

研究終了報告書

「ゲノム配列から自動で全細胞モデリングする技術の開発」

研究期間：2019年10月～2023年3月

研究者：海津 一成

1. 研究のねらい

ゲノムや細胞を合成する技術の急速な進歩に対して、ゲノムを設計する技術は未だ十分でない。細胞シミュレーションは、細胞のふるまいを予測し設計するための有効な技術の一つである。しかし、従来の細胞シミュレーションは細胞の一部のみを対象とし、ゲノム配列の一塩基レベルの変化などに柔軟に対応できなかった。そこで本研究は、大腸菌を対象として、ゲノム配列を入力として自動で細胞全体のモデルを構築しシミュレーションできる「全細胞モデリング」技術の開発を目的とする。全細胞モデリングは、遺伝子発現、代謝、ゲノム複製などを含め、一つの細胞を全ゲノム規模でまるごと計算機上にモデル化する。また、全ゲノム規模の巨大なモデルの構築と維持は人力では困難であり、我々はゲノム配列を解析して既存のデータベース上の情報と統合することでこれを自動化する。ゲノム配列を直接あつかうことで、モデルの数理的な詳細を理解しなくても、ゲノム配列から直接そのゲノムを有した細胞のふるまいを計算機上で再現することができる。またエージェントベース・シミュレーションによって、ゲノムは配列を有する DNA 分子として表現され一塩基レベルで計算される。ゲノムを遺伝子の集まりとしてあつかう従来の常微分方程式のモデルに対して、エージェントベース・モデルではゲノム上での遺伝子の位置や向き、オペロンなど、これまであつかえなかったゲノム配列構造が表現型に及ぼす影響を再現できる。これを利用して、様々なゲノム配列を仮想的に改変・設計し、計算機上でその表現型を予測することで、ゲノム配列構造の設計原理を明らかにする。「遺伝型から表現型の予測」は、生物学における対象の理解度を測る基準となるだけでなく、一塩基レベルでゲノムを自在に設計するためには必須な技術である。またその実現を目指す本技術は、ゲノム配列を解析するバイオインフォマティクス、配列から分子の構造や機能を予測する構造生物学、分子のはたらきから細胞のふるまいを再現するシステム生物学を統合したプラットフォームであり、さらにゲノム配列情報を介して生化学、遺伝学、分子生物学、生物物理学、合成生物学など異なる分野の知見を結び付ける。

2. 研究成果

(1) 概要

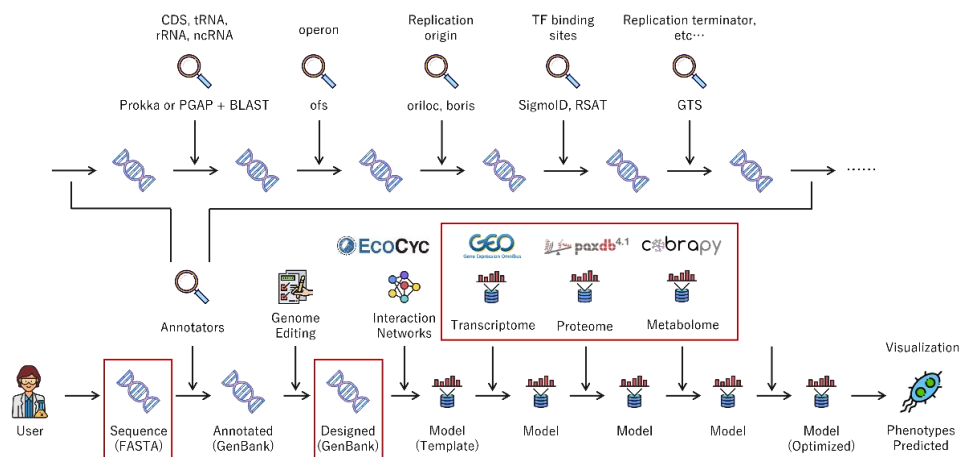
本研究では、原核生物、主に大腸菌のゲノム配列を直接の入力として自動でモデルを構築し、シミュレーションによって細胞のふるまいを計算、可視化するフレームワークを開発した。第一に、複数のバイオインフォマティクスツールとデータベースを統合して、ゲノムの配列を解析してアノテーションを付与し、データベース上の知見から反応ネットワークを構築、マルチオミクスデータから未知パラメータを決定してモデルを構築するワークフローを開発した。これらの過程はゲノムの塩基配列から自動で行われ、その結果と記録の記録はクラウド上に保存され、仮想化技術によって再現性が保証される。このモデルは、一塩基レベルのエージェントベースと全代謝系の常微分方程式を連成計算によってシミュレーションされる。シミュレーション

