

機関番号：12601

研究種目：特定領域研究

研究期間：2005～2010

課題番号：17017006

研究課題名（和文）生体メカニズムの物質変換構造を明らかにするための相互作用マップ

研究課題名（英文）Interaction map to elucidate structural transformations in biological systems

研究代表者

有田 正規（ARITA MASANORI）

東京大学・大学院理学系研究科・准教授

研究者番号：10356389

研究成果の概要（和文）：

基礎代謝物、脂質、植物二次代謝物（フラボノイド）の化合物情報、植物の分類情報、生薬・漢方情報を管理するポータルを構築した。ポータル上には酵素反応、代謝経路の情報のみならず文献情報、学会情報等も掲載し、メタボロミクス分野の統合ポータルとして機能する。同じ仕組を国内の複数学会で採用してもらい学会横断型のポータルとしても機能する。

研究成果の概要（英文）：

A portal site was created to host chemical information on basic metabolites, lipid, plant secondary metabolites (flavonoids), scientific classification of plant species, and crude drug and Kampo information. The site functions as an integrative portal for metabolomics by including information on biochemical reactions, metabolic pathways, literature, and academic societies. The same system is used by a few domestic societies to achieve a cross-societal network.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2005年度	11,600,000	0	11,600,000
2006年度	14,600,000	0	14,600,000
2007年度	14,800,000	0	14,800,000
2008年度	15,000,000	0	15,000,000
2009年度	15,000,000	0	15,000,000
総計	71,000,000	0	71,000,000

研究分野：バイオインフォマティクス

科研費の分科・細目：生体生命情報学・

キーワード：代謝、データベース、フラボノイド、バイオポータル

## 1. 研究開始当初の背景

特にバイオインフォマティクス分野では、生物の代謝はわかり難いとして敬遠される場合が少なくない。ATGCの文字列として扱えるゲノム情報や美しい立体構造を対象にするタンパク質科学に比較して、ベンゼン環に代表される化学構造はわかりづらいし、華々しさにも欠ける。代謝研究は古い学問である。多くの教科書やデータベースが存在し、学べき情報量も多い。いずれも論理的に整理されているのだが、分野の発展には更に分かり

やすくまとめることが1つのキーポイントであると考えた。つまり、どう表現すれば全体像や重要な点を一目でわからせるようにできるかを考えたかった。「すべては記号である」という情報科学の基本精神を存分にいかして、一見複雑でわかり難い世界の裏にあるメカニズムをシステムティックに表現する術を見出そうと試みたのが研究のきっかけである。

## 2. 研究の目的

当初は、代謝ネットワークにおける任意の炭素(窒素)原子をトレースでき、また部分ネットワーク全体の物質収支を自在に閲覧できるソフトウェアを作成することを目標とした。また同じ技術を用いて、インタラクティブなシグナル伝達のマップも作成できると考えた。これらのマップはグラフデータ構造を用いて表現し、これらをネットワーク越しに交換、検索、また重ね合わせる技術を開発する計画であった。しかし後述するように、この目的は研究3年目にフラボノイドの代謝マップブラウザを作成した時点で「先が見えた」と感じた。そうしたソフトウェアを短期間の研究成果として完成させることは可能だが、より難しいのは時代に合わせて内容をアップデートし、維持管理することだからである。そのため後半は「バイオポータル」の構築に取り組み、長期的に代謝情報・知識を維持管理できる仕組みの開発を心がけた。

## 3. 研究の方法

研究内容がソフトウェア開発であるため、研究成果部分にまとめて記述する。

## 4. 研究成果

### 2005-6年度の成果概要

#### フラボノイドの階層分類と分類用ソフトウェアの開発

植物二次代謝の代表的カテゴリーであるフラボノイドの化学構造およそ6000種を、その構造に基づいて4階層に分類した。母核構造、導入水酸基の位置、配糖体等の修飾パターンの順で生合成経路も加味しながら分類し、分類を反映した12桁のID番号付けも同時におこなった。(分類したフラボノイド分子数は2009年度には7000種に拡大した。)分類用のソフトウェアも同時に開発したがフラボノイド階層分類用に特別に作成したため、現在は利用していない。

