

Title	全ゲノムベースによる原核細胞シミュレーション環境G-Cellの開発
Sub Title	G-Cell : Genome based prokaryotic cell simulation platform
Author	高萩, 航
Publisher	慶應義塾大学湘南藤沢学会
Publication year	2015
Jtitle	生命と情報 No.22 (2015. ) ,p.75- 81
JaLC DOI	
Abstract	<p>ポストゲノム時代と呼ばれた近年、生命情報科学分野ではゲノム情報を解読するための手段やツールの開発が進展してきた。しかし、未だゲノムから情報を読み取り機能を予測することのできる全ゲノムベースのシミュレーションモデルは存在しない。ゲノムから読み取ったタンパク質の発現やその挙動という生命の複雑かつ巧妙なシステムの情報を統合しシミュレーションモデルとして提供することは、実験では予測や推定が困難な現象を抽象化して扱うことができるという点で有意義である。また、時間軸や現象と現象の"隙間"を埋める事を可能にするなど次世代の生命科学研究において重要な役割を担うと考えている。そこで本研究では、全ゲノムベース細胞シミュレータG-Cellの開発を行うとともに、G-Cell上に大腸菌K12株のゲノム配列を入力情報としたDNA複製モデルを実装し、複製の開始、DNAの伸長複製、それに関わる酵素複合体の形成に注目して現象の考察を行った。本モデルは大腸菌に関する既存の情報の統合という位置付けも担っており、大腸菌に関する主要なデータベースであるEcoCycに登録された全てのDNA複製関連タンパク、酵素、遺伝子を実装できるよう、1つの酵素・昨日を1つのコンポーネントとして実装し、必要に応じて機能を追加削除できる汎用的なシステムに設計した。この大腸菌K12株DNA複製モデルを用いて、DNA複製関連酵素複合体の形成にかかる時間、元となる単量体の量の変化のシミュレーション、複製関連タンパク質複合体であるプライモソームの挙動の調査を行った。その結果、本モデルを用いてDNA複製関連複合体形成の律速段階となる分子の量の変化を観察することができ、また複製に伴う塩基の複製ミスによる誤複製塩基の蓄積を時間経過に沿って観察することができた。本研究で構築したG-Cellプラットフォームと大腸菌DNA複製シミュレーションモデルは、入力とするゲノム情報を変更するだけでその情報を解釈し挙動の変化に反映することがある程度可能である。本モデルを合成生物学やゲノム設計といった分野で利用し実験結果と比較対照することで、実験では解明が難しい細胞内での複合体量の時系列変化や、DNA複製や転写に伴う1つ1つの塩基へのタンパク質の結合解離の状態について新たな知見をもたらすことができる。</p>
Notes	慶應義塾大学湘南藤沢キャンパス先端生命科学研究会 2015年度学生論文集 卒業論文ダイジェスト
Genre	Technical Report
URL	<a href="https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO92001004-00000022-0075">https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO92001004-00000022-0075</a>

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

# 全ゲノムベースによる原核細胞シミュレーション環境G-Cellの開発 G-Cell : Genome based prokaryotic cell simulation platform

慶應義塾大学 環境情報学部 高萩航

## 要旨

ポストゲノム時代と呼ばれた近年、生命情報科学分野ではゲノム情報を解読するための手段やツールの開発が進展してきた。しかし、未だゲノムから情報を読み取り機能を予測することのできる全ゲノムベースのシミュレーションモデルは存在しない。ゲノムから読み取ったタンパク質の発現やその挙動という生命の複雑かつ巧妙なシステムの情報を統合しシミュレーションモデルとして提供することは、実験では予測や推定が困難な現象を抽象化して扱うことができるという点で有意義である。また、時間軸や現象と現象の”隙間”を埋める事を可能にするなど次世代の生命科学研究において重要な役割を担うと考えている。そこで本研究では、全ゲノムベース細胞シミュレータG-Cellの開発を行うとともに、G-Cell上に大腸菌K12株のゲノム配列を入力情報としたDNA複製モデルを実装し、複製の開始、DNAの伸長複製、それに関わる酵素複合体の形成に注目して現象の考察を行った。本モデルは大腸菌に関する既存の情報の統合という位置付けも担っており、大腸菌に関する主要なデータベースであるEcoCycに登録された全てのDNA複製関連タンパク、酵素、遺伝子を実装できるよう、1つの酵素・機能を1つのコンポーネントとして実装し、必要に応じて機能を追加削除できる汎用的なシステムに設計した。この大腸菌K12株DNA複製モデルを用いて、DNA複製関連酵素複合体の形成にかかる時間、元となる単量体の量の変化のシミュレーション、複製関連タンパク質複合体であるプライモソームの挙動の調査を行った。その結果、本モデルを用いてDNA複製関連複合体形成の律速段階となる分子の量の変化を観察することができ、また複製に伴う塩基の複製ミスによる誤複製塩基の蓄積を時間経過に沿って観察することができた。本研究で構築したG-Cellプラットフォームと大腸菌DNA複製シミュレーションモデルは、入力とするゲノム情報を変更するだけでその情報を解釈し挙動の変化に反映することがある程度可能である。本モデルを合成生物学やゲノム設計といった分野で利用し実験結果と比較対照することで、実験では解明が難しい細胞内での複合体量の時系列変化や、DNA複製や転写に伴う1つ1つの塩基へのタンパク質の結合解離の状態について新たな知見をもたらすことができる。