ン結果は、一細胞レベルでの増殖速度、大きさ、重量やたんぱく質・RNAなどの存在比、また全ゲノム規模のマルチオミクス、ゲノムDNAの複製プロファイルなどと定量的に比較できる。本ワークフローは、大腸菌の野生株をはじめとして遺伝子欠損株や近縁株、最小ゲノムなどのゲノム配列を入力として検証された。第二に、開発したワークフローを用いて仮想的に設計されたゲノム配列の計算を行った。有する遺伝子を維持したままゲノム構造を大きく変化させるゲノムシャッフリング実験を計算機上でを行い、野生型と大きく異なる構造を持ちながら同等の増殖速度を示す配列を見出すなど新たな法則を解析中である。第三に、可視化技術を中心にゲノム設計を支援する複数の新規ソフトウェアを開発した。GenBank形式のゲノムファイルを編集するコマンドラインツールは、本ワークフローの入力となるゲノム配列を直観的かつ柔軟に操作・改変することを可能にする。また、データベース上のレイアウトをもとに反応経路を可視化する新規ライブラリにより、全細胞シミュレーションをはじめとしたデータのデジタルダッシュボードを容易に開発できるようにした。さらにこれを発展させ、VR技術を用いて代謝ネットワークの可視化、モデリング、シミュレーションを行い、複数の研究者が共同で新しい知見を探索しモデルを改善するための仮想空間を開発した。

(2) 詳細

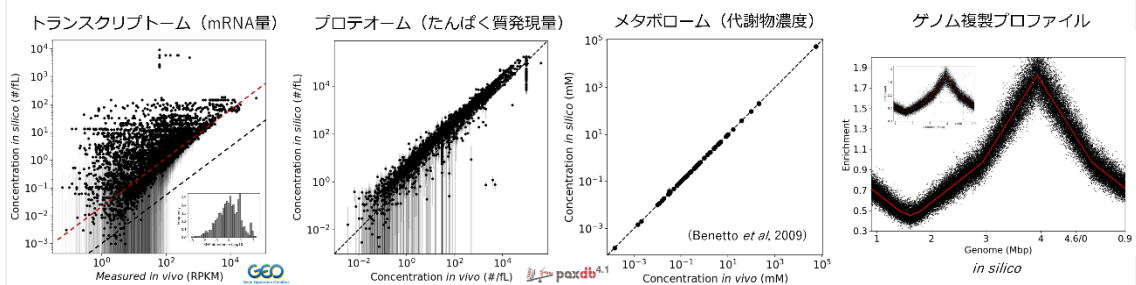
研究項目「ゲノム配列から自動でモデルを構築する」

原核生物を対象に、ゲノム配列を入力として、それを解析し、全細胞モデルを自動で構築するワークフローを開発した。これによりいわゆる野生型のモデルだけでなく、様々な遺伝型のモデルを自動で作成し、再現することができるようになった。



このワークフローはまず、複数のバイオインフォマティクスツールを用いてゲノム配列を解析しアノテーション情報を付与する。遺伝子(たんぱく質、rRNA、tRNA)は Prokka もしくは NCBI の PGAP によって同定する。次に遺伝子の情報をもとに、転写ユニットの予測やプロモーター領域の解析を行う。さらに、ゲノムの複製起点や複製フォーク阻害点も予測する。これらのツールを組み合わせることで、ゲノムの FASTA ファイルからアノテーション情報を付与した GenBank 形式のファイルを生成する。

