

## 研究課題別評価書

### 1. 研究課題名

分子スケール差を統合する代謝シミュレーション

### 2. 氏名

有田 正規

### 3. 研究のねらい

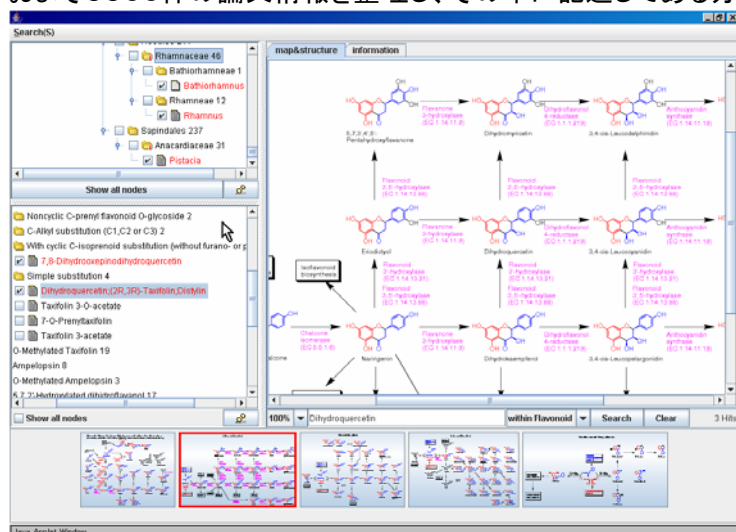
低分子化合物から脂質、糖鎖、二次代謝物といった中—高分子化合物まで、シームレスに物質変換経路を追跡できる機能を備えた電子代謝マップを提供する。

### 4. 研究成果①

#### ◇ フラボノイドデータベースの作成

フラボノイドとは植物に特有の二次代謝物のグループである。蕎麦に多いルチン(rutin)、お茶のカテキン類(catechins)、ブルーベリーやシソの色素であるアントシアニン類(anthocyanins)、大豆のイソフラボン類(isoflavones)、機能性飲料に入っているフラバン類(flavans)はいずれもフラボノイドに属する。フラボノイドの総数は1万種近くあるが、それぞれはどの植物に含まれ、どんな効能があるのだろうか。また効能を持つフラボノイドは互いに関連しているのだろうか。こうした科学的事実を分子構造に基づいて把握するためのデータベースを設計、作成した。現在収録する分子数は約 7000 種であり、その規模と精度は他に類を見ない。ただし、フラボノイドの重合体およびネオフラボノイドと呼ばれる分子種は登録していないため、世の中に知られているフラボノイドを網羅しているわけではない。以下にデータベースの内容を具体的に記す。

データベースは、化学構造とそれに付随する情報、植物種とそれに付随する情報、そしてどの分子がどの植物において同定されたかという文献情報の3カテゴリーから構成される。構造情報については、全構造に階層的分類を施し、大きく分けて母核となる構造、水酸基の位置、水酸基以外の修飾という3階層で表現した。この分類はすべて人手による検証を必要とするが、分類自体をすべて手作業で行うことは非常に煩雑で精度の低下を招く恐れがある。そこで分類作業を円滑にするためのソフトウェアツール作りから作業を開始した。作成したツールは化学分子それぞれをファイルとして扱い、ファイル集合をディレクトリ構造間で移動させることにより階層分類に仕上げていく仕組みになっている。このツールのおかげで、プログラミングやコンピュータになじみがない人でも数千規模のファイル集合を、分子構造を参照しながら分類、ラベリングすることが可能となった。植物種の情報については、およそ1600の植物種とそれらの系統分類をウェブ上の複数リソースを整理統合し、各生物種の上位概念を和英の両表記でまとめあげた。文献情報として、およそ5500件の論文情報を整理し、その中に記述してある分子と生物種の対応を整理した。



上記の結果は代謝マップをリンクさせた可視化ソフトウェア、フラボノイドビューワを用いて公開している。

(<http://metabolome.jp/software/>)ただしおよそ 7000 分子全てを網羅する代謝マップを描くことは不可能なため、代表的なマップをいくつか用意しておき、マップに無い分子は適当なマップに自動的にリンクする機能を用意する。

