(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2011-0100757
(43) 공개일자 2011년09월15일

(51) Int. Cl.
G06F 17/30 (2006.01) *G06F 17/00* (2006.01)
(21) 출원번호 10-2010-0019754
(22) 출원일자 2010년03월05일
심사청구일자 2010년03월05일

(71) 출원인
인하대학교 산학협력단
인천 남구 용현동 253 인하대학교
(72) 발명자
한경숙
인천시 남구 용현동 인하대학교 컴퓨터정보공학부
박병규
인천시 남구 용현동 인하대학교 컴퓨터 정보공학과
(74) 대리인
이원희

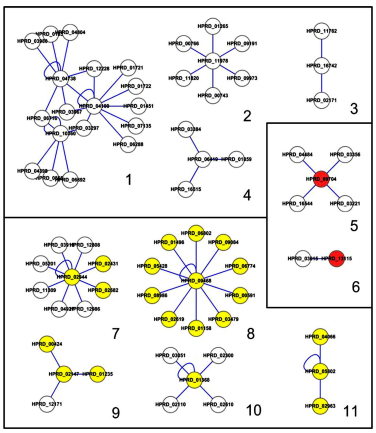
전체 청구항 수 : 총 14 항

(54) 온톨로지 기반의 단백질 상호작용 검색 엔진

(57) 요 약

본 발명은 온톨로지 기반의 단백질 상호작용 검색 엔진에 관한 것으로서, 더욱 상세하게는 변형된 괴델수에 의한 단백질 간의 생물학적 관계의 표현을 사용한 단백질 상호작용 검색 엔진은 특정 단백질과 상호작용하는 대부분의 단백질을 효율적으로 검색할 수 있고 유전자 온톨로지 용어 간의 계층구조를 반영하므로 더욱 특정한 용어를 사용하여 검색할 수 있으므로, 다양한 상호작용을 효율적으로 검색할 수 있는 검색 엔진으로서 유용하게 사용할 수 있다.

대 표 도 - 도3



이 발명을 지원한 국가연구개발사업
과제고유번호 2009-0066471
부처명 교육과학기술부
연구관리전문기관
연구사업명 기초연구사업 (여성과학자 지원사업)
연구과제명 온톨로지 기반의 단백질 상호작용 검색 기법
기여율
주관기관 인하대학교
연구기간 2009년 05월 01일~2012년 04월 30일

특허청구의 범위

청구항 1

단백질 정보를 입력받아 생물학적 데이터베이스 또는 유전자 온톨로지 데이터베이스에 적용하여 상기 단백질에 대응되는 분자적 기능(molecular function) 분류, 생물학적 과정(biological process) 분류 및 세포성분(cellular component) 분류의 유전자 온톨로지 용어를 할당하는 유전자 온톨로지 용어 할당 수단;

상기 유전자 온톨로지 용어 할당 수단에 의해 할당된 유전자 온톨로지 용어에 소수를 순차적으로 지정하는 알고리즘, 상기 지정된 소수를 로컬 DB에 저장하는 알고리즘, 및 생물학적 데이터베이스에 분류된 단백질 정보를 바탕으로, 상호작용하는 단백질의 유전자 온톨로지 용어에 지정된 소수 및 이의 루트 용어에 지정된 소수들을 곱하는 알고리즘을 통해 생성되는 변형된 괴델수(Godel number)로, 유전자 온톨로지 용어 간의 관계를 표현하는 표현 수단;

상기 표현 수단에 의해 생성된 변형된 괴델수의 표현을 바탕으로, 분석하고자 하는 단백질에 대한 단백질 상호작용 정보를 검색하는 단백질 상호작용 검색 수단; 및

상기 검색 수단에 의한 결과로부터 상기 분석하고자 하는 단백질과 상호작용하는 단백질을 판정하는 판정 수단을 포함하는 유전자 온톨로지 기반의 단백질 상호작용 검색 엔진.

청구항 2

제 1항에 있어서, 상기 유전자 온톨로지 데이터베이스는 GO(Gene Ontology)를 바탕으로 하는 것을 특징으로 하는 단백질 상호작용 검색 엔진.

청구항 3

제 1항에 있어서, 상기 단백질 상호작용 검색 수단은 유전자 온톨로지 용어 간의 상호작용 관계에 지정된 변형된 괴델수와 분석하고자 하는 단백질에 지정된 변형된 괴델수의 나눗셈 및 모듈로 연산(modulo operation)을 통해 수행되는 것을 특징으로 하는 단백질 상호작용 검색 엔진.

