

RCGAToolbox: 動力学モデルのパラメーター推定のための実数値遺伝的アルゴリズムソフトウェア

前田 和勲^{1,2,a)} Fred C. Boogerd³ 倉田 博之¹

概要: 代謝経路やシグナル伝達経路のような生化学ネットワークの挙動を理解するには動力学モデルが不可欠である。この動力学モデルを構築する上で、動力学パラメーターの推定がボトルネックとなっている。この問題を解決するために、我々はRCGAToolboxを開発した。これは数値計算環境MATLAB上で動作する動力学パラメーター推定ソフトウェアであり、アルゴリズムとしては実数値遺伝的アルゴリズムを提供する。さらに、グラフィカルユーザーインターフェースを持っており、並列計算などの高度なオプションも簡単に使用できる。この本稿では、RCGAToolboxの概要と遺伝子回路設計問題への応用について紹介する。RCGAToolboxはGitHubで入手できる。

1. はじめに

動力学モデルは複雑な生化学ネットワークの挙動を理解する上で不可欠である [1]。動力学モデルは微分方程式で記述されることが多く、その中には多くの動力学パラメーターが含まれる。動力学パラメーターとは、例えば、酵素反応速度式における V_{max} や K_m などのことである。これらの動力学パラメーターの多くは測定されていない。そこで、動力学モデルの挙動が実験データと一致するように、動力学パラメーターを推定することが必要になる。この動力学パラメーター推定問題は逆問題の1つであり、現在においても動力学モデルを構築する上でボトルネックになっている [2]。

動力学モデルを構築するには数値計算環境のMATLABが比較的良好に使用されている。MATLAB向けの動力学パラメーター推定ソフトウェアとしては、MEIGO [3]、AMIGO2 [4]、D2D [5]、PESTO [6]、CADLIVE Toolbox [7] などがある。しかし、これらのソフトウェアは使いやすさの点で問題がある。例えば、上記のソフトウェアのうちCADLIVE Toolbox以外はグラフィカルユーザーインターフェース (GUI) を持っていない。インストーラーやドキュメントが整備されていないという問題もある。また、並列計算のオプションを備えてない、あるいは、並列計算を行なう際にスクリプトの大幅な書き換えが必要という問

題もある。さらに、SBML [8] やPETab [9] といった、システム生物学のデータ標準への対応も重要である。また、動力学パラメーター推定の過程では、数万回のシミュレーションを繰り返す必要があるため、高速なシミュレーションエンジンを備えていることも重要である。

上記の点を考慮して、我々はRCGAToolboxを開発した [10]。RCGAToolboxは、動力学パラメーター推定問題を解決するためのアルゴリズムとしてUNDX/MGG、REX^{star}/JGGという2つの実数値遺伝的アルゴリズム (RCGA) を提供する。我々は、使いやすさを重視してRCGAToolboxを設計しており、GUIや詳細なドキュメントが付属している。本稿では、まず、RCGAToolboxについて紹介する。その後、その性能評価を行なう。最後に、RCGAToolboxの応用範囲の広さをデモンストレーションするために、遺伝子回路設計問題への応用について紹介する。

2. 方法

2.1 動力学パラメーター推定問題

動力学パラメーター推定問題は、次の制約付き最適化問題として扱うことができる。

$$\text{minimize } f(\mathbf{x}), \quad (1)$$

$$\text{subject to } \mathbf{g}(\mathbf{x}) \leq \mathbf{0}, \quad (2)$$

$$\mathbf{x}^{lb} \leq \mathbf{x} \leq \mathbf{x}^{ub} \quad (3)$$

ここで、 $\mathbf{x} = (x_1, x_2, \dots)$ は設計変数ベクトル、 f は目的関数、 $\mathbf{g} = (g_1, g_2, \dots)$ は制約関数ベクトルである。 $g_i > 0$ のとき、 i 番目の制約が満たされていないことを示す。 \mathbf{x}^{lb} と

