

平成 22 年 5 月 12 日現在

研究種目：若手研究（スタートアップ）  
 研究期間：2008～2009  
 課題番号：20800025  
 研究課題名（和文） マルチスケールシミュレーションのための統合データベース  
 研究課題名（英文） Database Integration for Multi-scale Simulation  
 研究代表者  
 木戸 善之（KIDO YOSHIYUKI）  
 大阪大学・臨床医工学融合研究教育センター・特任助教（常勤）  
 研究者番号：70506310

研究成果の概要（和文）：生体機能を解明するためには、ゲノム、化合物、タンパク質、細胞、器官を統合したマルチスケールシミュレーションがフィジオーム・システムバイオロジーとして必要不可欠となる。しかし実現するためには各分野間での専門用語の差異や、単語の意味の齟齬、データの互換性などの問題がある。そこで本研究ではマルチスケールシミュレーションのためのデータベースの統合を目指し、MeSH Term や Gene Ontology などのオントロジーを用いた他分野とのデータベース統合、連携を図った。

研究成果の概要（英文）：Multi-scale simulation is mainly aimed at integrating scientific knowledge through computational methods and across different physiological levels; ranging from molecular level, genomic level, cellular level, organ level up to human body part level. In this study, we aim to integrate among some databases using ontologies such as Gene Ontology, MeSH Term and other related ontologies.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2008年度	1,010,000	303,000	1,313,000
2009年度	720,000	216,000	936,000
年度			
年度			
年度			
総計	1,730,000	519,000	2,249,000

研究分野：情報学

科研費の分科・細目：生体生命情報学・生命体システム情報学

キーワード：マルチスケールシミュレーション

## 1. 研究開始当初の背景

計算機の目覚ましい発展に伴い、高速な計算機を用いたシミュレーションや大量データの解析を行なうことが可能となり、医学、生物学、薬学など様々な分野での応用研究が期待されている。その分野の1つとしてフィジオーム・システムバイオロジーが挙げられ

る。フィジオームとは細胞、器官、生体といった生体の生理機能の総体を指し、フィジオーム・システムバイオロジーは、ゲノムからタンパク質、細胞内器官、細胞、臓器、個体に至るまで生体の構造と機能を統合的に解明する新しい生命・生体科学である。生体機能を解明するためには、個々のスケールでの

シミュレーションでは不十分であり、ゲノム、化合物、タンパク質、細胞、個体を統合したマルチスケールシミュレーションが必要不可欠となる。つまりフィジオーム・システムバイオロジーでは分子レベルから個体レベルに至るまで様々なスケールを研究対象としており、生理機能を統合的に理解するためにマルチスケールシミュレーションを可能にする基盤が必要となる。またマルチスケールシミュレーションを行うためのデータは各スケールで出そろい始めている。例えば、分子レベルの化合物立体構造データベース、遺伝子データベース、細胞生理学データベース、医療疾患データベースなどが公共のデータベースサイトから得ることができるが、各スケールのデータベースは異なるフォーマット、異なる語彙を用いており統合的な利用は困難である。

## 2. 研究の目的

本研究ではマルチスケールシミュレーションの基盤構築を目的とし、フィジオーム・システムバイオロジーにおけるシミュレーションで統合利用するための生体分子データを細胞の構成要素として分類したデータベースの構築を目指す。まずフィジオームで重要な機能を果たすチャンネル・タンパク質に注目する。チャンネル・タンパク質とは細胞の生体膜にある膜貫通タンパク質のことであり、イオンを透過させることで細胞、個体に影響するシグナルを伝達する。例えばチャンネル・タンパク質と化合物の結合による変化があげられる。化合物がチャンネル・タンパク質に結合し、イオンチャンネルが開口し、イオンが細胞内へと流入する。それにより細胞が様々な機能を発現することとなる（図1参照）

