

細胞行動データベース CBDB: Cell Behavior Database

岩爪道昭*¹
Michiaki IWAZUME*¹

団まりな*² *³
Marina DAN*²

金子洋之*³
Hiroyuki KANEKO*³

浅利今朝男*⁴
Kesao ASARI*⁴

*¹ 情報通信研究機構
National Institute of Information
and Communication Technology*¹

*² 階層生物学研究ラボ
Hierarchical Biology Lab.*²

*³ 慶應義塾大学
Keio University*³

*⁴ (有)アズシステムズ
AZ Systems Inc.*⁴

A large amount of molecular level information on biological phenomena, such as embryogenesis, cell proliferation, cell differentiation, signal transduction, transcriptional regulation, cell response, etc. is rapidly increasing because of the progress of genome science in recent years. In order to understand high-order biological phenomena, it is necessary to bridge the gap between the amount of information in the molecular level and the amount of information in the cell functional level. However, the biological knowledge in the cell functional level has not been systematized well. We have been seeking a method to systematically describe and understand cellular behaviors in morphogenesis. Various morphogenetic processes were manually selected and described according to the following format: "cell(s)=(subject)" "does something=(output matter)" "under a certain condition=(input matter)". CBDB is a biological database on cell behaviors in the format with images and movies which indicate To sophisticate and make use of the CBDB, we propose an AI-based approach for more multiple searching of instances and knowledge discovery from the CBDB. As the first step, we attempted to construct the ontology for understanding morphogenetic phenomena. The ontology will also facilitate clearer definition of terms and concepts, and knowledge sharing among communities of various fields. In this study, we introduce the CBDB and the prototype of ontology-based information retrieval system.

1. はじめに

近年、ゲノム科学(ゲノム情報学)の進歩により遺伝子やたんぱく質など生命現象に関する膨大な情報が蓄積されつつある。生体分子間の相互作用を情報技術によって網羅的に解析することで、生命のよりよい理解はもとより、創薬、テラーメイド医療などの応用期待されている。しかし、現在蓄積されている情報は、意味が不明確で不均質なデータも少なくない[荒木 05]。また、分子レベルの知見の蓄積に対して、より上位である細胞や細胞集団の振る舞いや機能に関する知見は十分体系化されているとは言い難い[伊藤 01]。

我々は、多様な高次生命現象を有機的に理解するためには、生命の基本単位である「細胞レベル」に関する知識の体系的が必須である、という考えに基づき(図1参照)、形態形成現象における細胞の素過程の網羅的記述と体系化を試みている[岩爪 01]。その試みの一つとして、我々は「細胞行動」を記載する方法として「A細胞は、B の状況下で、C を行う」というフォーマットに基づき、学術雑誌や成書から個々の細胞行動を抽出し、現在約 1300 件のデータを持つ細胞行動データベースを構築している。本論文では、細胞行動データベースの概要と今後の展望について紹介する。

2. 細胞行動に関する知識の体系化

2.1 細胞レベル知識の体系化における課題

(1) 体系化の方法論が未確立

細胞レベルに関する知識を体系化における最大のボトルネックの一つは、そもそも細胞の視点から知識を記述する方法論や体系化された知識が確立していない点にある。流用可能な体系知やオントロジーもなく、ドメインエキスパートのコミュニティにおいてもさまざまな解釈が存在する。したがって、オントロジーをトップダウンに構築し、それに基づいて知識の記述・体系化するという方法は現実的でない。生のデータを収集しながら、試行

