

平成 22 年 5 月 31 日現在

研究種目：若手研究（B）

研究期間：2008～2009

課題番号：20710158

研究課題名（和文） オンライン細胞モデリング及びシミュレーション環境の構築

研究課題名（英文） Development of Online Cell Modeling and Simulation Environment

研究代表者

荒川 和晴（ARAKAWA KAZUHARU）

慶應義塾大学・政策・メディア研究科・講師

研究者番号：40453550

研究成果の概要（和文）：

生命の動態を理解するためには、既知の知見やデータから構成論的アプローチにより再構築した細胞モデルのシミュレーションが欠かせない。本研究では細胞のモデリング及びシミュレーションを支援するためのソフトウェアシステムとして、モデルのプロトタイピングのためのウェブアプリケーション、動的モデリングのために必要な式やパラメータなど定量情報のデータベースシステム、そしてシミュレーション実行及び解析のためのウェブサービスを開発し、インターネット上で公開した。

研究成果の概要（英文）：

In order to understand the dynamic nature of living cells, an integrative approach to reconstruct cell models with existing biological knowledge and huge masses of omics data for quantitative simulation is a necessary step for the advancement of systems biology. For this purpose, we have developed a modeling and simulation environment as online web services, including a software for cell model prototyping, a database system of quantitative data, and a set of web services for simulation and analyses.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	1,900,000	570,000	2,470,000
2009年度	1,400,000	420,000	1,820,000
年度			
年度			
年度			
総計	3,300,000	990,000	4,290,000

研究分野：複合新領域

科研費の分科・細目：ゲノム科学・ゲノム情報科学

キーワード：システムバイオロジー・細胞シミュレーション・細胞モデリング・バイオインフォマティクス・分子細胞生物学・E-Cell・G-language・GEM System

1. 研究開始当初の背景

500 以上の生物の完全長ゲノム配列が読まれ、

マイクロアレイによるトランスクリプトームや質量分析器によるプロテオームやメタボロームなど、細胞レベルでの網羅的かつ定

量的な測定が可能になった今、これまでに蓄積されてきている細胞内のさまざまな分子生物学的知見とこれらの情報を組み合わせ、統合的に生命現象を理解するシステムバイオロジーが注目されている(Kitano, 2002)。研究代表者らも本研究課題申請の前年 Science 誌に発表した論文において、トランスクリプトーム、プロテオーム、メタボローム、フラックソームを網羅的に測定することで、大腸菌内での生化学反応が遺伝子欠損にも関わらず、ロバストかつシステム的に制御されている様子を明らかにした(Ishii, et al., 2007)。研究代表者が所属する研究所は、このようなシステムバイオロジー研究のためにシミュレーションのためのソフトウェア E-Cell(Tomita, et al., 1999)を開発してきている。一方、大量のデータを理解し、シミュレーションを行うためには、それらを生体パスウェイのコンテキストに基づいてコンピュータ上で数学的及び情報学的に記述する「モデリング」作業が不可欠となる。モデリングを支援する環境は、主に細胞シミュレーションのためのソフトウェアに付随するツールとしてさまざまなものが開発されているもの(例えば Copasi, CellDesigner など: Hoops, et al., 2006; Kitano, et al., 2005)、これらのツールはパスウェイのネットワーク図を製図しそこに式やパラメータを当てはめることを目的とするものであり、あくまでそのモデルは研究者による人手の文献やデータベースの調査が前提となっている。このように人手による煩雑な作業を必要とする限り、モデリングが今後のシステムバイオロジーの発展において大きなボトルネックとなる。

2. 研究の目的

本研究ではこのためにモデリングを支援するソフトウェア環境、そしてそれを用いてシミュレーションを行うソフトウェアのオンライン版ウェブサービスを開発することを目的とする。具体的には、(1) ゲノム情報を元に KEGG/BioCyc と同等の精度で代謝パスウェイモデルを自動生成するソフトウェア、(2) 定量データを文献から収集しデータベース化し、これを用いて定量モデルを作成できるシステム、(3) 細胞シミュレーションソフトウェアとその解析ツールのオンラインウェブサービスを開発し、ウェブ上で公開する。