#### 生合成マップブラウザ “Flavonoid Viewer” の作成

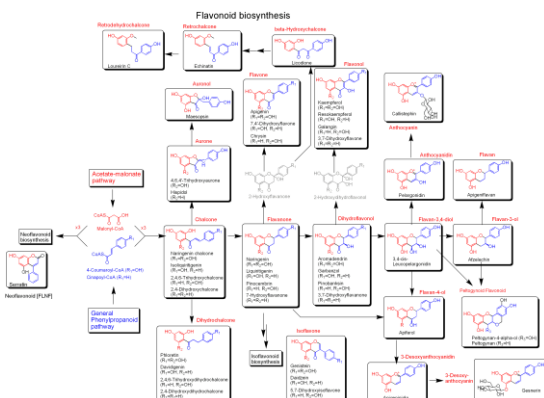


図1. フラボノイドビューワ (部分) 分子

構造描画で標準的な ChemDraw で記述した画像から文字列部分を自動認識し、フラボノイド名をクリックすると自動的に情報が出てくる仕組みになっている。

フラボノイドの生合成経路や構造間の関係を把握できるように、生合成経路の代謝マップブラウザを作成した。二次代謝に関与する物質には複雑な構造を持つものが多く、補酵素等の情報も得られない。そのため、人が見やすい形に描画したマップの断片を集めるだけで、リンク等の処理を人為的に施さなくてもマップ間を自由に渡り歩ける機能を実現した。

フラボノイドビューワでは通常のお絵かきエディタで作成した PostScript 画像から文字列を自動認識し、データベース中の代謝物情報と自動的にリンクする仕組みである。またマップ間の移動も可能なように設計した(図1)。

このマップブラウザと階層分類は現在も <http://www.metabolome.jp/software/> からアクセス、閲覧できる。

#### 脂質データベースの構築と学会による公式認定

これまで日本脂質生化学会が管理していた LipidBank データベースを全面的に更新した。脂質の構造データおよそ3000分子を独自に整理し、LipidBase データベースとして公開していたが、これは廃止し、LipidBank データベースに統合する作業をおこなった。これを「新 LipidBank」として日本脂質生化学会の公式データベースに認定してもらい、2007年に公開開始した(<http://lipidbank.jp/>)。

#### 研究期間前半の成果考察

研究期間前半において、インタラクティブな代謝マップの開発という目標にかなり近づいたが、この研究期間を通して得た知見は、データベースやインタラクティブな代謝マップは、作成よりもメンテナンスの方がずっと大変ということであった。構築時は目的意識も強く、ゼロから始めるために収集できるデータ量も多くなるが、更新作業は新規文献や新規物質と既収録データとの照合、追加による分類やマップレイアウトの再構築に多くの労力を費やす割には新規性を出しにくい。論文にもなりにくい。またシステムそのものが複雑化すると、研究の実施担当者が抜けてしまった時点でメンテナンス作業が停止してしまう。実際、フラボノイドビューワを作成してきた研究員が転職したため、ソフトウェアの整備はストップし、データの維持や精度向上は難しかった。

こうした状況を踏まえ、インタラクティブ性

を保ったまま回避する方策のひとつとして、代謝マップやデータベースそのものを Wiki 化することを考えた。そのため、2007 年度以降はそれまでの研究成果を引き継げる Wiki システムの開発を中心課題とし、当初の目的であったソフトウェアシステムを Wiki 上で実現することを試みた。

#### 2007-2008 年度の成果概要

#### 代謝の総合情報を発信するポータル metabolome.jp の構築

植物二次代謝や脂質代謝の解析は日本が世界をリードしている分野であり、日本発の研究成果を参照する研究者は圧倒的に日本人が多いと考えられる。研究に役立つソフトウェアの作成を目指し、自国の研究者に積極的に情報提供をするためには、研究内容やデータベースを日本語で情報発信することが重要である。そのため、若手研究者や一般の人々が代謝に興味を持てる話題を集めた日本語ポータルサイトを構築し、開発するソフトウェア、データも全てそこからダウンロードする形にした。

