

# 微生物培養培地オントロジーの開発

## Development of Microbial Growth Medium Ontology

川島 秀一<sup>\*1</sup>  
Shuichi Kawashima

岡本 忍<sup>\*1</sup>  
Shinobu Okamoto

<sup>\*1</sup> ライフサイエンス統合データベースセンター  
Database Center for Life Science

In biomedical research field, diverse and large amount of data has been produced, which has been resulted in lots of databases in the field. Database Center for Life Science in Japan has focused on building infrastructure to enable an effective use of such data in an integrated way by using Semantic Web technology. As part of the efforts, we have been developing Growth Medium Ontology (GMO), which describes nutritional and other components of growth media used for growing microorganisms. We have also created a linked open data of growth media based on the GMO. This paper presents the newly developed resources and discusses how to utilize them in integration with other life science-related LODs.

### 1. はじめに

大学共同利用法人情報システム研究機構 ライフサイエンス統合データベースセンター(以下 DBCLS)は、生命科学分野に蓄積されてきた大量のデータについて、セマンティックウェブ技術を活用することで、利用者がより効率的にデータを活用できる環境の構築を目指している。具体的には、既存データの RDF 化や、そのために必要なオントロジーの整備、RDF データ運用システムの開発等を行って、この目標を実現しようとしている [山口 2010]。

近年、国際的にも、生命科学系データベース(以下 DB)を RDF で提供しようとする動きが増えてきている。タンパク配列の DB である UniProt [The UniProt Consortium 2013] は、2008 年から RDF 版の公開を開始しており、2011 年にはタンパク質立体構造のリポジトリ DB である PDB の RDF 化がなされた。2013 年には、欧州バイオインフォマティクス研究所 (EBI) が、生物学プロセスの計算機モデルを収集した DB である BioModels、実験に使ったサンプルに関する DB である Biosamples、薬の生物活性の DB である ChEMBL、遺伝子発現情報の DB である ExpressioAtlas、分子パスウェイの DB である Reactome の 5 つのデータベースを RDF 化し公開した<sup>\*1</sup>。2014 年には、アメリカ国立生物工学情報センター (NCBI) が、公開されているものでは世界最大級の化学化合物データベース PubChem の RDF 版を公開した<sup>\*2</sup>。さらに現在、EBI は、ゲノムのデータベース Ensembl の RDF 化を準備中である。DBCLS でも、原核生物に関して NCBI のゲノム DB である RefSeq の情報を RDF 化を行っている。また DBCLS では、DB の RDF 化だけでなく、これらの利用可能な RDF データを統合して構築した、新しいゲノム情報検索サービスである TogoGenome<sup>\*3</sup> を構築してきた。現在、TogoGenome では、原核生物をターゲットにして、RefSeq のゲノム情報 RDF を中心として、適宜、UniProt のような既存 RDF データが統合されている。さらに、構造化された形で情報が利用できない対象についても、原核生物のゲノム情報に統合することで、有益な検索が行えると判断したものについては、共同研究者とともに、新規にオントロジーの開発や、RDF データの構築を行ってきた。具体的には、これまで、原核生物の生育環境に関

わるオントロジーや、表現型に関わるオントロジー等を開発し、これらは TogoGenome のファセット検索に活用されている。

本稿では、このゲノム配列を中心とした生物学データの統合の一環として我々が取り組んでいる、微生物の培養培地を RDF 化する際に必要となる、オントロジーの開発について報告したい。

