

물질대사경로예측을 위한 확률론적방법에 대한 연구

리석준, 염은철, 강주현

경애하는 김정은동지께서는 다음과 같이 말씀하시였다.

《새 세기 산업혁명은 본질에 있어서 과학기술혁명이며 첨단돌파에 경제강국건설의 지름길이 있습니다. 우주를 정복한 위성과과학자들처럼 최첨단돌파전을 힘있게 벌려 나라의 전반적과학기술을 하루빨리 세계적수준에 올려세워야 합니다.》

오늘날 세계적으로 물질대사경로정보가 인터넷을 통하여 광범히 배포되고있으며 이러한 정보들은 생명현상에 대한 종합적인 연구에 널리 리용[1-3]되고있다.

론문에서는 확률론에 기초하여 반응자료기지로부터 물질대사경로들을 예측하기 위한 한가지 방법을 밝혔다.

1. 확률적방법으로 반응을 선택하기 위한 선택알고리즘

물질대사경로예측은 출발물과 생성물이 주어지면 반응자료기지에 등록된 반응들을 조사하여 가능한 경로를 탐색하는 과정이다.

론문에서는 반응자료기지탐색단계에서 확률적인 선택을 진행하여 프로그램의 연산속도를 훨씬 높였다.

확률적인 선택이란 반응자료기지로부터 반응자료들을 탐색할 때 해당 반응의 상대적인 선택무게를 고려하여 선택의 우선권을 주는 방법이다.

생체반응에서는 흔히 한가지 반응물에 대하여 여러가지 생성물이 생겨난다. 반응선택의 확률은 해당 반응의 상대적무게에 대한 모든 반응의 무게의 합으로 주었다. 이것을 식으로 표시하면 다음과 같다.

$$P(R_i) = \frac{W_{R_i}}{\sum W_{R_j}}$$

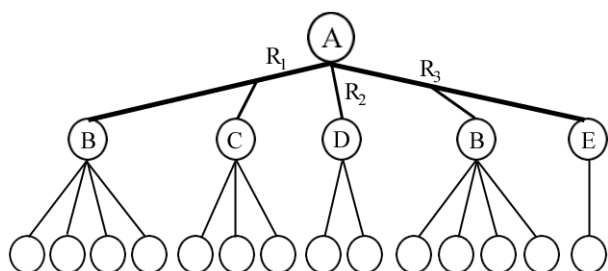


그림 1. 물질 A로부터 출발한 가상적인 반응

여기서 R_i 는 i 번째 반응, $P(R_i)$ 는 i 번째 반응의 선택확률, W_{R_i} 는 i 번째 반응의 선택무게이다.

실례를 들어 그림 1과 같은 가상적인 반응을 생각하자.

매개의 마디와 가지들은 반응에 참가하는 대사산물과 반응을 가상적으로 표시한것이다. 3개의 반응 R_1 , R_2 , R_3 에 대

하여 다음과 같이 정의할수 있다.



이러한 가정우에서 반응 1, 2, 3에 대하여 무게에 따라 선택순서를 정한다.

B부터 E까지의 물질대사산물들에 접속되는 물질들의 개수를 접속도(Degree of connectivity)라고 하면 $\deg(B)=4$, $\deg(C)=3$, $\deg(D)=2$, $\deg(E)=1$ 로 된다. 해당 반응들의 무게를 계산하면 $W_{R_1} = \min(4, 3)=3$, $W_{R_2} = 2$, $W_{R_3} = \min(4, 1)=1$ 로 된다. 이로부터 반응 1, 2, 3에 대한 선택확률은 각각 0.50, 0.33, 0.17로 된다. 해당 반응의 무게값이 크다는것은 그 반응의 가지가 많이 생긴다는것이며 가지가 많은 반응속에 목표대사산물이 포함되어있을 확률은 상대적으로 가지가 적은 반응속에 목표대사산물이 포함되어있을 확률보다 크다.

해당 반응에서 그 반응의 가지수중에 최소값을 선택한것은 최소값이 가지수전체를 대표하는 가장 적당한 값으로 되기때문이다.