¹ 九州工業大学大学院 情報工学研究院 生命化学情報工学研究系

² 科学技術振興機構 さきがけ

³ Department of Molecular Cell Biology, Faculty of Science, VU University Amsterdam, the Netherlands

a) kmaeda@bio.kyutech.ac.jp

\mathbf{x}^{ub} は、それぞれ設計変数の下限と上限を決めるベクトルであり、これらによって探索空間が決まる。式2を満たす空間は可能領域 (feasible region) と呼ばれる。制約なし問題では式2は省略される。制約付き最適化問題では制約が満たされているか調べるために、ペナルティ関数 ϕ が用いられる。

$$\phi(\mathbf{x}) = \sum_{i=1}^n [\max\{0, g_i(\mathbf{x})\}]^2 \quad (4)$$

ここで、 \mathbf{x} は設計変数ベクトルであり、 g_i は i 番目の制約関数、 n は制約関数の数である。 i 番目の制約が満たされているとき $g_i \leq 0$ となる。従って、 $\phi = 0$ のとき全ての制約が満たされていることを示す。制約付き最適化問題では、この ϕ と目的関数 f (式1参照) を同時に小さくする必要がある。

動力学パラメーター推定問題においては、設計変数ベクトル (\mathbf{x}) は動力学パラメーターベクトルに対応する。目的関数 (f) は、動力学モデルの挙動と実験データのずれを表す。また、制約関数 (g_i) は、例えば「代謝物濃度は 10 mM より小さい」といった動力学モデルが満たすべき生物学的な制約である。

2.2 実数値遺伝的アルゴリズム

遺伝的アルゴリズム (GA) は、生物の進化を模倣した最適化アルゴリズムの1つである [11]。実数値遺伝的アルゴリズム (RCGA) では、1つの設計変数ベクトル (実数) が「個体」と見なされる。RCGA では、まず、複数の個体をランダムに生成することで、初期集団を作成する。そして、集団に「選択」や「交叉」といった操作を繰り返すことで、目的関数を最小化する個体を得る。動力学パラメーター推定においては、個体は1セットの動力学パラメーターの値に対応する。RCGA を使うことで、動力学モデルが実験データに対してフィットするような動力学パラメーターセットを見つけ出すことができる。

2.3 遺伝子回路設計問題

「転写因子のコード領域」と「その転写因子で活性が制御されるプロモーター配列」を適切な順序で並べた DNA を作成し、それを細胞に入れることで、細胞に情報処理させることができる。このような情報処理可能な生化学ネットワークは、電子回路とのアナロジーから「遺伝子回路」と呼ばれる。遺伝子回路の設計を生物実験による試行錯誤で行うことは難しいので、コンピューターで自動的に遺伝子回路を設計することが求められている [12]。この遺伝子回路設計問題は、動力学パラメーター推定問題と似ており、動力学パラメーター推定に使われるアルゴリズムは、大幅な変更なく遺伝子回路設計問題に適用できる。動力学パラメーター推定問題と遺伝子回路設計問題で異なる点は次の

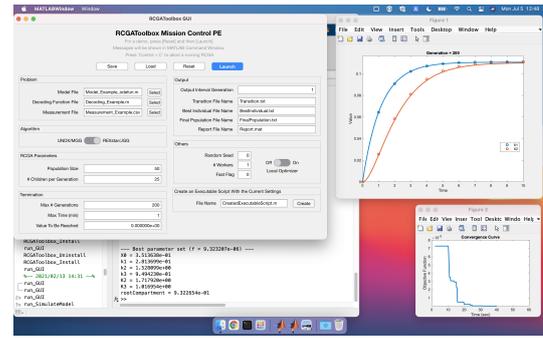


図1 RCGAToolbox のスクリーンショット。

2点である。

- 動力学パラメーター推定問題では、シミュレーションされた生化学ネットワークの挙動と実験データとのずれを最小化する。これに対して、遺伝子回路設計問題では、シミュレーションされた生化学ネットワーク (遺伝子回路) の挙動と「理想の挙動」とのずれを最小化する。
- 動力学パラメーター推定問題では、生化学ネットワークの構造は固定されており、動力学パラメーターのみが推定される。これに対して、遺伝子回路設計問題では、動力学パラメーターだけでなく、生化学ネットワーク (遺伝子回路) の構造も推定対象である。遺伝子回路の構造は、転写因子間の制御関係の有無で決まる。制御関係の有無を推定対象のパラメーター加えることで、遺伝子回路設計問題は動力学パラメーター推定問題と同様に扱える。

2.4 RCGAToolbox

RCGAToolbox は、MATLAB 上で動作する動力学パラメーター推定ソフトウェアである (図1) [10]。Windows, Mac, Linux の MATLAB で動作することを確認している。RCGAToolbox はドキュメントとともに GitHub で公開されており、誰でも利用できる。YouTube ではチュートリアル動画を公開している。

