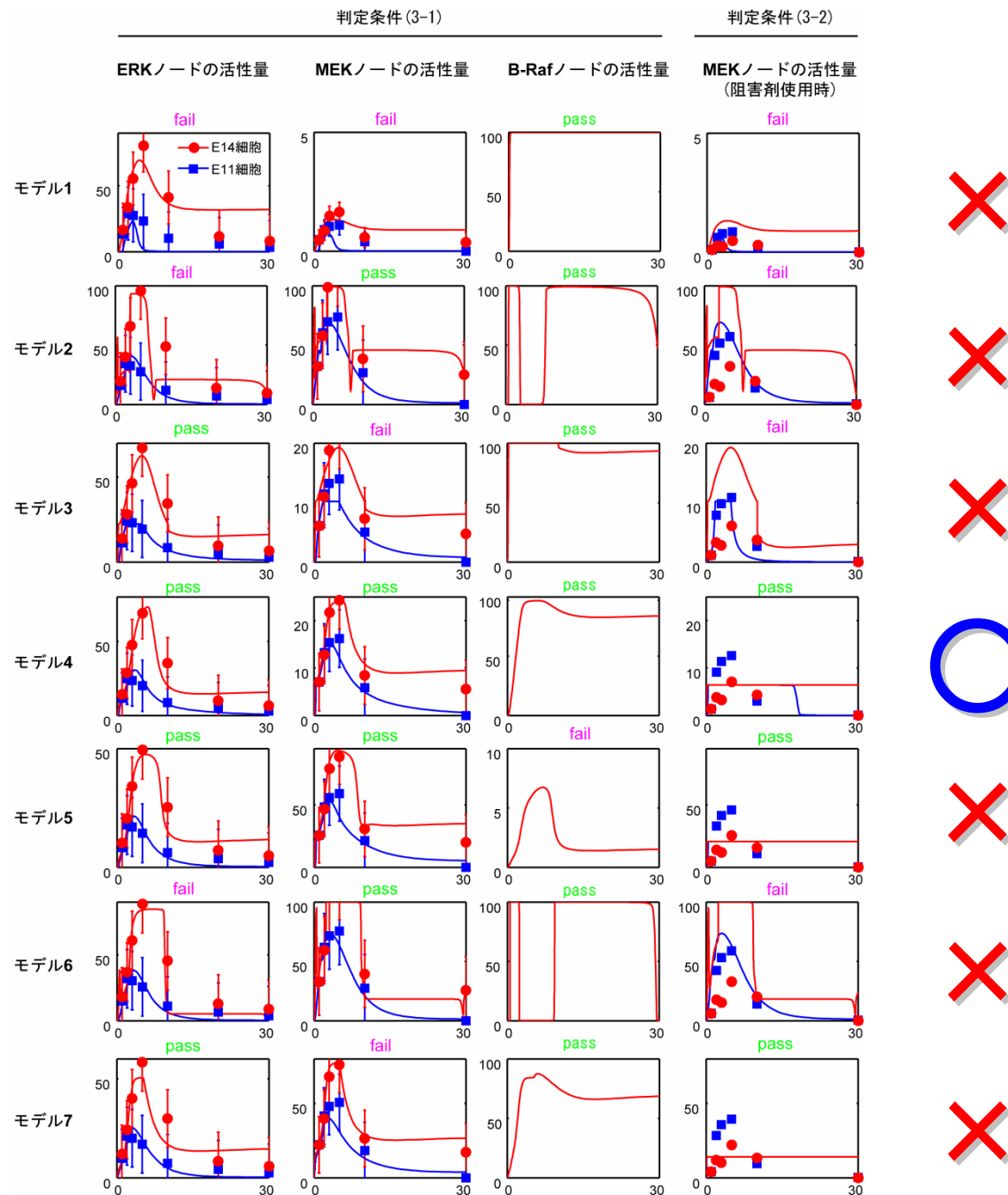


図3. 手順3の実行結果 (モデル1~7)

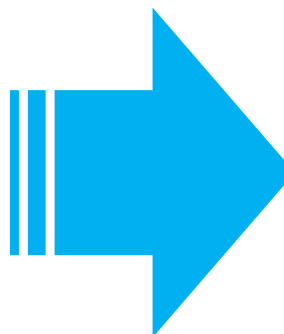
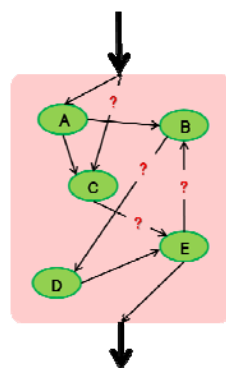
まとめ

- 細胞内シグナル伝達系において、Raf-1、B-Raf、MEK、ERKタンパク質の相互作用で構成されるネットワークのパスウェイ構造を予測する方法を提案した。
- その方法論のコンセプトはネットワークのパスウェイ構造予測問題をODEモデルのパラメータ推定問題に帰着させる点である。
- 本研究では以下の困難な点を大規模並列計算を用いて解決した。
 - ・非線形ODEモデルのパラメータ推定(非線形関数最適化)は多くの計算時間を要する。
 - ・推定は29種類のモデルに対して3試行を要する。
- アルゴリズムを適用した結果、ある特定の構造を持つモデルを予測結果として得ることができた。