Keywords : 全ゲノムベース, シミュレーション, G-Cell, 大腸菌, DNA複製

## 1. 序論

### 1.1. 研究背景

システム生物学の一分野として、これまで生命現象を数理モデルで記述しコンピュータ上で再現しようという試みが成されてきた。これは生命を分子の集合及びその個々の分子の挙動と捉えるのではなく、複雑な相互作用と分子ネットワークから構成される一つのシステムとして捉えるという考え方に基づいている。コンピュータによる生命プロセスのシミュレーションとして、1997年にE-Cellが発表された[1]。当初E-Cellプロジェクトが目指していたのは、自活できる最小限の要素を全てコンピュータ上で再現するというものであり、この時実装されたマイコプラズマ菌の細胞モデルは127個の遺伝子に基づくネットワークシミュレーションを軸としていた。E-Cell上にはこれまでに神経細胞の挙動を再現したモデルや、心筋細胞の極めて正確な電気生理学的挙動を再現した心筋モデル[2]など様々なモデルが構築されている。それまでの生命モデルは全てシミュレータの仕様に準じた”モデル”を入力としており、更新された知見を取り入れるためにはモデルへの反映を手動で行う必要があったため、常に最新の情報が反映されたモデルに保っておくことは容易ではなかった。さらに1つのモデルを作成するコストから、後続するモデルの作成が思うように進まないというのが現状である。この問題を解決すべく、2006年にゲノム情報を入力とした代謝シミュレーションシステムGEMが発表された[3]。GEMはシステム生物学分野における従来の手作業によるモデル作成のコスト問題を解決した上で、目紛しい速さで更新される生物データベースに自動で接続し、最新の情報に基づいた代謝モデルを生成する革新的なものであった。GEMはバイオインフォマティクスの観点からシステム生物学を見直し、生命プロセスの自動解釈という機能を加える事で同分野に新たな視点を示している。GenBank[4]やEMBL[5]といったデータベースやFASTAソフトウェア[6]を用いてゲノムの情報を読み取り、オープンリーディングフレームの予測、ゲノム上に存在する酵素の発現の確認を通して関連する代謝物のリストと代謝マップの生成まで

を行う。また2012年にはマイコプラズマ菌の全てのタンパク質情報が含まれ、コンピュータ内で遺伝子発現やゲノム複製、代謝の様子を再現した全細胞モデルが発表された[7]。この全細胞モデルは、マイコプラズマ菌の遺伝子型からタンパク質の発現や代謝変動といった表現型を予測することができる。開発者の1人であるスタンフォード大学のMarkus Covert博士は全細胞モデルに 1. 既存データセットの統合 2. 知識の限界を探る 3. 複合体及びマルチネットワーク表現型の予測 4. 新しい発見に繋がる実験の提案 5. 安全かつ効果的な遺伝子組換え生物デザインのフレームワークの提供 というコンセプトを与え、これらをまとめて全細胞モデルを作るべき理由としている[8]。

