

# 데이터 마이닝을 활용한 효소 대사물의 분석

정희택\* · 박춘구\*\*

Enzyme Metabolite Analysis Using Data Mining

Hyi-Thaek Ceong\* · Chun-Goo Park\*\*

요 약

최근 천연물로부터 신약 후보물질을 개발하려는 연구가 활발히 이루어지고 있다. 인체 내에서 천연물은 주로 효소에 의해 대사된다. 본 연구는 화합물의 인체내 대사반응과 주로 관련된 효소에 의한 대사반응의 특징을 연관규칙마이닝을 활용하여 분석한다. 화합물이 인체 내에서 효소 대사반응과 관련된 데이터를 BRENDA(BRAunschweig ENzyme DAtabase)로부터 수집하였다. 수집된 데이터를 효소대사반응의 기본 틀에 근거하여, 대사물들을 기질대사물, 생성대사물, 억제대사물, 그리고 활성대사물들로 구분한다. 이러한 대사물들로 이루어진 기질대사물 트랜잭션, 생성대사물 트랜잭션, 그리고 모든 대사물들을 포함한 효소반응트랜잭션들을 구성하였다. 또한 중 정보를 반영한 6개의 트랜잭션들로 구성하였다. 연관규칙 마이닝을 활용하여 6개의 트랜잭션에서 빈발대사물 및 패턴을 분석하였다. 또한 대사물들 사이의 관련성을 분석하였다. 그 결과 효소대사반응에 참여하는 대사물들의 분포와 패턴을 식별할 수 있었다. 더욱이 기질에만 속하는 순수 기질대사물들을 식별하였고 이들 대부분이 아주 낮은 지지도임을 확인할 수 있었다. 연구결과는 순수 기질대사물은 효과적인 대사변환 예측 모델 개발에 활용될 수 있다.

ABSTRACT

Recently, the researches to discovery drug candidates from natural herbs have received considerable attention. In human body, enzyme mostly metabolize the compounds of natural herbs. In this study, we analysis the enzyme interactions using assoication mining. We get this data from BRENDA(BRAunschweig ENzyme DAtabase) system. Based on enzyme interaction model, we divide the metabolites into substrate metabolites, product metabolites, inhibitor metabolites, and activating metabolites. We then compose substrate metabolite transaction, product metabolite transaction with each metabolites and enzyme interaction transaction with all metabolites. Also we take account of organism for each transactions. We mine frequent metabolites and patterns from six transactions using association rule mining. And we analysis the relationship among metabolites. As a result, we identify the distributions and patterns of metabolites consist in enzyme interactions. We found that metabolites include in only substrate are identified and have very low supports. This results can be useful to develop the effective metabolism prediction model for compounds of natural herbs.

키워드

Association Rule Mining, Enzyme Interaction Metabolite, Frequent Metabolite  
연관 규칙 마이닝, 효소 대사 반응, 대사물, 빈발 대사물

\*\* 전남대학교 생명과학기술학부(chungoo@jnu.ac.kr) · Received : Aug. 29, 2016, Revised : Oct. 13, 2016, Accepted : Oct. 24, 2016

\* 전남대학교 멀티미디어전공

· 접수일 : 2016. 08. 29

· 수정완료일 : 2016. 10. 13

· 게재확정일 : 2016. 10. 24

· Corresponding Author : Hyi-thaek Ceong  
Dept. of Multimedia, Chonnam National University  
Email : htceong@jnu.ac.kr

생물학, 분자생물학, 약동학 등과 관련된 데이터들이 기하급수적으로 축적되면서, 이들 데이터를 정보기술을 활용하여 효과적으로 분석하여 의미있는 정보의 추출 및 모델링 방안들이 활발히 연구되고 있다[1-6]. 최근에는 신약개발의 일환으로, 전통지식을 통해 효능이 경험적으로 알려진 동의보감과 같은 전통처방에 포함된 천연물에 대한 연구가 활발히 이루어지고 있다[7-8]. 즉, 천연물 소재가 인체 내 다중성분-다중표적(Multi-Component Multi-Target : MCMT) 작용 원리에 의해 유효물질의 효능 증가, 약물의 독성 방지, 그리고 물질의 내성제어가 알려진 신약개발에 이르는 생체 기작규명을 위한 연구가 활발히 이루어지고 있다.

본 연구에서는 약물의 인체내 대사반응과 주로 관련이되는 효소에 의한 대사반응의 특징을 분석하고자 한다. 일반적으로 약물의 약리 효능은 인체 내 조직에 의한 흡수, 전달, 독성, 대사, 배출에 의한 약동학적(pharmacokinetics) 특징의 복잡한 상호작용으로 나타나는데, 특히 천연물 기반 외인성 대사물질의 경우 인체 주요 장기에 존재하는 효소에 의해 제 3의 물질로 대사변환 될 수 있다[9-10]. 본 연구는 인체 내 천연물 기반 외인성 물질의 인체내 제 3의 유도체로의 변환기작, 변환산물 및 해당 물질의 약리 효과를 규명하기 위한 기반 지식의 축적 및 활용을 위해, 효소를 기반으로한 대사반응 특징을 분석하고자 한다. 분석된 지식베이스는 임의의 물질에 대한 대사변환 예측 모델의 학습과정에서 활용될 수 있다. 인체에서 중요한 효소 대사반응 정보의 분석을 통해 획득한 정보를 기반으로, 특정 물질이 인체에 들어왔을 때, 변환여부를 예측할 수 있다면, 신약 후보물질 개발에 효과적인 방안이 될 수 있다. 새로이 발견된 물질이 신약으로써 의미를 갖기 위해 실험을 통해 검증은 한다면, 많은 시간과 경비를 요한다. 그러나 본 연구에서 대상으로 하는 효소반응정보를 활용한 효소반응성을 예측할 수 있는 모델을 개발할 수 있다면, 보다 효과적인 접근이 가능하다. 이를 위해, 본 연구에서는 효소 대사반응 정보를 가장 많이 가지고 있는 BRENDA<sup>1)</sup>에서 효소별 대사물반응 정보를 획득하였고 이 데이터를 기반으로 대사물질들의 특징을 분석한다.

본 연구에서 대상으로하고 있는 효소 대사반응 데

이터들은 생체내(in vivo)와 생체외 실험(in vitro)에 의해 생성되고 축적된 데이터로 다양한 대사경로(metabolic pathway) 정보를 포함하고 있다. 이러한 데이터 전체를 대상으로 한 특징 분석은 거의 이루어지지 않았기 때문에 본 연구 2장에서는 효소 대사반응과 관련된 바이오 데이터베이스들을 기술한다. 3장에서는 획득한 BRENDA 데이터를 기반으로 빈발페턴의 분석을 통해 그 특징을 제시한다. 마지막으로 4장에서는 결론 및 향후 연구를 제시한다.

## II. 대사물질 변환 관련 바이오데이터베이스

전 세계에는 약물과 약물표적 정보를 제공하는 DrugBank<sup>2)</sup>, 단백질 정보를 기반으로한 다양한 정보가 있는 UniProtKB<sup>3)</sup>, 생물학과 관련된 화학물질 정보를 제공하는 ChEMBL<sup>4)</sup>, 생화학 반응 및 경로 데이터베이스인 KEGG<sup>5)</sup> 등 많은 바이오데이터베이스들이 존재한다. 본 연구는 효소를 기반으로한 대사변환 정보의 분석을 위해, 가장 많은 관련 정보를 가지고 있는 HMDB<sup>6)</sup>[11]와 BRENDA[12] 중심으로 기술한다.