3. 研究の方法

(1)ゲノム情報に基づいた、各種データベース利用による定性的代謝パスウェイネットワークモデルの自動生成

現状では各種公共データベースには重複した

情報が散在し、それぞれ独自の識別子やオントロジーに基づいて情報が整理されているため統合が困難である。一方で、分子生物学のデータの多くはセントラルドグマに則った形で保存されているため、何らかの形で対応する遺伝子の配列データへとリンクされている。この配列データに基づいた各種データベースの統合は、テキストマイニングや辞書の作成に依存しないため、比較的高速かつ汎用的に行う事ができる。本研究ではまず研究代表者らが既に作成している GEM System (Arakawa, et al., 2003)の配列ベースの統合を、高速な配列相同性検索が行える BLAT (Kent, 2002)などを用いてより高速化し、現在対応しているよりも多くの情報、特に酵素反応の活性化及び阻害の定性的情報などをもとに反応式を導き出す過程まで自動化する。

このように得られたプロトタイプモデルはそれだけで完全な細胞シミュレーションモデルとすることはまだまだ難しく、このドラフトモデルを土台として研究者がさらに情報を吟味し追加することでモデルを完成させていかなければならない。このような人手によるモデリング過程は非常に多くの生物学的知見とデータを横断的に検索・閲覧しながら行う必要がある。そのため、このような人手によるモデル作成時に素早くさまざまなオミクスに跨がる情報にアクセスできるような環境がセットになってはじめて自動生成のソフトウェアが意味を持つ。近年 Web 2.0 と呼ばれるパラダイムにより、非同期 XML 通信と Javascript (AJAX) などを用いた操作性に優れたウェブアプリケーションが開発されており、これらの技術を用いることで、インターネット上でもデスクトップ上とほとんどかわらない操作性で扱えるソフトウェアを開発することが可能である。よって、このような最新ウェブアプリケーション技術を使い、オンラインでモデリングを行える環境を構築する。

また、以上のように自動生成されたパスウェイモデルを視覚的に扱えるようにするため、SBML 形式で出力するとき SBML Layout Extension (Gauges, et al. 2006)の規格、あるいは CellDesigner の規格に基づき、座標情報やパスウェイオブジェクトの形式などを付加する。各々のオブジェクトノードの種類や関係は定性的モデルとして得られているが、視覚的にレイアウトされた座標情報は、代謝物質と酵素、さらにはその阻害物などを含めた大規模ネットワークに関しては現存しない。例えば KEGG データベースではパスウェイが細分化され、BioCyc の Omics Viewer や Reactome データベースの Sky Painter では反応が中心

となり代謝物質と酵素が同時に表示されない。このため、研究代表者らが作成しすでに公開・発表しているMEGUソフトウェア (Kono et al., 2006)などの座標をもとに、自動レイアウトなどを使い適切な可視化表示が行われるようにする。MEGUはKEGGパスウェイを繋げた統合マップを提供しており、これを利用する事で大規模パスウェイの多くはカバーすることができる。以上を自動化したソフトウェアとして開発する。

(2) 定量的モデルデータベースの作成

研究代表者が所属する慶應義塾大学先端生命科学研究所が中心となっている E-Cell プロジェクトは細胞シミュレーションを目的としてシミュレータやモデルの開発を 10 年近く行っている、世界的に見ても特に歴史の長いプロジェクトであり、現在までに数多くのモデルが論文や学会で発表されている。これらのモデル及びその部分を目的に応じて取り出し、組み合わせることができれば、定量モデル作成において重要な基盤データとなるが、これらのモデルはアーカイブすらされていないのが現状である。よって、E-Cell プロジェクトがこれまでに作成してきたモデルを BioModels データベースが提唱している MIRIAM (Le Novère, 2006)規格に含まれる内容などでアノテートし、容易に検索ができるような形式で公開できるようデータベース化する。