이러한 방법으로 반응의 선택확률을 계산하고 확률이 높은 순서로 탐색을 진행하여 계산속도를 높였다.

2. 반응경로연결을 위한 알고리즘

프로그램은 VC++의 재귀호출을 리용하여 작성하였는데 그 흐름은 다음과 같다.

먼저 기본수속에서는 질문대사산물로부터 시작하는 반응경로연결을 위한 알고리즘을 작성한다.

부분수속에서는 반응자료기지의 자료들을 적당한 조건에 맞게 탐색하면서 기본반응 경로에 연결시키기 위한 알고리즘을 작성한다. 부분수속과정은 재귀적호출과정으로 하였으며 이 과정은 일정한 조건 즉 경로길이가 사용자의 요구에 맞을 때까지 계속 재귀반복 하도록 하게 하였다.

여기서 경로길이(pathway length)는 해당 물질대사경로의 반응단계수이다. 실례를 들어 어떤 물질대사경로가 5개의 반응으로 되어있으며 이 5개의 반응들이 모두 하나로 연결되었다고 하면 이 물질대사경로의 반응길이는 5로 된다.

프로그램계산과정에서 반응길이가 일정한 값을 넘으면 연산과정이 지수함수적으로 길어지게 된다. 이것을 막기 위하여 프로그램에서는 반응길이제한을 설정할수 있도록 하였다.

일반적으로 경로길이제한이 n 이고 모든 반응들의 총가지수가 m 일 때 반응경로를 나타내기 위한 변수는 $n \times m$ 행렬로 된다. 여기서 행렬의 매 원소는 경로에 포함되는 대사산물들을 보관한다. 여기서 반응들의 총가지수 m 은 계산과정에 정해지게 되며 따라서 논문에서는 프로그램계산과정에 동적으로 변하는 행렬방식을 택하였다.

이러한 방법으로 계산된 반응경로는 다시 일정한 처리를 거쳐 나무가지모양의 반응 경로로 표시되게 하였다.

아래에 물질대사경로예측프로그램(MPPP)의 알고리즘을 서술하였다.

수속 1 전체 대사경로구축(in 목적대사물질, in 선택안, out 경로)

Begin

Call 수속 2(목적대사물질, 선택안, 경로)

경로에 대한 기초분석

If 생성된 물질대사경로 < 원래의 물질대사경로

경로 ← Null

End if

End

수속 2 부분대사경로구축(in 대사물질, in 선택안, out 경로)

Begin

If 경로길이 > 제한길이

경로 ← Null

Return

End if

For 대사물질을 포함하고있는 자료기지의 매개 반응 R

If R가 경로안에 존재

무게값 ← 0

else if

선택안에 기초하여 무게값선정

End if

End for

무게값에 기초하여 임의의 반응선택

선택된 반응을 경로에 추가

For 선택된 반응의 매개 반응물 m

If m이 경로에 포함

계속

Else if

Call 수속 2(m, 선택안, 경로)

End if

End for

End

3. 결과 및 논의

우리는 위에서 작성한 프로그램을 리용하여 당분해경로의 일부인 포도당-1린산으로부터 글리세르알데히드-3린산으로까지 되는 반응경로에 대하여 반응길이 7로 하여 예측하고 분석하였다. 계산결과는 그림 2와 같다.

그림 2에서 출발물질(C00103)과 목적물질(C00118)은 모두 강조체로 표시하였다. C00103은 포도당-1린산, C00118은 글리세르알데히드-3린산으로서 이 물질들을 출발물, 목적물로 하는 반응경로는 당분해경로의 일부이다.

프로그램의 계산결과는 나무구조의 자료로 얻어지며 이것을 대면부에 반영하여 그림 2와 같은 결과를 얻었다.

