

平成 29 年 6 月 1 日現在

機関番号：14501

研究種目：基盤研究(C) (一般)

研究期間：2014～2016

課題番号：26430198

研究課題名(和文) 酵素反応の選択性と多様性の情報解析

研究課題名(英文) Informatics analysis on selectivity and diversity of enzymatic reaction

研究代表者

荒木 通啓(araki, michihiro)

神戸大学・科学技術イノベーション研究科・特命准教授

研究者番号：40396867

交付決定額(研究期間全体)：(直接経費) 4,000,000円

研究成果の概要(和文)：新規代謝経路を構築していく上で、未知・既知を問わず代謝経路中に必要とされる酵素遺伝子候補の選択は重要な課題の一つである。本研究では、このような酵素遺伝子をより効率的に選択していくために、各データベースに拡散した情報を当該目的に適した形式で統合した新たな酵素反応データベースを再構築した。これをもとに、酵素遺伝子の選択性と多様性を考慮しつつ、代表的な遺伝子配列の選択をサポートできる生物・化学情報解析並びに機械学習方法を開発した。

研究成果の概要(英文)：In constructing a novel metabolic pathway, selection of enzyme gene candidates required for metabolic pathways, whether unknown or known, is an important task. In this research, in order to more efficiently select such enzyme genes, we reconstructed a new enzyme reaction database that integrated information diffused in each database in a form suitable for this purpose. Based on this, we have developed biological and chemical information analysis and machine learning methods that can support selection of representative gene sequences while considering selectivity and diversity of enzyme genes.

研究分野：合成生物学

キーワード：バイオインフォマティクス

### 1. 研究開始当初の背景

合成生物学・代謝工学の技術の成熟に伴い、新たな代謝経路の設計・構築は、医薬中間体・バイオ燃料・バイオプラスチック原料を生産していくうえにおいても、必要不可欠な要素技術の一つとなってきている。代謝経路は、要素としての化合物情報・酵素情報とそれらを有機的に結びつける酵素反応のペア情報により構成されているが、新たな代謝経路の設計・構築に関する研究は、属人的な酵素反応知識の拡大による研究が多く、必ずしも網羅的な知識を利用した研究ではない現状がある。

また、新たに設計された代謝経路には、未知・既知を問わず推定された酵素候補が複数出現することになり、代謝経路を実際に構築していく上で、酵素遺伝子の選択が最も重要な課題となってくる。しかしながらこの点に関して、現在のところ各人がKEGG・BRENDAといった各酵素反応データベースに拡散した情報、文献・特許情報をマニュアルで調査し、属人的な直観による意思決定がなされている状況である。実際に共同研究を推進する中でも、酵素遺伝子の効率的かつ信頼性の高い選択方法の開発が強く望まれている。

### 2. 研究の目的

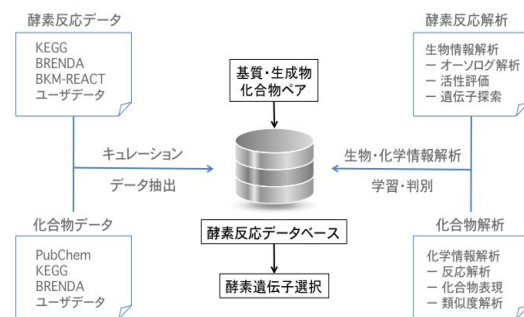
このような背景のもと、より効率的な酵素遺伝子選択を行うためには、各データベースに拡散した情報を一定の基準で統合した新たな酵素反応データベースを再構築し、それをもとに属人的な意思決定を再現できる、またはサポートできる生物・化学情報解析を行っていく必要があるとの着想に至った。現行の代謝経路設計基盤では、すでに類似反応を触媒する酵素を独自データベースより検索し、各酵素の本来の基質・生成物との化学構造の類似度をもとに酵素選択を行うという手法を開発している。本研究ではこれをさらに発展させ、酵素遺伝子選択という目的に特化した独自の酵素反応データベースを構築し、生物・化学情報解析を行っていく。このため本研究では、酵素反応の選択性と多様性にフォーカスした情報解析を行うために、既知データベースに含まれる網羅的な情報を統合した独自のデータベースを構築していく。これをもとに、生物・化学情報解析、機械学習の利用により、データベースを拡張していくとともに、選択性と多様性それぞれに重要な特徴量を明示的にスコア化し、任意の基質・生成物ならびに化合物ペアに対して、酵素遺伝子の選択を提案できるシステムの構築を目指す。

