

大規模な動的な生命分子ネットワークモデルのグローバルな最適化

前田和勲、吉田圭佑、倉田博之
九州工業大学大学院情報工学研究科

ダイナミックシミュレーションはタンパク質や代謝物、遺伝子などから成る生命分子ネットワークの動的挙動を解析する手段である。生体内の動的挙動を再現するためには、リバースエンジニアリングの手法を用いて、多数の速度パラメータの値を決定する必要がある。我々はモジュール分割統合と統計解析技術を組み合わせた最適化手法 DIS (Decomposition and Integration with Statistical analysis)を開発した。これと遺伝的アルゴリズム (GA)を組み合わせてことによって、通常の GA では最適化困難な大規模生命分子ネットワークのパラメータ最適化に成功した。さらに、この最適化手法によってグローバルな最適解が得られた。

Global Optimization of Large-scale Dynamic Biomolecular Network Model

KAZUHIRO MAEDA, KEISUKE YOSHIDA, HIROYUKI KURATA
Graduate School of Computer Science and Systems Engineering,
Kyushu Institute of Technology

Dynamic simulation is a way to analyze dynamics of biomolecular networks consisting of proteins, metabolites and genes. In order to simulate their intrinsic dynamics, it is necessary to optimize the values of many kinetic parameters by reverse-engineering. We developed a new optimization method, DIS (Decomposition and Integration with Statistical analysis) combining decomposition and integration method with statistical analysis. By using DIS and genetic algorithm (GA), we optimized the parameters of large-scale biomolecular network which conventional GA cannot optimize. In this article, we also show that this method can obtain global solutions.

1. 始めに

コンピュータ科学の急速な発展と膨大なオミックスデータの蓄積によって、生命をシステムとして捉えるシステムバイオロジーという新しい研究分野が生まれた。システムバイオロジーは個々の分子に着目するのではなく、複数の分子から成るシステムの構造や機能に着目する。システムバイオロジー研究に欠かせないのが生命分子ネットワークのダイナミックシミュレーションである。シミュレーション中で生体内の生命分子ネットワークの挙動を再現するためには多数の速度パラメータを適切に設定する必要がある。このためリバースエンジニアリングによって、測定困難な速度パラメータの値を決定しなければならない。

一般に大規模な生命分子ネットワークのパラメータ最適化は困難である。これは探索空間

が多峰性かつ広大であることに起因する。また、生体内の生命分子ネットワークの挙動を再現できるパラメータ(解)は広大な探索空間に分布している可能性がある。従って、できるだけ広い範囲から多様な解を見つけ出すことが必要である。

これらの問題を克服するために、我々は新しい最適化手法 DIS (Decomposition and Integration with Statistical analysis)を開発した。また DIS を大腸菌の窒素同化システムに適用し、DIS の有効性を検証した。加えて本研究では、生物学的な解の多様性という新しい概念を提唱する。

2. 最適化手法

DIS はモジュール分割統合法²⁾と統計解析技術を組み合わせた最適化手法である。DIS

はモジュールを分割統合することによる探索空間の狭小化を避けるために、分割した空間において多様な解を作成して、統合空間でそれらを一様に混合し、解の局所的収束を防ぐ。DISの手順は次の通りである。

1. モデルをいくつかのモジュールに分割する(図 1-a)
2. モジュールごとに最適化を行い、適合度がある程度良いパラメータベクトルを抽出する(図 1-b)
3. それらのパラメータベクトルをモジュールごとにいくつかのクラスターに分ける(図 1-c)
4. 各クラスターを構成するパラメータベクトルの中心パラメータベクトルを作る(図 1-d)
5. 各モジュールのクラスター中心パラメータベクトルを組み合わせる(図 1-e)
6. 組み合わせたパラメータベクトルの近傍を探索する(図 1-f)

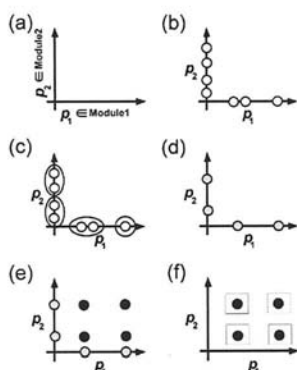


図 1 DIS のプロセス (2 次元の場合)

3. 生物学的な解の多様性

統計学的意味での解の多様性とは、得られた解の値のばらつきのことである。これとは別に、本研究では生物学的意味での解の多様性という概念を提唱する。生物学的な解の多様性とは最適化によって得られた解を用いたシミュレーションの挙動の多様性である。