### 2.1 HMDB

HMDB(Human Metabolome Database)는 인체에서 발견되는 작은 분자의 대사(metabolism) 정보를 포함하고 있다. 제공된 데이터에는 화학적 데이터, 임상 데이터(clinical data), 분자 생물/생화학 데이터를 포함하고 있다. 검색되는 대사물질에 대한 정보로는 화학식, SMILES(Simplified Molecular-Input Line-Entry System, 화학 물질 종의 구조를 ASCII문자열로 기술한 것), InChIKey(국제적인 화학물질 식별자 InChI에 대해, 검색을 쉽게 하기 위해 해쉬된 InChI), 물리적 특성, 생물학적 위치, 관련된 질병 정보들을 제공한다. 특히, 해당 물질이 관련된 효소와 대사되는 정보를 포함하고 있다. 이러한 대사과정에 대한 정보는 KEGG와 연계되어 제공한다. 효소물질과 반응하는

1) BRENDA, <http://www.brenda-enzymes.org/>

2) DrugBank, <http://www.drugbank.ca/>

3) UniProtKB, <http://www.uniprot.org/uniprot/>

4) ChEMBL, <https://www.ebi.ac.uk/chembl/>

5) KEGG, <http://www.genome.jp/kegg/>

6) Human Metabolome Database, <http://www.hmdb.ca/>

대사물질의 특성을 분석하기 위해 HMDB에서 총 388 개의 대사 정보를 획득하고, 이와 반응이 가능한 2207 개의 효소 정보를 모두 취합하였다. 그러나, 이는 의미있는 정보를 추출하기 위해, 데이터로서 충분하지 않기 때문에 초기 분석 대상으로서 활용하였다. 이 과정에서 특히 고려한 점은 하나의 대사물질에 많은 동의어(synonym)가 존재하는 것이다. 이를 위해, 본 연구에서는 InChiKey 변환을 기반으로 하였다. 다만, 하나의 물질명에 대해, 이를 InChiKey로 변환함에 있어 CIR<sup>7)</sup>과 ChemSpider<sup>8)</sup>를 널리 사용되나, 두 시스템이 동일한 이름도 다른 InChiKey를 생성하는 경우가 있어, 가장 널리 사용되고 많은 화학물질 정보를 가지고 있는 ChemSpider의 API를 활용하여 모든 대사물질들을 InChiKey로 변환하였다.

### 2.2 BRENDA

BRENDA( BRAunschweig ENzyme DAtabase)는 EC( Enzyme Commission number) 번호를 기준으로 가장 많은 효소 정보를 포함하고 있다. 효소와 반응하는 화합물 정보인 대사물(metabolite)을 기준으로 할 때, 185,000건 이상 포함하고 있으며, 6,886개(867개의 transferred 효소를 포함)의 효소 정보를 포함하고 있다. 이러한 데이터들은 관련 논문들을 전문가에 의해 정리되어 축적된 신뢰성있는 데이터이다.

각 효소를 기반으로, 효소에 대한 구조 정보, 종(organism)별 정보, 반응 정보, 관련된 대사경로 등을 제공한다. 특히 본 연구에서 대상이 되는 효소와 대사물의 반응(enzyme-metabolite interactions) 정보를 포함하고 있다. 효소와 대사물의 반응 정보는 크게 생체의 실험에 의한 대사반응정보와 생체내 대사반응정보를 모두 포함하고 있다. 각 반응에는 대사 대상이 되는 대사물 집합인 기질(substrate), 반응 결과에 의한 대사물 집합인 생성물(product), 관련 종(organism) 명, Uniprot 번호, 가역성(reversibility) 정보를 포함하고 있다. 또한 각 대사반응에 있어 각 중별 보조인자(cofactor), 억제 화합물(inhibitor), 활성 화합물(activaing compound) 정보를 포함하고 있다.

### III. 효소 대사반응을 이루는 대사물분석

효소 대사반응의 의미는, 특정 효소가 대사물 집합인 기질(S)에 반응하여 대사물 집합인 생성물(P)을 생성함을 의미한다. 이때, 이러한 대사반응을 억제제(i)에 의해 저해되어가 활성 화합물(a)에 의해 활성화된다. 이는 수식 (1)과 같다. 예로들면 세균 Thermus에서 알코올 탈수소효소(alcohol dehydrogenase; EC 1.1.1.1)에 의한 대사반응은 그림 1과 같다. 기질을 이루는 대사물은 ‘(1S,3S)-3-methylcyclohexano’과 ‘NAD’이며, 생성물을 이루는 대사물은 ‘(rac)-3-methylcyclohexanone’, ‘NADH’, 그리고 ‘H’이다. ‘Urea’는 대사물은 조건에 따라 억제나 활성화합물로서 역할을 수행한다. 이러한 효소 대사반응은, 세포에 에너지를 공급하는 반응 뿐만아니라 생체분자의 합성을 조절하고, 해독작용, 항염증작용 등을 촉매하는 중요한 반응이다.

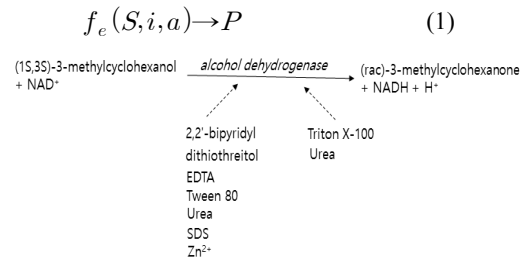


그림 1. 알코올 탈수소효소의 대사반응  
Fig. 1 Alcohol dehydrogenase Interaction

효소 대사반응 데이터의 분석을 위해, HMDB 데이터를 획득하여 그 기반 특성을 분석하였으나, 데이터 양의 문제로, BRENDA 데이터를 획득하여 분석하였다. HMDB에서 획득한 대사반응 정보는 BRENDA에서 획득한 정보에 모두 포함되어 있다. BRENDA 데이터는 웹 크롤링(Web Crawling)을 통해 획득하였다. 획득한 데이터에서 생체의 실험에 의한 반응정보, 생체내 반응정보, 억제 화합물, 활성 화합물 정보를 중심으로 분석 대상 데이터를 정리하였다. 표 1과 같이, 전체 효소에 대한 반응 정보는 374,420건이며, 생체외 반응정보는 308,341건이다. 생체외 반응정보가 생체내 반응정보를 모두 포함하고 있기 때문에 생체외 반응정보를 대상으로 분석하였다. 분석 대상 데이터에는

7) Chemical Identifier Resolver, <https://cactus.nci.nih.gov/chemical/structure>

8) ChemSpider, <http://www.chemspider.com/>

가역성 정보가 있다. 즉, 기질과 생성물의 역할이 반대가 되어 생성물이 기질이 되고 기질이 생성물이 될 수 있음을 의미한다. 308,341개의 반응에서 가역반응 23,791개가 추가되어 최종 생체의 반응정보인 332,132건을 분석 대상으로 하였다. 생체의 반응정보의 기질은 658,897개의 InChiKey 값이 존재하는 대사물로 이루어져있다. 생체의 반응정보에서 생성물은 567,903개의 대사물들로 이루어져있다. 기질과 생성물을 이루는 대사물들에서 중복성을 제거하여 기질에는 31,180개의 유일한 대사물을, 생성물에는 26,684개의 유일한 대사물을 식별하였다. 대사반응을 억제하는 억제 화합물은 중복을 포함한 200,794개의 대사물로 이루어져있으며, 대사반응을 활성화하는 활성 화합물은 중복을 포함한 23,793개의 대사물로 이루어져있다.