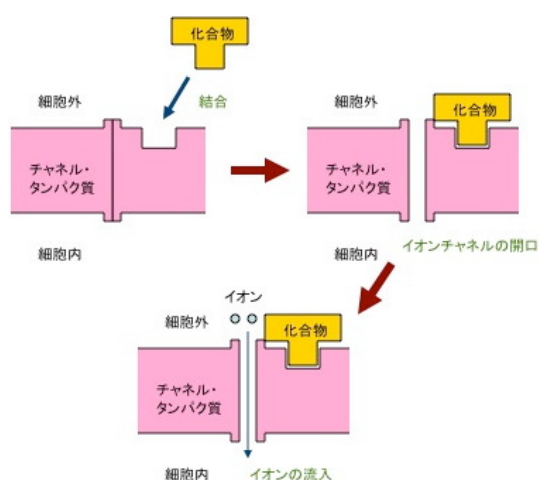


図1. 膜貫通イオンチャンネルタンパク質

このようにチャンネル・タンパク質にかかわる化合物、チャンネル・タンパク質の立体構造、イオンチャンネルの役割、細胞の生理機能変化

はそれぞれデータをデータベースから取得する必要がマルチスケールシミュレーションで求められている。そのためチャンネル・タンパク質の機能を既にある公共データベースから解き明かすために、代謝ネットワークデータベースや分子ネットワークデータベースを統合し、チャンネル・タンパク質の機能に関与する分子、化合物について構造や機能を検索できる統合データベースを目指す。また遺伝子やタンパク質のデータベースはインターネット上に公開され散逸しているため、ウェブサービスやグリッドコンピューティングの技術を用いた動的な検索ができるシステムを構築し、フィジオーム・システムバイオロジーのマルチスケールシミュレーションの実現に貢献する。

## 3. 研究の方法

フィジオーム・システムバイオロジーは、マルチスケールシミュレーションなどを用い、生理機能や疾患などへの影響を考察する学術分野である。そのため既存の研究分野を統合的かつ包括的に集約する研究分野でもある。本研究はマルチスケールシミュレーションの基盤を構築することが目的であるため、以下の課題をマイルストーンとして推進する。

- (1). チャンネル・タンパク質の機能に関連する分子の構造、機能を検索する統合データベースの設計
- (2). 統合データベース検索システムの構築およびマルチスケールシミュレーションツールとの連携

まず(1)に挙げたように、生理機能を解明するためのマルチスケールシミュレーションで重要なチャンネル・タンパク質の機能解明のためのタンパク質配列データベース(UniProt)とタンパク質立体構造データベース(PDB)、化合物(PubChem, LIGAND)と統合を試み、チャンネル・タンパク質から関連する化合物の機能、構造を検索するシステムを設計する。統合にはタンパク質や遺伝子の機能や注釈を体系立てて整理統合した語彙情報データベース、オントロジーを用いデータベース間を動的に統合するシステムを設計する。

次に、(1)で設計した統合データベース検索システムの実装を行い、チャンネル・タンパク質から関連する化合物の機能が検索できることを検証する。また別プロジェクトではあるが、申請者が従事している大阪大学臨床医工学融合研究教育センターのGlobal COEプログラムにて「医・工・情報学融合による予測医学基盤創成」において、マルチスケールシミュレーションで利用するための基盤として、生体の数理モデルと変数を表現するための ISML (in-Silico Markup

Language)と ISML を記述するツール ISIDE を開発している。(2)で構築した統合データベース検索の結果と ISML, ISIDE を連携し, チャンネル・タンパク質から関連する化合物, 細胞生理学の数理モデルを検索し, マルチスケールシミュレーションの基盤として検証実験を行い, 生理機能解明の基盤となりうることを明らかにする。

#### 4. 研究成果

本研究では, マルチスケールシミュレーションの基盤構築を目的とし, フィジオーム・システムバイオロジーにおけるシミュレーションにて利用するための異分野データベースの統合に関する研究を行った。本研究ではまず遺伝子およびタンパク質の関連に着目し, PDB を検索するために, 遺伝子関連語彙情報オントロジーである Gene Ontology への単語を機能アノテーションとした検索手法について開発を行った。PDB では既に Gene Ontology の語彙が付与されているが, 付与されていないエントリも数多くある。そのためタンパク質配列データベース UniProt を利用し, タンパク質配列とそれに対応した Gene Ontology の語彙の参照用データベースを構築し, 補完した機能アノテーションの妥当性について評価した (図2参照)。

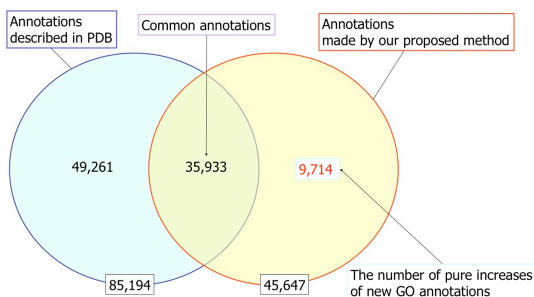


図2. 既存の PDB アノテーションと提案手法によるアノテーションの数の比較

また本研究ではマルチスケールシミュレーションで利用するためのデータ構造を定義する ISML の語彙情報, in-silico Ontology (ISMO) と複数のデータベースやオントロジーを統合検索するメタオントロジーの設計と提案を行った。メタオントロジーでは, 複数の言語体系であるオントロジーの共通項を探し出し, 関連する語彙を探索することを可能にする。これにより複数の分野にまたがった統合的なデータベース検索が可能となり, 例えばチャンネルイオンタンパク質に関連する遺伝子, タンパク質, 生体現象などの探索が可能になると考えられる。