(3) 細胞シミュレーションソフトウェアのオンラインウェブサービス

前述の通り細胞シミュレーションを行うプラットフォームとして、Copasi や E-Cell などのソフトウェア環境が存在するが、現状ではこれらのソフトウェアはインストールやメンテナンスに専門知識を要し、また、大規模なシミュレーションを行う場合には高性能の計算機を必要とする。そこで、本研究課題で開発したオンラインのモデリング環境と連動して利用できるように、E-Cell や Copasi をウェブサービスとして提供する。この時、バイオインフォマティクスの分野でウェブサービスの相互運用性を高めるために有効であると注目されている Representational State Transfer (REST) という HTTP と URI に基づいたプロトコルにて実装することで、ユーザが容易にブラウザから利用できるのみならず、他のバイオインフォマティクス及びシステムバイオロジーのソフトウェアから利用しやすい環境を提供する。

4. 研究成果

(1) ゲノム情報に基づいた、各種データベース利用による定性的代謝パスウェイネットワー

クモデルの自動生成

細胞のモデリングを行う際には、少なくとももゲノム、トランスクリプトーム、そして酵素反応のネットワークとしてのパスウェイを横断的に検索・閲覧し、網羅的な視点からデータにアクセスすることが不可欠である。このようなデータの階層性に加え、それぞれのオミクス階層においては多尺度性が存在している。ゲノムやパスウェイは二次元地図として表現されるが、これらの地図は数千から数万の要素を持つ極めて大規模なものであり、詳細に個別の要素を見る用途と、全体を俯瞰するように眺めることでシステムとしての挙動を把握する場合などでは異なる尺度でデータを閲覧する必要がある。よって、これら多角的かつ多尺度のデータを閲覧するために、4つのオミクス階層を切り替えながら、Google Maps API を用いた Zoomable User Interface (ZUI)の実装により、スムーズに多尺度的な閲覧を可能にするウェブアプリケーション Genome Projector を開発した(図1)。最新のインターネットアプリケーション開発技術を用いることにより、本ソフトウェアはデスクトップのアプリケーションと比較しても遜色がないユーザインタフェースをウェブアプリケーションながら実現している。

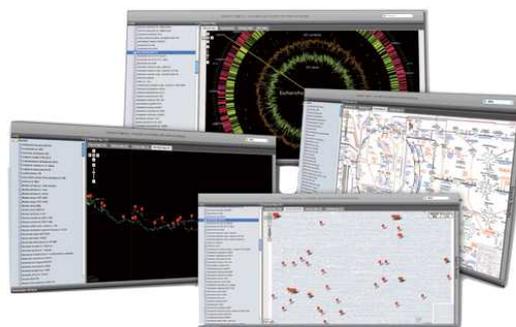


図1. Genome Projector

次に、このような情報を元に定性的代謝パスウェイのネットワークモデルを構築するために、同様の技術を用いてウェブアプリケーション Pathway Projector を開発した。Pathway Projector はKEGG Atlasをもとに酵素や遺伝子のノードを追加した大規模代謝マップをZUIで閲覧できるソフトウェアであり、この上で経路検索やデータのマッピング、パスウェイの編集などを行うことができるが、特に重要な点として、我々が開発したゲノム情報からさまざまな公共データベースの情報を統合することで自動的にパスウェイのモデルをプロトタイプ化することができる GEM System を前述の通り高速化したものが組み込まれている。よって、Genome Projector を元を選び出したゲノム、あるいは

は遺伝子セットを Pathway Projector にかけることでパスウェイモデルをオンラインで自動生成することが可能になり、従来はセットアップが非常に困難でその結果の解釈も難しかった GEM System によるモデリングが容易かつ可視化を伴って利用できるようになった。



図2. Pathway Projector

Pathway Projector や GEM System で得られたモデルを Cell Designer などのソフトウェアでさらに詳細にモデリングを行う場合、既に述べたようにそのレイアウトが研究者にとってわかりやすいことが重要である。そこで、SBML で定義されたモデルを Pathway Projector で改良した KEGG Atlas の座標情報を用いてレイアウトするためのオンラインツール METAL System も別途提供している (<http://ws.g-language.org/metal/>)。