作成したサイト <http://www.metabolome.jp/> は Plone というシステム上に構築されている。2007 年度当初のアクセス IP アドレス数が毎月 1500 弱であったのに対し、内容の充実とともに増え続け、2009 年度には毎月 4000 アドレスまで増加した。(同一機関や同一プロバイダからの複数ユーザアクセスは 1 件として集計している)。日本語によるウェブサイト構築は研究内容の国際的な発信という視点とは矛盾するが、生薬や漢方などそもそも日本語でしか入手できない情報も多い。

#### フラボノイドの情報整理

フラボノイドのデータベースにおける文献情報の整理を行った。フラボノイド 7000 分子について辞書等から収拾した文献情報は入力者や入力方法によってスタイルが異なったり、入力ミスがあったりする。当初 5700 文献 (文献に記される物質数は重複を含め 30078 件) 収集されていた文献情報のうち、①辞書などの孫引き情報を削除し、オリジナルの文献のみを記載、②重複を省く、処理を行った結果、19728 件まで減らせることができた。ほぼすべての文献を PubMed や Google Scholar 等で調べ、スペルミス等の間違いも修正した。また、全フラボノイドについて構造を反映させた名称を入力する作業を行った。また糖の立体構造表記が統一されていなかったため、1610 フラボノイドについて糖鎖部分を修正し、統一した書式とした。

#### データベースを MediaWiki に移行

これまで Java applet と MySQL によりデータベースを構築してきたが、全データを

MediaWiki と呼ばれる Wikipedia のために開発されたシステムに移植した。MediaWiki によるシステムは従来のデータベース構築に比較して修正やメンテナンスが格段に容易であり、外部ユーザの書き込みやディスカッションも可能である。ただ、基本はテキスト用のシステムであるため、分子構造や数値属性を扱うために様々な機能の追加を施した。2008 年度にはフラボノイドビューワの分子情報全てを移行し、新しく分子量、分子名、著者にもとづく索引、統計情報のページを作成した。統計情報や索引は通常の Wiki ページ内では作成できない内容にあたる。これらの作成法や新規性は論文や解説記事等で発表した。同様に、基礎代謝物 1850 分子も階層分類に基づいて MediaWiki に移行した (図 2)。このポータルを新しく取得した <http://metabolomics.jp/> というアドレスで公開し、2009 年に毎月 6000 アドレス、2011 年には 8000 アドレスからアクセスされるまでに発展した。

#### 脂質代謝物、基礎代謝物とのデータ統合

基礎代謝物はフラボノイドと同じメタボロミクス JP 上に整理したが、脂質分子は日本脂質生化学会との連携を考慮して <http://lipidbank.jp/wiki/> 上に MediaWiki をインストールして整理した。現在このサーバは <http://jcbl.jp/> という脂質生化学会サーバとしても機能する。中・長鎖脂肪酸の分類に引き続き、グリセロ脂質とスフィンゴ脂質の階層分類を日本脂質生化学会と協力して作成した。これらのカテゴリーには糖脂質も含まれるが糖鎖のレイアウトアルゴリズムは作成できなかった。

#### 2009-2010 年度

#### 基礎代謝物の構造情報、原子のマッピング情報、酵素反応、パスウェイ情報を Wiki 上に構築

およそ 1800 の基礎代謝物について分子構造に基づいた階層分類を施し、構造に基づく 12 桁の ID 番号を新規に振ってから Wiki 上に公開した。この ID を用いて、酵素反応における原子のマッピング情報を Wiki ページとして効率よく実現し、間違いを誰でもすぐに修正できる仕組みを整えた。さらにこのマッピング情報を組み合わせ、酵素反応の Wiki ページを 2100 種の EC 番号について作成した。この基本情報だけで基礎代謝のほぼ全てをカバーした。

#### 植物分類情報、薬用植物の解析

University of Development Alternative の Mohammed Rahmatullah 教授夫妻を招聘し、バングラデシュの薬用植物解析する手順を整えた。現在植物サンプルの試料提供手続き

(マテリアルトランスファー)を完了させ、富山大学と共同で解析をおこなっている。バングラデシュの植物については効能、用途を整理して Wiki 上に掲載する準備を整えたが、実際に公開を始めたのは 2009 年度になってからである。日本薬局方に記載される生薬全て(300 超)および 40 以上の文献に記載される漢方処方(350 超)については、富山大学と共同して Wiki データベース上に情報を蓄積した。