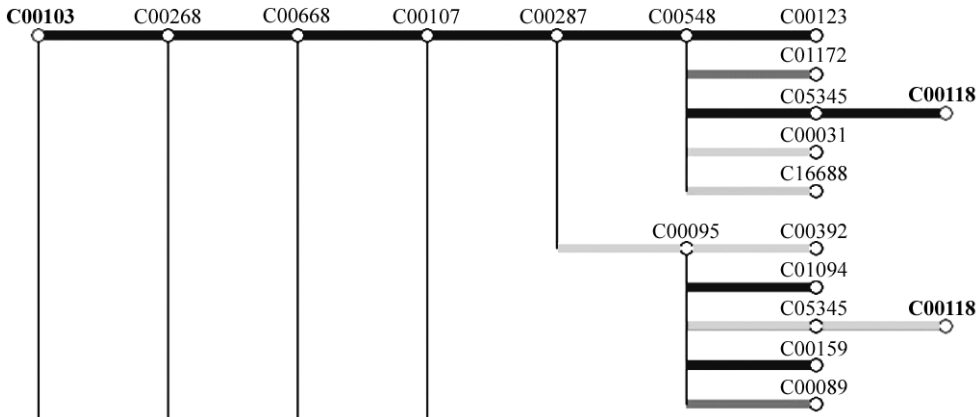


그림 2. MPPP의 계산결과

C00103을 출발물질로 하고 C00118을 목적물질로 하여 반응길이 7로 계산한 결과의 일부

그림 2에서 원으로 표시된것은 대사산물로서 그우에 대사산물의 등록번호를 주었다.

그림 2의 결과로부터 당분해경로의 일부인 포도당-1린산으로부터 글리세르알데히드-3린산까지 되는 반응경로에는 매우 많은 길이 있다는것을 알수 있다.

그림 2의 결과로부터 포도당-1린산으로부터 글리세르알데히드-3린산까지의 경로에 포함되는 반응가지들의 분포정형을 계산하였는데 그 결과는 표와 같다.

표. 포도당-1린산으로부터 글리세르알데히드-3린산까지의 반응경로에 포함되는 반응들의 분포정형

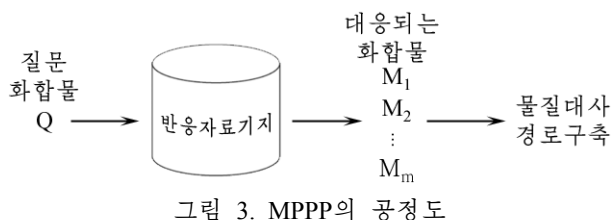
No.	대사경로	대사경로이름	반응개수 /개
1	당분해경로	Glycolysis/Gluconeogenesis	51
2	농마와 사탕대사	Starch and sucrose metabolism	336
3	갈락토즈대사	Galactose metabolism	284
4	아미노당과 핵산당대사	Amino sugar and nucleotide sugar metabolism	215
5	5탄당린산경로	Pentose phosphate pathway	188
6	과당과 만노즈대사	Fructose and mannose metabolism	91
7	5탄당, 글루쿠론산의 호상전환	Pentose and glucuronate interconversions	170
8	아스코르빈산대사	Ascorbate and aldarate metabolism	38
9	피리미딘대사	Pyrimidine metabolism	76
10	글리세롤기름질의 생합성	Glycerolipid metabolism	28
11	제아틴의 생합성	Zeatin biosynthesis	5
12	안싸미진의 생합성	Biosynthesis of ansamycins	7
13	스트렙토미친저항성	Streptomycin biosynthesis	44
14	이노시톨린산대사	Inositol phosphate metabolism	74
15	부티로신과 네오미친의 생합성	Butirosin and neomycin biosynthesis	23
16	폴리케티드의 당단위생합성	Polyketide sugar unit biosynthesis	44
17	방코미친기항생소의 생합성	Biosynthesis of vancomycin group antibiotics	7
18	메탄대사	Methane metabolism	35
19	리보플라빈대사	Riboflavin metabolism	20
20	빛합성생물에서 탄소고정	Carbon fixation in photosynthetic organisms	42
21	비타민 B ₆ 대사	Vitamin B ₆ metabolism	31
22	글리세로린기름질의 생합성	Glycerophospholipid metabolism	48