RCGAToolbox では、UNDX/MGG と REX^{star}/JGG という2つの RCGA が利用できる。一言で述べると、UNDX/MGG は保守的、REX^{star}/JGG はより積極的な RCGA である。UNDX/MGG は、探索の初期において個体が局所解にトラップされる問題 (初期収束) を避けるために、あえて集団の更新速度を落としている [13]。また、交叉では、親個体の性質を子個体が受け継ぐことに重点を置いている [14]。これに対して、REX^{star}/JGG では、集団の更新速度が速く、子個体がより有望な進化方向に向かうような交叉法が用いられる [15]。これらの RCGA は多くの研究で用いられた実績がある (e.g. [16], [17], [18], [19])。

RCGAToolbox は使いやすい GUI を持っており、ユーザーは、モデルファイル、実験データファイル、探索範囲

設定ファイルの3つを指定するだけで動力学パラメーター推定を実行できる。GUIでは並列計算のような高度なオプションも選択できる。また、ユニークな機能としてスクリプト自動生成機能が挙げられる。GUI上でオプションなどを設定し、「Create」ボタンを押すだけで動力学パラメーター推定を実行するスクリプトが作成できる。このため、高度な動力学パラメーター推定タスクを実行する際にもユーザーはゼロからスクリプトを書き始める必要がない。

RCGAToolboxではサードパーティによる高速なシミュレーションエンジン [20], [21] が利用可能である。MATLABではパッケージマネージャーが一般的ではないので、インストール時にトラブルが起きやすいが、RCGAToolboxのインストーラーは必要なサードパーティツールを自動的にインストールする。

RCGAToolboxはSBMLやPEtabといった標準データ形式をサポートしている。SBML [8]は、動力学モデルを記述する標準データ形式であり、BioModelsなど主要なデータベースで採用されている。RCGAToolboxのユニークな機能として、MATLAB ODEファイルとSBMLの相互変換機能がある。SBMLは業界標準ではあるが、XMLベースなので、人間が直接編集するには向いていない。このため、MATLABユーザーは動力学モデルをODEファイルとして扱うことが多い。RCGAToolboxは、MATLAB ODEファイルとSBMLの2つの形式を相互に変換できる。PEtab [9]は最近提案された、動力学パラメーター推定問題を定式化するための標準データ形式である。RCGAToolboxはPEtabをサポートしているので、様々な動力学パラメーター推定ソフトウェアと相互運用できる。つまり、動力学パラメーター推定問題では、対象とするモデルごとにベストなアルゴリズムが異なるが、PEtabで問題を記述しておけば、RCGAToolboxやその他のソフトウェアを簡単に適用できる。

3. 結果と考察

3.1 ベンチマーク実験

我々は、使いやすさに重点を置いてRCGAToolboxを開発したが、RCGAToolboxは動力学パラメーター推定ソフトウェアとしては性能は良いのだろうか？これを調べるために、RCGAToolboxが提供する2つのRCGA (UNDX/MGGとREX^{star}/JGG)のベンチマーク実験を行った。

ベンチマーク問題としては、6つの動力学パラメーター推定問題を採用した。表1は、それぞれの問題の特徴をまとめたものである。Three-stepとHIVは、動力学パラメーター推定アルゴリズムの性能評価において長い間使われてきた問題であるが、ここでは制約付き最適化問題に拡張したものを使っている [22]。B2からB6は、動力学パラメーター推定問題のベンチマーク問題として提案されたもので

ある [23]。

UNDX/MGGとREX^{star}/JGGの性能を評価する際の対照アルゴリズムとしては、MATLABのGlobal Optimization Toolboxに付属するGA (MATLAB GA), MEIGOのEnhanced Scatter Search (eSS) [24]を採用した。eSSは動力学パラメーター推定問題に置いて最も強力なアルゴリズムであることが知られている [25]。

表1に示したValues to be reachedまで目的関数が下がれば、それぞれの動力学モデルは実験データに対して十分にフィットする。そこで、ベンチマーク実験では、24時間以内にValues to be reachedまで目的関数を減少させることができれば動力学パラメーター推定は「成功」、そうでなければ「失敗」とみなした。