ポストゲノム時代と呼ばれた近年、生命情報科学分野ではゲノム情報を解読するための手段やツールの開発が進展してきたが、未だゲノムから情報を読み取り機能を予測することのできるゲノムベースの全細胞モデルは存在しない。ゲノムから読み取ったタンパク質の発現やその挙動という生命の複雑かつ巧妙なシステムの情報を統合しシミュレーションモデルとして提供することは、実験では予測や推定が困難な現象を抽象化して扱うことができるという点で有意義であり、時間軸や現象と現象の”隙間”を埋める事を可能にするなど次世代の生命科学研究において重要な役割を担うと考えている。E-Cell システムはモデル作成のコストが高く、またゲノムの解釈を前提としていない。GEM システムはその点を改善しているがモデル記述の自由度が低く、対象を代謝モデルに限定している。マイコプラズマ菌の全細胞モデルは細胞の遺伝子発現、代謝変動、ゲノム複製など幅広い生命現象を再現できるが、モデル作成に要求されるコストが高く汎用性が低い。私はゲノム情報のモデルへの反映、モデル作成コストの削減、また自由度の高いゲノムベースシミュレーションシステムの構築を目指して、ゲノム情報を入力として細胞の生物学的プロセスを記述できる G-Cell (Genome-based prokaryotic cell simulation platform) の構想に至った。G-Cell は全ゲノムベースの細胞シミュレーションプラットフォームの先駆けとなり、入力となるゲノムの変化を認識・解釈し、出力であるモデルに反映するシステムとなることが期待される。

## 1.2. 研究の目的

G-Cellプロジェクトでは、唯一の入力情報であるゲノムの情報解釈を行い、DNAの転写、複製、翻訳といったメカニズムの記述及びシミュレーションを目指す。G-Cell 0.0.1では、入力情報としてのゲノムから読み取るのは塩基配列と遺伝子の有無である。入力ゲノムに変異が入った場合や一部領域を欠失した場合、欠損した遺伝子に対応するタンパク質が発現せず、これによりゲノム情報の変化に応じて細胞システムの挙動が変化し、ゲノムベースによる細胞モデル構築及びシミュレーションが実現できる。本研究ではまずG-Cellにゲノムの読み込み及び解釈、モデルの記述、シミュレーションと結果の描画を行うことのできるシステムを構築し、これをG-Cell 0.0.1とする。その後計算アルゴリズムの動作チェック、サンプルデータの描画、計算速度のベンチマークを行い、シミュレーションシステムとしてのG-Cellを評価する。また大腸菌K12株のDNA複製モデルをG-Cell上に実装し、ゲノムベースシミュレーションがG-Cell上でどのように行えるかを示す。

## 2. 対象と手法

### 2.1. 研究対象

G-Cellシステムを構築する上で、私はゲノム情報を入力とした原核細胞のモデル構築、ゲノム情報の変化を自動で解釈し、シミュレーション結果に反映するシステムの構築、汎用プログラミング言語Python (<https://www.python.org>) によるモデル記述の負担軽減という3つの意義を与えた。G-Cellを用いたシミュレーションは、実験では解明が難しい細胞内での複合体量の時系列変化や、DNA複製や転写に伴う1つ1つの塩基へのタンパク質の結合解離の状態について新たな知見をもたらすことが期待される。

本研究では、G-Cellによる生命モデル作成の例として大腸菌K12株のDNA複製をG-Cell上で構築し、複製関連タンパク質の形成過程とその機能を再現することによりコンピュータ上でのDNA複製の様子をシミュレーションした。大腸菌などの原核生物におけるDNA複製とは、細胞分裂S期においてDNAが複製される段階・メカニズムのことを指す。DNA複製は生物学におけるセントラルドグマの一過程である。DNA複製は大分して3つの過程の集合と考えることができ、それぞれの過程は複製の開始、伸長、終結と呼ばれる。本研究では構築したG-Cell上でDNA複製の開始、伸長、翻訳に関わる酵素やタンパク質の形成、分解、挙動を実装し、入力ゲノムを生物学的なプロセスを経て複製する過程、その際消費されるATPの量、複製にかかる時間、また複製ミス塩基の蓄積といった情報を得ることができるモデルを構築した。本論文では、複合体の形成と変異ゲノムに対する応答について結果を示し、議論を行う。

### 2.2. 研究手法

G-Cell の計算アルゴリズムには Gillespie 法を選択した[9, 10]。Gillespie 法は複合体の形成など化学反応の確率論的記述をする上で用いられる。Gillespie 法では起こりうる全ての反応、分子の挙動をイベントとして定義し、全てのイベントの反応傾向を計算することによりどの化学反応が起こるかを選択する。複合体の形成や単量体への解離などの確率論的に起こる物理現象を独立なマルコフ過程であると仮定し、各々を