まとめ

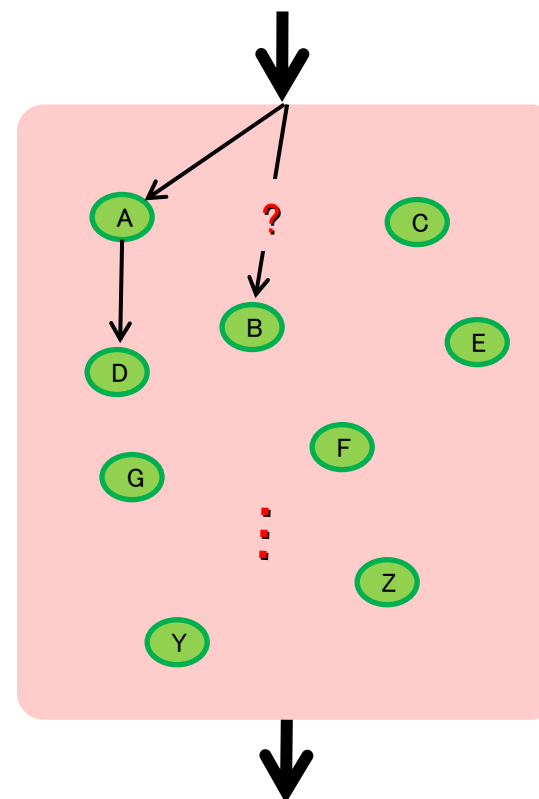
小規模なネットワーク

数十～百並列計算



大規模なネットワーク

数千～万並列計算



得られたネットワークを用いて、クリティカルパスウェイ、
薬剤効果の予測を行うことができるものと期待される。

謝辞

- CHO細胞の提供

横山茂之プロジェクトディレクター

(理研GSC タンパク質基盤研究グループ)

白水美香子上級研究員

(同グループ タンパク質システム研究チーム)

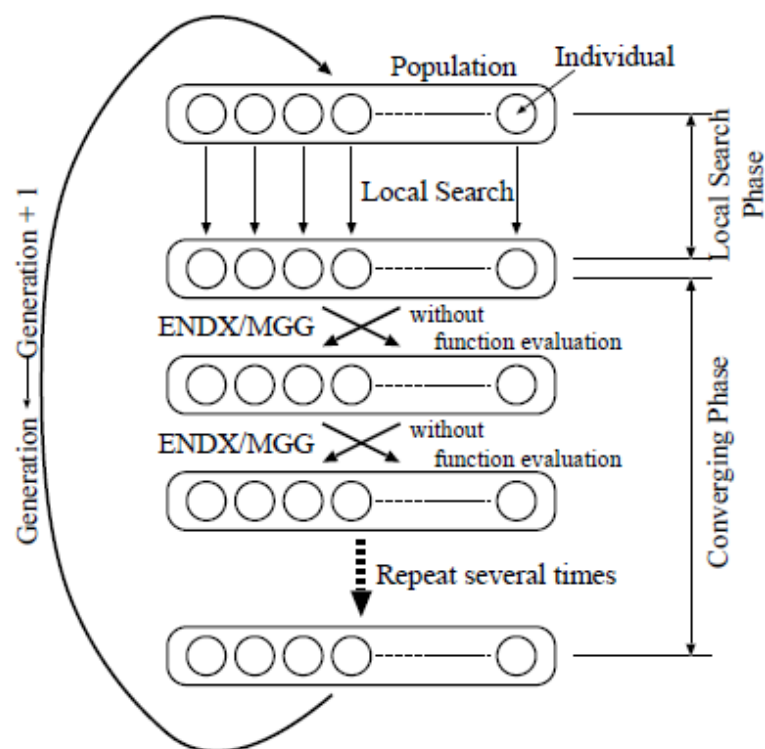
- 生化学実験

湯本範子氏

(理研GSC 情報伝達チーム)

● パラメータ推定 (推定器)

木村ら^[4]によって開発された遺伝的アルゴリズムを用いた推定器を使用:
GLSDC (Genetic Local Search with distance independent Diversity Control)



Algorithm

1. As an initial population, create n_p individuals randomly. Set $\text{Generation} = 0$ and the iteration number of the converging operations $N_{iter} = N_0$.
2. *Local Search Phase*
Apply a local search method to all individuals in the population.
3. Adaptation of N_{iter}
If the best individual in the population is not improved as compared with the previous generation, set $N_{iter} \leftarrow N_{iter} + N_0$. Otherwise, set $N_{iter} = N_0$.
4. *Converging Phase*
Execute the replacement of individuals according to ENDX/MGG N_{iter} times. There is no need to calculate the fitness value at this time.
5. Stop if halting criteria are satisfied. Otherwise, $\text{Generation} \leftarrow \text{Generation} + 1$ and go to step 2.