표 1. BRENDA 데이터  
Table 1. BRENDA data

# Record		Comment
374,420		In vivo and in vitro interaction data
308,341		In vitro data
332,132		Reversibility considered
Substrate	Product	Redundent metabolites
658,897	567,903	
31,180	26,684	Unique metabolites
200,794		Inhibitors
23,793		Activating Compounds

효소 대사반응을 분석하기 위해, 크게 세가지 방향으로 수행하였다. 첫째는 효소 대사반응에 있어 기질을 이루는 대사물(이하 기질대사물)들의 특성들을 분석하였다. 두 번째로는 효소반응에서 생성물을 이루는 대사물(이하 생성대사물)들의 특징을 분석하였다. 마지막으로 억제제와 활성 화합물을 포함한 기질, 생성물 모두의 대사물들에 대한 특성을 분석하였다. 각 분석에 있어, 효소 대사반응은 여러 종에서 동일한 형태로 발생하고 그 정보가 BRENDA에 포함되어 있다. 즉, 중복된 효소 대사반응 정보를 포함하고 있기 때문에, 중복 정보를 반영한 중복된 정보와 중복 정보를 제거한 정보를 대상으로 각각 분석하였다. 이러한 데이터의 분석을 위해 연관규칙생성[13]에서 활용하는

빈발항목을 활용하였다. 연관규칙은 전체 집합  $N$ 에서, 빈발항목으로부터 규칙을 발견하는 방법으로 수식 (2)와 같이, 지지도(support), 신뢰도(confidence), 향상도(lift)의 척도를 활용한다. 이는 데이터들의 공통된 통계적 특징을 분석하기 위한 방안이며 패턴분석을 위한 방안이다. 또한 대사반응 예측 모델 개발에 활용하기 위함이다. 빈발항목 분석은 시스템 R에서 패키지 arules<sup>9)</sup>[13]을 활용하였으며, 이를 시각화하기 위해 패키지 arulesViz<sup>10)</sup>를 활용하였다[14].

$$X \rightarrow Y \begin{cases} \text{Support} = \frac{\text{freq}(X,Y)}{N} \\ \text{Confidence} = \frac{\text{freq}(X,Y)}{\text{freq}(X)} \\ \text{Lift} = \frac{\text{Support}}{\text{Support}(X) * \text{Support}(Y)} \end{cases} \quad (2)$$

### 3.1 대사반응에서 기질대사물들의 분석

효소 대사반응에서 기질을 이루는 31,180개의 대사물들을 분석하기 위해, 하나 이상의 대사물을 포함한 기질 정보만을 대상으로 하였다. 이는 InChiKey 변환 과정에서 변환되지 않는 대사물만을 포함한 기질이 존재하기 때문이다. 332,132개의 트랜잭션에서 329,504개의 분석대상 최종 트랜잭션을 식별하였다. 대사반응에 참여하는 기질대사물을 분석하기 위해 두 개의 트랜잭션집합으로 구분하였다. 첫째, 다양한 종에서 확인된 효소 대사반응 정보를 중복 포함하고 있는, 329,504개의 트랜잭션으로 이루어진 기질대사물전체 트랜잭션집합이다. 둘째, 중복 구분을 제거한 효소 대사반응 정보만을 포함한, 53,022개의 기질대사물유일 트랜잭션집합이다. 전자는 다양한 종과 실험에서 축적된 전체 정보를 대상으로 분석한다는 의미를 가지며, 후자는 대사반응 정보 자체만을 대상으로 분석한다는 의미를 가진다.

9) Package 'arules', <https://cran.r-project.org/web/packages/arules/arules.pdf>, 2016.

10) Package 'arulesViz', <https://cran.r-project.org/web/packages/arulesViz/arulesViz.pdf>, 2016.

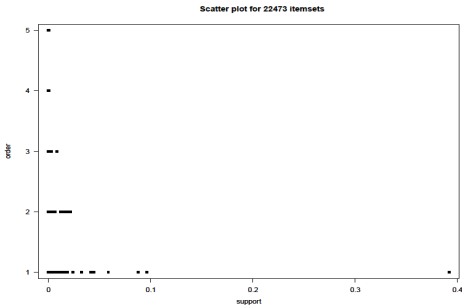


그림 2. 지지도 0.00001이상의 빈발 기질대사물 분포

Fig. 2 Frequent substrate's metabolites distribution with support  $\geq 0.00001$

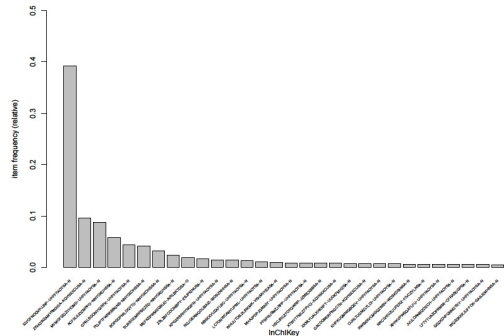


그림 4. 상위 30 기질대사물

Fig. 4 Top 30 substrate's metabolites

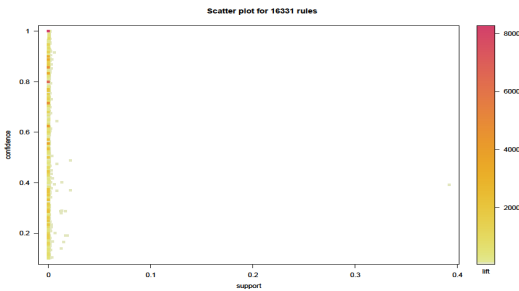


그림 3. 지지도 0.00001이상의 기질대사물 규칙 분포

Fig. 3 Substrate's metabolite rules distribution with support  $\geq 0.00001$

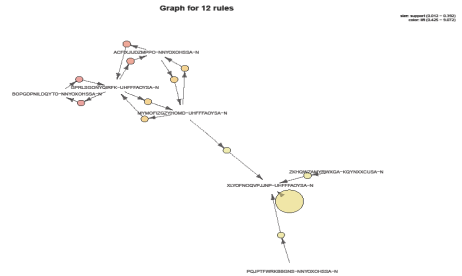


그림 5. 지지도 0.01에서 빈발 기질대사물들의 규칙 시각화

Fig. 5 Rules visualization for frequent substrate's metabolites with support  $\geq 0.01$

### 3.1.1 기질대사물전체 트랜잭션집합 분석

표 2. 기질대사물전체 트랜잭션 분석

Table 2. All substrate's metabolite transactions analysis

Support	Frequent metabolites	Single frequent metabolites	Rules
0.01	21	14	12
0.001	288	200	103
0.0001	4,017	1,581	2,892
0.00001	22,473	7,497	16,331
0.000007	28,728	9,510	20,634
0.000004	42,964	14,037	30,603
0.000003	96,611	31,180	68,835

먼저 기질대사물전체 트랜잭션들에 대한 분석결과 는 표 2와 같다. 신뢰도는 0.1에서 지지도를 달리하여

실험하였다. 지지도 0.0001에서 빈발기질대사물은 4,017개 였으며, 이때 생성되는 규칙은 2,892개 였다. 여기서 빈발기질대사물 중 대사물이 하나만 존재하는 경우는 1,581개였다.

기질대사물전체 트랜잭션에서 지지도 0.00001이상을 만족하는 빈발기질대사물의 특징은 그림 2와 같다. 빈발기질대사물은 최대 5개의 대사물로 이루어져 있으며 낮은 지지도를 확인할 수 있다. 또한 이러한 빈발기질대사물들로 이루어진 규칙의 분포는 그림 3과 같다. 이를 통해, 낮은 지지도에서 높은 신뢰도와 향상도를 확인할 수 있고, 많은 기질대사물이 넓게 분포되어 있지 않음을 확인할 수 있다. 이를 구체적으로 확인하기 위해, 지지도 상위 30개에 속하는 기질대사물들을 분석하였더니 그림 4와 같다. 즉, 대부분의 기

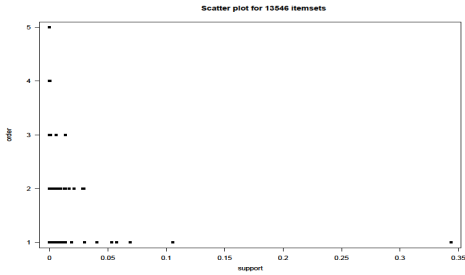


그림 6. 지지도 0.00003이상의 빈발기질대사물 분포

Fig. 6 Frequent substrate's metabolites distribution with support  $\geq 0.00003$

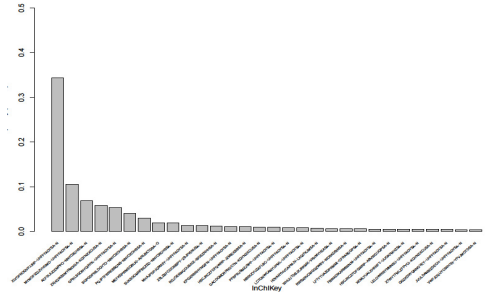


그림 8. 상위 30 기질대사물

Fig. 8 Top 30 substrate's metabolites

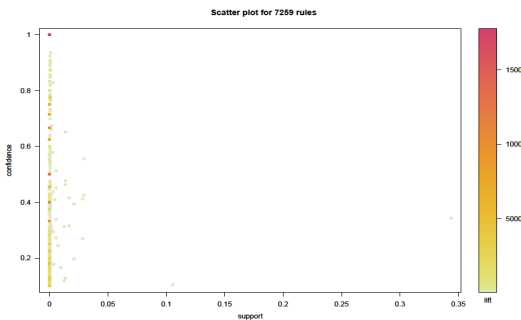


그림 7. 지지도 0.00003에서 기질대사물 규칙 분포  
Fig. 7 Substrate's metabolite rules distribution with support  $\geq 0.00003$