청구항 4

제 1항에 있어서, 상기 단백질 상호작용 검색 수단은 마이크로소프트 계열 프로그래밍 언어, 자바 언어 또는 컴퓨터 운영체제에 따라 제공되는 기타 프로그래밍 언어로 입력된 데이터를 이용하여 연산이 수행되는 것을 특징으로 하는 단백질 상호작용 검색 엔진.

청구항 5

제 1항에 있어서, 상기 판정 수단은 검색 수단을 통해 수행된 연산값 중에서, 유전자 온톨로지 용어 간의 상호작용 관계에 지정된 변형된 괴델수를 분석하고자 하는 단백질에 지정된 변형된 괴델수로 나눈 나머지가 0일 경우, 상기 유전자 온톨로지 용어가 할당된 단백질과 상기 분석하고자 하는 단백질이 상호작용하는 것으로 판단하는 것을 특징으로 하는 단백질 상호작용 검색 엔진.

청구항 6

분석하고자 하는 단백질과 상호작용하는 단백질을 검색 및 판정하는 제 1항의 검색 엔진; 및

상기 검색 엔진을 통해 선별된 단백질 상호작용 정보에 대한 시각화 데이터를 생성하는 시각화 수단을 포함하는

유전자 온톨로지 기반의 단백질 상호작용 검색 장치.

청구항 7

제 6항에 있어서, 상기 분석하고자 하는 단백질의 정보는 유전자 기호, 유전자 명칭, 단백질 명칭 및 단백질의 아미노산 서열로 구성된 군으로부터 선택되는 어느 하나인 것을 특징으로 하는 단백질 상호작용 검색 장치.

청구항 8

1) 컴퓨터 입력 수단을 통해 분석하고자 하는 단백질 정보를 입력하는 단계;

2) 유전자 온톨로지 용어 할당 수단을 통해 생물학적 데이터베이스 또는 유전자 온톨로지 용어 할당 수단을 통해 유전자 온톨로지 데이터베이스에 적용하여 분석하고자 하는 단백질에 대응되는 분자적 기능(molecular function) 분류, 생물학적 과정(biological process) 분류 및 세포성분(cellular component) 분류의 유전자 온톨로지 용어를 할당하는 단계;

3) 표현 수단을 통해 생물학적 데이터베이스에 분류된 단백질 정보를 바탕으로 유전자 온톨로지 용어에 소수를 순차적으로 지정하는 알고리즘, 상기 지정된 소수를 로컬 DB에 저장하는 알고리즘, 및 상호작용하는 단백질의 유전자 온톨로지 용어에 지정된 소수 및 이의 루트 용어에 지정된 소수들을 곱하는 알고리즘으로 변형된 괴델수(Godel number)를 생성하여, 상기 변형된 괴델수로서 유전자 온톨로지 용어 간의 관계를 표현하는 단계;

4) 단백질 상호작용 검색 수단을 통해 단계 3)의 변형된 괴델수의 표현을 바탕으로 상기 단계 1)에서 입력된 단백질에 대한 단백질 상호작용 정보를 검색하는 단계; 및

5) 판정 수단을 통해 상기 단계 4)의 단백질 상호작용 검색 결과로부터 상기 분석하고자 하는 단백질과 상호작용하는 단백질을 판정하는 단계를 포함하는 유전자 온톨로지 기반의 단백질 상호작용 검색 방법.

청구항 9

제 8항에 있어서, 단계 1)의 입력 수단은 자동완성기능(autocomplete functionality)을 가지는 것을 특징으로 하는 단백질 상호작용 검색 방법.

청구항 10

제 8항에 있어서, 단계 1)의 단백질 정보는 유전자 기호, 유전자 명칭, 단백질 명칭 및 단백질의 아미노산 서열로 구성된 군으로부터 선택되는 어느 하나인 것을 특징으로 하는 단백질 상호작용 검색 방법.

청구항 11

제 8항에 있어서, 단계 1)의 분석하고자 하는 단백질의 특성은 하나 또는 그 이상을 지정하는 것을 특징으로 하는 단백질 상호작용 검색 방법.