### 研究後半の成果考察

研究期間後半は、Wiki 型データベースの開発を中心課題とした。生物学の情報を Wiki 化する作業が盛んだが、通常の Wiki システムは情報を収集する手段を提供こそすれ、情報をまとめて知識にする手段を持たない。(知識という概念を定義することは難しいが、端的には概要を一目で把握できる表やテーブルが含まれる。こうした情報の「まとめ」をここでは知識と記述する。)知識を個別の情報ページから自動作成するには、検索コマンドとそれを処理するプログラム部分を備えた Wiki が必要である。これをふまえ、MediaWiki というシステムに自前のコマンドを追加したものを Wiki 型データベースと称することにした。現在、データの収集対象は生薬、植物種、基礎代謝、フラボノイドと多岐にわたり、互いのページが密接にリンクしてある。通常のハイパーリンクによる連携と大きく異なる特徴は、データページ間でコンテンツを検索して自動リンクする点である。ページ内容が相互に自動更新され、必要最小限の手間とコストでメンテナンスを可能にするため、研究期間前半で直面したような問題を回避できる。しかし、通常のデータベースに比較して不便な点もあり、今後も改良が必要である。

### <国内外での成果の位置づけ>

#### 国内学会横断型 Wiki の構築

Wiki 型データベースを開発する過程で様々なデータを掲載し、2010 年度時点で以下の学会の公式データベースとして Wiki 型を採用することが決定した。

日本脂質生化学会 (<http://lipidbank.jp/>) LipidBank はデータの再整理や再番号付けを含め研究期間を通して扱ってきたデータベースである。

また下記の 2 学会では学会の公式データベースとして Wiki 型を採用してもらおう。

#### 和漢医薬学会

(<http://wakandb.u-toyama.ac.jp/wiki/>) 富山大学との共同研究を通じて和漢薬情報を収集し、臨床、細胞活性情報も収録予定。

日本植物細胞分子生物学会 細胞の組織培養

情報等を収録予定。

上記学会の公式データベースを同じプラットフォームで扱うことはそのまま学会間の融合につながり、アウトリーチの観点からも意義ある成果といえる。

### 国内のメタボロミクスまとめサイト構築

メタボロミクスは世界的にも注目される分野であり、代謝データや利用するソフトウェアを統合する作業は非常に重要である。情報をメタボロミクス JP というポータルサイトに集約することで、これらのサイトへは定期的に多くのアクセスがある。

### メタボロームシンポジウムの開催

2006 年度より国内でメタボロームシンポジウムが開催されているが、その実行委員として毎年運営を行い、2007, 2009 年度は予稿集の編集やプログラム編成の中心的役割を果たした。作成したバイオポータルには公募情報も掲載しており、国内における研究コミュニティの形成という点で重要な役割を果たした。

### 国際的な共同研究の推進

米国 Louisville 大学, NMR Center とは、メタボロミクスシンポジウムでの招待講演等を通して定期的な交流を行っている。密な交流とまではいかないが、今後も定期的に情報交換やプログラムの提供を続ける予定である。同様に研究交流をおこなう相手としてフィンランド ヘルシンキ大学 E Ukkonen グループ、米国 LipidMaps プロジェクト (San Diego) 等があげられる。また、バングラデシュの University of Development Alternative とは薬用植物のデータベース作成を共同で構築した。

### 5. 主な発表論文等

[雑誌論文、学会予稿集] (計 24 件) (全て査読あり。)

- ① Chang CW, Lyu PC, Arita M "Reconstructing phylogeny from metabolic substrate-product relationships" BMC Bioinformatics (Proceedings 9th APBC in Korea) 12(Suppl 1):S27, 2011
- ② Takemoto K, Arita M "Nested structure acquired through simple evolutionary process" Journal of Theoretical Biology 264(3), 782-786, 2010
- ③ Hasegawa Y, Arita M "Bistable stochastic processes in the q-exponential family" Physica A 389(21):4450-4461, 2010
- ④ Arita M "A pitfall of wiki solution for biological databases" Briefings