一つのイベントとして定義することで Gillespie 法によって計算する。本研究では複合体の形成、分解、それぞれの酵素の起こりうる挙動を全て Gillespie 法のイベントとして定義し、原核生物の DNA 複製過程を確率論に基づく反応系のモデルとして構築することとした。G-Cell はスクリプト言語 Python 2.7 を用いて記述され、ゲノムの読み込み、Gillespie 法で計算するイベントの作成、シミュレーション、データの出力を行う関数を組み込んだ。これらの G-Cell 関数を組み合わせることでゲノムベースのモデルを容易に作成できる環境を構築し、これを G-Cell 0.0.1 とした。構築した G-Cell プラットフォームを用いて、大腸菌 K12 株の DNA 複製の数理モデルを構築し、DNA 複製関連酵素の形成、分解、酵素の挙動といったダイナミクスをシミュレーションした。大腸菌の DNA 複製モデルの構築にあたって、DNA 複製関連タンパク質と複合体の形成に関する情報は大腸菌に関する主要なデータベースである EcoCyc[11]、遺伝子とその塩基配列に関する情報は GenBank (Ver. 3) [4]から取得した。

### 3. 結果

#### 3.1. 全ゲノムベースシミュレーションシステムG-Cellの設計

G-Cell 上に実装されたモデルは、3つのクラスに分類された機能を用いる。ユーザはこれらのクラス中で定義された関数を用いることで、ゲノムの読み込み・解釈とシミュレーションの実行、データの可視化・保存を行うことができる。G-Cell ではまず、ユーザが単量体と複合体、その初期値を定義する。次に入力となるゲノムを読み込み、1塩基につき1列の配列を与える。各塩基につきユーザが定義した単量体と複合体の数に相当するサイズの配列によってゲノム上の配列の状態を表現する。この配列はユーザが定義した単量体と複合体の位置に対応しており、これらの分子のゲノム上への結合解離の情報を保存していく。次に複合体の形成、分解、酵素の挙動に関するイベントを定義する。このイベントは先述した Gillespie 法により計算されるため、ユーザはそれぞれのイベントに化学量論反応の反応速度定数を定義する。最後にシミュレーション時間、対象とするゲノム配列を設定し、シミュレーションを開始することができる。この一連の過程は G-Cell に同梱されているフォーマットモデルを拡張することにより作成可能で、ゲノムベースシミュレーションモデルを容易に構築できる (図1)。フォーマットモデルとは、G-Cell でのシミュレーションに必要な最低限の機能が実装された基本モデルである。

```
#import G-Cell module
from functions import *
from proteins import *

#setup
time, location, SubList, events = Reactions().setup()
target = Reactions().Target('sequence.txt')

#Generate Monomer&Complex
Reactions().Complex('monomer',100,SubList,target)
Reactions().Complex('complex',0,SubList)

#Events setting
Reactions().Events(Compose('complex',['monomer'],[3],1.0e-2),events)
Reactions().Events(Decompose('complex',['monomer'],[3],1),events)

#simulation
seq, mod = Reactions().Readseq(target,SubList)
logt, logd, t, tend = Showdata().logger(time, SubList, 0, 0.01)
Simulation().Run(t, tend, SubList, events, logt, logd, mod, location)

#showdata, make .png
Showdata().png(['monomer','complex'],logt, logd, SubList,'default')

#Finalize
Showdata().csv(SubList,'summary.csv')
Simulation().Save(mod,SubList)
Simulation().Makedata('default')
```

図1. G-Cell フォーマットモデル

G-Cell では、このフォーマットモデルをベースにモデルを拡張していく。Monomer と Complex のような分子、

それぞれの分子の機能をイベントに定義し、G-Cell の関数を用いてシミュレーションを行うことができる。またシミュレーション結果の可視化も png 形式、csv 形式それぞれに書き出すことが可能であり、ゲノムベースモデルの作成、シミュレーション、可視化までのワークフローを容易に行うことができる。計算に用いる関数やシステムが用いるパラメータは全て別のファイルに記述しており、基本的にモデル毎に変化しない。故にモデル作成時に変更するパラメータは全てフォーマットモデルのものであり、必要なパラメータを多数のファイルに記述するといった乱雑なモデリングを避けることができる。