청구항 12

제 8항에 있어서, 단계 4)의 단백질 상호작용 정보를 검색하는 단계는 유전자 온톨로지 용어 간의 상호작용 관계에 지정된 변형된 괴델수와 분석하고자 하는 단백질에 지정된 변형된 괴델수의 정수 나눗셈 및 모듈로 연산을 통해 수행되는 것을 특징으로 하는 단백질 상호작용 검색 방법.

청구항 13

제 8항에 있어서, 단계 5)의 분석하고자 하는 단백질과 상호작용하는 단백질을 판정하는 단계는 검색 단계를 통해 수득한 연산값 중에서, 유전자 온톨로지 용어들 간의 상호작용 관계에 지정된 변형된 괴델수를, 분석하고자 하는 단백질에 지정된 변형된 괴델수로 나눈 나머지가 0일 경우, 상기 유전자 온톨로지 용어가 할당된 단백질과 상기 분석하고자 하는 단백질이 상호작용하는 것으로 판단하는 것을 특징으로 하는 단백질 상호작용 검색 방법.

청구항 14

제 8항에 있어서, 상기 단계 5) 이후에

a) 시각화 수단을 통해 상기 단계 5)의 판정 수단을 통해 선별된 단백질 상호작용 정보에 대한 시각화 데이터를 생성하는 단계; 및

b) 출력 수단을 통해 상기 단계 6)의 시각화 데이터를 출력하는 단계를 추가적으로 포함하는 것을 특징으로 하는 단백질 상호작용 검색 방법.

명 세 서

기 술 분 야

[0001] 본 발명은 온톨로지 기반의 단백질 상호작용 검색 엔진에 관한 것이다.

배 경 기 술

[0002] 효모 이중-하이브리드(yeast two hybrid) 및 질량분석기 기법과 같은 고효율 상호작용 검출 방법에 있어서 최근의 진전은 몇몇 개체에 대한 단백질 간의 상호작용 데이터의 급속한 확대를 이끌고 있다. 예를 들면, 4,000개의 *S. 세레비지애*(*S. cerevisiae*) 단백질 간의 약 8,000개의 상호작용, 7,600개의 D. 멜라노가스터(*D. melanogaster*) 단백질 간의 약 24,000개의 상호작용, 및 2,900개의 예쁜꼬마선충(*C. elegans*) 단백질 간의 5,100개 이상의 상호작용이 있다(Chen J, et al., Bioinformatics 2006, 22(16):1998-2004). 또한, 137,000개 이상의 단백질 상호작용 및 60,000개의 유전적 상호작용이 여섯 개의 주요 모델 개체에서 확인되었다(Breitkreutz BJ, et al., Nucleic Acids Research 2008, 36:D637-D640). 몇몇의 데이터베이스가 방대한 양의 단백질 간 상호작용 데이터를 위해 구축되고 있고, 이러한 것에는 BOND, DIP, MIPS, HPRD, HPID 및 BioGRID가 포함된다. 대부분의 데이터베이스는 사용자가 쿼리에 지정된 조건을 만족시키는 단백질 간의 상호작용을 검색할 수 있도록 해준다. 키워드 매칭(keyword matching) 또는 아이디 매칭(ID matching)은 검색 방법으로 가장 흔하게 사용되는 것 중 하나이다. 검색의 이러한 형태는 쿼리에 지정된 키워드 또는 아이디를 포함하는 데이터베이스에서 모든 기록을 검색한다. 사용자는 AND, OR 및 NOT과 같은 불 연산자(boolean operator)를 사용하여 검색 결과를 바꿀 수 있다.

[0003] 그러나, 키워드 매칭 또는 아이디 매칭에 기초한 검색 방법은 전적으로 구문론적이고 상기 키워드 또는 아이디 사이의 생물학적 관계를 고려하지 않는다. 만약 사용자가 너무 많은 단백질 간의 상호작용을 입수한다면, 단백질 간의 상호작용의 선별은 전적으로 사용자의 결정에 의지해야만 하고, 만약 사용자가 너무 적은 단백질 간의 상호작용을 입수하거나 아무런 결과를 얻지 못한다면, 사용자는 다른 수단을 찾아보아야 할 것이다. 예를 들어, BOND는 'ATP 결합'의 키워드 쿼리에 대해서 5,100개의 단백질 간 상호작용을 나타내주는데 반해 '뉴클레오티드 결합'의 키워드 쿼리에 대해서는 오직 96개의 상호작용을 보여준다. 유전자 온톨로지(Gene Ontology, GO) 계층에 있어서, 상기 용어 '뉴클레오티드 결합'은 상기 'ATP 결합' 보다 상위 수준이지만(Harris MA, et al., Nucleic Acids Research 2008, 36: D440-D444), 'ATP 결합'보다 더 적은 검색 결과를 보여준다. 상기 BOND 검색 방법이 키워드 간의 생물학적 관계를 고려하지 않기 때문에 이러한 이례적인 검색 결과가 발생한다. 게다가, 데이터베이스를 검색할 때 BOND 및 다른 많은 단백질 간의 상호작용 데이터베이스가 "자동완성기능(autocomplete)"을 지원하지 않기 때문에 사용자들은 반드시 쿼리에 있어서 정확한 키워드 또는 아이디를 입력해야만 한다.

