



Title	メタボローム技術で得られる代謝経路不明な多数の化合物の組み合わせから経路を予測する手法の開発
Author(s)	小寺, 正明
Citation	2010年度科学技術振興機構ERATO湊離散構造処理系プロジェクト講究録. p.193-203.
Issue Date	2011-06
Doc URL	<a href="http://hdl.handle.net/2115/48442">http://hdl.handle.net/2115/48442</a>
Type	conference presentation
Note	ERATO 세미나 2010 : No.32. 2011年1月21日
File Information	32_all.pdf



[Instructions for use](#)

ERATO セミナ 2010 - No. 32

# メタボローム技術で得られる代謝経路不明な多数の化合物の組み合わせから経路を予測する手法の開発

小寺正明

京都大学バイオインフォマティクスセンター

2011/1/21

## 概要

生体内での存在が確認されているがその生合成・生分解経路の不明な化合物 (orphan metabolites) は多く知られており、近年のメタボローム関連技術の発達によりその数はますます増えることが予想される。過去の研究では、そのような orphan metabolites の組み合わせから可能な酵素反応式を構築し、それらに適切な酵素番号 (EC 番号) を推定する手法を開発した。KEGG データベース中に登録されている任意の 2 つの orphan metabolites に対して二次元化学構造比較 (化学アラインメント) を行い、既知の酵素反応式を構成する部分反応式との照合を行うことで、可能性の高い酵素反応式を構築した。既知の酵素反応式を用いたクロスバリデーションでは 77% の酵素反応式を再構築することができた。再構築できなかった式の多くは、具体的な化学構造を持たない化合物を含むものであった。構築された反応式の妥当性チェックと EC 番号推定にはランダムフォレスト (random-forest) が用いられ、クロスバリデーションで 98% の精度を得た。ここで開発した手法の欠点のひとつは、酵素反応式の構築方法が非効率であるため膨大な計算時間を要求することである。今回の発表では、この研究の背景や手法について説明した後、問題点やその解決案について紹介する。



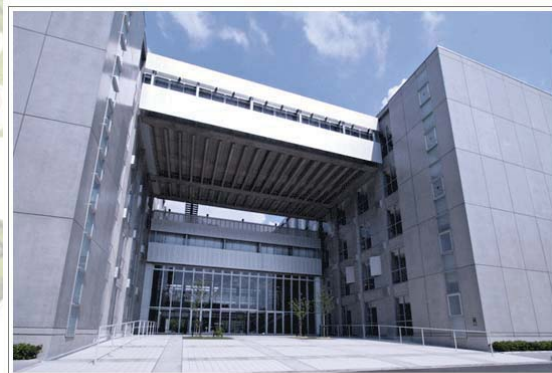
Elicitation of possible enzyme reaction equations involving orphan metabolites found in metabolomic study.

メタボローム技術で得られる代謝経路不明な多数の化合物の組み合わせから経路を予測する手法の開発



小寺正明 Masaaki KOTERA

## 京都大学化学研究所

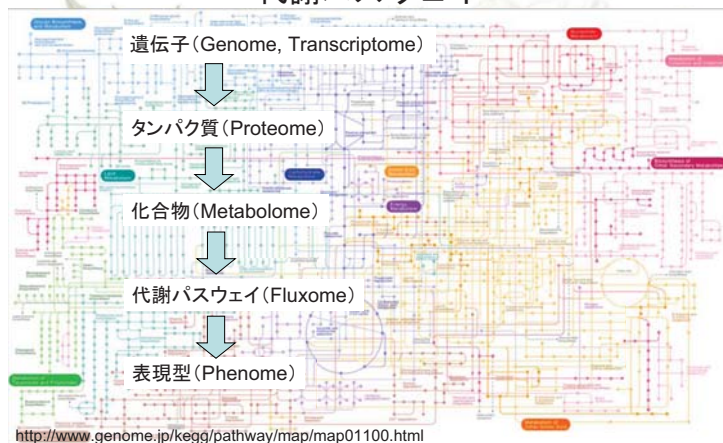


バイオインフォマティクスセンター正面入り口

## 本日の内容

- イントロ
- 2004 年の方法 (E-zyme)
- 2008 年の方法 (GREP)
- 2008 年の方法の改良へ向けて

## 代謝パスウェイ

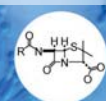


## メタボローム・メタボロミクス

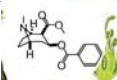
- 細胞の活動によって生じる特異的な分子を網羅的に解析すること
- 1. 分離技術の向上
  - ガスクロマトグラフィー (GC)
  - 高速液体クロマトグラフィー (HPLC)
  - キャピラリー電気泳動 (CE)
- 2. 検出技術の向上
  - 質量分析法 (MS)
  - 核磁気共鳴分光法 (NMR)
  - 薄層クロマトグラフィー (TLC)
- 解析対象の代謝化合物・代謝パスウェイ
  1. パスウェイ既知の物質 ... 主に一次代謝
  2. パスウェイ未知の物質 ... 主に二次代謝や環境物質分解

## 二次代謝産物

Penicillin, antibiotics from *Penicillium notatum*

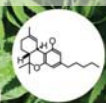


Cocaine, present in *Erythroxylon coca* the Coca plant



- 生物自身が生成し、生物が生ずる上で必要不可欠ではない(と考えられていた)代謝化合物
- 有名なものとして抗菌物質や色素など
- 様々な生物種が様々な産物を合成
- 様々な生理活性を持つ

Tetrahydrocannabinol, present in *Cannabis sativa*, protecting the plant from harmful UV-B radiation.