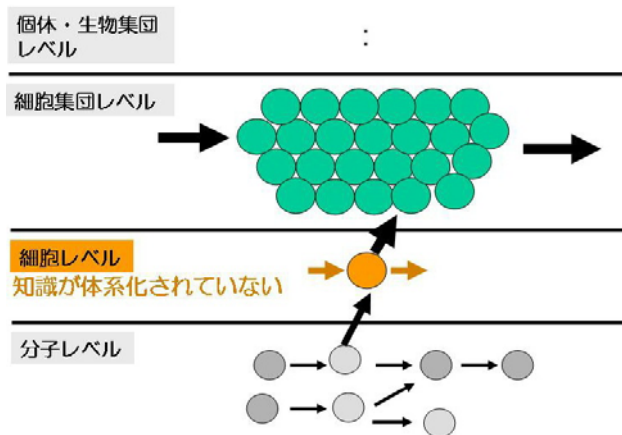


図1 細胞レベルに関する知識の体系化の必要性
錯誤的・漸進的に体系化を行うアプローチが必要となる。

(2) 知識の読み替えのボトルネック

もう一つの大きなボトルネックは、細胞レベルの視点で記述された知識が少ない点にある。現在、入手可能な情報(文献データ)の多くは、図1の最下層である分子レベルの視点から記述されており、細胞行動という観点から知識が明示的に書かれているものは非常に少ない。したがって、分子の視点で書かれた知識を、細胞の視点から読み替える(言い換える)必要がある。この読み替えには、高度な生物学の知識が必要なため、現在の自然言語処理技術では困難であり、ドメインエキスパートによる手作業に頼らざるを得ない。しかし、たった一人のドメインエキスパートによる作業は非効率であるため、研究コミュニティによる協調的な体系化作業が必要がある。また、協調作業を可能にするために、知識記述の指針を最低限共有する必要がある。

2.2 細胞行動素過程の記述フレームワーク

これらの問題を解決するために、我々は細胞の素過程を記述するための緩やかなフレームワークについて、試行錯誤しながら検討を行ってきた。以下では、これまでの試行錯誤の経緯を簡単に辿りながら、現在採用している知識記述方式について述べる。

連絡先: 岩爪道昭, 情報通信研究機構, 〒184-8795 東京都小金井市貫井北喜多町 4-2-1, TEL: 042-327-5295,
Fax: 042-327-560, e-mail: iwazume@nict.go.jp

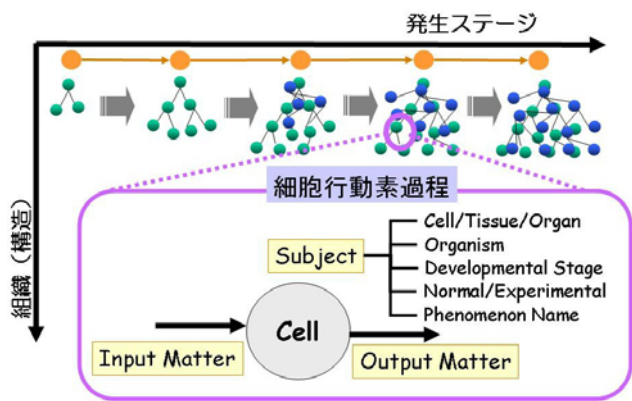


図2 細胞行動オントロジー

(1) Input/Output 表とその類型化の試み(第1段階)

形態形成現象と分子情報をつなぐキーとしての「細胞」の役割を表現する方法として、

「A (=Subject) 細胞は, B(=Input)という状況を感じて, C(=Output)を行う」

というフォーマットが提案され, 表形式のデータベースの構築が進められた[Kodama01, Iwazume01].

この作業にあたったドメインエキスパートのグループの関心が形態形成現象にあったため, 特に Input と Output の内容に興味の主軸がおかれ, それぞれの内容の分類項目を横軸と縦軸においたマトリックスが作られ, 細胞行動の類型化が試みられた[岩爪 01]. しかしながら, 「細胞」は決して1:1の因果律に従って行動するものでないため, Input と Output の内容の評価に多大な時間を要し, データの集積が進まないという問題が発生した.

(2) オントロジー工学的観点の導入(第2段階)

上記の問題を解決するために, 行動する主体としての「細胞」を, より広く, より単純な形で代えるためのフォーマットとして

「A 細胞は, B の状況下で, Cを行う」

という厳密性を問わない形式に修正された. また, オントロジー工学的な観点から, 細胞行動素過程の付帯事項として, Cell/Tissue/Organ (細胞名, 組織名), Organism(生物種), Developmental Stage (発生ステージ), Normal/Experimental (自然発生 or 実験条件下), Phenomenon Name (現象名)等, 最低限の情報を記述し, 形態形成過程の時空間構造(組織構造と発生過程)に位置づける, という方法を採用した(図2参照).

