

進化計算を用いた遺伝子回路設計

成瀬 裕紀¹・濱田 浩幸²・花井 泰三²・伊庭 斉志^{1*}

はじめに

合成生物学の分野では望ましい生体機能を実現することが重要な課題の一つであり、そのためには人工的な生体システムの設計図が必要となる。こうした設計図の構造パラメータの探索や実験系のパラメータ調整のために、情報学における最適化アルゴリズムや人工知能・機械学習の手法が盛んに用いられている。その目的は、生体システムの機能の頑強性や生体内で働く上での安定性を最適化することである。本稿では、合成生物学における、進化計算を用いた遺伝子回路設計について具体例をもとに説明する。

進化計算と最適化手法

最適化手法を用いた生体システムの設計では、生体内の反応モデルのパラメータを調整して目標となる出力や望ましい振る舞いを実現する。反応モデルは生体システムの要素やネットワーク構造によって決まり、モデルの係数は反応速度係数などのパラメータ値に対応する。つまり反応モデルとそのパラメータ値を全体的な目標に向かって調整することは最適化問題の一種であり、情報学で培ってきた手法を活用することができる。

最適化においてしばしば問題になるのは、複数の局所解がある場合に真の最適解（全域解）を効率よく探索しなくてはならないことである。生体システムの設計では、

反応モデルの種類やパラメータ値の範囲があらかじめ定まっており、この範囲を探索空間と呼ぶ。探索空間の中で目標とする値にある程度近い局所解が複数ある場合には、局所解で探索が止まり真の最適解（全域解）に到達できないことがある。たとえば図1のように最適解近辺の裾野が狭く、局所解近辺の裾野が広いときには、探索は局所解に収束しやすく最適解にたどり着くことは困難である。

局所解を複数持つ最適化問題に有効な手法として「進化計算」がある。進化計算は、生物の集団が世代を経て進化する過程をモデル化した最適化手法である。自然界の生物の個体は、環境の中で適応したもののほど生き残り多くの子を産むという過程を繰り返す。ただしその際に遺伝子の組換えによりランダムな変異が起こる。そしてより環境に適応したものが生き残ることで「進化」を実現する。進化計算においては、一個の個体（の遺伝子型）が最適化すべきパラメータの値や数学モデルを表現し、環境への適応とは目標値への到達度（部分解の正しさ、評価値）のことである。一世代を経るごとに集団内の個体の遺伝子を組み換えて新たな個体（子孫）を作り、評価値（適合度）が高い個体が次世代の集団に残る（世代交代）。

進化計算は確率的な最適化手法であり、集団による多様性を維持することで全体的な探索空間を効率的に探索する。そのためさまざまな実問題に適用され、その有効

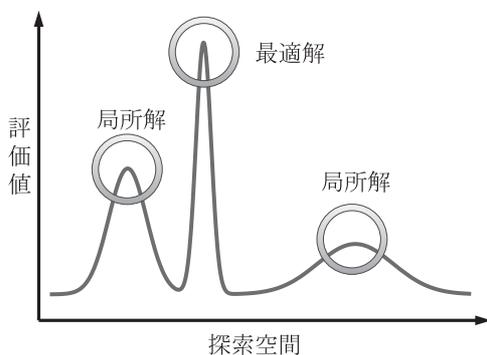


図1. 探索空間と評価値のイメージ図。一般的な最適化問題では解は一つとは限らず、もっとも評価値が高い最適解（全域解）とそうではない局所解が存在する。単純な勾配法による最適化では局所解に収束してしまい、全域解にたどり着くことは難しい。

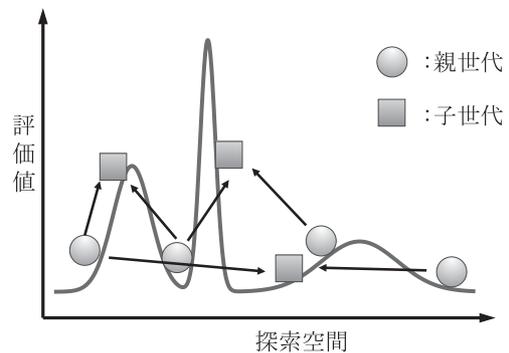


図2. 進化計算における世代交代。現在の世代（親世代）から二つの個体をランダムに選び、それぞれの持つパラメータ値の組合せで子世代の個体を作る。そして親世代と子世代の中から評価値が高い個体が次世代の個体となる。