Tetrodotoxin, a potent neurotoxin present in certain kinds of fish



## Orphan metabolites

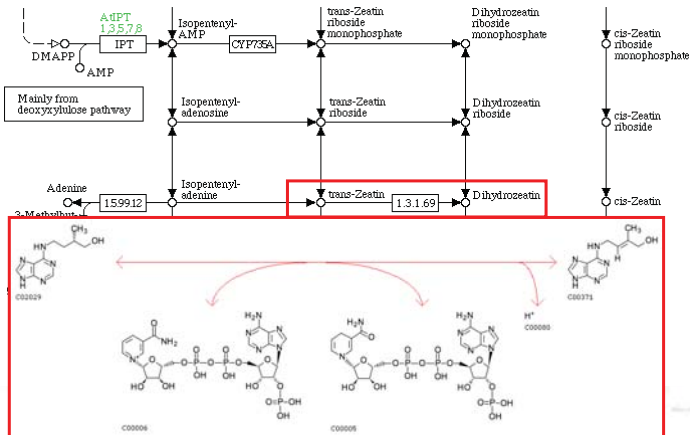
- いくつかの生物種が持つことが分かっているけれども生合成・分解経路が不明な物質
  - 植物や菌類などに多く見られる
    - 植物の自己防衛に働く例が多い
    - 分離・検出技術の向上により、さらに増える可能性
  - 医学・薬学・工学的にも役立つ可能性

## KEGG 中の orphan metabolites

(2008 年時点)

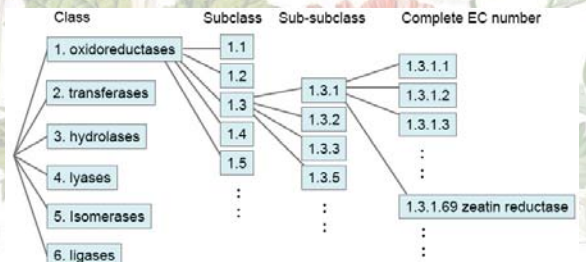
- 206 sesquiterpene lactones
- 124 isoquinoline alkaloids
- 122 indole alkaloids
- 107 diterpenoids
- 106 sesquiterpenoids
- 92 flavones and flavonols
- 89 quinones
- etc. (3108 compounds in total)

## 酵素遺伝子と酵素タンパク質と酵素反応



## EC 番号 (EC number)

- 国際生化学連合（現在の国際生化学分子生物学連合、IUBMB）の酵素委員会によって1961年に作られた、酵素分類と命名のためのID番号

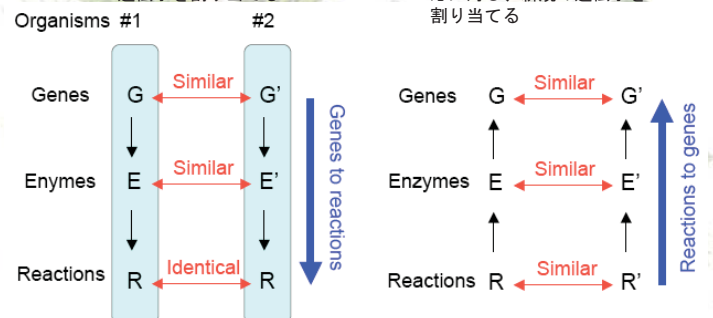


## アノテーション

- 元々の意味は、標本などを閲覧した際に、その標本に対するコメントなどを付与すること。
- 計算科学では：
  - あるデータに対して関連する情報（メタデータ）を注釈として付与すること、またはその付与されたメタデータやタグのこと。
- 生物情報科学 (Bioinformatics) では：
  - 遺伝子配列やタンパク質配列などの生物学的データに対して、実験的に確認された情報、または計算機的に予測された情報を付与すること、またはその付与された情報のこと。

## 2方向のアノテーション

- 通常のゲノムアノテーション
  - 既知の反応に対し、新規の遺伝子を割り当てる
- 逆方向のアノテーション
  - 可能性のある（未知の）反応に対し、新規の遺伝子を割り当てる

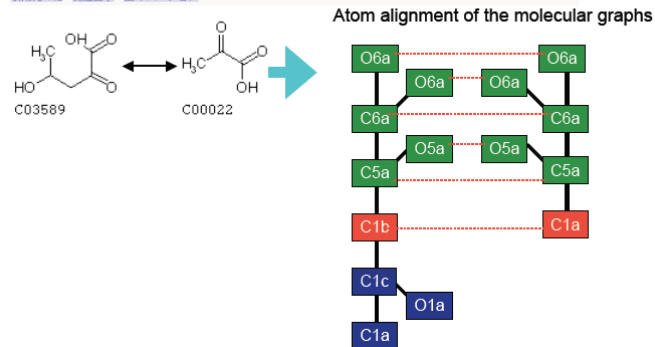


## 本日の内容

- イントロ
- **2004 年の方法 (E-zyme)**
- 2008 年の方法 (GREP)
- 2008 年の方法の改良へ向けて

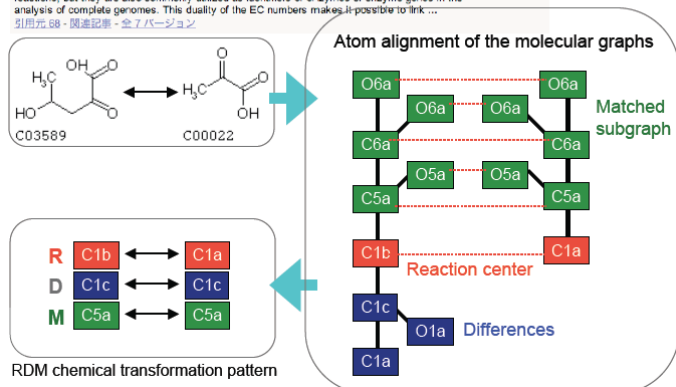
### Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways

M Hattori, Y Okuno, S Goto, M Kanehisa - J. Am. Chem. Soc., 2003 - ACS Publications  
 Cellular functions result from intricate networks of molecular interactions, which involve not only proteins and nucleic acids but also small chemical compounds. Here we present an efficient algorithm for comparing two chemical structures of compounds, where the chemical ...  
 引用元:112 - 関連記事 - 全4バージョン