### 3. 2. 大腸菌K12株DNA複製シミュレーションモデルの構築

DNA 複製の開始において、複製フォークの形成には二本鎖 DNA を解いてプライマーRNA と新しい DNA 合成の鋳型となる ssDNA を作る必要があり、この過程によって作られた ssDNA は DNA ヘリカーゼの結合部位となり、複製が開始される。複製の開始の際に最初に ssDNA となる特定の配列を複製起点と呼び、大腸菌では OriC (Origin of Chromosome) と呼ばれる配列がそれにあたる [12, 13]。大腸菌の OriC はイニシエータータンパクである DnaA が結合する 9 塩基長の結合部位 5 個 (DnaA-box) と DNA の解きほぐしが開始される 13 塩基長の反復配列を 3 個もつ [13, 14]。DnaA が OriC の DnaA-box 部位に結合してさらに 13 塩基長の部分を解離した後、ssDNA と DnaA が DNA ヘリカーゼである DnaB とヘリカーゼ装着タンパク質である DnaC の複合体を引き寄せる。DnaB は通常 DnaT, PriA, PriB, PriC, DnaG と複合体を形成しており、これを DNA 複製の開始を担う酵素複合体であるプライモソームと呼ぶ (表 1)。本ダイジェストでは、G-Cell 上で実装したプライモソームの形成ダイナミクスのシミュレーション結果を示す (図 2-1)。プライモソームの形成に必要な単量体の初期値は全て 600 個に設定した (標準的な大腸菌のサイズにおいて約  $1\mu\text{M}$  に相当)。DNA 複製時のリーディング鎖での DNA 合成を再現するために、酵素複合体プライモソームの機能を DnaA によるイニシエーションシグナル後の 2 本鎖 DNA を開環と設定し、DnaA による DNA 複製の発火後 2 本鎖 DNA が開環され、DNA 複製の準備が行われるかを検証した。その結果、DnaA-box への DnaA の結合後、プライモソームはリーディング鎖方向の 2 本鎖 DNA を開環した。プライモソームの他にも、本モデルは DNA 複製に関連するタンパク質単量体を 52 種類、酵素複合体を 31 種類定義しており、それぞれの酵素複合体の形成過程を同時にシミュレーションしている。

表 1. 酵素複合体プライモソームは 6 種類 14 個の分子で形成されている

名称	分子数	機能	参照
DnaB	6	DNA ヘリカーゼ. 二本鎖 DNA を開環する.	[15]
DnaT	3	複製開始関連因子. プライモソームの再構成に必要.	[16]
DnaG	1	DNA プライマーゼ. ラギング鎖の RNA プライマーを合成する.	[17, 18]
PriA	1	複製フォークの再開に用いるタンパク質.	[19]
PriB	2	プライモソーム再構成因子.	[20]
PriC	1	プライモソームの再構成, DNA 複製の再開に必要.	[21]

図 2-1 はプライモソームを構成する単量体である DnaB, DnaT, PriA, PriB, PriC, DnaG がそれぞれ消費され、プライモソーム複合体が形成されていく過程を表している。プライモソームは最終複合体で、DnaB は 6 量体で Replicative DNA Helicase, DnaT は 6 量体で Replicative Protein DnaT と前駆複合体が存在しており、本研究で構築したモデルではまず前駆複合体が形成され、それが新たな素材となって最終複合体であるプライモソームが形成される過程を示している (図 2-2)。