### 3. 研究の方法

酵素反応データベースの構築

酵素反応データとして KEGG・BRENDA 等

を利用し、化合物データとして主として PubChem データベースを利用して、独自の酵素反応データベースを構築した。



#### 酵素反応の生物・化学情報解析

酵素反応の基質・生成物の選択性に関する知識を集積されたデータから総合的に判断していくために、酵素配列解析による生物種・遺伝子解析といった酵素そのものに関する生物情報解析に加えて、化学構造・反応の類似度比較といった化学情報解析をもとに、各特徴について数値化を行った。これをもとに、任意の基質・生成物ならびに化合物ペアに対して酵素遺伝子を選択していく上で、属人的な知識と直観を再現しうる重み付けならびにスコア関数を定義した。また、生物・化学情報解析により、既知データ以外にも推定データを含めた酵素反応データベースの拡張を行い、ユニークな酵素に対しても新たな情報を付加していた。

#### 酵素反応の機械学習方法の開発

基質・生成物・酵素情報の各データを抽出し、化学構造・物性指標・アミノ酸配列データをもとに記述子を定義し、各記述子に対する値をもとに基質・生成物・酵素情報の特徴を数値(ベクトル)化し、一連のベクトルをマージすることで、一つの酵素反応の特徴ベクトルとして定義した。得られる特徴ベクトルを利用して、サポートベクターマシン(SVM)等の機械学習法を適用することで、当該酵素反応の判別器を作成した。

### 4. 研究成果

#### 酵素反応データベースの構築

酵素反応情報としては、EC 番号、生物種、遺伝子情報、基質・生成物、補酵素情報、活性情報(kcat・Km)、熱力学情報といった各基本情報のうち酵素遺伝子選択を行うために必要な情報を集積し、SQLにてデータベースを構築した。また、酵素反応データ拡張に伴って、データの質を確保していくために、宿主情報、酵素遺伝子情報の有無、熱力学的基準による反応方向性の判定などを含む独自のキュレーション基準を設定し、マニュアルキュレーションを実施した。一般に、酵素反応データベースは、EC 番号をエン트리

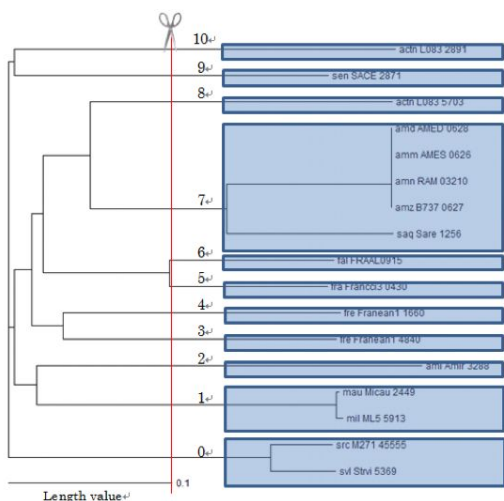
として関連する反応・遺伝子情報を記述したフラットファイルにより構成されているが、本研究では代謝設計に有用な形式として、化合物ならびに化合物ペアに対して各情報を関連付けたデータ形式をもとにウェブベースのユーザインターフェースを構築

Key ID	EC Num	Substrate	Structure1	Product	Structure2	comment_1
4714	3.1.1.1	CID16533		C03648		Ester_hydrolysis
8536	2.3.1.196	C12537		C03485		Ester_hydrolysis
14366	3.1.1.1	C03648		CID16533		Esterification
4892	2.3.1.196	C03485		C12537		Esterification

した。

### 酵素反応の生物・化学情報解析

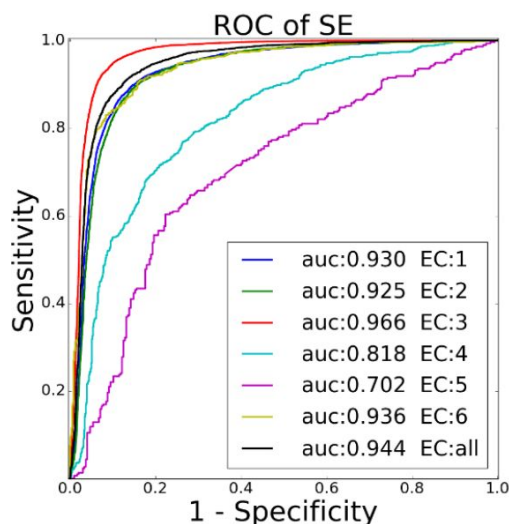
酵素反応の基質・生成物の選択性に関する知識を各種のデータから総合的に判断していくために、オーソログ解析による生物種・遺伝子解析、酵素活性依存度の解析、補酵素要求性などの酵素そのものに関する生物情報解析に加えて、化学構造・反応の類似度比較といった化学情報解析をもとに、各特徴について数値化を行った。特に配列解析では、アラインメント、クラスタリングを利用して、多様性を確保しつつ各クラスタリングの代表配列を選択する手法の開発を中心に行った。これをもとに、任意の基質・生成物ならびに化合物ペアに対して酵素遺伝子を選択していく上で、スコア関数を定義した。また、生物・化学情報解析により、既知データ以外にも推定データを含めた酵素反応データベースの拡張を行い、ユニークな酵素に対してモアノテーション情報を付加した。