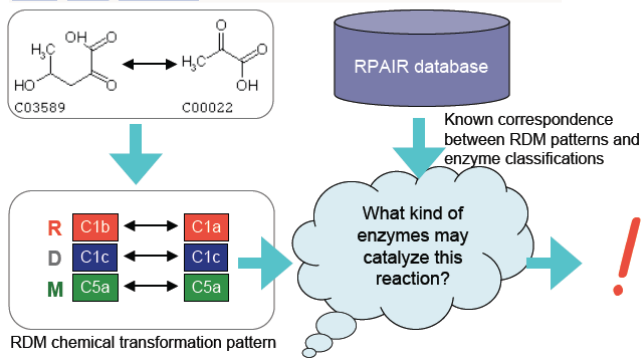
### Computational assignment of the EC numbers for genomic-scale analysis of enzymatic reactions

M Kotera, Y Okuno, M Hattori, S Goto, M ... - J. Am. Chem. Soc., 2004 - ACS Publications  
 The EC (Enzyme Commission) numbers represent a hierarchical classification of enzymatic reactions, but they are also commonly utilized as identifiers of enzymes or enzyme genes in the analysis of complete genomes. This quality of the EC numbers makes it possible to link ...  
 引用元:68 - 関連記事 - 全7バージョン



### E-zyme: predicting potential EC numbers from the chemical transformation pattern of substrate-product pairs

Y Yamashita, M Hattori, M Kotera, S Goto, M ... - 2009 - Oxford Univ Press  
 Results: In this article we propose a new method to predict the potential EC numbers to given reactant pairs (substrates and products) or uncharacterized reactions, and a web-server named E-zyme as an application. This technology is based on our original biochemical ...  
 引用元:10 - 関連記事 - 全11バージョン



## 本日の内容

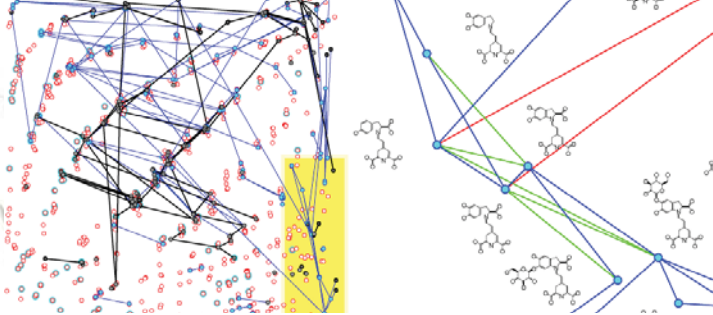
- イントロ
- 2004 年の方法 (E-zyme)
- **2008 年の方法 (GREP)**
- 2008 年の方法の改良へ向けて

## ダブリン大学 (アイルランド)



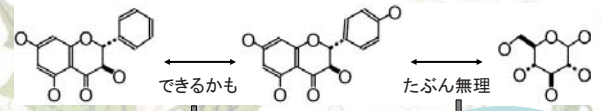


## Orphan metabolites 間の経路を結ぶ



Masaaki Kotera, Andrew G. McDonald, Sinead Boyce, Keith F. Tipton. **Eliciting Possible Reaction Equations and Metabolic Pathways Involving Orphan Metabolites.** *J. Chem. Inf. Model.* 2008, 48(12):2335-2349.

1回の酵素反応で変換できそうか？



既に知られている化合物だけで反応式が作れるか？

作れない

作れる

その反応式は既知の反応式に似ているか？

どれかに似ている

どれにも似ていない

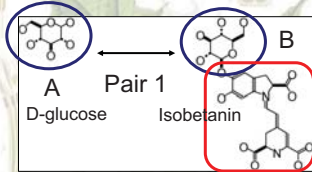
その経路はEC x.x.x.xに相当する酵素で代謝されるかも



## GREP ... Generator of Reaction Equations & Pathways

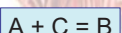
- 化合物の組み合わせから酵素反応式を作る (Kotera et al., 2008)

手順1、化合物ペアを作る

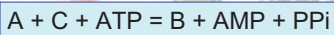


Pair 1 + Pair 2

手順2、化合物ペアの組み合わせで、部分的な反応式を作る

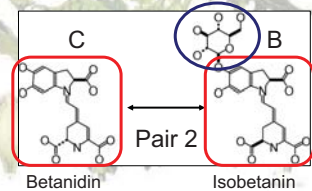


手順3、必要な化合物を補って、完全な反応式を作る



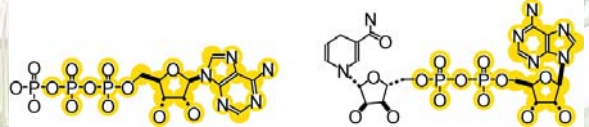
手順4、EC番号を自動的に割り当てる

EC6.--- (Ligase)



## 手順1、化合物ペアを作る

- 化合物2次元アラインメント

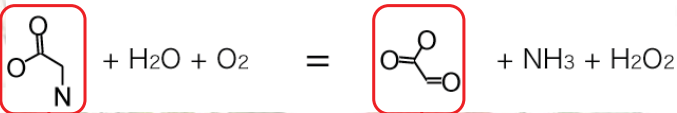


約14,000 化合物 (2008年当時)