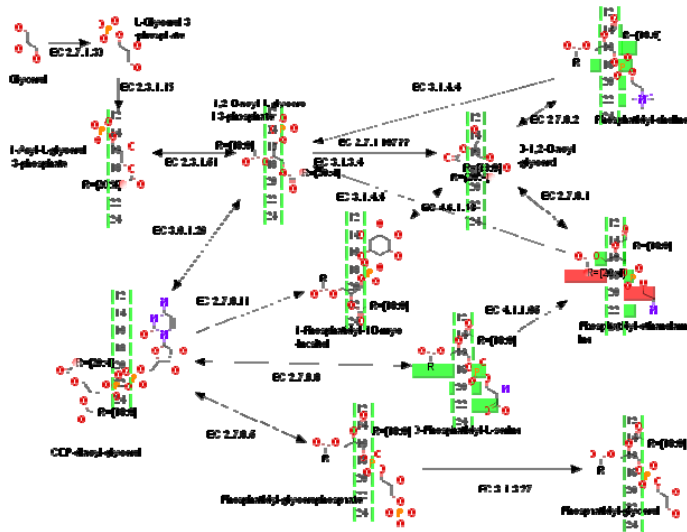
#### ◇ 脂質、糖質データベースの作成

フラボノイドは既存の良いデータベースが存在しなかったために独自に作成をおこなったが、脂肪酸や糖鎖は状況が異なる。例えば、日本脂質生化学会のメンバーが構築、維持してきたLipidBankは10年の歳月をかけて7000近くの脂質分子構造を網羅した価値ある研究資源である。しかし本研究の観点から見直すと、まだ改良すべき点が多く残されていた。例えば当初は全カテゴリーの分子に対して検索をする機能がなかった。これはデータベースの構造が28カテゴリーそれぞれで異なっており、相互に関係がないことが原因であった。また、分子量やLipidBank-IDによる検索も実現できていなかった。ウェブデータベースが全盛の時代にこのままでは、有用なデータを研究に活かすことができないのが現状である。そのため、日本脂質生化学会に参入して脂質データベース構築委員会を結成、さきがけ研究のテーマとしてLipidBankの全データを再整理、データベースの構造自体を一新する作業を進めた。具体的には、文献情報の検証、構造情報、名前情報の検証、データベースシステムそのものの改良を行ったため、データベースを新規に作成することに相当する労力が必要だった。以下にデータベースの内容を具体的に記す。

データベースは構造情報と文献情報の2カテゴリーから構成され、構造情報に付随して増すスペクトルや生理活性等の情報が付加されている。構造情報はすべて画像として格納されていたため、今回この情報をすべてMOLフォーマット形式で描きなおし、同時に複数の辞書やCASデータベースを利用して、構造と名前の再確認をおこなった。既に糖脂質や巨大な分子構造を除くおよそ4000近い分子構造の描き直しを終了したが、構造が一意に決まらない場合、検証する文献やデータが入手できない場合は修正が不可能であった。またおよそ3000件登録されていた文献は大幅に重複があり、入力者によって異なる様々な書式で記述されていた。これをすべて統一し、PubMed番号をつけられるものにはすべてウェブリンクを付加することでおよそ3000件の文献を整理した。以上の改良を施したLipidBankのアップデート版は、2007年度より日本脂質生化学会の公式データベースとして採用され、<http://lipidbank.jp/>において公開している。また学会内にデータベース構築委員会というグループを作り、その中心メンバーとしてアップデートに取り組んでいる。

#### ◇ 描画エディタと脂質分子発現量の視覚化ツール

多くのユーザはデータベースを眺めた後で、興味ある部分の経路検索や自分のデータの重ね合わせをおこないたい。またその結果を代謝マップ上で表示したい。そうしたニーズを満たすために、代謝物の量や分子種を表現できるブラウザを開発した。例えば、リン脂質クラス特異的な同定法により検出されたマウス網膜の質量分析スペクトラムを、作成した代謝マップのブラウザで表示させた様子は以下ようになる。この代謝マップではフォスファチジル基につく2種の長鎖脂肪酸のパターンがヒストグラムで示されており、それらの組み合わせに対応する分子種の発現量が色で示されている。2種の脂肪酸分子種が構造によって変化していることが一目でわかる。こうした表示はユーザが



選択した脂肪酸のパターン毎にオンデマンドで表示できる。現在は個々の分子からLipidBankのデータにリンクを張る作業をおこなっていないが、データベース側は既に稼働しているため、それほど労力をかけずにリンクを張ることは可能である。現在はグリセロ脂質の代謝マップについて作成してあるが、フラボノイドを含む他の代謝マップにも拡張可能である。

## 5. 自己評価

今回、データベースの作成作業に思った以上の時間や労力を割いてしまい、マップに描いたパスウェイの代謝量を推論する機能を汎用的な形で実装することができなかった。代謝マップエディタ開発の難しさを改めて実感した。エディタとして高機能を目指すのと同時に、どうすればユーザ自身が代謝マップを描く気になってくれるのか試行錯誤を重ねたが、個人研究で達成できる領域を少々超えていた感も否めない。しかし、少なくとも以下の項目については近いうちに実現する。

### ◇ フラボノイドと脂質データベースの統合

現在、別の研究予算で MassBank (<http://massbank.jp/>)という質量スペクトルデータベースを構築しており、これと今回のデータベースを相互にリンクさせた統合データベースを完成させる。単一のデータベースで、物理化学的性質、文献情報、生合成経路、質量スペクトルをアクセスできるようになれば当初の計画内容である分子スケール差の克服につながる。