### 2. 微生物培養培地のオントロジー構築

#### 2.1 微生物培養培地

微生物を研究するには、まず研究材料としての当該微生物を研究室で培養できることが必要になる。微生物を培養するには、培地と呼ばれる、微生物の生育に必要な栄養源が含まれる液体、またはそれを寒天などで固めたゲル状固形物が用いられる。微生物に必要な栄養素は微生物種によって異なるため、種に応じて様々な培地が存在している。培地は、合成培地と天然培地に大別することができる [駒 2011]。前者は、培養に必要な栄養源が、すべて化学薬品からなる培地である。合成培地は、化学組成が明らかでないため、そこで増殖する微生物に必要な栄養の十分条件は明らかとなる(必要条件とは限らない)。このような培地は、比較的栄養要求性の低い、すなわち、限られた栄養素を用いて、自らが増殖するのに必要な物質を全て合成できる代謝系を備えた微生物の培養に用いられる。一方、後者は肉エキスや、タンパク質を酵素で分解して作成したポリペプチドの集まりであるペプトン等が添加されている培地である。天然培地は、その成分が明らかではないが、多様な栄養素を含んでいることから、様々な微生物の培養に応用可能である。もちろん、栄養が過多であればいいというわけではなく、不要な栄養が含まれると増殖を阻害することもあるので、微生物毎に、増殖に適切な成分が含まれた培地を使用する必要がある。

#### 2.2 培養培地をオントロジー化する意義

このように、培地成分は、微生物が環境中からどのような栄養を吸収して増殖するかについての手がかりを与えてくれる貴重な情報である。吸収された栄養源は、様々な代謝経路を経て、生命活動を維持するために必要な分子が合成される。代謝経路に関する情報、すなわち、細胞内で酵素により分子がどの

連絡先: 川島秀一

ライフサイエンス統合データベースセンター 〒277-0871 千葉県柏市若柴 178-4-4 東京大学柏の葉駅前研究棟 6 階

<sup>\*1</sup> <http://www.ebi.ac.uk/rdf/>

<sup>\*2</sup> <https://pubchem.ncbi.nlm.nih.gov/rdf/>

<sup>\*3</sup> <http://togogenome.org/>

ように変換されていくかという情報については、様々な DB が開発されてきた。しかし、その代謝経路へ渡される分子の情報、すなわち研究室においては培地の成分情報については、これまで RDF 化はもちろん、成分での検索が行えるレベルでの電子化も行われてこなかった。従って、培地情報を RDF 化することは、これまでできなかった、培地の成分情報と、ゲノム情報や、代謝経路情報、表現型情報を統合した検索を可能にし、生命現象の理解を深める上で有用であると考えられる。

培地そのものは、名前のついた一部の有名なものを除くと、その都度目的に応じて成分が調整されて作られるので、オントロジーの概念として整理するには適していない。それよりも、我々は、培地作成に利用される物質や、どういった生物の培養に利用されるのか、といった情報を記述する OWL オントロジー (GMO: Growth Medium Ontology) を構築し、そのオントロジーを利用して培地の情報の RDF 化を行った。

## 2.3 培地成分の語彙の整理

培地成分は文献等にも記述されているが、今回、製品評価技術基盤機構 NBRC および理化学研究所 JCM から、それぞれウェブでも公開されている培養培地の情報を提供頂けたため (それぞれ、649 および 917 培地)、そこから成分を抽出し、マニュアルで統制語彙化を行った。化学化合物に関しては、PubChem や ChEBI 等、かなり網羅的なデータベースが公開されており、全ての培地成分が化学化合物だけであれば、その URI を使うことで培地成分を記述する RDF を作成することができるため、新たなオントロジーを構築する必要性は少ない。

しかし、

- 実際の培地の成分には、PubChem や ChEMBL 等にも記載されていない化合物が少なからず存在する
- 培地に利用される天然物には、肉エキスやペプトン等よく利用されるもの以外にも、トマトジュースや乾燥エビ等様々な材料が含まれ、そういった多種多様な概念に対応させられる既存の URI が存在しない
- 多種多様な概念に分類しうる成分を培地成分としての観点から、階層的に分類したい