(2) 定量的モデルデータベースの作成

GEM System では定性的なネットワークモデルに公共データベースで存在するパラメータや以下の一般化ミカエリスメンテン式

$$V = KcF[E] \prod_{i=0}^{nSubs} \frac{[S_i]^{N_i}}{(KmS_i^{N_i} \prod_{k=0}^{nInh} (1 + \frac{[I_k]}{KIS_k})^{N_i} + [S_i]^{N_i})} - KcR[E] \prod_{j=0}^{nProds} \frac{[P_j]^{N_j}}{(KMP_j^{N_j} \prod_{p=0}^{nInh} (1 + \frac{[I_p]}{KIP_p})^{N_j} + [P_j]^{N_j})}$$

で自動導出した微分方程式を持たせることが可能であるが、まだまだ酵素反応の定量情報を持つ公共データベースは少ない。そこで、文献情報からこれらのデータを収集した in silico model (ism) データベースを開発・公開した。現在 ism データベースには 410 化合物、748 反応における 1297 の定量パラメータが含まれている。また、ism データベースは「ショッピングカート」機能を有しており、研究者はこのデータベースを検索しながら適切な酵素を選びカートに追加して行くことで、システムが自動的に必要な細胞システムや化合物などを補完し、SBML で記述されたモデルを作成することも可能である (図3)。

図3. Ism データベース



(3) 細胞シミュレーションソフトウェアのオンラインウェブサービス

細胞シミュレーションソフトウェアとして世界的に広く使われている Copasi と、多解法・多次元尺度によるシミュレーションが可能な点で他のソフトウェアよりも複雑なモデリング・シミュレーションが可能な E-Cell に関して、REST によるウェブサービスを開発・公開した。特に Copasi では、シミュレーションに加え、パラメータ推定、代謝制御解析、感受性解析、定常状態解析などのモデル解析もウェブサービスとして利用することができる。これらのサービスは全て URI をベースとしており、例えば E-Cell であれば <http://soap.g-language.org/ecell/> という URI に SBML か EML に基づくモデルを POST によりアップロードし、そこで発行される ID を

[http://soap.g-language.org/ecell/run/\[Session ID\]](http://soap.g-language.org/ecell/run/[Session ID]) という URI にアクセスすることでシミュレーションを行うことができる。このように、インターネットにアクセスすることができる環境さえあれば REST のサービスは容易に扱うことが可能であり、ユーザがブラウザ上で解析を行うことも、コンピュータプログラムから API として利用する場合も、同じサービスを共有することが可能である。

このように、Rich Internet Application (RIA) として実装された Genome Projector や Pathway Projector で細胞の定性的モデリングを行い、ism データベースで定量情報を付加し、さらにそのモデルを REST ウェブサービスで実行する環境が本研究により整い、オ

ンラインで研究者がシステムバイオロジーのためのモデリング及びシミュレーションを行うための基盤が構築できたと言えるだろう。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計6件)

1. Kazuharu Arakawa, Nobuhiro Kido, Kazuki Oshita, Masaru Tomita, "G-language Genome Analysis Environment with REST and SOAP Web Service Interfaces", Nucleic Acids Research, 査読有, Vol.38, 2010, in press.
2. Nobuaki Kono, Kazuharu Arakawa, Ryu Ogawa, Nobuhiro Kido, Kazuki Oshita, Keita Ikegami, Satoshi Tamaki, Masaru Tomita, "Pathway Projector: Web-based Zoomable Pathway Browser using KEGG Atlas and Google Maps API", PLoS One, 査読有, Vol.4, No.11, 2009, e7710.
3. Kazuharu Arakawa, Kazuki Oshita, Masaru Tomita, "A web server for interactive and zoomable Chaos Game Representation images", Source Code for Biology and Medicine, 査読有, Vol.4, 2009, 6.
4. Kazuharu Arakawa, Satoshi Tamaki, Nobuaki Kono, Nobuhiro Kido, Keita Ikegami, Ryu Ogawa, Masaru Tomita, "Genome Projector: zoomable genome map with multiple views", BMC Bioinformatics, 査読有, Vol.10, 2009, 31.
5. Kazuharu Arakawa, Haruo Suzuki, Masaru Tomita, "Computational Genome Analysis Using The G-language System", Genes, Genomes and Genomics, 査読有, Vol.2, No.1, 2008, pp.1-13.
6. Yukino Ogawa, Kazuharu Arakawa, Kazunari Kaizu, Fumihiko Miyoshi, Yoichi Nakayama, Masaru Tomita, "Comparative study of circadian oscillatory network models of Drosophila", Artificial Life, 査読有, Vol.14, No.1, 2008, pp.29-48.