図2に目的関数の値の推移を示す。制約を持つThree-stepとHIVにおいては、MATLAB GAは初期世代から探索が進まなかったため結果を示していない。MATLAB GAでは、Three-step, HIV, B2, B4において全ての試行で失敗した。RCGAToolboxが提供するUNDX/MGGはすべての問題で失敗した。これに対して、eSSは、HIVを除くすべての問題に成功した。REX^{star}/JGGは、B4を除くすべての問題で成功した。

まとめると、RCGAToolboxが提供するREX^{star}/JGGの性能は高く、MATLABで提供されているGAを超える。また、強力な動力学パラメーター推定アルゴリズムとして知られているeSSと同程度の性能を持っていると言える。

3.2 遺伝子回路設計問題への応用

RCGAToolboxの応用範囲の広さをデモンストレーションするために、RCGAToolboxを遺伝子回路設計問題に応用した。今回は、大腸菌で機能する遺伝子回路を設計することを想定した。また、遺伝子回路を構成する転写因子としてはVoigtらが提案した転写因子ライブラリを想定した [12], [26]。遺伝子回路がとりうる全ての構造を探索することは現実的ではないので、「1つのオペロンに1つのコード配列を持つ」、「1つのオペロンに最大で2つのプロモーター配列を持つ」という制約のもとで遺伝子回路設計を行った。アルゴリズムとしては、RCGAToolboxのREX^{star}/JGGを用いた。

最初に静的な遺伝子回路の設計を試みた結果、OR, NOR, AND, NANDといった論理ゲートが設計できた。次に、動的な挙動を示す遺伝子回路の設計を行った。その結果、First-In First-Out (FIFO,あるいはキュー), Last-In First-Out (LIFO,あるいはスタック)の挙動を示す遺伝子回路が設計できた。このように、RCGAToolboxは遺伝子回路設計問題にも適用可能である。

4. おわりに

本稿では、我々が開発した動力学パラメーター推定ソフ

表 1 ベンチマーク問題.

Problem ID	Three-step	HIV	B2	B4	B5	B6
Organism	Generic	Human	<i>E. coli</i>	Chinese hamster	Generic	<i>D. melanogaster</i>
Description level	Metabolism Transcription Translation	PPI	Metabolism	Metabolism	Signal Transduction	GRN (spatial)
Unknown parameters	36	20	116	117	86	37
Dynamic states	8	9	18	34	26	108 - 212
Constraint functions	24	12	0	0	0	0
Values to be reached	1.00×10^{-3}	0.04	468	91.4	6.15×10^3	2.17×10^5