* 著者紹介 東京大学大学院情報理工学系研究科（教授） E-mail: iba@iba.t.u-tokyo.ac.jp
¹東京大学大学院情報理工学系研究科、²九州大学大学院農学研究院

性が確認されている。合成生物学やシステム生物学でも、遺伝子ネットワークの推定や設計の問題に盛んに用いられている。

従来の最適化設計方法

2000年代から最適化アルゴリズムを用いたコンピュータによる生体システムの設計の研究が発表されてきた。Francisら¹⁾は進化計算によってスイッチ回路や発振回路となる生体システムを自動生成し、その生成されたシステムとすでに発見されていた生体システムとの構造の類似性を報告している。

またBarnesら²⁾は、ベイズ確率論的手法によって生体システムを設計する手法を提案した。目標の出力を決めると、ベイズ確率によりシステムのネットワーク構造がどれだけ適しているかを事後分布として評価する。そこで候補となる複数のネットワーク構造を考えて、その中から最適な構造とそのパラメータ値を事後分布の比較によって見つけることができる。ベイズ確率論的手法での事後分布はネットワーク構造が複雑であるほど小さくなるため、この手法によって目標出力と構造の複雑さのバランスが取れた結果を得ることができる。ただし候補となるネットワーク構造をあらかじめ選んで入力しなければならない。

既存研究での問題点 以上で紹介したように生体システム設計の研究はこれまでも行われてきたが、これらのシステムの設計結果を実際にウェットとして実現することは難しい。この理由は主に二つある。一つ目は、既存研究では最適化にS-system³⁾やHill式⁴⁾と呼ばれる近似的に作った単純な反応モデルを用いていた点である。単純なモデルはタンパク質のみを変数とした微分方程式で表される。しかし実際にはmRNA、ポリメラーゼやリボソームなど、その他の多くの生体分子もすべて変数に記述した化学反応速度式を用いて設計しなければならない。しかし反応速度式では方程式の数が単純なモデルと比べて約8倍も多くなり、効率的な最適化手法を用いても計算時間がかかりすぎてしまう。さらに方程式の数に伴って最適化すべきパラメータ数も多くなるために局所解の数が増大し、膨大な時間を費やしても真の最適解が得られる可能性が低くなる。

二つ目の理由として、既存研究ではネットワークの頑強性が評価されてなかったことがあげられる。遺伝子ネットワークの頑強性とは周りの環境やネットワーク内部の変動（摂動）に対してネットワークが安定に機能する度合いのことを指す。実世界ではさまざまなパラメータが一定でなく揺らぎを持つため、ある程度のパラメー

タ変動が生じてても所望の機能を果たすことが求められる。そのため、生体システムの設計段階では、システムの出力だけでなく頑強性も評価した最適設計が必要である。

遺伝子ネットワークによる発振回路の設計

本研究では遺伝子ネットワークの発振回路を進化計算によって設計し、その遺伝子ネットワークをウェットで実現することを目指す。そのためには遺伝子ネットワークのネットワーク構造と反応速度式のモデルでのパラメータ値をとともに最適化しなければならない。上述したように反応速度式のモデルでこれらを最適化することは困難であり、効率的に最適化するには探索空間を狭めなければならない。そこでネットワーク構造とそのパラメータ値とを二段階に分けて最適化する手法を提案する。またウェットでの再現を見据えて、この最適化において遺伝子ネットワークの発振特性だけでなくその頑強性も評価する。本研究では最適化したパラメータ値に変動を加えたときの遺伝子ネットワークの出力を評価することで頑強性を調べる。

提案する二段階最適化に置いて、一段階目では既存研究にも用いられたS-systemモデルを用いて遺伝子ネットワークのネットワーク構造とパラメータ値を最適化する。この最適化で得たパラメータ値は反応速度式のモデルとは関係しないが、フィードバックループなどのネットワーク構造は実際の遺伝子ネットワークにも有効だと考えられる。

そして二段階目では、一段階目で得たネットワーク構造を反応速度式に置き換え、パラメータ値を最適化する。反応速度式に記述する生体分子とその反応の流れを図3に示す。遺伝子は、抑制因子、促進因子およびポリメラーゼと結合することで合計4つの状態を持つ。そしてポリメラーゼが結合した遺伝子によってmRNAが生成され、mRNAとリボソームからタンパク質単量体が作られ、