### ◇ シミュレーション

現在、脂質全体や糖鎖全体の代謝マップを用意できていないため、個々の原子をシームレスにトレースする機能が実現されていない。現在、ユーザが描いたポストスクリプトファイルから座標を取得する機能を開発しているため、近いうちに通常のお絵かきエディタで描いた代謝マップに代謝物情報等を重ね合わせる機能が実現できるはずである。

## 6. 研究総括の見解

有田研究者は、低分子化合物から脂質、糖鎖、二次代謝物といった中—高分子化合物まで、シームレスに物質変換経路を追跡できる機能を備えた電子代謝マップの構築を行った。これまでユーザがウェブ等で物質の代謝の仕組みを検索する場合、非常に専門的な結果や健康食品等の結果が検索され適当な結果が見つからないことが多かったが、本研究により脂質やフラボノイド等の二次代謝物の検索が代謝マップを介して簡単にかつ使いやすいデータベースシミュレーションシステムを構築した。また日本脂質生化学会の脂質データベース構築委員会を結成し、脂質データである LipidBank の全データを再整理、データベースの再構築を行ったことを高く評価する。

以上のようにメタボロームという分野を広め、国内の研究組織をまとめるデータベースの構築を行ったが、今度はこれら研究成果の広報にも注力し多くのユーザ利用が実現するよう研究を進めてほしい。

## 7. 主な論文等

### A さきがけの個人研究者が主導で得られた成果

#### (1) 論文(原著論文)発表

1. Arita, M. and Tokimatsu, T. "Detection of monosaccharide types from coordinates" Proceedings of the 18th International Conference on Genome Informatics, 3-14, 2007.

#### (2) 特許出願

なし

#### (3) 受賞

なし

#### (4) 解説論文、書籍

1. Arita, M., Fujiwara, Y. and Nakanishi, Y. "Map Editor for the Atomic Reconstruction of Metabolism (ARM)" Chapter II.3 In Plant Metabolomics (Eds. K Saito, RA Dixon, L Willmitzer) Biotechnology in Agriculture and Forestry Vol.57, Springer Verlag, 129-140, 2006.
2. Arita, M., Robert, M. and Tomita, M. "All systems go: launching cell simulation fueled by integrated experimental biology data." Current Opinion in Biotechnology, 16(3), 344-349, Elsevier, 2005.
3. Arita, M. "Scale-freeness and biological networks." Journal of Biochemistry, 138, 1-4, Oxford Univ Press, 2005

(5)招待講演 (国際のみ)

1. Arita, M. "Software Tools to Study Flavonoids", Systems Biology and Bioinformatics Symposium (SBBS07), March 27-29 Taipei Taiwan, 2007.
2. Arita, M. "Databases for Flavonoids and Measurement in Arabidopsis", 2nd International Symposium on Environmental Metabolomics, March 24-25, University of Louisville(KY), USA, 2007.
3. Arita, M. "Map Editor for the Atomic Reconstruction of Metabolism", 1st Louisville Symposium on Environmental Metabolomics, November 5-6, Louisville (KY) USA, 2005.

B その他の主な成果

(1)論文(原著論文)発表

1. Horai, H., Arita, M., Nishioka, T. "Comparison of ESI-MS Spectra in MassBank Database" Proceedings of the International Conference on BioMedical Engineering and Informatics (BMEI2008), Hainan China, 2008 accepted
2. Okada, K., Asai, K., Arita, M. "Flow Model of the Protein-protein Interaction Network for Finding Credible Interactions." Proceedings of 5th Asia Pacific Bioinformatics Conference (APBC2007), Hong Kong, 2007
3. Kusano, M., Fukushima, A., Arita, M., Jonsson, P., Moritz, T., Kobayashi, M., Hayashi, N., Tohge, T., Saito, K. "Unbiased characterization of genotype-dependent metabolic regulations by metabolomic approach in Arabidopsis thaliana" BMC Systems Biology, 1 :53, 2007
4. Tuji, H., Altaf-Ul-Amin, Md., Arita, M., Nishio, H., Shinbo, Y., Kurokawa, K., Kanaya, S. "Comparison of Protein Complexes Predicted from PPI Networks by DPCLus and Newman Clustering Algorithms" IPSJ Transactions on Bioinformatics, 47 (SIG17), 31-41, 2006 (B)

(2)特許出願

なし

(3)受賞

なし

(4)解説論文、書籍

1. Shinbo, Y., Nakamura, Y., Altaf-Ul-Amin, M., Asahi, H., Kurokawa, K., Arita, M., Saito, K., Ohta, D., Shibata, D., Kanaya, S. "KNApSACk: A Comprehensive Species-Metabolite Relationship Database." Chapter II.6 In Plant Metabolomics (Eds. K Saito, RA Dixon, L Willmitzer) Biotechnology in Agriculture and Forestry Vol.57, Springer Verlag, 165-184, 2006

(5)招待講演 (国際のみ)

1. Arita, M. "Metabolic Pathway Chart for Flavonoid", 2nd Taiwan-Japan Bilateral Symposium on Bioinformatics, Nov 7-9, Tainan Taiwan, 2006

(6)国内発表

14件