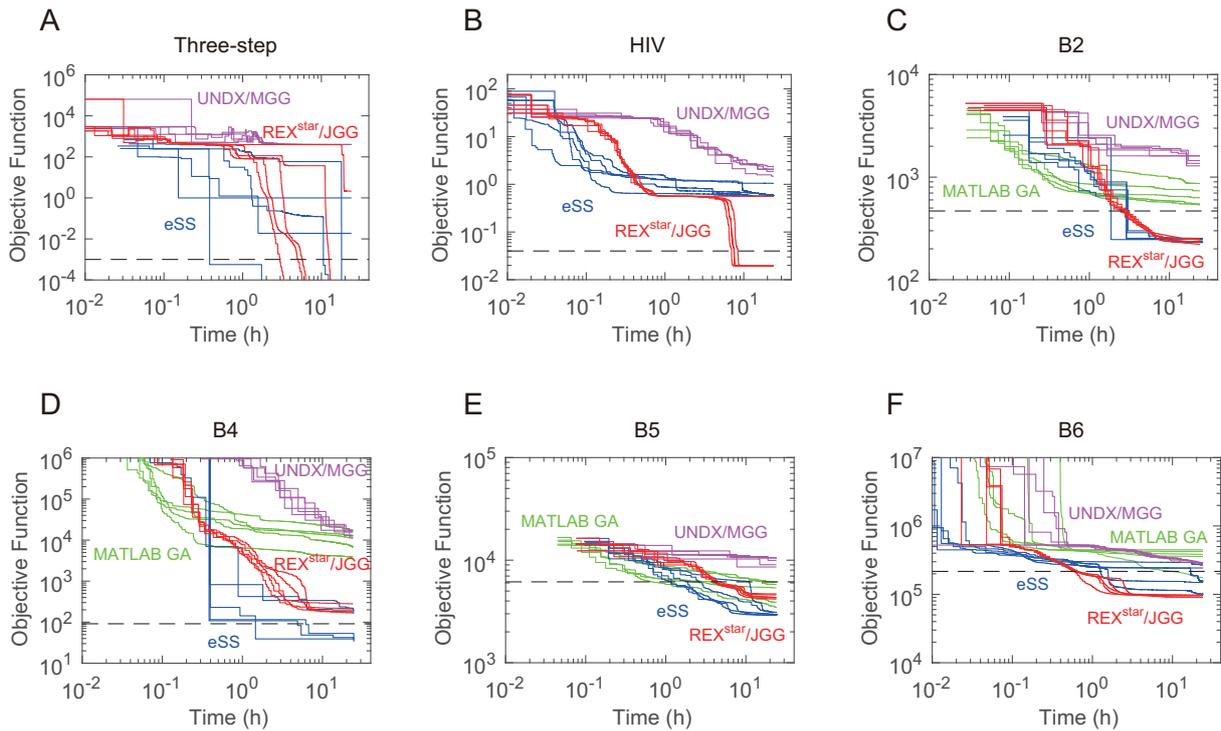


図 2 ベンチマーク実験の結果. UNDX/MGG と REX^{star}/JGG が RCGAToolbox で利用できるアルゴリズム. MATLAB GA と eSS は比較対象のアルゴリズム. 各アルゴリズムについて試行回数は 5 回.

トウェア RCGAToolbox を紹介した. 生命科学分野のソフトウェアの多くは, 使いやすさの点で問題があることが指摘されている [27], [28], [29]. この問題は, 動力学パラメーター推定ソフトウェアにおいても存在する. これに対して, RCGAToolbox は使いやすい GUI や豊富なドキュメントを備えている. これによって, 初学者でも簡単に高度な動力学パラメーター推定が行えるようになった.

また, 本稿では, ベンチマーク実験を通して, RCGAToolbox が提供する REX^{star}/JGG が優れた性能を持つことを示した. さらに, 遺伝子回路設計問題への応用を示して, RCGAToolbox の応用範囲の広さをデモンストレーションした.

謝辞 本研究は, 科研費 若手研究 (18K18153), 学術変革領域 (B) (20H05743), 基盤研究 (B) (19H04208), JST さきがけ (JPMJPR20K8) の支援を受けたものである.

参考文献

- [1] Kitano, H.: Systems biology: A brief overview, *Science*, Vol.295, No.5560, pp.1662–1664 (2002).
- [2] Banga, J.R.: Optimization in computational systems biology, *BMC Syst. Biol.*, Vol.2, No.1, p.47 (2008).
- [3] Egea, J.A. et al.: MEIGO: An open-source software suite based on metaheuristics for global optimization in systems biology and bioinformatics, *BMC Bioinformatics*, Vol.15, p.136 (2014).
- [4] Balsa-Canto, E., Henriques, D., Gabor, A. and Banga, J.R.: AMIGO2, a toolbox for dynamic modeling, optimization and control in systems biology, *Bioinformatics*, Vol.32, No.21, pp.3357–3359 (2016).
- [5] Raue, A. et al.: Data2Dynamics: A modeling environment tailored to parameter estimation in dynamical systems, *Bioinformatics*, Vol.31, No.21, pp.3558–3560 (2015).
- [6] Stapor, P. et al.: PESTO: Parameter ESTimation TOolbox, *Bioinformatics*, Vol.34, No.4, pp.705–707 (2018).
- [7] Inoue, K., Maeda, K., Miyabe, T., Matsuoka, Y. and Ku-