표계속

No.	대사경로	대사경로이름	반응개수 /개
23	피루빈산대사	Pyruvate metabolism	38
24	니코틴산과 니코틴산아미드대사	Nicotinate and nicotinamide metabolism	11
25	에테르기름질의 생합성	Ether lipid metabolism	8
26	글리신, 세린, 트레오닌대사	Glycine, serine and threonine metabolism	6
27	페닐알라닌, 티로신, 트립토판생합성	Phenylalanine, tyrosine and tryptophan biosynthesis	8
28	레몬산순환	Citrate cycle (TCA cycle)	5
29	부탄산대사	Butanoate metabolism	6
30	원핵생물에서 탄소고정경로	Carbon fixation pathways in prokaryotes	4
31	발린, 로이신, 이소로이신의 생합성	Valine, leucine and isoleucine biosynthesis	4
32	류황대사	Sulfometabolism	4
33	알라닌, 아스파라긴산, 글루타민산대사	Alanine, aspartate and glutamate metabolism	2
34	타우린, 히포타우린대사	Taurine and hypotaurine metabolism	2
35	시스테인, 메티오닌대사	Cysteine and methionine metabolism	1
36	D-알라닌대사	D-alanine metabolism	1
37	글리옥실산과 디카르복실산대사	Glyoxylate and dicarboxylate metabolism	7
38	C5-2염기산대사	C5-branched dibasic acid metabolism	2
39	판토텐산과 CoA생합성	Pantothenate and CoA biosynthesis	7
40	테르페노이드골격의 생합성	Terpenoid backbone biosynthesis	1
41	글리코실포스파티딜이노시톨(GPI)- 닻모양체생합성	Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	7
42	β -알라닌대사	beta-Alanine metabolism	6
43	푸린대사	Purine metabolism	3
44	산화적リン산화	Oxidative phosphorylation	2
45	빛합성	Photosynthesis	3
46	노보비오신생합성	Novobiocin biosynthesis	13
47	12-, 14-, 16-마클로라이드 생합성	Biosynthesis of 12-, 14- and 16-membered macrolides	4

표에서 보는바와 같이 포도당-1린산으로부터 글리세르알데히드-3린산까지 이르는 반응경로에 포함되는 중간산물들은 47가지의 대사경로에 참가하며 그중 많은 비중을 차지하는것은 당분해경로, 농마와 사탕대사, 갈락토즈대사, 아미노산과 핵산당대사, 5탄당린산경로, 과당과 만노즈대사, 5탄당, 글루쿠론산의 호상전환, 아스코르빈산대사, 피리미딘대사, 글리세롤기름질의 생합성, 이노시톨린산대사, 메탄대사, 빛합성에서의 CO₂대사, 글리세로린기름질의 생합성, 피루빈산대사라는것을 알수 있다.

프로그램의 연산공정을 도식으로 보면 그림 3과 같다.



맺 는 말

우리는 확률적방법에 의한 반응선택알고리즘과 반응경로연결알고리즘을 작성하고 그에 기초하여 물질대사경로예측프로그램을 개발하였다.

참 고 문 헌

- [1] M. Yousofshahi et al.; Metabolic Engineering, 13, 435, 2011.
- [2] Y. Moriya et al.; Nucleic Acids Research, 38, Web Server Issue, 10~20, 2010.
- [3] D. K. Peter et al.; Briefings in Bioinformatics, 11, 1, 40, 2009.

주체104(2015)년 4월 5일 원고접수

Research on Probabilistic Methods for Metabolic Pathway Prediction

Ri Sok Jun, Yom Un Chol and Kang Ju Hyon

We have already constructed metabolic pathway database and metabolic pathway search program.

In this paper we have designed selection algorithm of reactions using probabilistic method and reaction pathway connection algorithm, and developed metabolic pathway prediction program based on them.

Our program showed a good result in prediction of pathway from glucose-1-phosphate to glyceraldehydes-3-phosphate.

Key words: metabolic pathway prediction, reaction database, metabolic pathway database