4. 窒素同化システムのパラメータ最適化

4.1. 窒素同化システム

大腸菌は環境中のアンモニアと TCA 回路の α -ケトグルタル酸からグルタミン酸、そしてグルタミンを合成する。窒素同化システムはグ

ルタミン合成酵素(GS)の活性と発現量を調節し、グルタミン(N)と α -ケトグルタル酸(C)の濃度比(N/C)を一定に保つように働く。環境中のアンモニア濃度が減少するとグルタミン合成速度が低下して、N/C が低下する。このとき、窒素同化システムは GS を活性化、また GS の発現量を増加させてグルタミン合成速度を上昇させ、N/C を増加させる。窒素同化システムを単純化したネットワークマップを図 2 に示す。図 2 において角が丸い長方形はタンパク質、角が尖った長方形は遺伝子、三角形は遺伝子制御領域を示す。矢印($\rightarrow$)は結合、物質変化、転写・翻訳を示す。円を持つ線($\circ$)は触媒作用、垂直な線を持つ線($\perp$)は転写・翻訳の抑制を示す。

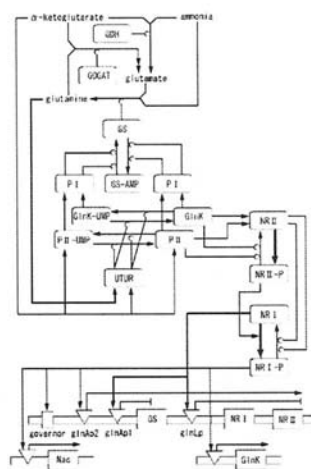


図 2 窒素同化システム

窒素同化システムは GS 制御の観点から、GS 活性制御モジュールと GS 合成制御モジュールに分割することができる。

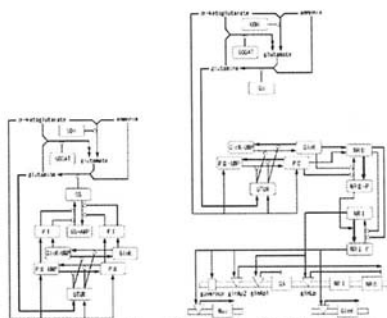


図 3 GS 活性制御モジュール(左)と GS 合成制御モジュール(右)

4.2. 数理モデル作成

遺伝子層とタンパク質層を二相分離法³⁾、代謝層をMichaelis-Menten式で数理モデル化した。この結果、代数方程式が53式、微分方程式が21式、変数(分子種の種類)が94個の数理モデルが得られた。

4.3. 評価関数

既に述べたように窒素同化システムはN/Cを一定に保つように働く。本研究では、シミュレーション開始から1000分後にアンモニア濃度を減少させ、その後、定常に達したN/Cがアンモニア濃度減少前のN/Cと等しくなったとき、最良の適合度0を返すように定義した。適合度は小さいほど良い。本研究では 10^{-2} より小さい適合度を与えるパラメータベクトルを解とした。

$$\text{Fitness} = 1 - \frac{N/C_{\text{AmmoniaLimiting}} - \text{minimum}N/C}{N/C_{\text{AmmoniaRich}} - \text{minimum}N/C}$$

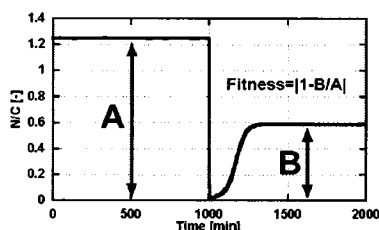


図4 アンモニア濃度減少時のN/Cの挙動

4.4. 生物学的な解の多様性の検証方法

今回は生物学的な解の多様性の検証にノックアウトシミュレーションを用いた。ノックアウトシミュレーションとは特定遺伝子を破壊した場合の生命分子ネットワークの挙動をシミュレートすることである。

大腸菌の窒素同化システムに関して様々なノックアウト実験の結果が報告されている⁴⁾(表1、表2)。大腸菌の野生株(Wild Type)はアンモニアリッチ環境からアンモニア制限環境へ移されると、総GS濃度を増加させる(表1・①)。またアンモニアリッチ環境下での $\Delta glnG$ 株の総GS濃度は野生株のものよりも小さい(表2・⑦)。最適化で得られた解を用いてノックアウトシミュレーションを行い、表1、表2に示した挙動を再現できるかを調べる。その結果を扱いやすくするためにKSV(Knockout Simulation Vector)という11次元ベクトルを定義した。KSVの各要素は0か1という値を

とり得る。次にKSVの例を示す。

①②③④⑤⑥⑦⑧⑨⑩⑪
10111001011

このKSVはノックアウトシミュレーションでの挙動がノックアウト実験の挙動と表1、表2の①、③、④、⑤、⑧、⑩、⑪について一致し、②、⑥、⑦、⑨について一致していないことを示す。最適化によって得られた解を用いて様々なノックアウトシミュレーションを行い、得られたKSVの種類が多ければ、最適化によって得られた解は生物学的に多様であると言える。

表1 NH₃濃度に対する総GS濃度の変化

	NH ₃ -rich vs. NH ₃ -limiting
①Wild Type	+
② $\Delta glnG$ (NR I)	-
③ $\Delta glnL$ (NR II)	+
④ $\Delta glnD$ (UTUR)	+
⑤ $\Delta glnB$ (P II)	+
⑥ $\Delta glnK$ (GlnK)	+