約14,000 x 14,000 / 2 = 98,000,000 化合物ペア

## 意味のある化合物ペアを見つけたい

EC1.4.3.19



Reactant pair

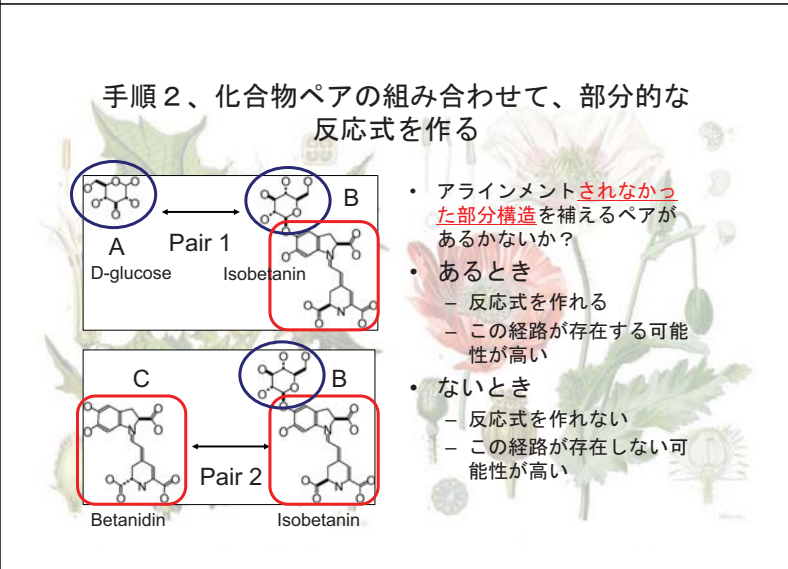
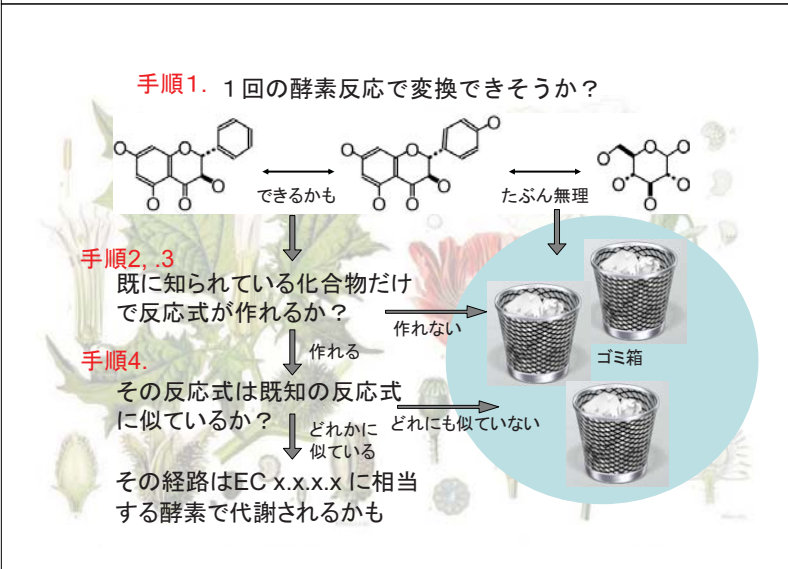
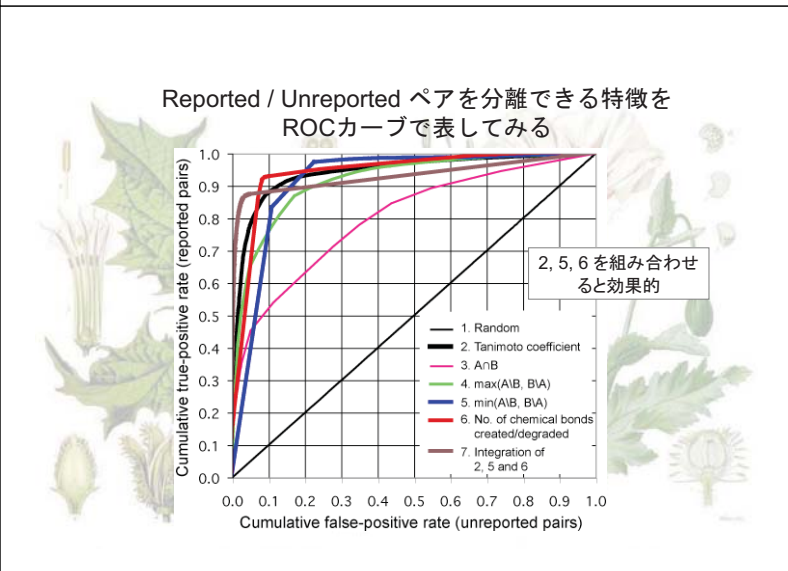
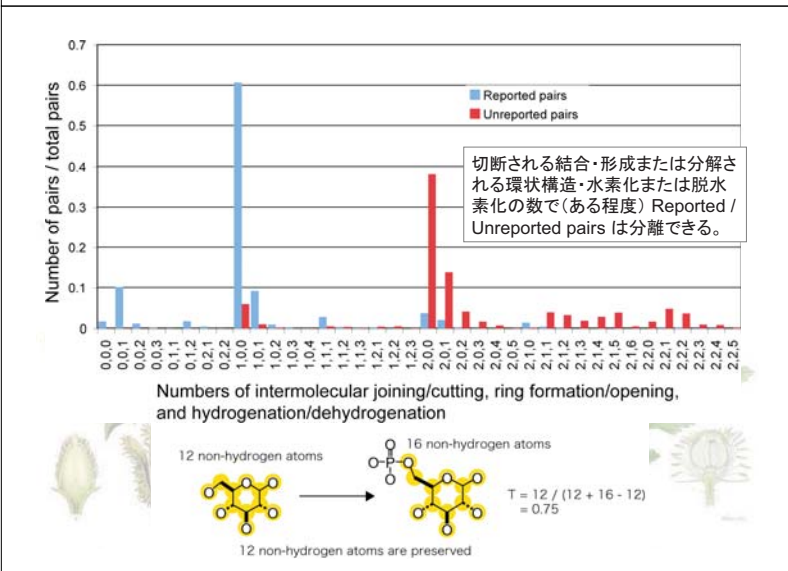
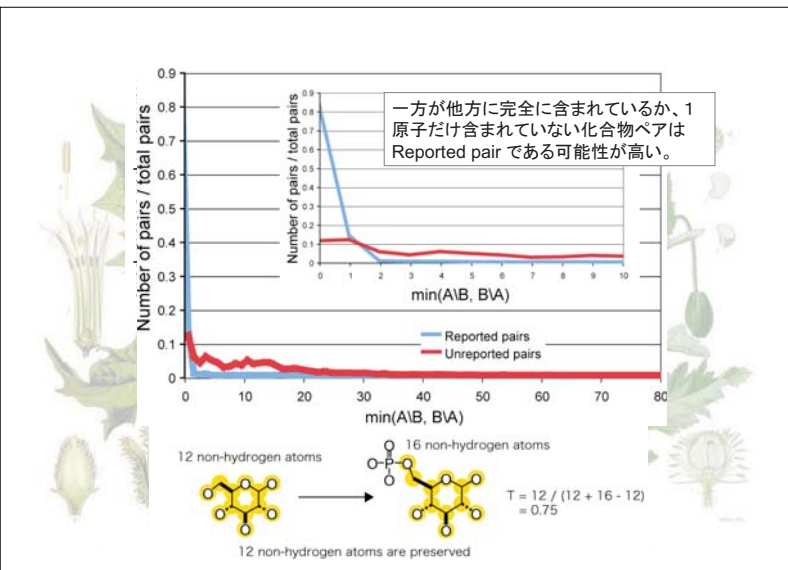
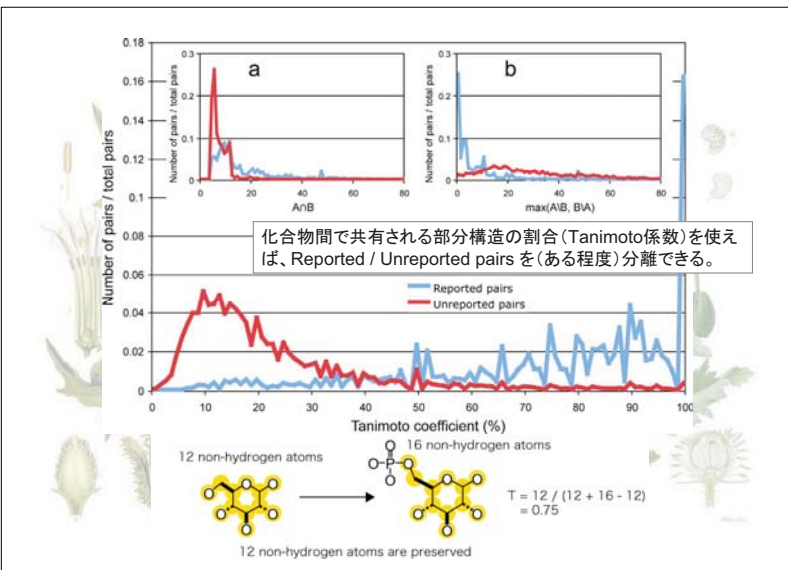
化合物ペアの中で、酵素反応式の左辺と右辺を構成し、(水素原子を除く)原子グループを保存するペアを、「Reactant pair」と呼ぶことにする。