等の理由から、新規にオントロジーを構築することにした。また化学化合物名は一般に多数の同義語が存在し、さらに化学式で書かれている場合もあるため (例えば、Sodium phosphate と  $\text{NaH}_2\text{PO}_4$ )、機械的に名寄せすることは困難なため、すべてマニュアルで統制語彙への整理を行った。成分の中には、商品として売られている培地等、複数の物質からできているものもあるが、こういったものもひとつの概念として、クラスを定義した。この結果、現在、培地成分を、653 の概念 (OWL クラス) として整理した。各クラスに、PubChem, ChEBI, Wikipedia, SNOMED-CT, MESH term 等で該当する URI がある場合、rdfs:seeAlso プロパティを用いて、その URI を記述した。

## 2.4 オントロジーの階層構造

今回、整理した培地の語彙をいくつかの階層に整理した。大きな分類として、化学化合物のように成分が明確に定義されるもの (Defined component) と、天然物のように成分が不明なもの (Undefined component) に大別した。また、水関係と、寒天 (Agar) 関係や pH 調整のための緩衝液等は、栄養としての成分ではないため、別の階層に分類した。図1に、培地の成分を表す概念 Component 以下の階層図を示す。

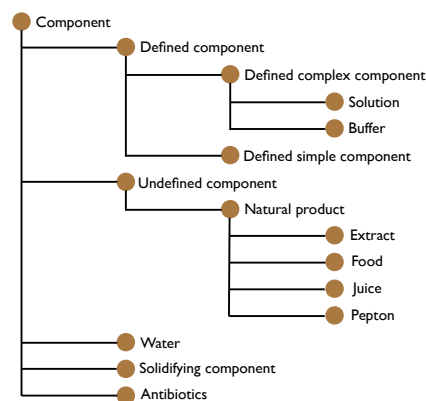


図1 GMOにおける培地成分に関する概念階層

## 2.5 培地 RDF データ

次に、統制語彙化した培地成分を利用して NBRC で提供頂いた、649 培地の RDF 化を行った。この RDF データは TogoGenome にインストールされており、生物種を検索した際、関連情報の一つとして可視化されている。GMO で全ての成分を、Defined または Undefined component の下位概念として定義したため、それぞれの培地を合成培地 (全ての成分が Defined component に分類されるもの) か天然培地 (1 つ以上の成分が Undefined component に分類されるもの) に分類することができ、その情報も含めて RDF 化を行った。

## 3. まとめ

今回、微生物培養培地の成分を記述するオントロジーを構築し、649 種の培地について RDF 化を行った。さらにこれらを TogoGenome にインストールすることで、培地の栄養成分情報とゲノム情報や微生物表現型情報等と RDF として統合したシステムが構築できた。今後は、構築した RDF グラフを利用し、培地の栄養源から、生物の代謝ネットワークを介して、遺伝子情報へとつながる知識を発見するシステムの構築を計画している。

## 謝辞

本研究は、DBCLS のメンバーおよび、JST NBDC 統合化推進プログラム「ゲノム・メタゲノム情報を基盤とした微生物 DB の統合」(代表 黒川顕東工大教授)メンバーの方との共同作業のなかでなされたものです。また、培養培地の情報は、製品評価技術基盤機構 NBRC および理化学研究所 JCM からご提供頂きました。関係者の皆様に感謝の意を表します。

## 参考文献

- [山口 2010] 山口敦子, 片山俊明: 我が国のデータベース構築・統合戦略(第2回) データベースを統合利用するための基盤としてセマンティックウェブ技術, 細胞工学 Vol. 30, No.11, pp. 1210-1215, 学研メディカル秀潤社, 2011.
- [The UniProt Consortium 2013] The UniProt Consortium: Update on activities at the Universal Protein Resource (UniProt) in 2013, Nucleic Acids Research Vol. 41(Database issue), pp D43-D47, 2013.
- [駒 2011] 駒大輔, 山中真人, 森芳邦彦, 大本貴士: 培地の成分知っていますか?, 生物工学 Vol. 89(4) 195-199, 2011.