このアノテーション情報に加えて既存のデータベース(知識ベース)上の反応や制御などの分子機能の知見を統合する。とりわけ、これらのデータベースは異なる識別子(ID)を用いているため、相互参照情報や BLAST を用いた配列の相同性検索によってこれを統一する。最後に、構築されたモデルをもとにシミュレーションを実行し、データベース上の各種オミクスデータ(トランスクリプトーム、プロテオーム)により未知パラメータを推定する。この自動モデリングは従来研究者の手によって人力でなされてきたモデリングを繰り返し再現可能かつ第三者によって解釈可能なものにした。このワークフローを用いて、遺伝子欠損株や近縁株、最小ゲノムなどから実際にモデルを構築し、シミュレーションを行えることを確認した。



研究項目「全代謝モデルを自動で構築する」

遺伝子発現やゲノム複製に対して、代謝物とその反応は細胞内の分子数が多く、反応速度も速い。そのため我々の全細胞モデルでは、代謝系は酵素反応速度論に基づく常微分方程式 (ODE) によって計算し、その他のエージェントベースと連成する。全代謝の ODE モデルの自動構築手法には既にいくつかの報告がある。既存の手法の問題点を改善し、前述のワークフローと組み合わせた。まず、ゲノム上の酵素遺伝子とデータベース上の反応ネットワークの情報をもとに、代謝流束解析 (FBA) を実施し、各代謝反応の流束を求める。各反応の速度式には一般ミカエリスメンテン式を用いる。次にメタボローム解析による代謝物濃度や流束と FBA で推定された流束から、反応式の未知パラメータを推定する。この未知パラメータ決定では、ヤコビ行列を用いた安定性解析によって定常状態の安定性を保証する。この方法により、パラメータ最適化なしに、メタボロームを再現する定常モデルを自動で構築できる。酵素による反応制御は BiGG データベース上の制御遺伝子の論理式と EcoCyc 上の複合体の情報から自動的に決定する。加えて、複合体構成分子によっては欠損してもその機能を失わない場合があるが、実験による遺伝子の必須性に基づき自動的に複合体形成反応を修正できるようにした。

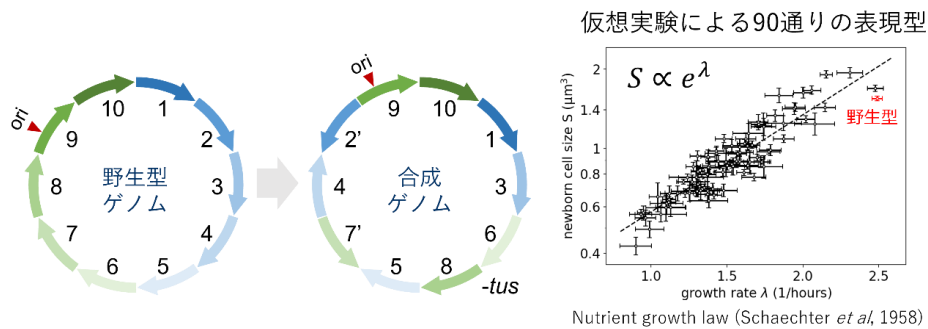
研究項目「仮想実験によりゲノム配列構造原理を明らかにする」

本研究で開発した全細胞自動モデリング技術によって、仮想的なゲノム配列を有する細胞のふるまいを再現し、その性質について調べた。

まず、一分子計測により測定された大腸菌の複製機構のキネティクスをもとにゲノム上の RNA ポリメラーゼや複製フォーク阻害分子との衝突をモデルに新たに実装した。その結果は複製起点・終始点を複数の改変ゲノム株の実験データと定量的に比較・検証し

た。次に、ゲノムファイル編集ツールを開発した。GTS (Genome Transformation Subprograms) は GenBank 形式などのフラットファイルの編集を目的として、アノテーションの選択や追加、アノテーション情報に基づく配列の切り出しや挿入などを、スクリプトを書くことなく自在かつ厳密、そして高速に行うことのできるコマンドラインツールである。これにより本ワークフローの入力となるゲノム配列ファイルを自由かつ直観的に操作・改変できるようになった。

野生型ゲノムを分割してその向きや位置を入れ替えた新たなゲノムを作成する仮想実験 (ゲノムシャッフリング) を実施した。本仮想実験では、ゲノムを構成する遺伝子は野生型と同一であるがゲノムの配列構造が異なる。90 の新たなゲノム配列を作成しその表現型を計算したところ、大腸菌の持つ基本的な法則性を保ちながらも多様な表現型を得た。中には野生型と大きく異なる構造を持ちながら同等の増殖速度を示すものもあり、特に複製起点を起点とした遺伝子の位置の対称性を中心として現在その法則性を解析中である。