さらに Resource Namespace Service (RNS) を利用した Web サービス, 数理モデルの探索を提案した。RNS ではリソース

(数理モデルや Web サービス) の物理位置情報を木構造データベースで保持し, ユーザは任意の名前とリソースを紐付けて管理することができ, ファイルシステムを操作するようにリソースを扱うことが可能となる。それにより複数のリソースについて位置情報を意識することなく, 扱うことができる。RNS を利用することでユーザは遠隔地のホスト, 自身のローカルホストのそれぞれのファイルを単一の RNS を参照することでデータを扱うことが可能となり, 広域分散環境でのユーザビリティ向上につながると考えられる。

また分散ファイルシステムを利用した画像情報, データの管理にも取り組んだ。超高圧電子顕微鏡から出力される高解像度画像ファイルとそのメタデータ (撮影日, 試料の詳細) を分散ファイルシステムで管理し, 組織にまたがったユーザ間でのデータ共有を行うためのシステムを構築した。データ共有の際, 最も重要視されるのはセキュリティである。仮想組織 (Virtual Organization) 内のユーザにはデータを共有し, 他の VO ユーザからはデータを隠蔽化する排他的処理を行うため, 分散ファイルシステムとして Gfarm-VO を利用した。また Gfarm-VO にはメタデータが付随するが, システム専用のメタデータとして機能しており, ユーザからは読み書きすることができない。そこで新たなユーザが読み書き可能なメタデータを付加し, 本システムに導入した (図3参照)。

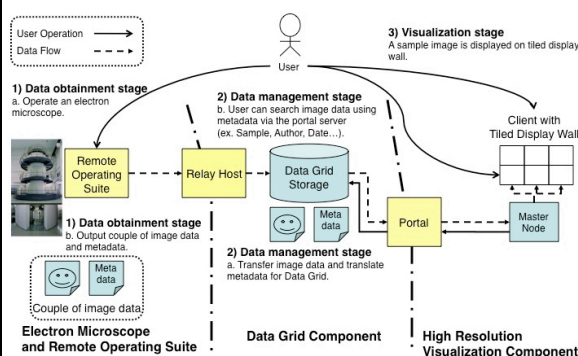


図3. システム概要図

#### 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3件)

- ① Eisaku Sakane, Yoshiyuki Kido, Susumu Date, Shinji Simojo, A Data Management and Visualization System towards Online Microscopic Imaging, Proc of MIPRO, 査読有, Vo. 1, 2009, pp139-148

- ② Junya Seo, Yoshiyuki Kido, Shigeto Seno, Yoichi Takenaka, Hideo Matsuda, A Method for Efficient Execution of Bioinformatics Workflows, Genome Informatics, 査読有, Vol.23, 2009, pp.139-148
- ③ Yoshiyuki Asai, Yasuyuki Suzuki, Yoshiyuki Kido et al. (員数: 10名中3番目), Specifications of insilico ML 1.0: A Multilevel Biophysical Model Description Language, The Journal of Physiological Sciences, 査読有, Vol.58, No.7, 2008, pp.447-458

[学会発表] (計7件)

- ① 木戸 善之, マルチスケールシミュレーションのための Web サービスとデータ探索に関する研究, 情報処理学会研究報告バイオ情報学研究会, 2010年3月5日, 石川
- ② Yoshiyuki Asai, Yoshiyuki Kido et al., insilicoML and integrated environment, XXXVI International Congress of Physiological Sciences, 2009年7月27日-8月1日, 京都
- ③ Yoshiyuki Kido, Discovery of relation between modeling data using in-silico modeling ontology and meta-ontology, XXXVI International Congress of Physiological Sciences, 2009年7月27日-8月1日, 京都
- ④ Masahiko Hamada, Yoshiyuki Kido et al., A Method for Making a New Reference Set of PDB Entries for Retrieving Protein 3D Structures with Structural Annotations, The 2008 Annual Conference of JSBi, 2008年12月15日, 大阪
- ⑤ Takashi Shimizu, Yoshiyuki Kido et al., A Method for Reducing Bounds of Compound Search by Dividing Structure Key, The 2008 Annual Conference of JSBi, 2008年12月15日, 大阪
- ⑥ Yoshiyuki Kido, A Biological Analysis System using Data-Staging with GridFTP, The 5<sup>th</sup> Korea-Japan Grid Symposium, 2008年10月10日, Jeju, Korea
- ⑦ 木戸 善之, ISMO: in-Silico Modeling Ontology の提案, 生体医工学シンポジウム 2008, 2008年9月19日, 大阪

## 6. 研究組織

### (1) 研究代表者

木戸 善之 (YOSHIYUKI KIDO)

大阪大学・臨床医工学融合研究教育センター・特任助教 (常勤)

研究者番号: 70506310

### (2) 研究分担者

なし

### (3) 連携研究者

なし