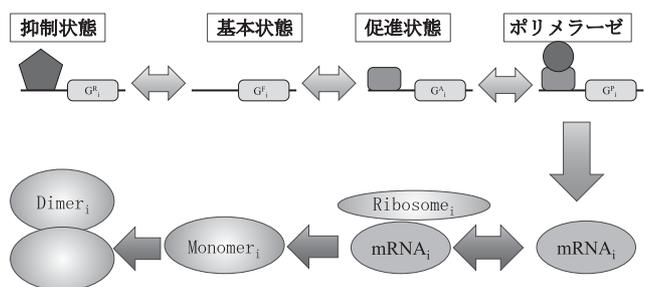


図3. 二段階目で考える化学反応の流れ。それぞれの生体分子や各状態の遺伝子が微分方程式の変数として記述される。

さらに二量体が形成される。この二量体が遺伝子の抑制因子または促進因子として働く。このときネットワーク内の一種類の遺伝子につき合計8種の生体分子がある。したがってモデルである微分方程式も一つの遺伝子ごとに8つ必要になる。

たとえば、図3における各分子の反応速度式モデルは以下のような微分方程式で表される。

$$\begin{aligned} \frac{d[G_i^R]}{dt} &= k_1[G_i^F][Repressor] - l_1[G_i^R] \\ \frac{d[G_i^A]}{dt} &= k_2[G_i^F][Activator] - l_2[G_i^A] \\ \frac{d[G_i^P]}{dt} &= k_3[G_i^A][Polymerase] - (l_3 + k_4)[G_i^P] \\ \frac{d[G_i^F]}{dt} &= -k_1[G_i^F][Repressor] + l_1[G_i^R] \\ &\quad - k_2[G_i^F][Activator] + l_2[G_i^A] \\ \frac{d[mRNA_i]}{dt} &= k_4[G_i^P] - k_5[mRNA_i] \\ \frac{d[Rib_i^{mRNA}]}{dt} &= k_6[mRNA_i][Rib] - (l_6 + k_7)[Rib_i^{mRNA}] \\ \frac{d[Monomer_i]}{dt} &= k_7[Rib_i^{mRNA}] - k_8[Monomer_i] \\ &\quad - k_9[Monomer_i]^2 - k_{10}[Dimer_i] \\ \frac{d[Dimer_i]}{dt} &= k_9[Monomer_i]^2 - k_{10}[Dimer_i] \\ &\quad - k_x[Dimer_i][G_j^F] + k_x[G_j^X] \end{aligned}$$

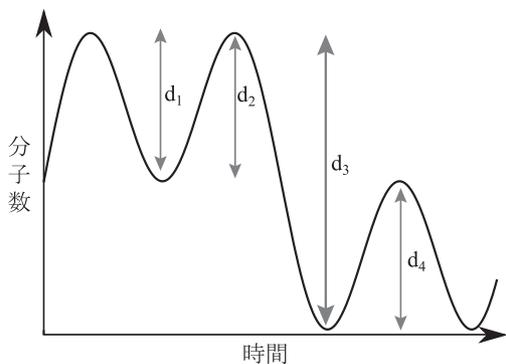


図4. 発振回路となる遺伝子ネットワークの出力例と評価値に使用する値。グラフ内の各振幅を用いることで定常振動であるかを評価する。

ここで、添え字*i*は遺伝子の番号、Gの添字は遺伝子の状態と番号を示す。またActivatorとRepressorは遺伝子*G_i*に結合する抑制因子や促進因子となるタンパク質であり、*G_j*はDimer_{*i*}が結合する他の遺伝子を表す。二段階目の最適化ではネットワーク構造からActivator, Repressor, *G_j*を決定して微分方程式に書き表し、その係数を最適化する。

評価関数 本研究では固定された時系列データの出力を求めるのではなく、定常振動の特性を持つ出力を期待する。そこで次のような定性的振る舞いを考慮した評価値を用いる。たとえばネットワークの出力の時系列データが図4のようであったとする。このとき、グラフでの極大値と極小値を求めて、各点での振幅値を計算する。この各振幅の比を用いることで定常振動の度合いを評価する。

さらに、遺伝子ネットワークの頑強性を評価するため、ネットワークの各パラメータ値にランダムな変動を加えた上で出力を計算して定常振動の度合いを求める。これを数十～数百回繰り返し平均を取ることで発振回路の頑強性を評価し、パラメータ変動に対してどれだけ安定して振動し続けられるかを推計する。

一段階目の結果 一段階目の最適化で得られたネットワーク構造を図5に示す。どのネットワークも負のフィードバック構造を含み、またフィードバック構造は複数であるものが比較的多かった。得られたネットワークはすべて3遺伝子からなるフィードバック構造を持ち、さらに2遺伝子のフィードバック構造を持つものも多かった。