[0004] 최근에 소수의 온톨로지 기반의 정보 검색 방법이 생물학적 문헌 또는 데이터베이스를 위해 개발되었지만(Corsi C, et al., BMC Bioinformatics 2007, 8(Suppl 1); Milward D, et al., Comparative and Functional Genomics 2005, 6:67-71; Muller HM, et al., Plos Biology 2004, 2:1984-1998), 데이터베이스에 있어서 단백질 상호작용에 대한 온톨로지 기반의 검색을 보고한 것은 거의 없다.

[0005] 이에, 본 발명자들은 분석하고자 하는 단백질의 의미상으로 관련된 모든 단백질-단백질 상호작용을 검색하는 유전자 온톨로지(Gene Ontology, GO) 및 검색 엔진의 표현(representation) 방법을 개발하기 위해 연구한 결과, 변형된 괴델수를 사용한 신규한 온톨로지 기반의 검색 엔진은 특정 단백질과 상호작용하는 단백질을 기존의 ID-매칭 검색 방법보다 더욱 효율적이고 다양하게 검색할 수 있고, 용어의 계층구조를 반영할 뿐만 아니라 다중 용어를 사용함에 있어서 더욱 많은 단백질 상호작용의 검색이 가능함을 밝힘으로써 본 발명을 완성하였다.

[0006]

발명의 내용

해결하려는 과제

[0007] 본 발명의 목적은 온톨로지 기반의 단백질 상호작용 검색 엔진을 제공하는 것이다.

과제의 해결 수단

[0008] 상기 목적을 달성하기 위하여, 본 발명은

[0009] 단백질 정보를 입력받아 생물학적 데이터베이스 또는 유전자 온톨로지 데이터베이스에 적용하여 상기 단백질에 대응되는 분자적 기능(molecular function) 분류, 생물학적 과정(biological process) 분류 및 세포성분(cellular component) 분류의 유전자 온톨로지 용어를 할당하는 유전자 온톨로지 용어 할당 수단;

[0010] 상기 유전자 온톨로지 용어 할당 수단에 의해 할당된 유전자 온톨로지 용어에 소수를 순차적으로 지정하는 알고리즘, 상기 지정된 소수를 로컬 DB에 저장하는 알고리즘, 및 생물학적 데이터베이스에 분류된 단백질 정보를 바탕으로, 상호작용하는 단백질의 유전자 온톨로지 용어에 지정된 소수 및 이의 루트 용어에 지정된 소수들을 곱하는 알고리즘을 통해 생성되는 변형된 괴델수(Godel number)로, 유전자 온톨로지 용어 간의 관계를 표현하는 표현 수단;

[0011] 상기 표현 수단에 의해 생성된 변형된 괴델수의 표현을 바탕으로, 분석하고자 하는 단백질에 대한 단백질 상호작용 정보를 검색하는 단백질 상호작용 검색 수단; 및

[0012] 상기 검색 수단에 의한 결과로부터 상기 분석하고자 하는 단백질과 상호작용하는 단백질을 판정하는 판정 수단을 포함하는 유전자 온톨로지 기반의 단백질 상호작용 검색 엔진을 제공한다.

[0013] 또한, 본 발명은

[0014] 분석하고자 하는 단백질과 상호작용하는 단백질을 검색 및 판정하는 상기의 검색 엔진; 및

[0015] 상기 검색 엔진을 통해 선별된 단백질 상호작용 정보에 대한 시각화 데이터를 생성하는 시각화 수단을 포함하는 유전자 온톨로지 기반의 단백질 상호작용 검색 장치를 제공한다.