## 意味のある化合物ペアを見つけたい

全化合物ペア  
Unreported pairs

Proposed pairs

既知の Reactant pairs  
= "Reported pairs"



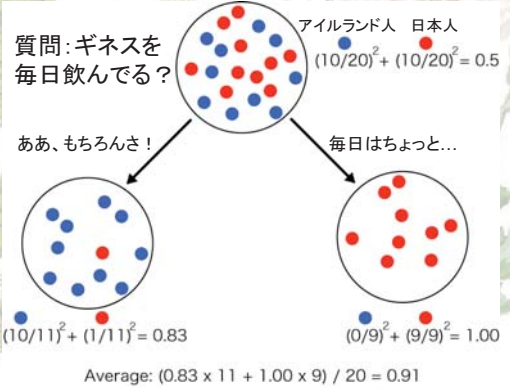


### 手順3、必要な化合物を補って、完全な反応式を作る

- 左右辺の差分が“H2”の場合
  - 水素化・脱水素化反応
    - “NADH:NAD”などを補う
- 左右辺の差分が“CH2”の場合
  - メチル化・脱メチル化反応
    - “S-Adenosyl-L-methionine:S-Adenosyl-L-homocysteine”などを補う
- 左右辺の差分が“NH3:O”の場合
  - 酸化的脱アミノ反応またはアミノ転移反応
    - “H2O+O2:NH3+H2O2”などを補う
- etc

### 手順4、EC番号を自動的に割り当てる

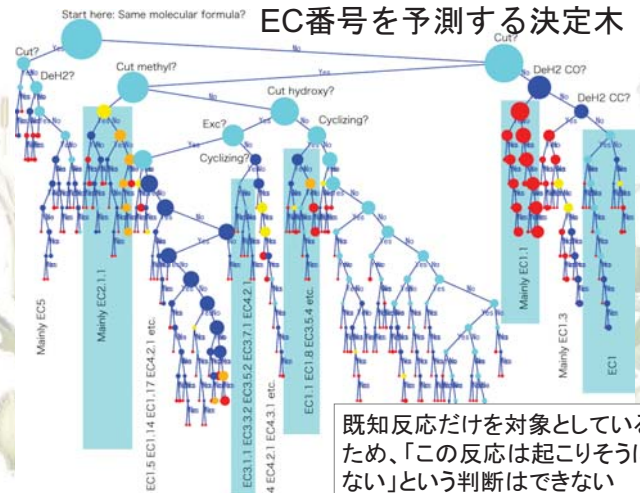
- 決定木とジニ係数



### 質問集

- 化合物数
  - 左右辺それぞれ化合物の数はいくつか?
- 分子式
  - 左右辺で原子の数は釣り合っているか?
- 化学結合
  - どんな化学結合が切断または生成されるか?
- 官能基・部分構造
  - どんな官能基や部分構造が切断・生成されるか?
  - どんな官能基や部分構造が保存されるか?
- etc