表2 遺伝的状態に対する総GS濃度の変化

	NH ₃ -rich
⑦Wild Type vs. $\Delta glnG$ (NR I)	-
⑧Wild Type vs. $\Delta glnL$ (NR II)	+
⑨Wild Type vs. $\Delta glnD$ (UTUR)	-
⑩Wild Type vs. $\Delta glnB$ (P II)	+
⑪Wild Type vs. $\Delta glnK$ (GlnK)	+

4.5. 最適化

DISを使って、窒素同化システムのパラメータ最適化を行った。その手順を次に示す。

1. 窒素同化システムをGS活性制御モジュールとGS合成制御モジュールに分割した。
2. GS活性制御モジュールとGS合成制御モジュールのそれぞれについて40000回のランダムサーチを行い、適合度が0.99より小さいパラメータベクトルを抽出した。
3. 抽出したパラメータベクトルをモジュールごとに4つのクラスターに分割した。
4. 各クラスターに含まれるパラメータベクトルの中心のパラメータベクトルを作成した。
5. 各モジュールのクラスター中心パラメータベクトル各4個を組み合わせて16個のフルモデルのパラメータベクトルを作成した。
6. 作成したフルモデルのパラメータベクトルの近傍を集中的にGAで探索した。

5. 結果

DIS を用いて窒素同化システムのパラメータ最適化を行った結果、141 個の解が得られた。解に含まれる個々のパラメータの最大値と最小値を図 5 に示した。全パラメータの探索範囲を $10^2 \sim 10^{10}$ としたことを考慮すれば、多くのパラメータは広い範囲の値をとっていると言える。このことから、最適化によって統計的に多様な解が得られたと言える。

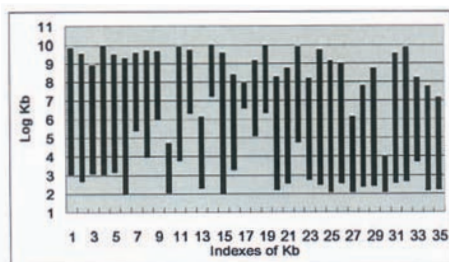


図 5 解が見つかった範囲

得られた解を用いてノックアウトシミュレーションを行った結果、ほとんどの解は 6 つの挙動を再現できる vii 型の KSV を示した(表 3)。しかし、その中で、7 つの挙動を再現できる viii 型の KSV が得られたことは生物学的に意義がある。8 種類もの KSV が得られたことから、最適化によって生物学的に多様な解が得られたと言える。

表 3 KSV の種類と解の個数

KSV	解の個数
i 10000101010	1
ii 10000101110	3
iii 10000101111	1
iv 10000111110	4
v 10010101010	2
vi 10010101110	1
vii 10010111010	123
viii 10010111110	3

6. 終わりに

モジュール分割統合と統計解析技術を組み合わせた最適化手法 DIS を開発し、それを用いることによって、大腸菌窒素同化システムのパラメータをグローバルに最適化することができた。さらに生物学的な解の多様性という新しい概念を提唱した。

現在、生命分子ネットワークの制御構造の理解に基づいて、合理的に有用細胞を設計することが求められている。有用細胞の合理的設計を実現するために、解糖系、TCA 回路、ペントースリン酸経路などを含む中央代謝システムの解析が必要である。このような大規模な生命分子ネットワークの解析においても、シミュレーションは重要な研究手段である。本研究で示したように、大規模な生命分子ネットワークのパラメータ最適化では、DIS のようなシステムを分解するアプローチが有効である。従って、DIS の有用性は今後増していくものと考えられる。

参考文献

- 1) G. Koh, *et al.*: A compositional approach to parameter estimation in path way modeling: a case study of the Akt and MAPK pathways and their crosstalk, *Bioinfo.*, Vol.22, pp.271-280 (2006).
- 2) H. Kurata, *et al.*: CADLIVE dynamic simulator: Direct link of biochemical networks to dynamic models, *Genome Res.*, Vol.15, pp.590-600 (2005).
- 3) H. Kurata, *et al.*: Two-phase partition method for simulating a biological system at an extremely high speed, *Genome Info.*, Vol.11, pp.185-195, (2000).
- 4) A. Ninfa, *et al.*: Integration of antagonistic signals in the regulation of the nitrogen assimilation in Escherichia coli, *Curr. Top Cell Regul.*, Vol.36, pp.31-75 (2000).
- 5) 小野, 佐藤, 小林: 単峰性正規分布交叉 UNDX を用いた実数値 GA による関数最適化, 人工知能学会誌, Vol.14, pp.1146-1155 (1999).
- 6) 佐藤, 小野, 小林: 遺伝的アルゴリズムにおける世代交代モデルの提案と評価, 人工知能学会誌, Vol.12, No.5, pp.734-743 (1996).