[0016] 아울러, 본 발명은

[0017] 1) 컴퓨터 입력 수단을 통해 분석하고자 하는 단백질 정보를 입력하는 단계;

[0018] 2) 유전자 온톨로지 용어 할당 수단을 통해 생물학적 데이터베이스 또는 유전자 온톨로지 용어 할당 수단을 통해 유전자 온톨로지 데이터베이스에 적용하여 상기 단계 1)에서 입력된 분석하고자 하는 단백질에 대응되는 분자적 기능(molecular function) 분류, 생물학적 과정(biological process) 분류 및 세포성분(cellular component) 분류의 유전자 온톨로지 용어를 할당하는 단계;

[0019] 3) 표현 수단을 통해 생물학적 데이터베이스에 분류된 단백질 정보를 바탕으로 유전자 온톨로지 용어에 소수를

순차적으로 지정하는 알고리즘, 상기 지정된 소수를 로컬 DB에 저장하는 알고리즘, 및 상호작용하는 단백질의 유전자 온톨로지 용어에 지정된 소수 및 이의 루트 용어에 지정된 소수들을 곱하는 알고리즘으로 생성되는 변형된 괴델수(Godel number)로 유전자 온톨로지 용어 간의 관계를 표현하는 단계;

[0020] 4) 단백질 상호작용 검색 수단을 통해 단계 3)의 변형된 괴델수의 표현을 바탕으로 상기 단계 1)에서 입력된 단백질에 대한 단백질 상호작용 정보를 검색하는 단계; 및

[0021] 5) 판정 수단을 통해 상기 단계 4)의 단백질 상호작용 검색 결과로부터 상기 분석하고자 하는 단백질과 상호작용하는 단백질을 판정하는 단계를 포함하는 유전자 온톨로지 기반의 단백질 상호작용 검색 방법을 제공한다.

발명의 효과

[0022] 본 발명의 변형된 괴델수에 의한 단백질 간 생물학적 관계의 표현을 사용한 단백질 상호작용 검색 엔진은 기존의 ID-매칭 검색 방법보다 더욱 정확하고 효율적이며 원하는 상호작용 검색 결과를 얻을 수 있고, 유전자 온톨로지 용어 간의 계층구조를 반영함으로써 생물학적 정보(단백질 기능, 세포내 소기관, 단백질 종류)를 검색 조건으로 할 수 있고, 더욱 특정한 용어를 사용하여 구체적으로 검색할 수 있으며, 다중 검색 조건을 제시하여 검색의 범위를 한정할 수 있을 뿐만 아니라 단백질 상호작용 검색 과정이 기본적인 정수 연산으로 가능하게 하여 효율적이므로, 유전자 온톨로지의 체계를 검색에 직접 이용함으로써 단백질 상호작용을 체계적으로 검색할 수 있는 검색 방법으로서 유용하게 사용할 수 있다.

도면의 간단한 설명

[0023] 도 1은 온톨로지 기반의 검색 엔진의 사용자 인터페이스에 있어서, GO 용어에 대한 자동완성기능을 사용한 예시를 나타낸 그림이다.

도 2는 유전자 온톨로지 계층의 예를 나타낸 그림이다.

도 3은 뉴클레오티드 결합 및 ATP 결합으로 검색된 인간 단백질의 상호작용 네트워크를 나타낸 그림이다:

1 내지 11: HPRD 데이터로부터 "뉴클레오티드 결합"의 쿼리에 의해 온톨로지 기반의 검색으로 검색된 70개의 단백질 간 상호작용, 특히, 5 및 6은 "뉴클레오티드 결합"의 쿼리에 의해 ID-매칭 검색에서도 검색된 5개의 단백질 간 상호작용;

7 내지 11: 상기 "뉴클레오티드 결합"보다 더욱 특정화된 "ATP 결합"의 쿼리에 의해 온톨로지 기반의 검색 또는 ID-매칭 검색으로 검색된 31개의 상호작용:

노란색 노드: "ATP 결합"으로 분명하게 주석이 달린 인간 단백질;

붉은색 노드: "뉴클레오티드 결합"으로 분명하게 주석이 달린 인간 단백질; 및

흰색 노드: "ATP 결합" 및 "뉴클레오티드 결합"의 분명한 주석이 없는 인간 단백질.

발명을 실시하기 위한 구체적인 내용

[0024] 이하, 본 발명을 상세히 설명한다.

[0025] 본 발명은 유전자 온톨로지 기반의 단백질 상호작용 검색 엔진을 제공한다.