### EC番号を予測する決定木



### 「起こりそうにない」も判断する決定木

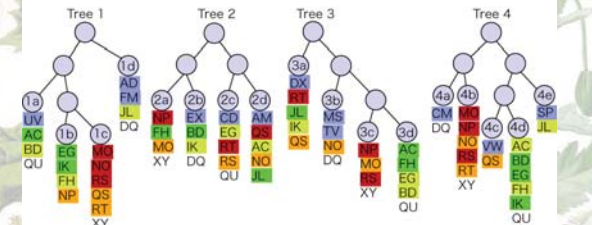
1. 全 reported pairs からランダムに選んだ95%を教師セットに、残り5%をテストセットに使う。
  - 教師セットの reported pairs にはその EC 番号をラベルとして与え、テストセットの reported pairs にはラベルを与えない。
2. 全 unreported pairs からランダムに選んだ数に「EC99」（そんなEC番号はありません）というラベルを与え、残りの unreported pairs にはラベルを与えない。
  - 「EC99」を与える割合：1~5%
3. ラベルを最もよく分離する決定木を作成。
4. ステップ1~3を100回繰り返す。

### 多数の決定木による多数決

Reported pairs: EC1: AC, BD, EG, FH, IK, JL; EC2: MO, NP, NO, QS, RS, RT

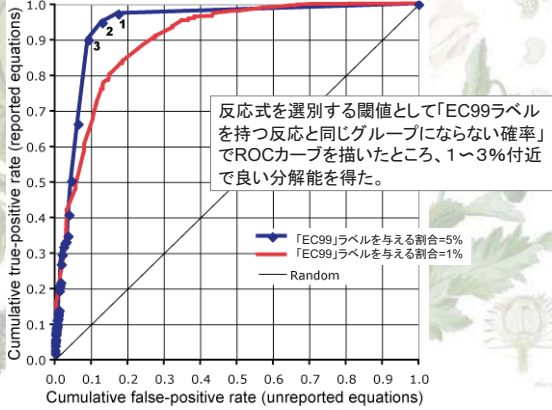
Unreported pairs: AB, AD, AC, ..., YZ

Test 1: EC1=(AC, EG, IK), EC2=(MO, NO, RS), EC99=(AD, FM, UV)  
 Test 2: EC1=(BD, FH, JL), EC2=(NP, QS, RT), EC99=(AM, CD, EX)  
 Test 3: EC1=(AC, FH, JL), EC2=(NP, RS, RT), EC99=(DX, MS, TV)  
 Test 4: EC1=(BD, EG, IK), EC2=(MO, NP, RS), EC99=(CM, VW, SP)



- ペアXYを含むグループから reported pair を任意に選んだ時に
  - 「EC1」である確率 = 0.125
  - 「EC2」である確率 = 0.875
  - 「EC99」である確率 = 0.000
- ペアDQを含むペアから reported pair を任意に選んだ時に
  - 「EC1」である確率 = 0.125
  - 「EC2」である確率 = 0.000
  - 「EC99」である確率 = 0.875

### 「EC99」を分離できる能力をROCカーブで表してみる



Each orphan compound is emphasized with a link. Click [Fig] on the left hand, and the image of the reaction is shown. [Suggested reactions] [Reported reactions]

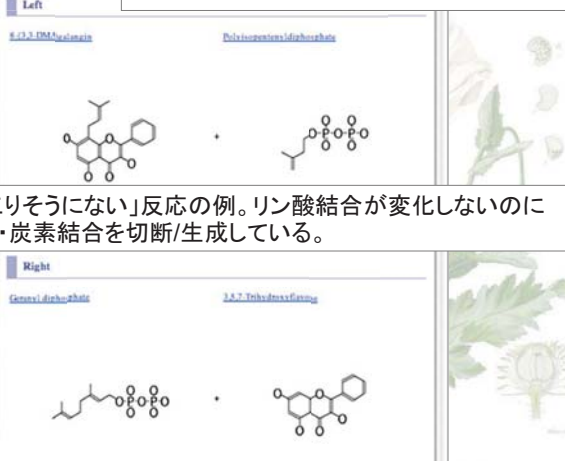
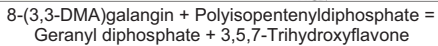
#### Suggested reactions