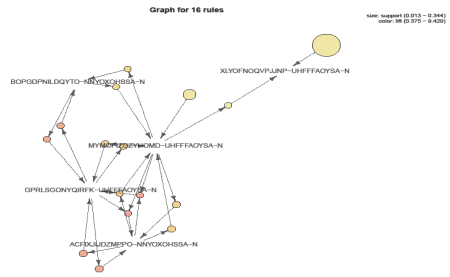


그림 9. 지지도 0.01에서 빈발 기질대사물 규칙 시각화

Fig. 9 Rules visualization for frequent substrate's metabolites with support  $\geq 0.01$

질대사물은 낮은 지지도임을 알 수 있다. 그림에서 제시된 기질대사물은 InChiKey 값들로 이들 중 상위 10 개에 속하는 화합물 명칭은 H<sub>2</sub>O, ATP( Adenosine Triphosphate), O<sub>2</sub>, NADPH, H, NAD<sup>+</sup>, NADP<sup>+</sup>, S-(5'-Adenosyl)-L-methionine conjugate acid, Acetyl-CoA, α-Ketoglutaric acid, Coenzyme A 순이다. ATP, NADH 등은 화학에너지는 대사에 있어 생체 에너지 대사의 중심이 되는 대사물이다. S-(5'-Adenosyl)-L-methionine conjugate acid는 메틸기 전송, 황전환작용(transsulfuration) 그리고 아미노프로필레이션(aminopropylation)에 참여하는 대사물이며, α-Ketoglutaric acid는 생체내에서 유기산과 아미노산의 대사에 중요한 역할을 수행하는 알파케토글루타르산이다[10].

두 개 이상의 기질대사물 정보를 포함한 빈발기질대사물들의 패턴을 일부를 기술하면 다음과 같다. {(E)-Ferulic acid, O<sub>2</sub>}, {Ethanol, Stearic acid}, {5-Amino-1-(5-O-phosphono-β-D-ribofuranosyl)-1H-imidazole-4-carboxylic acid, ATP}, {L-(+)-Aspartic acid, O<sub>2</sub>, H<sub>2</sub>O}, {NADPH, H, O<sub>2</sub>, Arachidonic acid}, {Carbon dioxide, Ammonia, H<sub>2</sub>O, ATP}, {reduced ferredoxin, H., Acetylene, H<sub>2</sub>O, ATP}. 마지막으로 이러한 빈발기질대사물들에 대한 규칙을 arulesViz를 활용하여 시각화하였다. 지면상의 이유로 지지도 0.01을 만족하는 빈발대사물 21개에 의한 12개의 규칙에 대해서만 수행하였다. 그림 5에 표현된 노드들은 기질대사물의 InChIKey이고 원은 향상도(lift)를 비례적으로 표현한 것이다.

### 3.1.2 기질대사물유일 트랜잭션집합 분석

기질대사물유일 트랜잭션 집합에 속하는 53,022개의 트랜잭션들에 대한 빈발기질대사물 및 규칙은 표 3과 같다. 그림 6에서 빈발기질대사물은 최대 5개의 대사물로 이루어져 있으며 낮은 지지도를 확인할 수 있다. 빈발기질대사물들로 이루어진 규칙의 분포는 그림 7과 같고, 지지도 상위 30개에 속하는 기질대사물들을 분석하였더니 그림 8과 같다. 즉, 대부분의 기질대사물은 낮은 지지도를 가짐을 알 수 있다. 상위 10개에 속하는 화합물 명칭은, 기질대사물전체 트랜잭션 집합 분석과 달리 H<sub>2</sub>O, O<sub>2</sub>, NADPH, ATP, H, NADH, NAD<sup>+</sup>, S-(5'-Adenosyl)-L-methionine conjugate acid, NADP<sup>+</sup>, 과산화수소(Hydrogen peroxide), Acetyl-CoA, Coenzyme A 순이었다. 두 개 이상의 기질대사물을 포함 빈발기질대사물은, 기질대사물전체 트랜잭션 분석 결과와 지지도에서만 차이가 있을뿐 동일한 내용을 포함하고 있었다. 이는 두 트랜잭션이 기질대사물 31,180개를 동일하게 포함하고 있기 때문이다. 마지막으로 규칙을 시각화 하면 그림 9와 같다.

표 3. 기질대사물유일 트랜잭션 분석  
Table 3. Unique substrate' metabolite transactions analysis

Support	Frequent metabolites	Single frequent metabolites	Rules
0.01	21	14	16
0.001	160	134	37
0.0001	1,922	1,462	403
0.00005	5,432	3,489	1,839
0.00003	13,546	6,478	7,259
0.00001	96,611	31,180	74,547

### 3.2 대사반응에서 생성대사물 분석

효소 대사반응에서 26,684개의 대사물로 이루어진 생성대사물들의 특성을 분석하기 위해, 기질대사물 트랜잭션분석과 동일하게 하나 이상의 대사물을 포함한 생성물 정보만을 대상으로 하였다. 332,132개의 트랜잭션에서 287,917개의 분석대상 최종 트랜잭션을 식별하였다. 생성대사물 트랜잭션들을 분석하기 위해 두 개의 생성대사물 트랜잭션집합으로 구분하였다. 첫째,

다양한 중에서 확인된 효소 대사반응 정보를 중복 포함하고 있는, 287,917개의 트랜잭션으로 이루어진 생성대사물전체 트랜잭션집합이다. 둘째, 종별 구분을 제거한 효소 대사반응 정보만을 포함한, 43,592개의 생성대사물유일 트랜잭션집합이다.

### 3.2.1 생성대사물전체 트랜잭션집합 분석

생성대사물전체 트랜잭션들 279,881개에 대한 분석 결과는 표 4와 같다. 비교를 위해, 대사반응에 있어 기질대사물 분석방법과 동일하게 신뢰도는 0.1에서 지지도를 달리하여 실험하였다. 지지도 0.0001에서 빈발 생성대사물은 4,510개였으며, 이때 생성되는 규칙은 3,683개였다. 여기서 빈발생성대사물 중 대사물이 하나만 존재하는 경우는 1,673개였다.

생성대사물전체 트랜잭션에서 지지도 0.00001이상을 만족하는 빈발생성대사물의 특징은 그림 10과 같다. 빈발생성대사물은 최대 8개의 대사물로 이루어져 있으며 낮은 지지도를 확인할 수 있다. 또한 이러한 빈발생성대사물들로 이루어진 규칙의 분포는 그림 11과 같다. 이를 통해, 낮은 지지도에서 높은 신뢰도와 향상도를 확인할 수 있었다. 상위 30개에 속하는 생성대사물들을 분석하였더니 그림 12와 같다. 기질대사물들과 비교하면, 대부분의 생성대사물들이 보다 낮은 지지도임을 알 수 있다. 상위 10개에 속하는 화합물 명칭은 인산(Phosphoric acid), ADP(Adenosine Diphosphate), H<sub>2</sub>O, NADP<sup>+</sup>, 피로인산(Pyrophosphoric Acid), 암모니아(Ammonia), 이산화탄소(Carbon dioxide), NADH, Coenzyme A, H<sup>+</sup> 순이다. 앞서 분석한 빈발기질대사물과 달리 인산, 피로인산, 암모니아, 이산화탄소 등이 빈발생성대사물로 식별되었다.