3. 細胞行動データベース

現在, 我々が構築を進めている細胞行動データベース

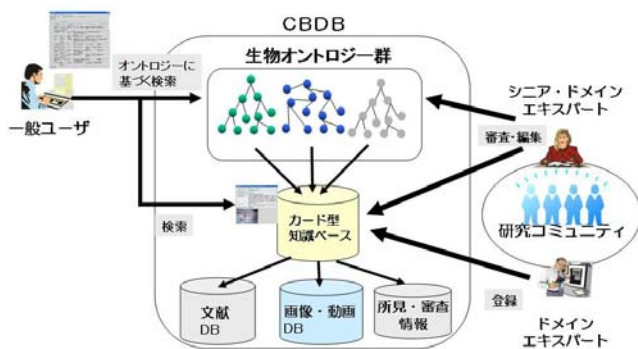


図3 細胞行動データベース概要

(CBDB: Cell Behavior Database)の概要を図3に示す(以下, CDDB と略記). 現在のCDDBの仕様は, (i)細胞行動素過程を記述したカード型知識ベース, (ii)収録知識の理解を補助する画像, 動画, 簡潔な説明文(テキスト), 文献情報, (iii)カード型知識間の構造化・統合化を行うオントロジー群(iv)レフリー(ベテランのドメインエキスパート)による審査・所見情報によって構成されている.

3.1 細胞行動素過程に関するカード型知識ベース

本研究では, 細胞行動素過程の収録するためにカード型知識と呼ばれる記述方式を採用している[久保田 03]. カード型知識とは, テキストや画像, 動画など異なるメディアレベルの情報断片から構成される知識メディア[Stefik86]の一種である.

知識ベース応用システムにおけるボトルネックは, ドメインエキスパートからの知識獲得と知識のメンテナンスであるが, カード型知識は, 知識工学に関するリテラシーを持たないドメインエキスパートでも知識入力容易であり, 上記のボトルネックを緩和するメリットがある.

一方, カード型知識ベースは, ネット上に流通しているロコモ情報や個人の日常知のような厳密性・無矛盾性を要求されないドメインでは有効な方法の一つであるが, 客観性が要求される科学知や工学知の体系化と利用においては, 完全にフリーな形式のカード情報では不十分である. そこで, 我々は, 2.2 節で述べた細胞行動素過程のフォーマットに基づくカード型知識を採用した(図4).

3.2 協調オーサリングおよびレフリー支援

従来のフォーマット(Input/Output 表)では, 博士レベル以上のドメインエキスパートが知識入力を行っていたが, カード型知識入力の導入より, 現行CBDBでは, 生物学の学部・修士レベルの学生から教員レベルのドメインエキスパートによって協調的に知識入力が行えるようになってきている. しかし, 約 1300 件のデータを収録していく過程で, シニアエキスパートから見ると細胞行動とはいえないものも収録される事例が見られるようになった. この問題への対応策として, シニアエキスパートによるレフリー制度を取り入れている. 各ドメインエキスパートが入力した知識は, レフリーによって所見・修正が加えられ, 入力者にフィードバックされる. 最終的にレフリーによって採択された知識のみが, 公開用のデータベースに登録されるようになってきている.

以上のようにCBDBは単なるデータベースではなく, 協調オ

慶応義塾大学 Cell - Behavior DataBase - Microsoft Internet Explorer					
ファイル(F)	編集(E)	表示(V)	お気に入り(S)	ツール(T)	ヘルプ(H)
1. ID	635				
2. Cell/Tissue/Organ	一次運動神経の MP 細胞				
3. Organism	ゼブラフィッシュ				
4. Normal/Experimental	E				
5. Developmental Stage	体筋形成期				
6. Input Matter	元々ある MP 細胞の隣に移植される				
7. Output Matter	元々ある MP 細胞と関係なく背側体幹筋に軸索を伸ばし支配する				
8. Explanation	背側一次運動神経の MP 細胞は, 各半体筋あたりに 1 個ある. MP 細胞は背側体幹筋の運動に軸索を伸ばし支配する. 赤い色素で標識された MP 細胞 (図) が元々ある MP 細胞 (図: 肌色細胞) に隣接して移植されているが, 移植によって 2 個の MP 細胞が同じ体筋内にある場合も, 異なる筋に注射することもある. 斜めの点線は体筋の境界を示し, 水平方向の点線は背側と背索の境界を示す.				
9. Reference	Eisen, J. S. and Melancon, E. (2001) "Interactions with identified muscle cells break motoneuron equivalence in embryonic zebrafish" Nature Neuroscience. 4: 1065-1070.				
10. Phenomenon Name	軸索伸張				

図4 カード型知識

ーサリングを支援するCMSと見ることが出来る。

4. 細胞行動に関するオントロジー構築と利用へ向けて

生物学分野では、現在様々なオントロジーが構築されてきている[Ashburner00, Rzhetsky00, Schulze-Kremer97, Stevens00, Takai-Igarashi00].