二段階目の結果 一段階目で得られたネットワークを反応速度式モデルに記述して最適化した。その結果得られた発振回路の中でもっとも評価が高いものを図6に示す。このネットワークは正のフィードバックを二つと

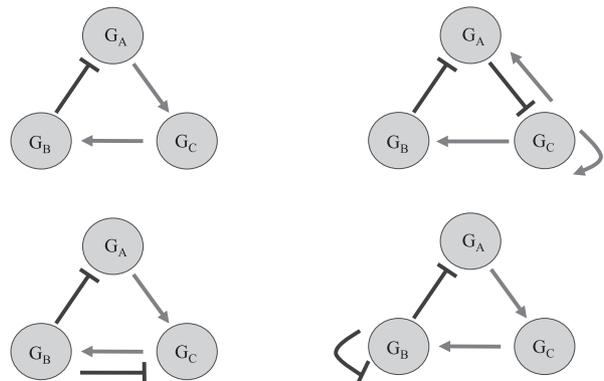


図5. 一段階目の最適化で得られたネットワーク構造

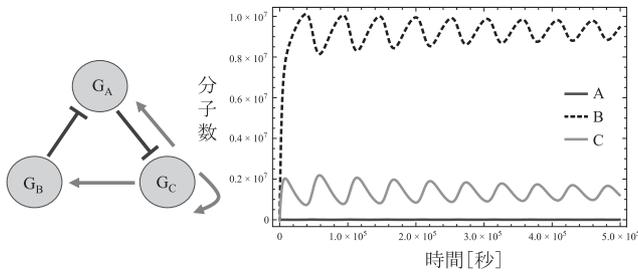


図6. 二段階目で最適化されたネットワーク構造と最適化されたネットワークの出力。

負のフィードバックを持つ。負のフィードバックだけでなく正のフィードバックも安定的な発振回路に必要であると考えられる。この遺伝子ネットワークの頑強性を評価するために、パラメータの最適値を変動させたときの振動数の変動率を調べた(図7)。グラフの軸の目盛りは元の値に対する比を表す。横軸や縦軸の“1”は、それぞれ元のパラメータ値と振動数を示す。つまり、横軸のパラメータ値変動率に対して縦軸の振動数変動率が

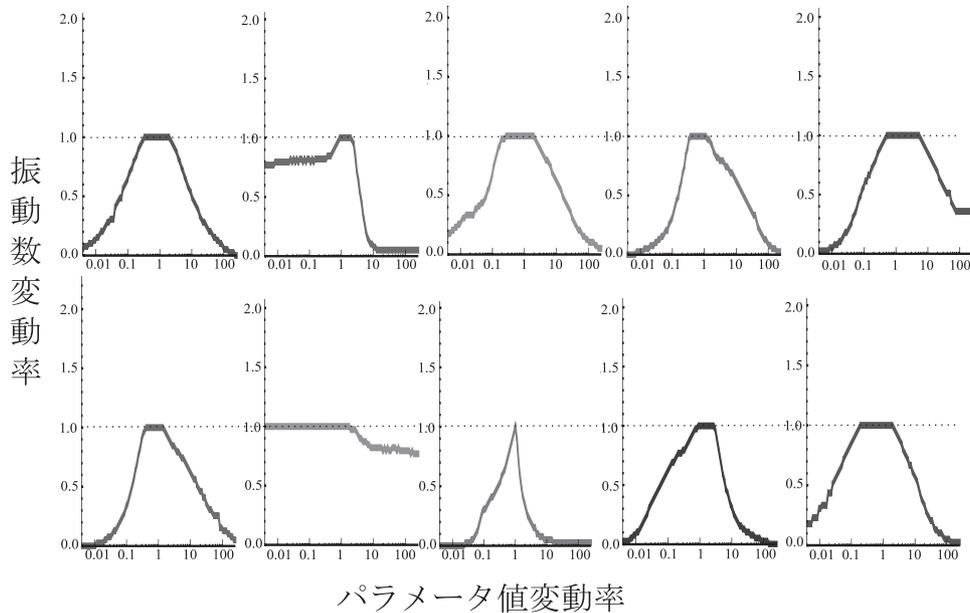


図7. 二段階目で最適化されたネットワークのパラメータ値変動率と振動数変動率のグラフ。横軸の“1”は元のパラメータ値を、縦軸の“1”は元の振動数を表す。パラメータ値変動率に対して振動数変動率の値が一定であればネットワークの頑強性が高い。