표 4. 생성물전체 트랜잭션 분석  
Table 4. All product's metabolite transactions analysis

Support	Frequent metabolites	Single frequent metabolites	Rules
0.01	27	23	8
0.001	323	225	103
0.0001	4,510	1,673	3,683
0.00001	26,765	8,686	22,052
0.000007	39,464	12,584	33,959
0.000004	39,464	12,584	33,959
0.000003	130,741	26,684	375,995

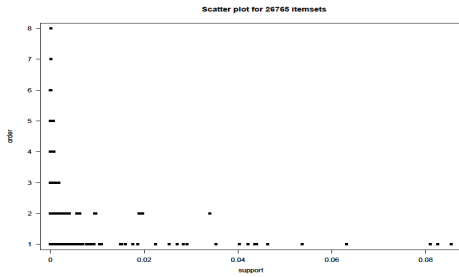


그림 10. 지지도 0.00001에서 빈발생성대사물 분포  
Fig. 10 Frequent product's metabolites distribution with support  $\geq 0.00001$

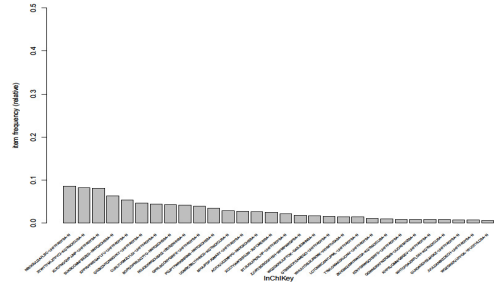


그림 12. 상위 30 생성대사물  
Fig. 12 Top 30 product's metabolites

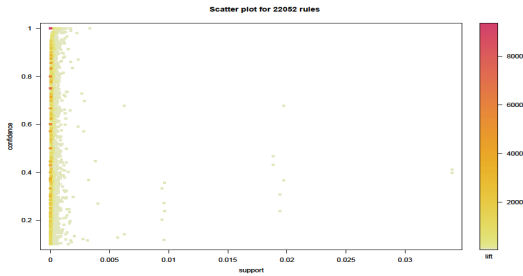


그림 11. 지지도 0.00001에서 생성대사물 규칙 분포  
Fig. 11 Product's metabolite rules distribution with support  $\geq 0.00001$

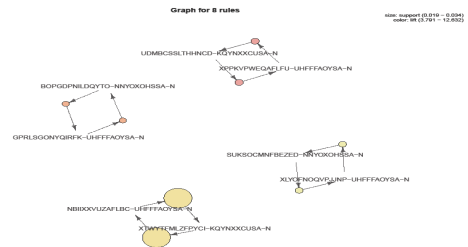


그림 13. 지지도 0.01에서 빈발 생성대사물의 규칙 시각화  
Fig. 13 Rules visualization for frequent product's metabolites with support  $\geq 0.01$

두 개 이상의 대사물 정보를 포함 빈발생성대사물 일부를 기술하면 다음과 같다. {Galactose 1-phosphate, ADP}, {L-threo-2,3-Hexodiulosonic Acid g-Lactone, DL-Noradrenaline}, {Carbon dioxide, 2-(6-Hydroxy-1,3 - benzothiazol-2-yl)-1,3-thiazol-4(5H)-one, AMP, H<sub>2</sub>O, Pyrophosphoric Acid}, {Formic acid, L-(-)-methionine, (4-Amino-2-methyl-5-pyrimidinyl)methyl dihydrogen phosphate, carbon monoxide, 5'- Deoxyadenosine}. 마지막으로 그림 13과 같이 지지도 0.01을 만족하는 빈발생성물 27개에 의한 8개의 규칙을 시각화 하였다. 기질대사물 분석과 달리 각 규칙들이 분리되어 있음을 알 수 있다.

### 3.2.2 생성대사물유일 트랜잭션집합 분석

표 5. 생성대사물 유일 트랜잭션 분석  
Table 5. Unique product's metabolite transactions analysis

Support	Frequent metabolites	Single frequent metabolites	Rules
0.01	24	20	8
0.001	195	169	28
0.0001	2,379	1,723	725
0.00005	5,103	3,321	2,035
0.00003	13,018	6,188	8,531
0.00001	130,741	26,684	381,470

생성대사물유일 트랜잭션 집합에 속하는 43,592개의 트랜잭션에 대한 빈발생성대사물 및 규칙은 표 5와 같다. 그림 14에서 빈발생성대사물은 최대 8개의



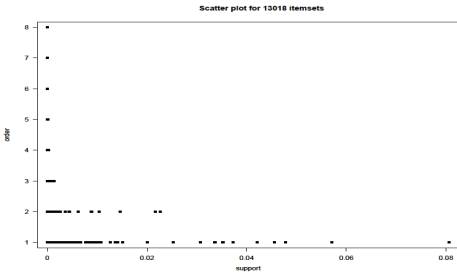


그림 14. 지지도 0.00003에서 빈발 생성대사물의 분포  
 Fig. 14 Frequent product's metabolites distribution with support  $\geq 0.00003$

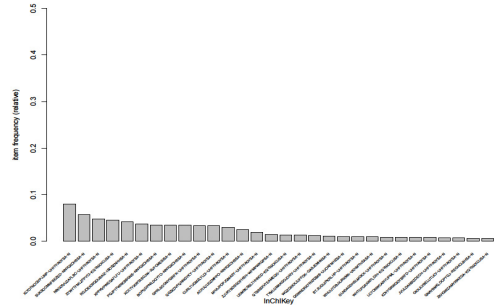


그림 16. 상위 30 생성대사물  
 Fig. 16 Top 30 product's metabolites

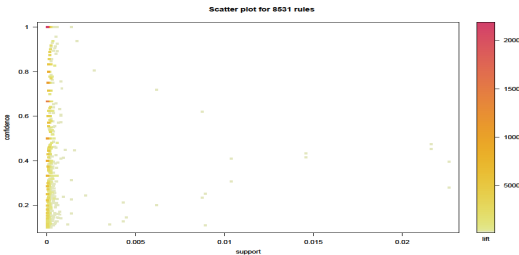


그림 15. 지지도 0.00003에서 생성대사물 규칙 분포  
 Fig. 15 Product's metabolite rules distribution with support  $\geq 0.00003$

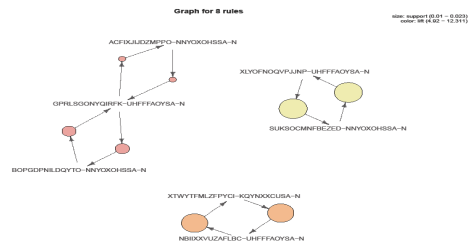


그림 17. 지지도 0.01에서 빈발 생성대사물의 규칙 시각화  
 Fig. 17 Rules visualization for frequent product's metabolites with support  $\geq 0.01$

대사물로 이루어져 있으며 낮은 지지도를 확인할 수 있다. 빈발생성대사물들로 이루어진 규칙의 분포는 그림 15와 같고, 지지도 상위 30개에 속하는 생성대사물들을 분석하였더니 그림 16과 같다. 기질대사물과 비교해도 대부분의 생성대사물들은 보다 낮은 지지도임을 알 수 있다. 상위 10개에 속하는 화합물 명칭은, 생성물전체 트랜잭션집합 분석과 달리 H<sub>2</sub>O, NADP<sup>+</sup>, 인산, ADP, Coenzyme A, 피로인산, UDP(uridine 5'-diphosphate), NAD<sup>+</sup>, NADH, H<sup>+</sup> 순이었다. 두 개 이상의 생성대사물로 이루어진 빈발대사물 패턴은 생성대사물전체 트랜잭션 분석 결과와 지지도에만 차이가 있을뿐 동일한 내용을 포함하고 있었다. 이는 두 트랜잭션 모두가 생성대사물 26,694개를 동일하게 포함하고 있기 때문이다. 마지막으로 규칙을 시각화 화면 그림 17과 같다.