しかしながら、これらのオントロジーは、工学的なオントロジーの成果を生物にそのまま適用したものが多く、時間変化にともなう構造・基本性質の変化、状況依存的な分子や細胞の機能発現など生物特有の複雑な現象を記述するだけの方法論が確立しているとは言い難い[岩爪 01]. 例えば、個体発生過程では、当初は「1個の細胞=1個体」と見なせるものが、時間の経過とともに複雑さのレベルを上げ、「複数の細胞=1個体」なる[団 96, 団 99]. このような過程を記述するためには、1つのビューによる静的な構造の記述だけでは不十分である。現在、我々はデータの拡充と平行して、以下の課題に取り組んでいる。

4.1 細胞行動に関するオントロジーの構築

細胞行動現象を明示的に記述し、体系化するためには、オントロジーの整備は不可欠である。しかし、現時点では細胞行動の視点にたった基本語彙の整理は十分になされているとは言えない。我々は、現在の約 1300 件のデータの中から、比較的数量多くのデータが収録されている脊椎動物と棘皮動物の種名検索を完遂できる種名オントロジーの構築を進めている。

4.2 多角的・横断的・発見的検索手法の開発

収載された多数のカード型知識群において、内容的に関連があるものを多角的・横断的に検索するための手法や知識ベース内に暗黙的に含まれている新たな知識を発見するための手法についても検討を進めている。

4.3 データベース連携と細胞行動バイオセマンティックウェブ構築

他のバイオデータベース、バイオオントロジーとの連携・相互運用を想定した知識の形式化・記述方式を検討している。また、近年注目されているセマンティックウェブ技術の動向をも踏まえ、細胞行動に関するバイオセマンティックウェブの構築も検討している。

5. 考察と今後の課題

5.1 現行CBDBの用途

研究への応用: 細胞が自殺するアポトーシス現象の収録データが実際の研究にどのように活用され得るかドメインエキスパートからの所見を得た。その結果、研究遂行のための基礎知識の習得のみならず、新たな研究課題の設定や、実験計画の立案にも活用できることが判明している。

教材としての応用: 動物の発生過程における誘導現象を例に、中学、高校レベルから大学教養レベル迄の教育現場で試験的に利用を進めている。画像や動画を含む知識カードが学生たちの興味を直截的に引き起こすことが明らかになった。

教員への支援: 複数の高校教諭との議論を通して、CBDBが、教科内容への疑問の解決や、教科内容を補強するための最新学問情報の取得に、大きな可能性を持つことが示された。現在、上記の用途に関して、評価実験を計画している。

5.2 今後のロードマップ

(1) 4000 件の画像付きデータの収録

「高次生命現象理解のための細胞行動データベース(現在約 1300 件収録)」を 4000 件の知識ベースに成長させることをとおして、内容の充実を図りながら、一般公開できる網羅性を獲得する。そのために、従来行っていたランダムなデータ収集に加え、全ての細胞の体内における位置取りや、細胞行動として捕らえうる限りでの細胞機能の収録を行う。

(2) 500 件の動画データの収録

細胞行動を最もヴィジュアルに表示するものとして、動画データを 500 件収録する。各データ当たり、0.5 - 2 分間の短編映画とし、ナレーション付きの動画をコンピュータ上で提示できるように加工する。

(3) 枠組みの拡大への取り組み

これまで細胞行動の枠外として収録を見合わせていた例(例えば、興奮性細胞が活動電位を発生する現象や、発生過程の細胞が示す領域による遺伝子発現など)を収録することにより、狭義の「細胞行動」だけでなく、「細胞の電気生理学的性質」や「細胞の認識能」などに関する現象の明示化を図る。