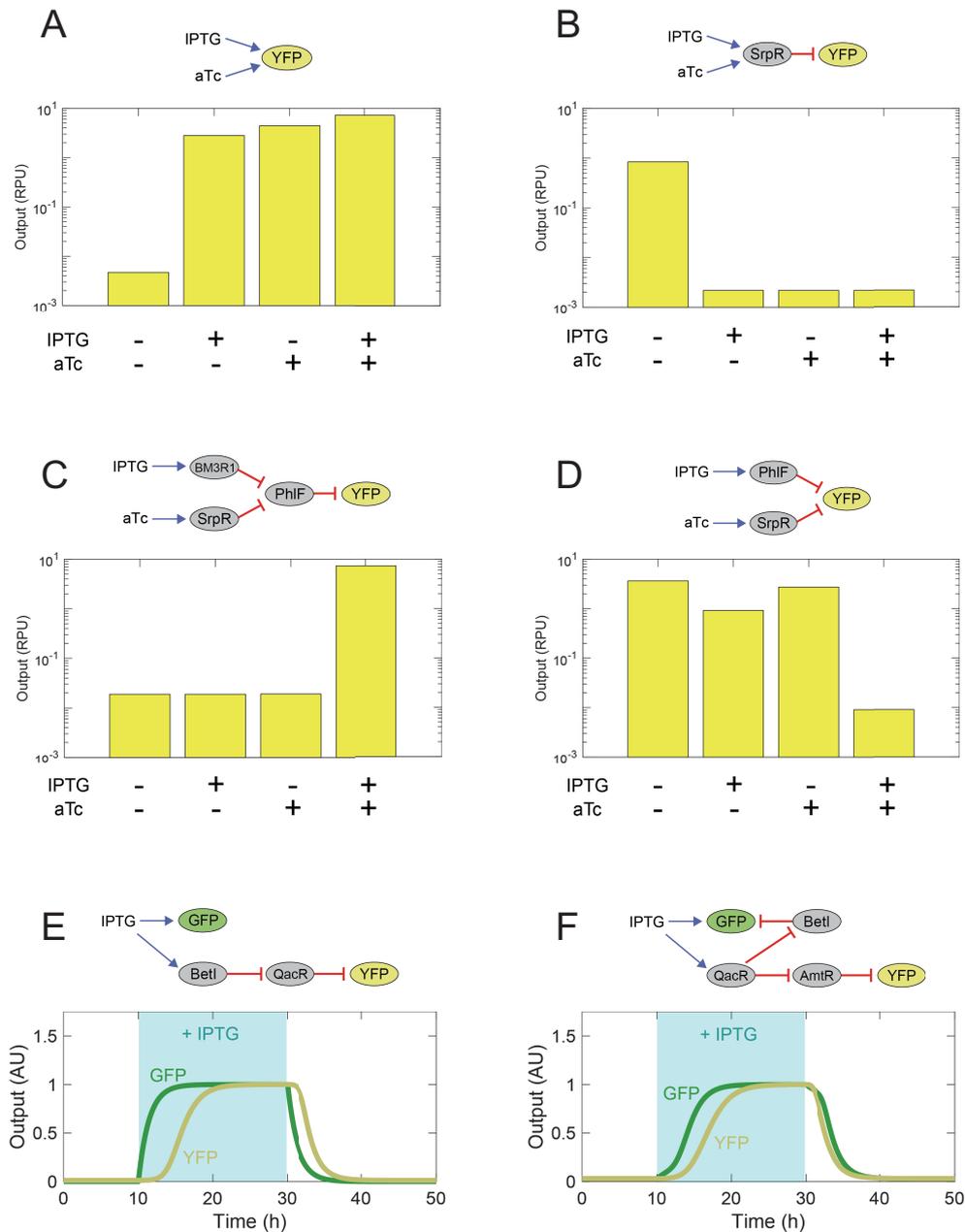


図 3 RCGAToolbox を使って設計した遺伝子回路。A: OR ゲート, B: NOR ゲート, C: AND ゲート, D: NAND ゲート, E: First-In First-Out, F: Last-In First-Out.

rata, H.: CADLIVE toolbox for MATLAB: Automatic dynamic modeling of biochemical networks with comprehensive system analysis, *Bioprocess Biosyst. Eng.*, Vol.37, No.9, pp.1925–1927 (2014).

[8] Keating, S.M. et al.: SBML Level 3: An extensible format for the exchange and reuse of biological models, *Mol. Syst. Biol.*, Vol.16, No.8, p.e9110 (2020).

[9] Schmiester, L. et al.: PETab-Interoperable specification of parameter estimation problems in systems biology, *PLoS Comput. Biol.*, Vol.17, No.1, p.e1008646 (2021).

[10] Maeda, K., Boogerd, F.C. and Kurata, H. RCGAToolbox: A Real-coded Genetic Algorithm Software for Parameter Estimation of Kinetic Models. *IPSJ Transactions on Bioinformatics 2021*;14(0):30-35.