[0026] 상기 검색 엔진은,

[0027] 단백질 정보를 입력받아 생물학적 데이터베이스 또는 유전자 온톨로지 데이터베이스에 적용하여 상기 단백질에 대응되는 분자적 기능(molecular function) 분류, 생물학적 과정(biological process) 분류 및 세포성분(cellular component) 분류의 유전자 온톨로지 용어를 할당하는 유전자 온톨로지 용어 할당 수단;

[0028] 상기 유전자 온톨로지 용어 할당 수단에 의해 할당된 유전자 온톨로지 용어에 소수를 순차적으로 지정하는 알고

리즘, 상기 지정된 소수를 로컬 DB에 저장하는 알고리즘, 및 생물학적 데이터베이스에 분류된 단백질 정보를 바탕으로, 상호작용하는 단백질의 유전자 온톨로지 용어에 지정된 소수 및 이의 루트 용어에 지정된 소수들을 곱하는 알고리즘을 통해 생성되는 변형된 괴델수(Godel number)로, 유전자 온톨로지 용어 간의 관계를 표현하는 표현 수단;

[0029] 상기 표현 수단에 의해 생성된 변형된 괴델수의 표현을 바탕으로, 상기 분석하고자 하는 단백질에 대한 단백질 상호작용 정보를 검색하는 단백질 상호작용 검색 수단; 및

[0030] 상기 검색 수단에 의한 결과로부터 상기 분석하고자 하는 단백질과 상호작용하는 단백질을 판정하는 판정 수단을 포함할 수 있다.

[0031] 상기 유전자 온톨로지 용어 할당 수단은 특정 유전자의 기능이 유전자 온톨로지에서 정의하는 기능 중 어떠한 용어에 해당되는지를 결정하고 해당되는 유전자 온톨로지 용어를 유전자에 할당하는 역할을 한다. 유전자가 복수의 기능을 수행할 경우, 복수의 유전자 온톨로지 용어가 유전자에 할당될 수 있다.

[0032] 상기 유전자 온톨로지 용어 할당 수단은 네트워크를 통한 생물학적 DB로부터 특정 유전자를 동정하고 유전자와 관련되는 유전자 온톨로지 용어를 획득할 수 있다. 본 발명의 실시 예에 의하면 네트워크를 통해 액세스할 수 있는 생물학적 데이터베이스는 HPRD, BOND, DIP, MIPS, HPID, BioGRID, Unigene, LocusLink, Swiss-Prot, MGI(Mouse Genome Informatics), UniProt, EMBL(European Molecular Biology Laboratory) 및 IPI(International Protein Index) 등이 있다. 위의 데이터 베이스들 중 대부분은 유전자의 기능과 관련된 유전자 온톨로지 용어를 제공하고 있으며, 유전자와 관련된 유전자 온톨로지 용어를 제공하지 않더라도 상기 생물학적 데이터베이스에서 제공하는 유전자의 기능 정보를 이용하여 유전자 온톨로지 데이터 베이스를 이용하여 해당 유전자 온톨로지 용어를 할당할 수 있다. 네트워크를 통해 액세스할 수 있는 유전자 온톨로지 데이터 베이스로는 GO(Gene Ontology)가 대표적이라 할 수 있다.

[0033] 상기 단백질 상호작용 검색 수단은 유전자 온톨로지 용어 간의 상호작용 관계에 지정된 변형된 괴델수와 분석하고자 하는 단백질에 지정된 변형된 괴델수의 나눗셈 및 모듈로 연산(modulo operation)을 통해 수행될 수 있다.

[0034] 상기 단백질 상호작용 검색 수단은 마이크로소프트 계열 프로그래밍 언어, 자바 언어 또는 컴퓨터 운영체제에 따라 제공되는 기타 프로그래밍 언어로 입력된 데이터를 이용하여 연산이 수행될 수 있다.

[0035] 상기 판정 수단은 상기 검색 수단을 통해 수행된 연산값 중에서, 유전자 온톨로지 용어 간의 상호작용 관계에 지정된 변형된 괴델수를 분석하고자 하는 단백질에 지정된 변형된 괴델수로 나눈 나머지가 0일 경우, 상기 유전자 온톨로지 용어가 할당된 단백질과 상기 분석하고자 하는 단백질이 상호작용하는 것으로 판단할 수 있다.

[0036] 또한, 본 발명은 유전자 온톨로지 기반의 단백질 상호작용 검색 장치를 제공한다.