(4) 日・英二言語対応

これまで日本語で作成していたデータを英語に翻訳することで、細胞行動知識ベースの英語バージョンを構築し、既存のデータベースとの連携を図る。

6. まとめ

細胞レベルに関する知識の体系化の試みとして、細胞行動データベースの概要と今後の展望について紹介した。生物系の情報データベースとしては、ある生命現象に限定されたもの(細胞内シグナルデータベース、DNA データベース等)、生物種に限った分類型のもの、特定の研究室で得られた実験結果を収録したもの等が存在している。それらに対し、CBDBは、広範な高次生命現象の基礎となる細胞行動を明示的に記述し、体系化・共有化するための新たな試みである。基本的な方法論や枠組みはほぼ確立しており、本データベースが知的資産として研究および教育に効果を発揮できるとの感触を得ている。今後は、漸進的なデータの拡充とオントロジーの構築を進める予定である。

謝辞

本研究は、私立大学等経常費補助金特別補助による「高次生命現象理解のための細胞行動データベース」の一環として行われたものである。

参考文献

- [荒木 05] 荒木, 川本, 小林, 永田, 出宮: バイオポータルプロジェクトにおけるオントロジー構築と利用について, 雑誌名, 人工知能学会全国大会, 2005.
- [伊藤 01] 伊藤隆司: ゲノムからインタラクティブへ, 細胞工学, Vol.20, No1, 30-37, 2001
- [岩爪 01] 岩爪, 団, 本田, 門谷, 金子, 児玉, 望月, 中島, 田中: "個体発生オントロジー——形態形成現象における細胞の振る舞いの明示的記述と理解への試み——" 2001 年度人工知能学会全国大会, 2B2-06, 2001.

- [金子 04]金子洋之, 団まりな “細胞行動データベース(2) – 投稿データの審査と内容的特徴 –” 日本動物学会第 75 回大会(神戸) 2004.
- [久保田 03] 久保田, 黒橋, 西田:知識カードを用いた分身エージェント, 電子情報通信学会論文誌「ソフトウェアエージェントとその応用論文特集」, volJ86-D-I, No.8, pp.600-607, 2003.
- [団 96] 団まりな: 生物の複雑さを読む, 平凡社, 1996.
- [団 99] 団まりな: 生物現象の諸階層構造～上位階層の新機能はどのようにして生じるのか, 第 13 回人工知能学会全国大会特別講演, 1999.
- [Ashberger00]1. M. Ashberger et al., “Gene Ontology: tool for the unification of biology. Nature Genetics” , Vol.5, 25-29, Nature America Inc. , 2000.
- [Stefik86] Mark Stefik. The next knowledge medium. AI Magazine, Vol. 7, No. 1, pp. 34-46, 1986.
- [Stevens00] R. Stevens, C. A. Goble and S. Bechhofer, “Ontology -based Knowledge Representation for Bioinformatics”, Journal Briefings in bioinformatics, 2000.
- [Kodama01] R. Kodama et al.: Repertoire of cell behaviors in morphogenesis: (I) What do cells sense and how do cells respond?, ICDB2001, 2001.
- [Iwazume01] M. Iwazume et al.: Repertoire of cell behavior in morphogenesis: (II) AI-based method for knowledge sharing, ICDB2001, 2001.
- [Schulze-Kremer97] Schulze-Kremer: Adding Semantics to Genome Database: Towards an Ontology for Molecular Biology , the Proc. of ISMB97, 1997.
- [Stevens00] R. Stevens, C. A. Goble and S. Bechhofer: Ontology -based Knowledge Representation for Bioinformatics, Journal Briefings in bioinformatics, 2000.
- [Rzhetsky00] A. Rzhetsky et al., A Knowledge Mode for Analysis and Simulation of Regulatory Networks, the Third Annual Workshop on Bio-Ontologies, 2000.
- [Gennari00] J. Gennari: Ontology for Knowledge Sharing in Neuroanatomy, the Third Annual Workshop on Bio-Ontologies, 2000.
- [Takai-Igarashi00] T. Takai-Igarashi and T. Takagi: Cell Signaling Ontology, the Third Annual Workshop on Bio-Ontologies, 2000.