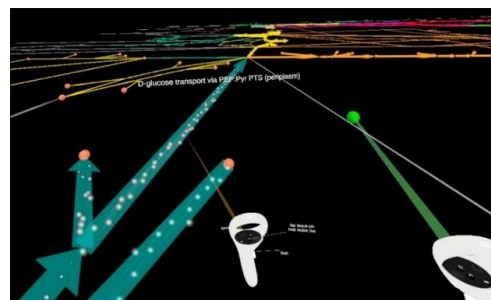
研究項目「可視化技術によって全ゲノム規模のモデリングを支援する」

全細胞モデルは、数千の遺伝子によって構成され、遺伝子発現や代謝、ゲノム複製など大きく異なる細胞機能を含む巨大かつ複雑なシステムである。この巨大なネットワークを把握し、シミュレーションの結果から有意義な差異や新規な発見を見出すことは容易ではない。従って、可視化技術は全細胞モデリングにおいて補助的であるのみならずほとんど不可欠である。我々はまず伝統的な「代謝マップ」の可視化を行うライブラリ Dash-Pathway を開発した。Dash-Pathway

A) ダッシュボード (Dash-Pathway)



B) VR空間 (ECellDive)



は KEGG や WikiPathways といったデータベースから反応経路のレイアウトを取得し、インタラクティブに可視化できる。また代謝物や反応の各種データベース ID と値のテーブルを入力として

経路上に表示できる。

また、さらなるモデリング支援環境として VR デバイスを用いた可視化、モデリング、シミュレーション空間を開発している。近年、VR や AR を中心にクロスリアリティ(XR)技術の進展が著しい。そこで、VR ヘッドセット Meta Quest2 を用いて仮想空間中で代謝ネットワークの可視化や反応ネットワークのモデリング、代謝流束解析やシミュレーションの実施などを行うソフトウェアを開発した。本ソフトウェアでは複数の研究者が同時に一つの VR 空間に接続し、共同でデータの閲覧やモデルの変更を行うことができる。過去に行われたモデルの変更をサーバーを介して共有し、共同で編集することが可能である。科学領域における XR 技術を利用した可視化はまだ模索段階ではあるが、全細胞モデリングでの活用をもとに、今後様々な科学分野への応用が期待できる。

3. 今後の展開

本研究を通して、ゲノム配列から遺伝子発現、ゲノム複製、代謝を含む全細胞モデリングを自動で行うワークフローを開発した。現在ワークフローの対象となっているのは大腸菌ゲノムであるが、ガンマプロテオバクテリア綱の近縁種のゲノムにも応用可能であることを確認しており、大腸菌以外の原核生物への対応も *Bacillus subtilis* などデータベースが充実したモデル生物を中心に行う計画である。本技術は出芽酵母やヒト培養細胞など高等細胞への応用も可能だが、原核細胞と比較するとモデル化が困難な部分も多いため、実験研究者と議論して適切なモデル表現について検討中である。*in vitro* 再構成系や最小ゲノムなど人工ゲノムへの応用も可能だが、実用的には後述する実験技術の開発を伴う必要がある。本技術を使った仮想実験ではゲノムシャッフリングをはじめゲノム構造に関するデータを得られる段階であり、その解析を行っている。ただし現在は一部の小規模な変更を除いて仮想実験で行う大規模な変更ゲノムをハイスループットに合成・測定する技術に欠けるため、実現可能な実験を検討中である他、既存の株については既報のデータを用いるほか自身でも実験測定を進めている。

本研究では主としてモデル生物である大腸菌の野生型に関する実験データによるモデリングを行った。大規模なゲノム変更や他種への応用といったより実用的な自動モデリングには均質で広範囲の実験データを必要とする。そのため所属研究室と共同で大腸菌の遺伝型と表現型の関係をハイスループットに測定するための自動実験システムの開発を既に進めている。今後 3 から 5 年を目途に、現在計算機上での情報処理に限定された自動化を測定実験の自動化まで統合した広義のモデリングの自動化を実現する。寄せ集めのデータに基づくのではなく、将来的には、プラスミド合成を足掛かりにしたゲノム変更や細胞合成の実験自動化技術と結びつけた「全細胞モデリング」が王道だと考えており、引き続き本領域のゲノム合成、人工細胞合成技術開発と連携してゆきたい。