[学会発表](計9件)

1. Kazuharu Arakawa, Nobuhiro Kido, Kazuki Oshita, Masaru Tomita, "G-language Bookmarklet: オンライン生命情報リソースへのお手軽ゲー

トウェイ", 第13回オープンバイオ研究会, ポスター, 2010年3月6日, 石川.

2. Kazuharu Arakawa, Nobuhiro Kido, Kazuki Oshita, Masaru Tomita, "REST/SOAP Web Service API for G-language System", Genome Informatics Workshop, ポスター, 2009年12月14日, 横浜.
3. Kazuharu Arakawa, Masaru Tomita, "G-language Genome Analysis Environment Version 2: Integrated workbench for computational genome sequence analysis", Intelligent Systems for Molecular Biology, ポスター, 2009年6月29日, Stockholm, Sweden.
4. Kazuharu Arakawa, Nobuhiro Kido, Kazuki Oshita, Masaru Tomita, "Web Service Interface for G-language Genome Analysis Environment", Bioinformatics Open Source Conference, 2009年6月28日, 口頭, Stockholm, Sweden.
5. 荒川和晴, 富田勝, "仮説推論・検証型生命情報科学のためのソフトウェアとインタフェース - G-language・GEM System・Genome Projector・E-Cell 3D -", 第一回生命情報科学若手の会, 口頭, 2009年4月25日, 三島.
6. 荒川和晴, 木戸信博, 大下和希, 富田勝, "G-language Project:BioHackathon 2009のサマリー", 第11回オープンバイオ研究会, ポスター, 2009年3月28日, 石川.
7. 荒川和晴, 富田勝, "ゲノム解析ワークショップ G-language GAE v.2", BMB2008 (第31回日本分子生物学会年会・第81回日本生化学会大会 合同大会), ポスター, 2008年12月12日, 神戸.
8. Kazuharu Arakawa, Masaru Tomita, "G-language Genome Analysis Environment Version 2: Integrated workbench for computational genome sequence analysis", Genome Informatics Workshop, ポスター, 2008年12月2日, Gold Coast, Australia.
9. Kazuharu Arakawa, "E-Cell 3D: 3-Dimensional Visualization of Dynamic Cell Simulation", Bioinformatics Open Source Conference, 2008年7月18日, 口頭, Toronto, Canada.

〔その他〕

1. G-language Genome Analysis Environment
<http://www.g-language.org/>
2. GEM System
<http://www.g-language.org/gem/models/static.cgi>
3. Genome Projector
<http://www.g-language.org/GenomeProjector/>
4. Pathway Projector
<http://www.g-language.org/PathwayProjector/>
5. In silico model database
<http://ism.iab.keio.ac.jp/>
6. METAL System
<http://ws.g-language.org/metal/>
7. E-Cell Simulation Environment 3D
<http://ecell3d.iab.keio.ac.jp/>
8. REST Web Service for Copasi
<http://www.g-language.org/wiki/copasi>
9. REST Web Service for E-Cell
<http://www.g-language.org/wiki/ecell>

6. 研究組織

(1) 研究代表者

荒川 和晴 (ARAKAWA KAZUHARU)
慶應義塾大学・政策・メディア研究科・講師
研究者番号：4 0 4 5 3 5 5 0

(2) 研究分担者

()

研究者番号：

(3) 連携研究者

()

研究者番号：