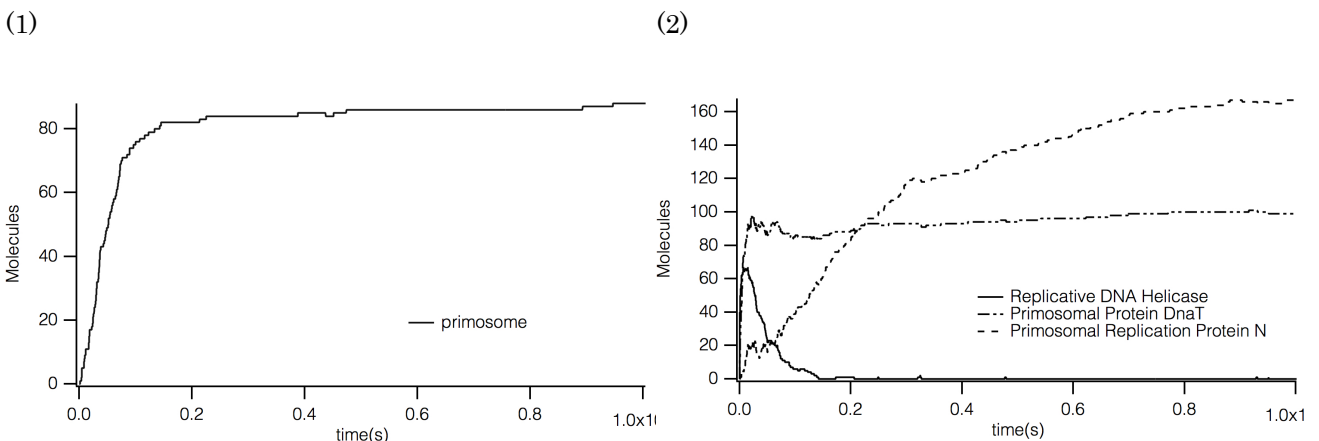


図 2. G-Cell によるシミュレーション結果

プライモソームの形成過程。縦軸は分子数、横軸は時間を表す。

(1) 酵素複合体プライモソームの形成 (2) プライモソーム前駆複合体の形成

次に、ターゲットとして読み込むゲノムが変異の入ったものであった場合、ユーザが定義したタンパク質を発現する遺伝子の有無を G-Cell が判断し結果に反映できるかを検証した。本研究では、大腸菌 K12 株ゲノムにそれぞれ 0.0001% から 1% まで変異が入ったゲノム配列をランダムに作成し、変異ゲノムを読み込んだ際に DNA 複製モデルで定義した 52 個の遺伝子のうち、いくつ欠損するかを調査した (図 3-1)。それぞれの変異率で 10 回シミュレーションを行っている。

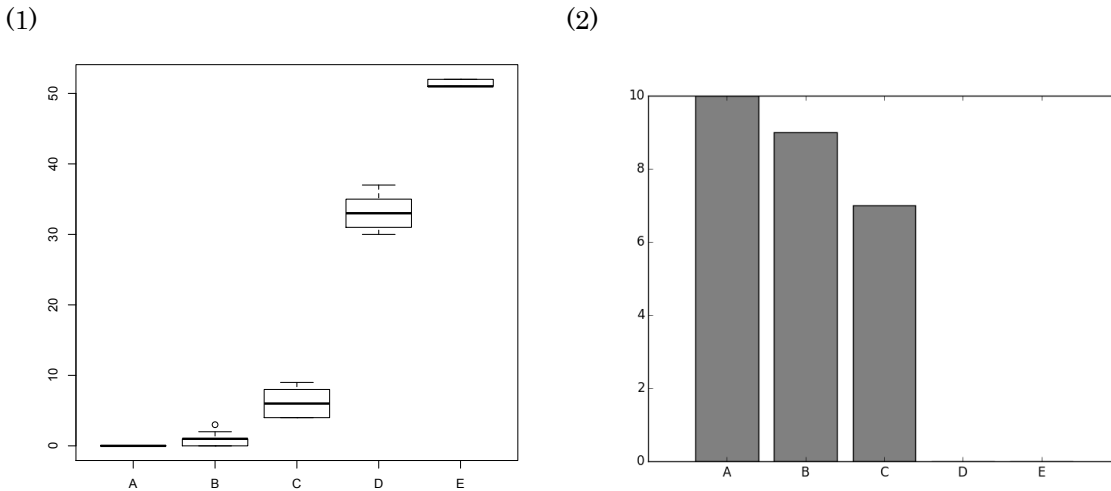


図 3. G-Cell によるシミュレーション結果

(1) ゲノムに変異が入った場合の欠損遺伝子数. A:0.0001%, B:0.001%, C:0.01%, D:0.1%, E:1%

(2) プライモソームが正しく形成された回数

ゲノムに変異を入れることにより、ゲノム全体の 1% に変異が入った状態では DNA 複製に関わる全ての遺伝子が欠損し大腸菌が生存不能になったためそれ以上シミュレーションを継続することが不可能であった。またゲノム全体の 0.0001% に変異が入った状態では、欠損した遺伝子は最大 1 つにとどまった。プライモソームが 14 個の分子で正しく形成された数は欠損遺伝子数に反比例したが、ゲノム変異率が 0.01% から 0.1% の間でプライモソームの形成能力が著しく低下した (図 3-2)。

#### 4. 議論

##### 4.1. G-Cell におけるモデル作成とシミュレーション

私が設定したゲノムベースシミュレータの要件は、ゲノム情報を入力とした原核細胞のモデル構築、ゲノム情報の変化を自動で解釈し、シミュレーション結果に反映するシステムの構築、モデル記述の負担軽減及び汎用性の確保という3つのである。現行のG-Cell 0.0.1システムはこれらの要件を満たしており、ゲノムベースシミュレーションに必要な最低限な機能を備えていると言える。