- Bioinformatics 10(3), 295-296, 2009
- ⑤ Takemoto K, Arita M "Heterogeneous distribution of metabolites across plant species" *Physica A* 388(13), 2771-2780, 2009
- ⑥ Horai H, Arita M, Ojima Y, Nihei Y, Kanaya S, Nishioka T "Traceable analysis of multiple-stage mass spectra through precursor-product annotations" *Proceedings of German Conference in Bioinformatics (GCB'09) (GI-Edition Lecture Notes in Informatics)*, Halle, Germany, 157:173-178, 2009
- ⑦ Arita M, Suwa K, Yoshimoto M, Hirai A, Kanaya S "Ontology checking and integration of crude-drug and kampo information" *Proceedings of the 2nd International Conference of Biomedical Engineering and Informatics (BMEI2009)*, Tianjin China, 3:1304-1307, 2009
- ⑧ Hasegawa Y, Arita M "Properties of the maximum q-likelihood estimator for independent random variables" *Physica A* 388, 3399-3412, 2009
- ⑨ Fukushima A, Kanaya S, Arita M "Characterizing gene coexpression modules in *Oryza sativa* based on a graph-clustering approach" *Plant Biotechnology* 26, 485-493, 2009
- ⑩ Morioka R, Arita M, Sakamoto K, Kawaguchi S, Tei H, Horimoto K "Period-phase map: two-dimensional selection of circadian rhythm-related genes" *IET System Biology* 3(6), 487-495, 2009
- ⑪ Sato S, Arita M, Soga T, Nishioka T, Tomita M "Time-resolved metabolomics reveals metabolic modulation in rice foliage" *BMC Systems Biology* 2:51, 2008
- ⑫ Arita M, Suwa K "Search extension transforms Wiki into a relational system: a case for flavonoid metabolite database" *BMC BioData Mining* 1:7, 2008
- ⑬ Fukushima A, Wada M, Kanaya S, Arita M "SVD-based anatomy of gene-expression for correlation analysis in *Arabidopsis thaliana*" *DNA research* 15(6), 367-374, 2008
- ⑭ 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
- ⑮ Arita M, Tokimatsu T "Detection of monosaccharide types from coordinates" *Proceedings of the 18th International Conference on Genome Informatics (Genome Informatics Series Vol.19)*, Singapore, 3-14, 2007
- ⑯ Nagasaki H, Arita M, Nishizawa T, Suwa M, Gotoh O "Species-specific variation of alternative splicing and transcriptional initiation in six eukaryotes" *Gene* 364, 53-62, 2005
- ⑰ Nagasaki H, Arita M, Nishizawa T, Suwa M, Gotoh O "Automated classification of alternative splicing and transcriptional initiation and construction of visual database of classified patterns" *Bioinformatics* 22(10), 1211-1216, 2006

[学会発表] (計 40 件) 英語による招待講演のみ記載

- ① Arita M "Knowledge Management using a Wiki-based System and its Application to Mass Spectra" *EBI Seminar*, April 19, EBI, Hinxton UK, 2011
- ② Arita M "Wiki-based platform for PKS in *Aspergilli*" *Pacificchem*, Dec 19, Honolulu (Hawaii), 2010
- ③ Arita M "Metabolic Pathway Analysis using Wiki", *The Sixth International Aspergillus Meeting (Asperfest 6)*, March 15 (15-17), Monterey USA, 2009
- ④ Arita M "Data Management of Cross-disciplinary Information using Wiki", *16th International Conference on Cytochrome P450*, June 21 (21-25), Okinawa Japan, 2009
- ⑤ Arita M "Web-based Resources for Metabolomics", *Tutorial at the CBI-KSBSB joint conference (Bioinfo2009)*, Nov 4 (4-6), Haeundae Busan Korea, 2009
- ⑥ Arita M "Integration of Flavonoid Information through Wiki" *Satellite Symposium of 4th International Conference on Polyphenols and Health (ICPH2009)*, Dec 11 (7-11), Yorkshire England, 2009
- ⑦ Arita M "Computational Analysis of Metabolism" *iPlant Mechanistic modeling grand challenge workshop*, November 9 (7-10), Biosphere2 Arizona USA, 2008
- ⑧ Arita M "Wiki-based Database of Secondary Metabolites", *JSPS-KOSEF Joint Seminar -Pioneering researches on fungal molecular biology-*, November 20 (19-20), Kanazawa Japan, 2008
- ⑨ Arita M "Databases for Flavonoids and