### 3.3 기질, 생성물, 억제제, 활성 화합물의 대사반응 대사물 분석

수식 (1)과 같이 효소에 의한 대사반응에 있어, 반응을 억제하거나 활성화하는 것에는 억제제와 활성화 화합물이 중요한 역할을 한다. 여기서는 생체의 반응 정보 332,132개에 억제제와 활성 화합물 정보를 포함한 데이터에 대한 분석을 수행한다. 먼저, 억제제와 활성 화합물 정보를 해당 효소 대사반응에 정확히 반영하기 위해, 대사반응 정보에서 문헌정보, 종 정보, Uniprot 번호 정보를 활용하여, 효소대사물 반응 정보 278,785개를 생성하였다. BRENDA가 실험 논문에 대한 결과를 수집 가공하여 제공하고 있기 때문에, 이러한 정보를 활용하여 효소대사물 반응 데이터를 생성하였고, InChIKey 획득이 가능한 기질대사물과 생성대사물 모두를 포함한 대사반응만을 대상으로 하였다.

대사물 79,724개로 구성된, 효소대사물 반응정보

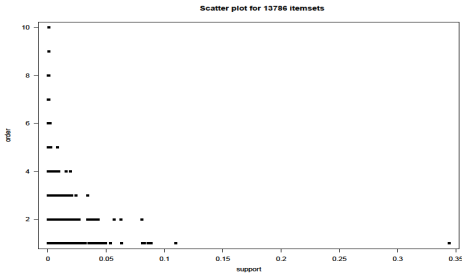


그림 18. 지지도 0.0005에서 빈발 대사물 분포  
Fig. 18 Frequent metabolites distribution with support  $\geq 0.0005$

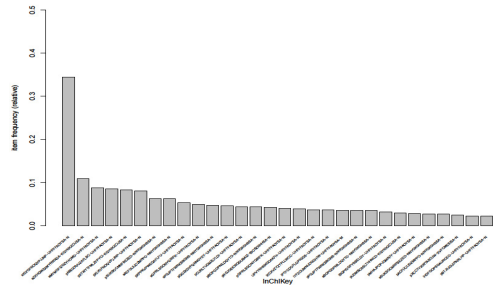


그림 20. 상위 30 대사물  
Fig. 20 Top 30 Metabolites

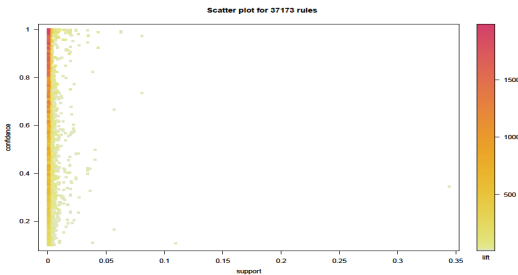


그림 19. 지지도 0.0005에서 대사물 규칙 분포  
Fig. 19 Metabolite rules distribution with support  $\geq 0.0005$

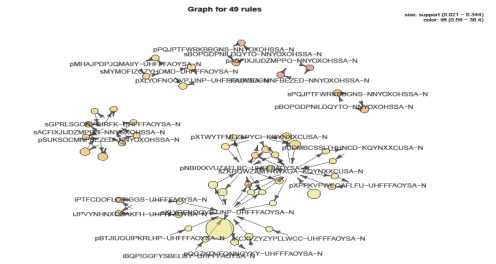


그림 21. 지지도 0.02에서 빈발대사물의 규칙 시각화  
Fig. 21 Rules visualization for frequent metabolites with support  $\geq 0.02$

278,785개는, 기질대사물 25,401개, 생성대사물 26,535개, 억제제를 이루는 대사물(이하 억제대사물) 41,151개 그리고 활성 화합물을 이루는 대사물(이하 활성대사물) 2,680개로 이루어져 있다. 효소대사물 반응 중에서 가장 많은 대사물을 갖는 대표적인 경우는, 효소 3-oxo-5 $\alpha$ -steroid 4-dehydrogenase(EC 1.3.99.5)에서 인류(homo sapiens)중에 대한 대사반응으로, 억제대사물이 대부분으로 350개 대사물로 이루어져 있다.

패턴분석의 일관성을 위해, 효소대사반응을 두 개의 효소대사반응 트랜잭션집합으로 구분하였다. 첫째, 다양한 종에서 확인된 효소 대사반응의 중복을 포함하고 있는, 278,785개의 트랜잭션으로 이루어진 효소 대사반응전체 트랜잭션집합이다. 둘째는, 종별 구분을 제거한 효소 대사반응만으로 구성된, 124,005개의 효소대사반응유일 트랜잭션집합이다.

### 3.3.1 효소대사반응전체 트랜잭션집합 분석

표 6. 효소대사반응전체 트랜잭션 분석  
Table 6. All enzyme interaction transactions analysis

Support	Frequent metabolites	Single frequent metabolites	Rules
0.02	56	31	49
0.01	128	59	146
0.001	3,728	658	7,514
0.0007	6,886	948	15,503
0.0005	13,786	1,323	37,173

효소대사반응전체 트랜잭션들 278,785개에 대한 분석 결과는 표 6과 같다. 많은 대사물 개수를 고려하여 신뢰도는 0.1에서 지지도를 달리하여 실험하였다. 대사물들은 기질대사물, 생성대사물, 억제대사물, 활성대사물 모두를 포함한다. 지지도 0.0005에서 빈발대사물은 13,786개였으며, 이때 생성되는 규칙은 37,173개

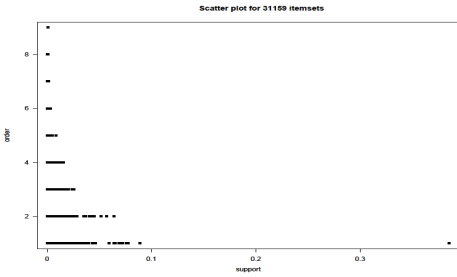


그림 22. 지지도 0.0005에서 빈발 대사물 분포  
Fig. 22 Frequent metabolites distribution with support  $\geq 0.0005$

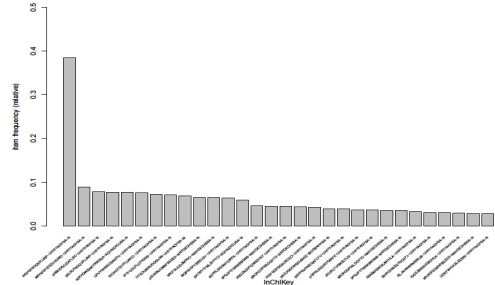


그림 24. 상위 30 대사물  
Fig. 24 Top 30 Metabolites

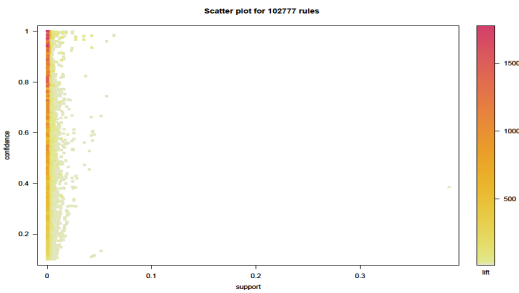


그림 23. 지지도 0.0005에서 대사물 규칙 분포  
Fig. 23 Metabolite rules distribution with support  $\geq 0.0005$

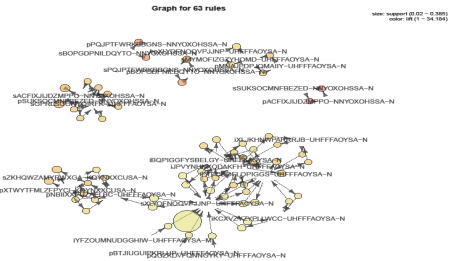


그림 25. 지지도 0.02에서 빈발대사물의 규칙 시각화  
Fig. 25 Rules visualization for frequent metabolites with support  $\geq 0.02$

였다. 여기서 빈발대사물 중 대사물이 하나만 존재하는 경우는 1,323개였다.