[0037] 상기 검색 장치는,

[0038] 분석하고자 하는 단백질과 상호작용하는 단백질을 검색 및 판정하는 상기의 검색 엔진; 및

[0039] 상기 검색 엔진을 통해 선별된 단백질 상호작용 정보에 대한 시각화 데이터를 생성하는 시각화 수단을 포함할 수 있다.

[0040] 상기 검색 장치는,

[0041] 분석하고자 하는 단백질에 대한 정보를 입력하는 컴퓨터 입력 수단; 및

[0042] 상기 시각화 수단에 의한 데이터를 출력하는 출력 수단을 추가적으로 포함할 수 있다.

[0043] 상기 분석하고자 하는 단백질의 정보는 유전자 기호, 유전자 명칭, 단백질 명칭 및 단백질의 아미노산 서열로 구성된 군으로부터 선택되는 어느 하나일 수 있다.

[0044] 상기 입력 수단은 자동완성기능(autocomplete functionality)을 가질 수 있다.

[0045] 상기 입력 수단은 키보드, 스캐너, 바코드입력기, 마우스, 타블렛, 트랙볼, 전자펜 및 디지털카메라로 구성되는 군으로부터 선택되는 어느 하나일 수 있다.

[0046] 상기 출력 수단은 모니터, 프린터 및 플로터로 구성되는 군으로부터 선택되는 어느 하나일 수 있다.

[0047] 상기 단백질 상호작용 검색 장치는 마이크로소프트 계열 프로그래밍 언어, 자바 언어 또는 컴퓨터 운영체제에 따라 제공되는 기타 프로그래밍 언어로 입력된 데이터를 이용하여 계산이 수행될 수 있다.

[0048] 상기 검색 장치는 단백질 상호작용 검색 수단 및 판정 수단으로써, http://ontology.hpid.org/의 검색 엔진을 사용할 수 있다.

[0049] 본 발명의 실시 예에 있어서, 원래의 괴델수 표현으로부터는 유전자 온톨로지 용어간의 정확한 관계를 추론할 수 없기 때문에, 자연수 대신 소수를 각 용어에 지정하고 각각의 용어에 대한 소수들을 곱한 변형된 괴델수에 의해 용어 간의 관계를 표현함으로써 변형된 괴델수 표현으로부터 유전자 온톨로지 용어 간의 관계의 추론을 가능하게 하였다. 또한, 상기 변형된 괴델수는 유전자 온톨로지 계층에 있어서, 루트 용어를 포함하여, 용어 간의 관계 및 조상 용어를 표현하는 소수들을 곱한 결과물이다.

[0050] 본 발명자들은 상기 온톨로지 기반의 단백질 상호작용 검색 장치의 검색 엔진 사용자 인터페이스(http://ontology.hpid.org/)를 구축하였다. 상기 검색 엔진은 C# 프로그래밍 언어로 수행되었고, 에라토스테네스의 체(Sieve of Eratosthenes)(http://mathworld.wolfram.com/SieveofEratosthenes)를 사용하여 26,000개 이상의 소수를 생성하였으며, 상기 숫자를 저장하고 그것의 곱셈 및 모듈로 연산(modulo operation)을 실시하기 위하여 JAVA Biginteger class를 사용하였다. 사용자는 Web Services Description Language(WSDL)를 사용한 검색 엔진의 웹 서비스로부터 단백질 간의 상호작용의 결과를 얻을 수도 있다. 상기 검색 엔진에 의해 확인된 단백질 상호작용 데이터가 표시되고, PSI-MI 포맷(Ceol A, et al., Nature Biotechnology 2004, 22(2):177-183) 또는 XML 스타일 시트를 가진 PSI-MI 포맷에 저장된다.