G-Cell 0.0.1では、入力情報としてのゲノムから読み取るのはモデルを構築した生物の塩基配列と遺伝子の有無である。遺伝子領域中の塩基に1つでも変異が入っていると、その遺伝子は発現せず、対応するタンパク質は存在しないと解釈される。そのため現行のG-Cellシステムでは、特定の塩基に変異が入った場合の酵素活性の低下や複合体の形成速度の低下といったメカニズムを表現することはできない。また、本研究で作成した大腸菌K12株のDNA複製モデルはDnaAタンパク質による1回の発火と複製のみ表現している。しかし実際にはDNA複製は進行中に次の発火が起こり、最初のDNAの複製終了を待たずに別のDNAが複製されていく。本モデルをより実現象に近づけるためには、例えばGillespie法のイベントをDNA複製の進行具合によって自動で追加し、2回目以降の発火も同時に計算を進めるなどのシステムの実装が必要である。また、DNA複製と遺伝子の転写・発現は干渉し、互いの進行を制御するという現象も報告されている。今後G-Cellをより汎用的なモデル構築システムに改良し、転写や翻訳といった生命現象も実装することで、ゲノムの転写・複製・翻訳という細胞が増殖するのに必要なメカニズムの挙動や相互干渉を検証することも可能になると考えている。

##### 4.2. DNA複製関連酵素の形成ダイナミクスと変異の入ったゲノムを読み込んだ際のモデルの応答

本研究では、タンパク質単量体52種類の消費過程及び酵素複合体31種類の形成過程を時系列で観察することができた。これはDNA複製の進行に伴う大腸菌内の分子の存在バランスやエネルギー消費に言及することができ、DNA複製における律速段階の予測や分子局在の解明に繋がると考えられる。欠損する遺伝子は遺伝子が持つ塩基配列の長さやゲノムの全長から予測することが可能であるが、遺伝子が欠損した結果細胞システムにどのような影響を与えるかは予測することが不可能であった。本研究では、読み込むゲノムに0.0001%から1%までの変異が入っていた場合、酵素複合体プライモソームが構成元の6種類14分子から正しく形成されるかを観察した。正しいプライモソームの形成能はゲノム変異率0.01%から0.1%の間で著しく低下したが、これは、紫外線などによる遺伝子損傷でゲノムに0.01%以上の変異が入った場合、複製に関連する重要な因子の一つであるプライモソームが形成されずDNA複製を開始できないことを示している。プライモソームの他にもDNAポリメラーゼIIIホロ酵素やDnaA分子についても同様に閾値を調べることで、DNA複製に必要な最低限の機能を保つ限界のDNA変異率、つまりは生存閾値の探索を行うことが可能であることが示唆された。今後はDNA複製モデルで挙動を再現できるタンパク質の種類を追加していくことで、DNA複製の調節機構や複製ミス塩基対の修復機構を検証し、生化学的な反応速度定数の知見や実験データとの比較を行っていく。

#### 4.3. 結論と展望

本研究で構築したG-Cellシステムを利用することで、ゲノムを入力とし、ゲノムの変化に応答できる生命モデルの作成を容易に行うことができる。しかし現状、ある特定の塩基に変異が入った場合、対応する酵素が欠損ではなく機能が変化するなどのシミュレーションは行うことができない。この問題は、酵素の機能や複合体の形成を全てモデル上で定義している点にある。今後、統一データベースからこのような情報を自動で採り機能を書き加えるシステムを構築することで、唯一入力とするゲノムを読み込ませるだけでコンピュータ上で細胞モデルを構築することが可能となることが期待される。これは生命シミュレーションが行き着く究極の課題であり、ゲノムに書いてあるはずの情報のみで生命が構築されるという仮説の証明にもつながる。本研究で構築したG-Cellはゲノムベースシミュレーションシステムのプロトタイプとして最小限の機能が備わっており、今後G-Cellデータベースの自動作成やモデル記述の簡略化、他のシミュレーションシステムとの比較を行えば、G-Cellは世界初の全ゲノムベースシミュレーションシステムとしての位置を確立できると確信している。ゲノム情報を読み取りDNA複製や転写といった生命のメカニズムを記述しモデルに反映するシステムは、実験では把握の難しい細胞周期ごとの分子数の局在やゲノム上に結合しているタンパク質の各々の状態や結合解離といった情報を予測することにつながり、既知の知見や実験データの時系列を埋める事を可能にするなど次世代の生命科学研究において重要な役割を担うと考えている。