- [Fig] EC2.4.1.-: D-Glucose 1-phosphate + Betanidin = Isobetandin + Orthophosphate
- [Fig] EC2.4.1.-: 2-Amino-2-deoxy-D-glucose + Betanidin + L-Glutamate = Isobetandin + H<sub>2</sub>O + L-Glutamine
- [Fig] EC2.4.1.-: beta-D-Glucose 1-phosphate + Betanidin = Isobetandin + Orthophosphate
- [Fig] EC2.7.- or EC2.3.1.-: Brachiose + Betanidin = Isobetandin + alpha-D-Glucose
- [Fig] EC2.7.- or EC2.4.1.-: D-Fructose 6-phosphate + Betanidin = Isobetandin + Orthophosphate
- [Fig] EC3.1.1.- or EC3.2.1.-: D-Glucose + Betanidin = Isobetandin + H<sub>2</sub>O
- [Fig] EC3.1.1.- or EC3.2.1.-: D-Galactose + Betanidin = Isobetandin + H<sub>2</sub>O
- [Fig] EC3.1.1.- or EC3.2.1.-: Carbinose + Betanidin = Isobetandin + H<sub>2</sub>O
- [Fig] EC3.1.1.- or EC3.2.1.-: beta-D-Glucose + Betanidin = Isobetandin + H<sub>2</sub>O
- [Fig] EC3.1.1.- or EC3.2.1.-: alpha-D-Glucose + Betanidin = Isobetandin + H<sub>2</sub>O
- [Fig] EC3.1.1.- or EC3.2.1.-: Glucose + Betanidin = Isobetandin + H<sub>2</sub>O
- [Fig] EC3.2.1.-: D-Fructose + Betanidin = Isobetandin + H<sub>2</sub>O
- [Fig] EC3.2.1.-: D-Myo-Inositol + Betanidin = Isobetandin + H<sub>2</sub>O
- [Fig] EC3.2.1.-: L-Sorbose + Betanidin = Isobetandin + H<sub>2</sub>O
- [Fig] EC6.1.1.-: D-Fructose + Betanidin + ATP = Isobetandin + ADP + Orthophosphate
- [Fig] EC6.1.1.-: D-Fructose + Betanidin + ATP = Isobetandin + Diphosphate + 5'-AMP
- [Fig] EC6.1.1.-: D-Myo-Inositol + Betanidin + ATP = Isobetandin + ADP + Orthophosphate
- [Fig] EC6.1.1.-: D-Myo-Inositol + Betanidin + ATP = Isobetandin + Diphosphate + 5'-AMP
- [Fig] EC6.1.1.-: D-Myo-Inositol + Betanidin + ATP = Isobetandin + Diphosphate + 5'-AMP
- [Fig] EC6.1.1.-: D-Glucose + Betanidin + ATP = Isobetandin + Diphosphate + 5'-AMP
- [Fig] EC6.1.1.-: D-Galactose + Betanidin + ATP = Isobetandin + ADP + Orthophosphate
- [Fig] EC6.1.1.-: D-Galactose + Betanidin + ATP = Isobetandin + Diphosphate + 5'-AMP
- [Fig] EC6.1.1.-: D-Glucose + Betanidin + ATP = Isobetandin + ADP + Orthophosphate
- [Fig] EC6.1.1.-: D-Glucose + Betanidin + ATP = Isobetandin + Diphosphate + 5'-AMP
- [Fig] EC6.1.1.-: D-Glucose + Betanidin + ATP = Isobetandin + ADP + Orthophosphate
- [Fig] EC6.1.1.-: D-Glucose + Betanidin + ATP = Isobetandin + Diphosphate + 5'-AMP
- [Fig] EC6.1.1.-: D-Glucose + Betanidin + ATP = Isobetandin + ADP + Orthophosphate
- [Fig] EC6.1.1.-: D-Glucose + Betanidin + ATP = Isobetandin + Diphosphate + 5'-AMP
- [Fig] EC6.1.1.-: D-Glucose + Betanidin + ATP = Isobetandin + ADP + Orthophosphate
- [Fig] EC6.1.1.-: D-Glucose + Betanidin + ATP = Isobetandin + Diphosphate + 5'-AMP

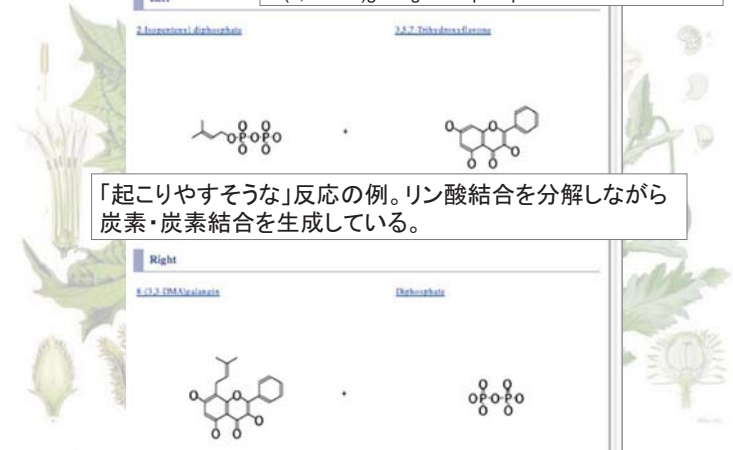
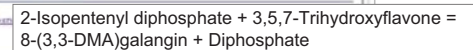
通常、ひとつの化合物ペアあたり複数の式が生成される。

#### Reported reactions

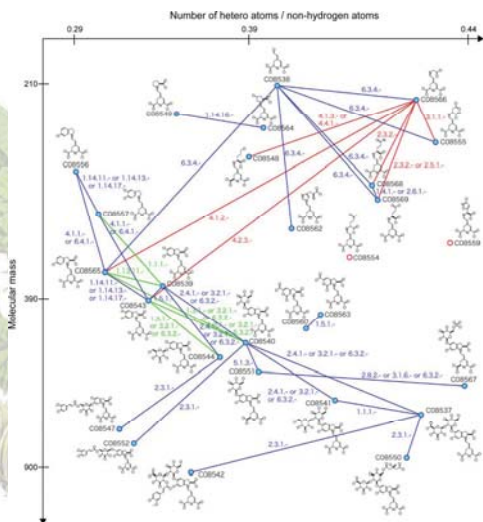
#### Suggested Reaction EC2.2.1.-



#### Suggested Reaction EC2.3.1.-



生成されたパスウェイの例



### 結論

- 3108 個の orphan metabolites 中 199 個について 490 反応を生成し、可能性のある酵素の EC 番号を推定した。
  - 既知の反応パターンと完全に一致しなくても扱える。
- 計算時間のかかるステップが多いため最後まで終了していない。
- 得られた反応の起こる可能性や生物種・遺伝子・タンパク質との関連づけについてさらなる解析が必要