효소대사반응전체 트랜잭션에서 지지도 0.0005이상을 만족하는 빈발 대사물의 특징은 그림 18과 같다. 빈발대사물은 최대 10개의 대사물로 이루어져 있으며, 기질대사물 및 생성대사물 분석과 유사한 낮은 지지도를 확인할 수 있다. 또한 이러한 빈발대사물들로 이루어진 규칙의 분포는 그림 19와 같다. 기질대사물 및 생성대사물 분석보다 상대적으로 높은 지지도를 확인할 수 있고, 많은 대사물이 넓게 분포되어 있음을 확인할 수 있다. 이는 억제대사물의 결과이다. 이를 구체적으로 확인하기 위해, 지지도 상위 30개에 속하는 대사물들을 분석하였다. 그림 20과 같다. 대부분의 대사물은 낮은 지지도 값을 알 수 있다. 상위에 속하는 대사물로부터 기질과 생성물에 속하는 대사물은 앞선 분석과 거의 동일한 순서였으며, 억제대사물

은 Cu(II), Ethylenediaminetetraacetic acid, Zinc 2+, (4-Carboxyphenyl)(chloro)mercury,mercury(II) cation 순이었으며, 활성대사물은 D-1,4 - dithiothreitol, 29-[4-(2,4,4-Trimethyl-2-pentanyl)phenoxy]-3,6,9,12,15,18,21,24,27-nonaoxanonacosan-1-ol, mercaptoethanol, Ethylenediaminetetraacetic acid, L-(+)-Cysteine 순이었다.

두 개 이상의 대사물 정보를 포함 빈발대사물들 일부를 기술하면 다음과 같다. {<sup>sm</sup>NADPH, <sup>im</sup>carbon monoxide, <sup>pm</sup>NADP<sup>+</sup>}, {<sup>sm</sup>H<sub>2</sub>O, <sup>im</sup>Cu(II), <sup>im</sup>NICKEL(II) ION, <sup>im</sup>cobalt(II) cation, <sup>pm</sup>Ammonia}, {<sup>sm</sup>[(2R,3S,4R,5R)-5-(6-Amino-9H-purin-9-yl)-3,4-dihydroxytetrahydro-2-furanyl]methyl(2R,3R,4S,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)tetrahydro-2H-pyran-2-yl dihydrogen diphosphate, <sup>am</sup>Pyridoxal 5'-phosphate, <sup>im</sup>AMP, <sup>pm</sup>ATP}. 대사물의 윗첨자는 기질대사물(sm),

생성대사물(*pm*), 억제대사물(*im*), 그리고 활성대사물(*am*)을 의미한다. 그림 25는 지지도 0.02을 만족하는 빈발항목 56개에 의한 49개 규칙의 시각화다.

### 3.3.2 효소대사반응유일 트랜잭션집합 분석

표 7. 효소대사반응유일 트랜잭션 분석

Table 7. Unique enzyme interaction transactions analysis

Support	Frequent metabolites	Single frequent metabolites	Rules
0.02	71	41	63
0.01	224	76	316
0.001	8,262	794	21,960
0.0007	16,257	1,118	48,204
0.0005	31,159	1,569	102,777

효소대사반응유일 트랜잭션 집합에 속하는 124,005 개의 트랜잭션에 대한 빈발대사물 및 규칙은 표 7과 같다. 그림 22에서 빈발대사물은 최대 9개의 대사물로 이루어져 있으며 낮은 지지도를 확인할 수 있다. 빈발 대사물들로 이루어진 규칙의 분포는 그림 23과 같고, 지지도 상위 30개에 속하는 대사물들을 분석하였다니 그림 24와 같다. 즉, 대부분의 대사물은 낮은 지지도 를 가짐을 알 수 있다. 상위 10개에 속하는 화합물 명칭은, 앞서 효소대사물반응전체 트랜잭션 분석에서 파악한 결과와 거의 유사하였다. 몇 개의 대사물만이 지지도에 따라 앞뒤로 순서가 바뀌는 정도였다.

두 개 이상의 대사물 정보를 포함 빈발대사물들 일부를 기술하면 다음과 같다. {<sup>sm</sup>H<sub>2</sub>O, <sup>im</sup>AMPe, <sup>im</sup>ADP, <sup>im</sup>ATP, <sup>pm</sup>AMP}, {<sup>sm</sup>NAD<sup>+</sup>, <sup>sm</sup>H<sub>2</sub>O, <sup>am</sup>mercaptoethanol, <sup>pm</sup>NADH, <sup>pm</sup>H<sup>+</sup>}, {<sup>sm</sup>NAD<sup>+</sup>, <sup>am</sup>D-1,4-dithiothreitol, <sup>im</sup>Cu(II), <sup>pm</sup>NADH, <sup>pm</sup>H<sup>+</sup>}. 마지막으로 지지도 0.02에서 생성된 대사물들의 규칙 시각화는 그림 25와 같다.

### 3.4 대사물들사이의 관련성 분석

앞서, 기질대사물들로만 구성된 기질대사물전체 트랜잭션집합과 기질대사물유일 트랜잭션집합, 생성대사물들로만 구성된 생성대사물전체 트랜잭션집합과 생성대사물유일 트랜잭션집합, 그리고 억제대사물, 활성대사물, 기질대사물, 생성대사물들 모두로 구성된 효소대사반응전체 트랜잭션집합과 효소반응유일 트랜잭션집합의 분석을 통해, 각 대사물의 특징과 패턴들을

분석하였다. 여기서는 대사물들 사이의 관련성을 분석하고자 한다.

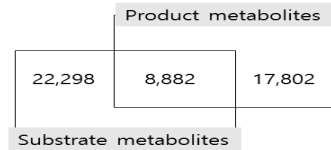


그림 26. 기질대사물과 생성대사물의 관계  
Fig. 26 Relationship between substrate's and product's metabolites

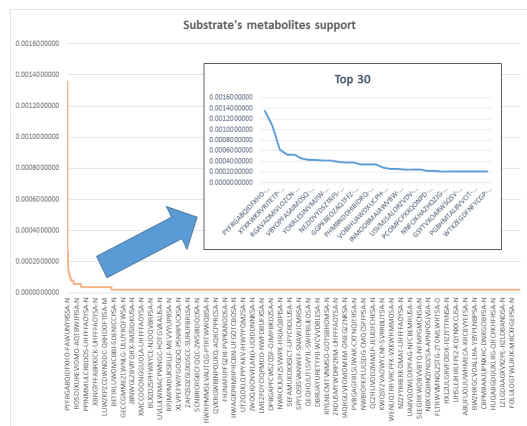
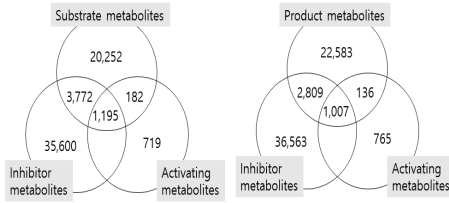


그림 27. 순수 기질대사물의 지지도  
Fig. 27 Supports of pure substrate's metabolites

먼저, 기질대사물과 생성대사물 트랜잭션 분석에서, 기질대사물과 생성대사물을 이루는 대사물은 각각 31,180개와 26,684개였으며, 공통된 대사물은 8,882개였다. 그림 26에 정리한바와 같이 기질에만 속하는 순수 기질대사물은 22,298개였다. 이것은 물질의 대사반응 여부를 판단하는 효소반응성 예측 모델 개발에 활용될 수 있다. 즉, 22,298개의 기질 대사물만을 모델 개발의 학습데이터로 활용함으로써 보다 효과적인 학습이 가능하다. 또한 생성대사물에만 속하는 17,802개를 활용하여 개발 모델을 검증할 수 있다. 한편 순수 기질대사물 22,298개의 지지도 분포는 그림 27과 같이, Top 30 이후에 있는 대사물들은 변화가 거의 없는 아주 낮은 지지도 를 가지고 있었다. 구체적으로 전체 22,298개의 순수 기질대사물 중에서 대사반응에 두 번 이상 참여하는 대사물은 681개(3%)뿐이었고 대부분 (21,617개)이 하나의 대사반응에만 속했다.