#### 謝辞

本研究の遂行にあたっては、研究メンターである海津一成研究員、アドバイザーである高橋恒一チームリーダーをはじめ国立研究開発法人理化学研究所 QBiC 生化学シミュレーションチームの協力無くしては実現不可能であった。特に海津研究員には研究の構想やモデルの細部に至るまで熱心な指導を賜った。また、E-Cellグループの方々にはミーティングを通じたアドバイスをいただき、大変お世話になった。最後に、このような研究環境と素晴らしい機会を提供して下さった富田勝教授に心より感謝申し上げたい。

## 参考文献

1. Tomita M, Hashimoto K, Takahashi K, Shimizu T, Matsuzaki Y, Miyoshi F, et al. E-CELL: Software Environment for Whole Cell Simulation. *Genome Inform Ser Workshop Genome Inform.* 1997;**8**:147-55.
2. Matsuoka S, Sarai N, Kuratomi S, Ono K, Noma A. Role of individual ionic current systems in ventricular cells hypothesized by a model study. *Jpn J Physiol.* 2003;**53**(2):105-23.
3. Arakawa K, Yamada Y, Shinoda K, Nakayama Y, Tomita M. GEM System: automatic prototyping of cell-wide metabolic pathway models from genomes. *BMC Bioinformatics.* 2006;**7**:168.
4. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. GenBank. *Nucleic Acids Res.* 2006;**34**(Database issue):D16-20.
5. Stoesser G, Baker W, van den Broek A, Camon E, Garcia-Pastor M, Kanz C, et al. The EMBL Nucleotide Sequence Database. *Nucleic Acids Res.* 2002;**30**(1):21-6.
6. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. *Proceedings of the National Academy of Sciences of the United States of America.* 1988;**85**(8):2444-8.
7. Karr JR, Sanghvi JC, Macklin DN, Gutschow MV, Jacobs JM, Bolival B, Jr., et al. A whole-cell computational model predicts phenotype from genotype. *Cell.* 2012;**150**(2):389-401.
8. Carrera J, Covert MW. Why Build Whole-Cell Models? *Trends Cell Biol.* 2015;**25**(12):719-22.
9. Cai X. Exact stochastic simulation of coupled chemical reactions with delays. *J Chem Phys.* 2007;**126**(12):124108.
10. Gillespie DT. Stochastic simulation of chemical kinetics. *Annu Rev Phys Chem.* 2007;**58**:35-55.
11. Karp PD, Riley M, Paley SM, Pellegrini-Toole A, Krummenacker M. EcoCyc: Encyclopedia of Escherichia coli genes and metabolism. *Nucleic Acids Res.* 1998;**26**(1):50-3.
12. Messer W, Egan B, Gille H, Holz A, Schaefer C, Woelker B. The complex of oriC DNA with the DnaA initiator protein. *Res Microbiol.* 1991;**142**(2-3):119-25.
13. Messer W, Blaesing F, Jakimowicz D, Krause M, Majka J, Nardmann J, et al. Bacterial replication initiator DnaA. Rules for DnaA binding and roles of DnaA in origin unwinding and helicase loading. *Biochimie.* 2001;**83**(1):5-12.
14. Fuller RS, Funnell BE, Kornberg A. The dnaA protein complex with the E. coli chromosomal replication origin (oriC) and other DNA sites. *Cell.* 1984;**38**(3):889-900.
15. Michel B, Ehrlich SD, Uzest M. DNA double-strand breaks caused by replication arrest. *EMBO J.* 1997;**16**(2):430-8.
16. Allen GC, Jr., Dixon NE, Kornberg A. Strand switching of a replicative DNA helicase promoted by the E. coli primosome. *Cell.* 1993;**74**(4):713-22.
17. Benz EW, Jr., Reinberg D, Vicuna R, Hurwitz J. Initiation of DNA replication by the dnaG protein. *J Biol Chem.* 1980;**255**(3):1096-106.
18. Rodina A, Godson GN. Role of conserved amino acids in the catalytic activity of Escherichia coli primase. *J Bacteriol.* 2006;**188**(10):3614-21.
19. Cadman CJ, McGlynn P. PriA helicase and SSB interact physically and functionally. *Nucleic Acids Res.* 2004;**32**(21):6378-87.
20. Sandler SJ. Requirements for replication restart proteins during constitutive stable DNA replication in Escherichia coli K-12. *Genetics.* 2005;**169**(4):1799-806.
21. Zavitz KH, DiGate RJ, Mariani KJ. The priB and priC replication proteins of Escherichia coli. Genes, DNA sequence, overexpression, and purification. *J Biol Chem.* 1991;**266**(21):13988-95.