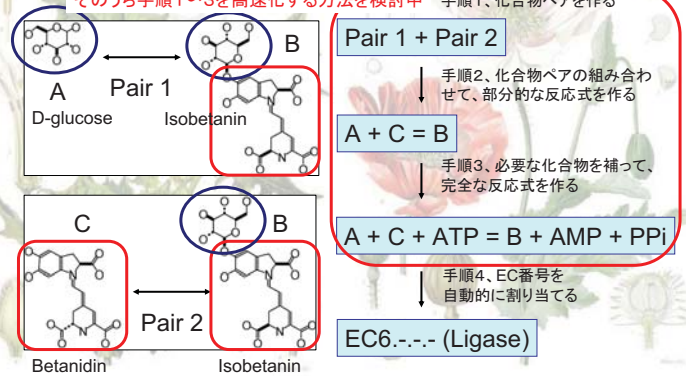


# 本日の内容

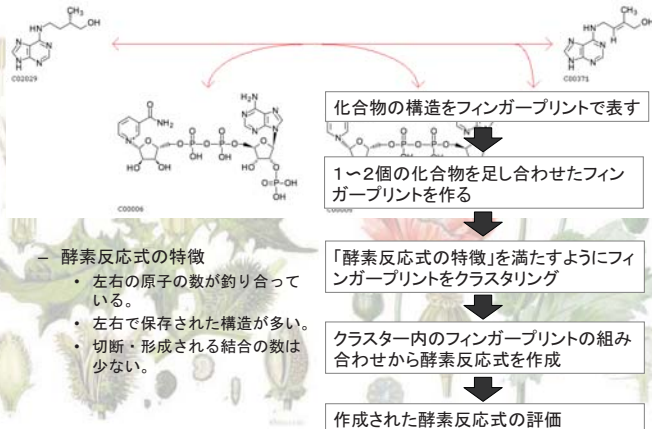
- イントロ
- 2004 年の方法 (E-zyme)
- 2008 年の方法 (GREP)
- 2008 年の方法の改良へ向けて

## GREP の改良へ向けて

- 手順 1、2、4 の計算効率が悪い  
そのうち手順 1~3 を高速化する方法を検討中



## フィンガープリントを用いた高速化の試み



## どんなフィンガープリントを設計するか？

- ATOM (原子)
  - どんな元素がいくつあるか
- BOND (結合)
  - どんな結合 (炭素-炭素など) がいくつあるか
- VICI (近傍)
  - どの原子の近傍にどの原子がいくつあるか
- RING (環)
  - どんな環状構造がいくつあるか
- SKEL (炭素骨格)
  - どんな骨格がいくつあるか

## KEGG 16,372 化合物から作成したフィンガープリント

16372 化合物上 100 化合物以上に現れるもの + 全ての Atom = 216 次元のベクトル

1. Atom_C 14001	:
2. Bond_CC 13861	183. Ring_CCNCO 106
3. Atom_O 13247	184. Skel_C2C2C2C2 100
4. Bond_CO 12955	185. Ring_CCCCCCCCCCCCCCCCCCO 100
5. Vici_OC 12371	186. Atom_* 95
6. Vici_CCC 11878	187. Atom_Na 78
7. Vici_CCCO 10068	:
8. Vici_CCCC 9250	215. Atom_Cu 2
9. Ring_CCCCC 8153	216. Atom_Cd 2
10. Vici_CC 7724	

最大 2 つまでの化合物でフィンガープリントを作るとして  
16,732 + 16,732 x 16,732 / 2 = 139,996,644 個のベクトル

## ベクトルの例 (抜粋)

↓ 化合物ID	↓ 216次元
C00002	10 6 13 5 2 0 3 0 0 0 1 5 10 0 0 1 1 3 0 1 0 1 0 0 1 0 1 0 0 0 0
C00003	21 15 14 11 5 2 6 1 0 0 2 7 14 0 1 2 3 2 3 0 2 0 0 2 0 2 0 0 0
C00004	21 15 14 11 5 2 6 1 0 0 2 7 14 0 1 2 3 2 3 0 2 0 0 2 0 2 0 0 0
C00005	21 15 17 11 4 2 6 1 0 0 2 7 14 0 1 2 3 2 3 0 2 0 0 2 0 2 0 0 0
C00006	21 15 17 11 4 2 6 1 0 0 2 7 14 0 1 2 3 2 3 0 2 0 0 2 0 2 0 0 0
C00007	0 0 2 0
C00008	10 6 10 5 2 0 3 0 0 0 1 5 10 0 1 1 3 0 1 0 1 0 0 1 0 1 0 0 0 0
C00009	0 0 4 0 1
C00010	21 14 16 9 4 1 4 0 0 2 1 7 14 0 1 2 5 2 3 0 1 0 1 1 0 1 0 0 0 0
C00011	1 0 2 2 2 0
↓ 化合物ID+化合物ID	↓ 216次元 ↓
C17168C17207	21 20 6 7 5 6 4 2 1 0 1 3 6 1 4 1 3 2 0 0 0 1 0 0 0 0 0 0 0 0
C17169C17169	22 22 0 0 0 8 0 4 2 0 0 4 8 0 4 0 4 4 0 0 0 0 0 0 0 0 0 0 0 0
C17169C17202	31 28 3 3 3 18 0 2 1 1 0 4 7 0 3 0 3 2 1 0 0 0 0 0 0 0 0 0 0
C17169C17203	21 21 1 0 0 9 0 4 2 0 0 4 7 0 3 0 3 4 0 0 0 0 0 0 0 0 0 0 0 0
C17169C17204	21 21 2 0 0 9 0 4 2 0 0 4 7 0 3 0 3 4 0 0 0 0 0 0 0 0 0 0 0 0
C17169C17205	14 13 6 3 2 4 1 2 1 1 0 2 4 1 2 0 2 2 0 0 0 0 0 0 0 0 0 0 0 0
C17169C17206	16 14 3 3 3 4 0 3 1 0 0 4 8 0 2 1 4 3 1 0 0 0 0 0 0 0 0 0 0 0
C17169C17207	17 16 6 7 5 4 4 2 1 0 1 2 4 1 2 1 2 2 0 0 0 0 0 1 0 0 0 0 0 0
C17202C17202	40 34 6 6 6 28 0 0 0 2 0 4 6 0 2 0 2 0 2 0 0 0 0 0 0 0 0 0 0 0
C17202C17203	30 27 4 3 3 19 0 2 1 1 0 4 6 0 2 0 2 2 1 0 0 0 0 0 0 0 0 0 0 0

← 139,996,644 個のベクトル →