Substrate metabolites  $\cap$  product metabolites  $\cap$  inhibitor metabolites  $\cap$  Activating metabolites = 959  
 Substrate metabolites  $\cap$  product metabolites  $\cap$  inhibitor metabolites  $\cap$  Activating metabolites = 15,089

그림 28. 효소대사반응에서 대사물들의 관련성  
 Fig. 28 Metabolites relationship in enzyme interactions

다음으로, 효소대사반응전체 트랜잭션집합과 효소 대사반응유일 트랜잭션집합의 분석에서 획득한 대사물의 특징을 분석하였다. 이들 트랜잭션들을 이루는 대사물로는 기질대사물 25,401개, 생성대사물 26,535개, 억제대사물 41,151개, 그리고 활성대사물 2,680개였다. 이들 사이의 관계를 정리하면 다음 그림 28과 같다. 기질대사물, 생성대사물, 억제대사물, 활성대사물 모두에 속하는 대사물은 959개였으며, 기질대사물에만 속하고 생성대사물, 억제대사물, 활성대사물에 속하지 않는 대사물은 15,089개였다. 15,089개의 순수 기질대사물만을 대상으로 효소반응성 예측 모델 개발에 활용될 수 있다. 예측 모델 개발에 있어 그림 26과 28을 비교할 때, 과적합(overfitting) 문제를 해결하기 위한 방안으로써 의미를 갖는다.

#### IV. 결론 및 향후 연구

본 연구는 천연물 기반 외인성 물질의 인체내 제 3의 유도체로의 변환기작, 변환산물 및 해당 물질의 약리 효과를 규명하기 위한 기반 지식의 구축을 위해, 약물의 인체내 대사반응과 주로 관련이되는 효소 대사반응의 특징을 분석하였다. 화합물이 인체 내에서 효소 대사반응과 관련된 데이터를 BRENDA로부터 수집하였다. 수집된 데이터를 효소대사반응의 기본 틀에 근거하여, 기질대사물들에 의한 트랜잭션집합, 생성대사물들에 의한 트랜잭션집합, 기질대사물, 생성대사물, 억제대사물, 그리고 활성대사물 전부를 포함한 트랜잭션집합으로 구분하였다. 각각의 트랜잭션집합에서 중별 차이를 허용하는 경우와 중별 차이를 제거한, 전체 트랜잭션 및 유일 트랜잭션으로 구분하였다. 구

분된 6종의 트랜잭션집합에 대한 대사물의 특징 및 패턴을 각각 분석하였다. 또한 대사물들 사이의 관련성을 분석하였고, 순수 기질대사물들을 식별하였다. 분석 결과 ATP 및 환원형 전자운반체(NADH, NADPH 등)를 제외한 많은 대사물들이 아주 낮은 지지도를 가짐을 확인할 수 있었다. 이러한 데이터는 신경망(Neural network)이나 SVM 방법[15-16]들을 활용한 대사변환 예측 모델 개발에 있어 학습 데이터 및 검증 데이터로서 활용할 수 있다. 또한 예측 모델 개발에 있어 특징 선택(feature selection)[17]이전에 효과적인 학습 데이터 선정에 있어 의미가 있다.

향후 연구에서는 최종 대사변환 예측 모델 개발에 앞서, 기질대사물들의 화학적 특성을 분석할 계획이다. 즉, 31,180개, 22,298개, 그리고 15,089개의 대사물들의 화학적 특성을 추출하고 이를 군집화(clustering)하여, 이들 기질대사물들의 특징을 분석하고자 한다. 본 연구에서 분석한 결과와 군집화 분석 결과를 기반으로 최종 대사변환 예측 모델을 개발하고자 한다.

#### 감사의 글

This work was supported by the Bio-Synergy Research Project(NRF-2015M3A9C4075820) of the Ministry of Science, ICT and Future Planning through the National Research Foundation.

#### References

- [1] A. Tarca, V. Jarey, X. Chen, R. Romero, and S. Drăghici, "Machine Learning and Its Applications to Biology," *J. of Public Library of Science(PLOS) Computational Biology*, vol. 3, issue 6, 2007, pp. 953-963.
- [2] K. Park, D. Kim, S. Ha, and D. Lee, "Predicting pharmacodynamic drug-drug interactions through signaling propagation interference on protein-protein interaction networks," *J. of Public Library of Science(PLOS) ONE*, vol. 10, e0140816, 2015.
- [3] J. Seum, S. Yoo, and H. Nam, "Prediction of Compound-Target Interactions of Natural

- Products Using Large-scale Drug and Protein Information," *In Proc. of the ACM Ninth Int. Workshop on Data and Text Mining in Biomedical Informatics*, Melbourne, Australia, Oct. 2015, pp. 15.
- [4] G. Heok and S. Hark, "Probabilistic model for bio-cells information extraction," *J. of The Korea Institute of Electronic Communication Sciences*, vol. 6, no. 5, 2011, pp.649-656.
- [5] G. Jim and H. Lee, "The Development of Liver cancer Vital Sign Information Prediction System using Aptamer Protein Biochip," *J. of The Korea Institute of Electronic Communication Sciences*, vol. 6, no. 6, 2011, pp.965-971.
- [6] S. Toon and G. Kim, "Personal Biometric Identification based on ECG Features," *J. of The Korea Institute of Electronic Communication Sciences*, vol. 10, no. 4, 2015, pp.521-526.
- [7] B. Kevin and Y. Yamanishi. "Supervised prediction of drug - target interactions using bipartite local models," *Bioinformatics*, vol. 25, issue 18, 2009, pp. 2397-2403.
- [8] Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, and M. Kanehisa, "Prediction of drug - target interaction networks from the integration of chemical and genomic spaces," *Bioinformatics*, vol, 24, issues 13, 2008, pp. 232-240.
- [9] M. Yang and U. Gwag, *Drug Metabolism*, Seoul, Sinilbooks, 2015.
- [10] Y. Kim, *Essentials of Enzymology*, Seoul, Worldscience, 2015.
- [11] D. Wishart, T. Jewison, A. Guo, M. Wilson, C. Knox, Y. Liu, and S. Bouatra, "HMDB 3.0 - The Human Metabolome Database in 2013," *Nucleic Acids Research*, vol. 41, Issue D1, 2013, pp. 801-807.
- [12] A. Chang, I. Schomburg, S. Placzek, L. Jeske, and D. Schomburg, "BRENDA in 2015: exciting developments in its 25th year of existence," *Nucleic Acids Research*, vol. 43, Issue D1, 2015, pp. 1-8.
- [13] M. Hahsler, S. Chelluboina, K. Hornik, and C. Buchta, "The arules R-Package Ecosystem: Analyzing Interesting Patterns from Large Transaction Data Sets," *J. of Machine Learning Research*, vol. 12, no. 1, 2011, pp. 2021-2025.
- [14] M. Hahsler and S. Chelluboina. "Visualizing association rules: Introduction to the R-extension package arulesViz," *R project module*, Dec. 2011, pp. 223-238.
- [15] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene Selection for Cancer Classification using Support Vector Machines", *Machine Learning*, vol. 46, issue 1, 2002, pp. 389 - 422.
- [16] J. Bedo, C. Sanderson, and A. Kowalczyk, "An Efficient Alternative to SVM Based Recursive Feature Elimination with Applications in Natural Language Processing and Bioinformatics," *In Australian Joint Conference on Artificial Intelligence*, Hobart, Australia, Dec. 2006, pp. 170-180.
- [17] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *J. of Machine Learning Research*, vol. 3, no. 1, 2003, pp. 1157-1182.

저자 소개

**정희택(Hyi-Thaek Ceong)**



1992년 전남대학교 전산통계학과 졸업(이학사)  
 1995년 전남대학교 대학원 전산통계학과 졸업(이학석사)  
 1999년 전남대학교 대학원 전산통계학과 졸업(이학박사)  
 1999년 ~ 현재 : 전남대학교 멀티미디어전공 교수  
 ※ 관심분야 : 데이터마이닝, 생물정보학, 기계학습

**박춘구(Chun-Goo Park)**



2000년 인하대학교 전자계산공학과 졸업(공학사)  
 2002년 광주과학기술원 정보통신공학과 졸업(공학석사)  
 2010년 펜실베이니아주립대학교 대학원 생물학과 졸업(이학박사)  
 2013년~현재 : 전남대학교 생명과학기술학부 조교수  
 ※ 관심분야 : 생물정보학, 의료정보학