Measurement in Arabidopsis”, 2nd International Symposium on Environmental Metabolomics, March 24-25, University of Louisville(KY), USA, 2007

- ⑩ Arita M “Software Tools to Study Flavonoids”, Systems Biology and Bioinformatics Symposium (SBBS07), March 27-29 Taipei Taiwan, 2007
- ⑪ Arita M “Metabolomics in Japan and the Need for a Metabolic Map Editor”, 1st Japan-Taiwan Bilateral Symposium on Bioinformatics, March 13-17, Tokyo Japan, 2006
- ⑫ Arita M “Atomic-level analysis of metabolism and mass spectral database”, Invited Lecture Course, May9-14, Center for Regulatory, Environmental, Analytical Metabolomics, University of Louisville(KY), USA, 2006
- ⑬ Arita M “Metabolic Pathway Chart for Flavonoid”, 2nd Taiwan-Japan Bilateral Symposium on Bioinformatics, Nov 7-9, Tainan Taiwan, 2006
- ⑭ Arita M “Map Editor for the Atomic Reconstruction of Metabolism”, 1st Louisville Symposium on Environmental Metabolomics, November 5-6, Louisville (KY) USA, 2005
- ⑮ Arita M “ARM Database for Interactive Metabolic Maps”, 1st International BioPAX Symposium, November 18, Tokyo Japan, 2005

〔図書、総説〕(計 11 件) Current Opinion 誌は簡単な査読有

- ① 有田 正規 「代謝産物の検索に役立つデータベース：化合物の素性・素行を調べるウェブサイト」 実験医学 9 月号 Vol.28(14), 2301-2303, 2010
- ② Fukushima A, Kusano M, Redestig H, Arita M, Saito K “Integrated omics approaches in plant systems biology” Current Opinion in Chemical Biology 13, 532-538, 2009
- ③ Arita M “What can metabolomics learn from genomics and proteomics?” Current Opinion in Biotechnology 20, 610-615, 2009
- ④ 福島敦史, 草野都, 有田正規, 斉藤和季 「植物メタボロミクス—代謝プロフィールからシステム解析へ」 実験医学, 1 月号, 羊土社, 2008
- ⑤ 有田正規, 諏訪和大 「Wiki によるフラボノイドのデータベース」実験医学増刊, 「バイオデータベースとソフトウェア最前線」, 4 月, 1148-1154, 羊土社, 2008

- ⑥ 有田正規 「生体ネットワークをどう研究すべきか」 数理学 5 月号 (527)79-83, サイエンス社, 2007
- ⑦ Arita M, Fujiwara Y, 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)
- ⑧ Shinbo Y, (4 authors), Arita, M., Saito, K., Ohta, D., Shibata, D. and 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)
- ⑨ 時松 敏明, 有田 正規 「代謝マップビューワで見るフラボノイド」 細胞工学, 25(12), 1388-1393, 2006.
- ⑩ 小林 徹也, 有田 正規, 森下 喜弘, 合原 一幸 「細胞内現象のシステムの理解—今理論に何が求められているのか?」 システム/制御/情報 50(8), システム制御情報学会誌 2006.
- ⑪ 杉峰伸明, 大塚一路, 有田正規, 合原一幸 「ネットワーク的思考で生命現象をよみとく」(特集「意識・脳・身体への接続へ」) 科学 76(3), 岩波書店 2006

〔産業財産権〕

- 出願状況 (計 0 件)
- 取得状況 (計 0 件)

〔その他〕

ソフトウェア

Flavonoid Viewer : フラボノイドの構造、分類データベース

<http://www.metabolome.jp/software/FlavonoidViewer/>

ポータル・データベース

本研究費により作成したが、サイトをクローズする予定のポータル

<http://metabolome.jp/>

本研究費により作成した研究成果公開用バイオポータルおよび研究室ウェブページ

<http://metabolomics.jp/>

6. 研究組織

(1) 研究代表者

有田 正規 (ARITA MASANORI)

東京大学・大学院理学系研究科・准教授

研究者番号 : 10356389