その他に開発したソフトウェアとして、ゲノムフラットファイル編集ツール GTS や反応経路可視化ライブラリ Dash-Pathway はオープンソースで公開されており誰でも利用可能である。また、システム生物学を中心とした研究支援のための VR プラットフォーム(ESCellDive)もアルファ版をオープンソースで公開しており、継続して開発を続ける。

4. 自己評価

研究目的の達成状況については、当初計画していたゲノム配列解析によるモデル自動生成

ワークフローの開発は達成し、全代謝モデルの実装も当初の計画は完了した。ゲノム設計への応用へ向けて、近縁種ゲノムや最小合成ゲノムのシミュレーションも確認した。ゲノム構造に関わる複製機構は実装済で、複数の実験で検証し仮想実験も実施した。仮想実験の解析とゲノム設計原理の解明は現在実施中で未達である。可視化プラットフォームは、反応経路可視化ライブラリを開発しデジタルダッシュボードを実装した。現在は VR 空間を利用した可視化に留まらないモデリング環境の開発に移行し、概ね完了している。研究実施体制は、海外から新たに全細胞シミュレーションを目指す留学生の受入れを行う他、領域内共同研究も開始し、国内外含め計算システム生物学のコミュニティ形成を進めている。本研究開始から後、全細胞モデリングの研究において進展もみられたが、依然として本研究の目指す自動モデリングは類例がなく、国際会議などの議論やコミュニティ内の情報共有でもその必要性が益々認識されつつある。現状では実験領域との連携は十分とは言えないが、プラスミド設計について共同研究をはじめた他、さきがけの支援を得て新たに自動実験システムの構築に着手しており自らモデリングの基礎となる実験を実施できるよう進めている。全細胞モデリングの実用化には基礎となる実験データの取得が不可欠である。今後は本研究領域で先端的な技術開発が進められているゲノム改変や合成技術を自動実験システムに導入する等、情報科学と実験科学のクローズドループに基づく自動モデリング、細胞設計へと展開する。本研究にて情報科学の面で全細胞モデリング技術は大きく進展・達成しており、実験計測を伴う実用的なモデリングの実現に向けて発展させる。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

該当なし

(2) 特許出願

該当なし

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

- (口頭発表)海津一成、西田孝三、高橋恒一、ゲノム配列とオミクスをつなぐ大腸菌全細胞モデリング、第42回日本分子生物学会年会、2019年12月
- (招待講演)海津一成、細胞モデリングを自動化する、情報計算化学生物学会2020年大会、2020年10月
- (誌上発表)西田孝三、海津一成、高橋恒一、パスウェイ可視化のためのダッシュボードコンポーネント、可視化情報学会誌、vol. 40 no. 156 pp. 8-13、2020年1月
- (ポスター発表)Kazunari Kaizu, Kozo Nishida, Koichi Takahashi, Automated whole-cell modeling from genomic sequence and multi-omics data, Intelligent Systems for Molecular Biology and European Conference on Computational Biology 2021, 2021年7月 (Annual award for an outstanding presentation: shared 1st prize in the poster category on 6th annual SysMod meeting @ ISMB/ECCB - Virtual)
- (口頭発表)Kazunari Kaizu, Kozo Nishida, Elliott Jacopin, Koichi Takahashi, Modeling a

whole bacterial cell from a genomic sequence, 第 44 回日本分子生物学会年会, 2021 年 12 月

- (口頭発表) Kazunari Kaizu, Whole-cell modeling of a bacterial cell from its genomic sequence, RIKEN BDR Symposium 2022 "Emergence in Biological Systems: Challenges to Bridging Hierarchies", 2022 年 3 月
- (口頭発表) 海津一成、西田孝三、高橋恒一、ゲノム配列から大腸菌をまるごとモデリングする、第 22 回日本蛋白質科学会年会、2022 年 6 月
- (口頭発表) 海津一成、Elliott Jacopin, 高橋恒一、大腸菌をゲノム配列からまるごとモデリングする、日本遺伝学会第 94 回大会、2022 年 9 月