### 酵素反応の機械学習方法の開発

Dragon Descriptor を利用して、基質・生成物の化学構造・物性指標から記述子を定義し

た。酵素情報に関しては、アミノ酸配列データをもとに記述子を定義し、各記述子に対する値をもとに基質・生成物・酵素情報の特徴を数値(ベクトル)化し、一連のベクトルをマージすることで、酵素反応の特徴ベクトルとして定義した。得られる特徴ベクトルを利用して、当該酵素反応あるいは酵素反応クラスの学習データを作成し、サポートベクターマシン(SVM)等の機械学習法を適用することで、酵素反応の判別器を作成した。評価はクロスバリデーションにより行い、各酵素クラスにおける予測精度の評価を実施した。



## 5. 主な発表論文等

〔雑誌論文〕(計9件)

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K.Y., Shimatani, Z. and Kondo, A. "Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems" *Science*, 353, 6305 (2016).

【論文・査読有】

DOI: 10.1126/science.aaf8729

Ye, X., Morikawa, K., Ho, S.H., Nishida, K., Araki, M., Hasunuma, T., Hara, K.Y., Kondo, A. "Evaluation of genes involved in oxidative phosphorylation in yeast by developing a simple and rapid method to measure mitochondrial ATP synthetic activity" *Microbial Cell Factories*, 14 (1), 56 (2015).

【論文・査読有】

DOI: 10.1186/s12934-015-0239-z

\*Araki, M., Cox III, R.S., Makiguchi, H., Ogawa, T., Taniguchi, T., Miyaoku, K., Nakatsui M., Hara, K.Y., Kondo, A. "M-path: A Compass for Navigating Potential Metabolic Pathways"

Bioinformatics, 31(6) 905-911 (2015).

【論文・査読有】

DOI: 10.1093/bioinformatics/btu750

Hara, K.Y., Aoki, N., Kobayashi, J., Kiriya, K., Nishida, K., Araki, M., Kondo, A. "Improvement of oxidized glutathione fermentation by thiol redox metabolism engineering in *Saccharomyces cerevisiae*" Appl. Microbiol. Biotechnol. 99: 9771-9778 (2015).

【論文・査読有】

DOI: 10.1007/s00253-015-6847-z

〔学会発表〕(計 30 件)

荒木通啓 「代謝経路デザインの限界」第 11 回日本ゲノム微生物学会年会 (2017.3.3 慶応義塾大学、藤沢)【招待講演】

牧口 大旭、小川 哲平、荒木通啓 「バイオプロセスデザインプラットフォームの構築」第 39 回日本分子生物学会 (2016.12.1 パシフィコ横浜、横浜)

Ogawa, T., Makiguchi, H. and Araki, M. "Development of Design Platform for Metabolic Pathways" 生命医薬情報学連合大会 2016 年大会 (2016.9.29-30 東京国際交流館プラザ平成、東京)

荒木通啓 「In Silico 代謝設計ツールの開発とその展開」農芸化学会 2016 年度大会 (2016.3.30 札幌コンベンションセンター、札幌)【招待講演】

江口晃一、工藤基徳、荒木通啓、石井純、山本浩明、近藤昭彦「大腸菌における 1,4-ブタンジオール人工合成経路の構築」農芸化学会 2016 年度大会 (2016.3.30 札幌コンベンションセンター、札幌)

荒木通啓、牧口大旭、小川哲平、近藤昭彦「インシリコ代謝経路設計基盤の展開」日本薬学会第 136 回年会(2016.3.28 横浜)

荒木通啓 「代謝デザインの展開」第 1 回デザイン生命工学研究会 (2016.3.8 東京工業大学)【招待講演】

Michihiro Araki 「Expanding the Scope of Metabolic Pathway In Silico」The 7th International Symposium of Innovative BioProduction Kobe (iBioK) (2016.1.27 神戸大学、神戸)【招待講演】

Michihiro Araki 「 Computational

platform for designing synthetic metabolic pathways」7th International Congress on Biocatalysis (2014.9.3 Hamburg University of Technology, Hamburg)【招待講演】

〔図書〕(計 3 件)

石井純、荒木通啓、中津井雅彦、崎濱 由梨、柘植 陽太、蓮沼 誠久、近藤 昭彦「合成生物学によるモノづくり微生物のデザインに向けて」生物工学会誌 Vol.93, 9 (2015)

〔その他〕

ホームページ等

<https://sites.google.com/site/michihiroaraki/>

<http://bp.scitec.kobe-u.ac.jp/m-path/db/>

<http://bp.scitec.kobe-u.ac.jp/pathpod/>

6. 研究組織

(1) 研究代表者

荒木 通啓 (ARAKI, Michihiro)

神戸大学・大学院科学技術イノベーション研究科・特命准教授

研究者番号：40396867