[11] Sun, J.Y., Garibaldi, J.M. and Hodgman, C. Parameter Estimation Using Metaheuristics in Systems Biology: A Comprehensive Review. *IEEE/ACM Trans Comput Biol*

Bioinform 2012;9(1):185-202.

[12] Nielsen, A.A., et al. Genetic circuit design automation. *Science* 2016;352(6281):aac7341.

[13] Satoh, H., Yamamura, M. and Kobayashi, S.: Minimal generation gap model for GAs considering both exploration and exploitation, *Proc. Int. Conf. Fuzzy Logic, Neural Networks and Soft Computing*, pp.494–497 (1997).

[14] Ono, I. and Kobayashi, S.: A real-coded genetic algorithm for function optimization using unimodal normal distribution crossover, *Proc. 7th Int. Conf. Genetic Algorithms*, pp.246–253 (1997).

[15] Kobayashi, S.: The Frontiers of Real-coded Genetic Algorithms, *J. Jpn. Soc. Artif. Intell.*, Vol.24, No.1, pp.147–162 (2009).

[16] Maeda, K., Minamida, H., Yoshida, K. and Kurata, H.: Flux module decomposition for parameter estima-

- tion in a multiple-feedback loop model of biochemical networks, *Bioprocess Biosyst. Eng.*, Vol.36, No.3, pp.333–344 (2013).
- [17] Maeda, K., Westerhoff, H.V., Kurata, H. and Boogerd, F.C.: Ranking network mechanisms by how they fit diverse experiments and deciding on *E. coli*'s ammonium transport and assimilation network, *NPJ Syst. Biol. Appl.*, Vol.5, No.1, p.14 (2019).
- [18] Tohsato, Y., Ikuta, K., Shionoya, A., Mazaki, Y. and Ito, M.: Parameter optimization and sensitivity analysis for large kinetic models using a real-coded genetic algorithm, *Gene*, Vol.518, No.1, pp.84–90 (2013).
- [19] Kimura, S., Sato, M. and Okada-Hatakeyama, M.: An Effective Method for the Inference of Reduced S-system Models of Genetic Networks, *Inform. Media. Tech.*, Vol.10, No.1, pp.166–174 (2015).
- [20] Serban, R.: SUNDIALSTB, a MATLAB Interface to SUNDIALS, Lawrence Livermore National Lab. (LLNL), Livermore, CA (United States) (2005).
- [21] Schmidt, H.: SBaddon: High performance simulation for the Systems Biology Toolbox for MATLAB, *Bioinformatics*, Vol.23, No.5, pp.646–647 (2007).
- [22] Maeda, K., Boogerd, F.C. and Kurata, H.: libRCGA: A C library for real-coded genetic algorithms for rapid parameter estimation of kinetic models, *IPSI Trans. Bioinformatics*, Vol.11, pp.31–40 (2018).
- [23] Villaverde, A.F., et al. BioPreDyn-bench: a suite of benchmark problems for dynamic modelling in systems biology. *BMC Syst Biol* 2015;9:8.
- [24] Egea, J.A., Balsa-Canto, E., Gracia, M.-S.G. and Banga, J.R.: Dynamic Optimization of Nonlinear Processes with an Enhanced Scatter Search Method, *Industrial & Engineering Chemistry Research*, Vol.48, No.9, pp.4388–4401 (2009).
- [25] Villaverde, A.F., et al.: Benchmarking optimization methods for parameter estimation in large kinetic models, *Bioinformatics*, Vol.35, No.5, pp.830–838 (2019).
- [26] Stanton, B.C., et al.: Genomic mining of prokaryotic repressors for orthogonal logic gates *Nature chemical biology* 2014;10(2):99–105.
- [27] Mangul, S., Martin, L.S., Eskin, E. and Blekhman, R.: Improving the usability and archival stability of bioinformatics software, *Genome Biol.*, Vol.20, No.1, p.47 (2019).
- [28] Pavelin, K. et al.: Bioinformatics meets user-centred design: A perspective, *PLoS Comput. Biol.*, Vol.8, No.7, p.e1002554 (2012).
- [29] List, M., Ebert, P. and Albrecht, F.: Ten Simple Rules for Developing Usable Software in Computational Biology, *PLoS Comput. Biol.*, Vol.13, No.1, p.e1005265 (2017).