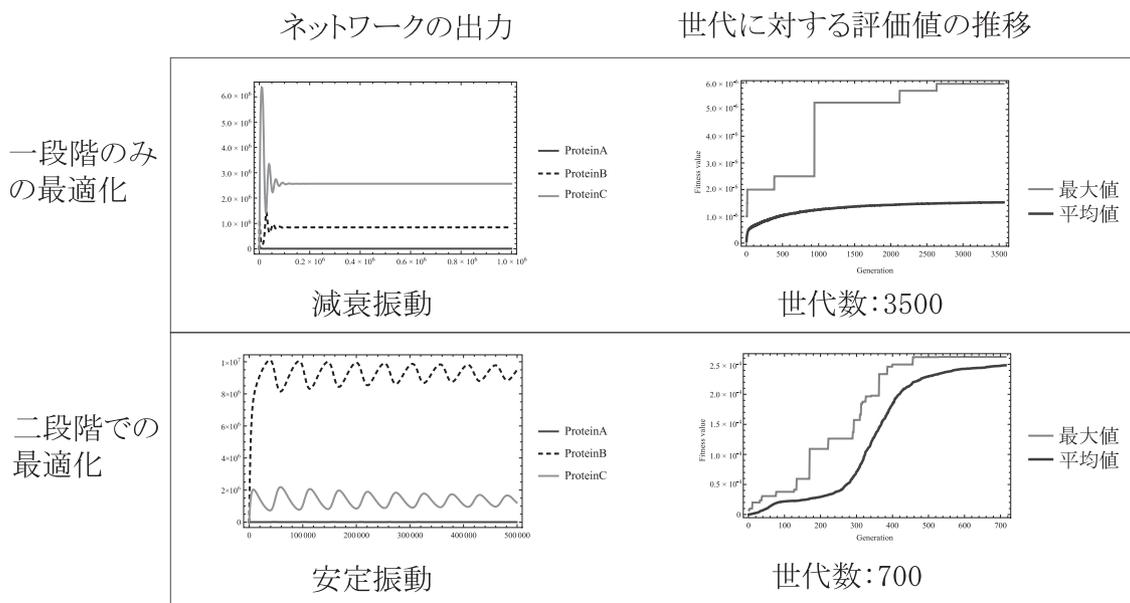


図8. 一段階のみの最適化と二段階での最適化の比較

一定である範囲が広いほどそのパラメータに関してネットワークの頑強性が高い。このグラフから、一つを除いてすべてのパラメータの頑強性が高く、中心付近に振動数一定の範囲を持っていることが分かる。つまり、この図から実際の生体内の実験においてどのパラメータの調整が頑強性を保つために重要であるかを知ることができる。

また本研究での二段階最適化の有効性を調べるため、反応速度式のモデルによる遺伝子ネットワークの設計を一段階で行ったものと二段階で行ったものを比較した。この結果を図8に示す。二段階の最適化の結果、図6のとおり安定振動する結果が得られ、その最適化にかかった世代数（進化計算における最適化のステップ数）は約700世代であった。一段階のみの最適化では、反応速度式のモデルを用いてネットワークの構造とそのパラメータを同時に最適化した。この最適化に3500世代もの世代数を費やしたにもかかわらず、減衰振動する遺伝子ネットワークしか進化しなかった。これらの結果から、反応速度式のモデルではネットワーク構造を固定することで最適化の効率が大きく変わり、二段階最適化が有効であることがわかる。

おわりに

本稿では、合成生物学のための遺伝子回路設計に進化計算を用いる手法について説明した。この他にも進化計算は遺伝子回路の設計や推定のためにさまざまに用いられている。たとえば、DNAツールキットに基づく生化学反応のための遺伝子回路設計を進化計算で行う手法も提案されている。また遺伝子ネットワークを用いたEvo-Devoアプローチ（進化発生生物学）を逆にAIや最適化に応用する試みも行われている。これらの詳細は文献5を参照されたい。

文 献

- 1) François, P. and Hakim, V.: *Proc. Natl. Acad. Sci. USA*, **101**, 580 (2004).
- 2) Barnes, C. P. *et al.*: *Proc. Natl. Acad. Sci. USA*, **108**, 15190 (2011).
- 3) Savageau, M. A. and Voit, E. O.: *Math. Biosci.*, **87**, 83 (1987).
- 4) Weiss, J. N.: *FASEB J.*, **11**, 835 (1997).
- 5) Iba, H. and Noman, N. (eds.): *Evolutionary Computation in Gene Regulatory Network Research*, Wiley (2016).