[0051] 본 발명자들은 HPRD(Prasad TSK, et al., Nucleic Acids Research 2009, 37:D767-D772)의 상호작용 데이터에서 온톨로지 기반의 검색 엔진을 시험하였고 그것을 ID-매칭 검색 방법과 비교하였다. 그 결과, ID-매칭 검색에 비해 온톨로지 기반의 검색 엔진에서 검색된 상호작용의 개수가 더욱 많았다. 한편, ID-매칭 검색은 일반적인 용어보다 더 특정적인 용어를 사용할 때 더욱 많은 검색 결과를 나타냈고, 하위수준의 용어로는 검색되었으나 그보다 더 상위 수준의 용어로는 검색되지 않은 데이터가 있었던 반면에, 상기 온톨로지 기반의 검색은 특정적인 GO 용어에 의한 상호작용뿐만 아니라 상기 GO 용어보다 더욱 특정화된 용어에 의한 상호작용도 검색하였다(표 4 및 도 2 참조). 상기 ID-매칭 검색 방법이 순전히 구문론적인 방법이고 GO 용어의 관계를 고려하지 않기 때문에 이러한 검색 변칙이 발생한 것으로 여겨진다. 또한, 인간 단백질의 상호작용을 상기 두 방법으로 검색하여 cytoscape(Shannon P, et al., Genome Research 2003, 13:2498-2504)를 사용하여 상호작용 네트워크를 알아본 결과, ID-매칭 검색을 통해서는 "뉴클레오티드 결합"에 의한 검색 결과과 더욱 특정화된 하위 용어인 "ATP 결합"보다 오히려 검색 결과가 적게 나타났으나, 온톨로지 기반의 검색 결과에서는 "ATP 결합"에 의한 검색 결과가 모두 "뉴클레오티드 결합"에 의한 검색 결과 내에 포함되는 것으로 확인되었다(표 5 및 도 3 참조). 또한, 두 가지의 단백질을 사용하여 다중 조건으로 검색한 결과, 온톨로지 기반의 검색이 ID-매칭 검색보다 더욱 많은 단백질 상호작용을 검색하였다(표 6 참조).

[0052] 따라서, 상기 온톨로지 기반의 검색 장치는 단백질 간의 계층적 관계를 고려하여 더욱 다양한 단백질 상호작용을 검색할 수 있는 단백질 상호작용 검색 엔진 및 이를 포함하는 검색 장치로서 유용하게 이용될 수 있다.

[0053] 아울러, 본 발명은 유전자 온톨로지 기반의 단백질 상호작용 검색 방법을 제공한다.

[0054] 상기 검색 방법은,

[0055] 1) 컴퓨터 입력 수단을 통해 분석하고자 하는 단백질 정보를 입력하는 단계;

[0056] 2) 유전자 온톨로지 용어 할당 수단을 통해 생물학적 데이터베이스 또는 유전자 온톨로지 용어 할당 수단을 통해 유전자 온톨로지 데이터베이스에 적용하여 분석하고자 하는 단백질에 대응되는 분자적 기능(molecular function) 분류, 생물학적 과정(biological process) 분류 및 세포성분(cellular component) 분류의 유전자 온톨로지 용어를 할당하는 단계;

[0057] 3) 표현 수단을 통해 생물학적 데이터베이스에 분류된 단백질 정보를 바탕으로 유전자 온톨로지 용어에 소수를 순차적으로 지정하는 알고리즘, 상기 지정된 소수를 로컬 DB에 저장하는 알고리즘, 및 상호작용하는 단백질의 유전자 온톨로지 용어에 지정된 소수 및 이의 루트 용어에 지정된 소수들을 곱하는 알고리즘으로 변형된 괴델수(Godel number)를 생성하여, 상기 변형된 괴델수로서 유전자 온톨로지 용어 간의 관계를 표현하는 단계;

[0058] 4) 단백질 상호작용 검색 수단을 통해 단계 3)의 변형된 괴델수의 표현을 바탕으로 상기 단계 1)에서 입력된 단백질에 대한 단백질 상호작용 정보를 검색하는 단계; 및

[0059] 5) 판정 수단을 통해 상기 단계 4)의 단백질 상호작용 검색 결과로부터 상기 분석하고자 하는 단백질과 상호작용하는 단백질을 판정하는 단계를 포함할 수 있다.

[0060] 상기 방법에 있어서, 단계 5) 이후에

[0061] a) 시각화 수단을 통해 상기 단계 5)의 판정 수단을 통해 선별된 단백질 상호작용 정보에 대한 시각화 데이터를 생성하는 단계; 및

[0062] b) 출력 수단을 통해 상기 단계 6)의 시각화 데이터를 출력하는 단계를 추가적으로 포함할 수 있다.

[0063] 상기 방법에 있어서, 단계 1)의 단백질 정보는 유전자 기호, 유전자 명칭, 단백질 명칭 및 단백질의 아미노산 서열로 구성된 군으로부터 선택되는 어느 하나일 수 있다. 유전자 온톨로지의 체계를 검색에 직접 이용함으로써, 분석하고자 하는 단백질 또는 상기 단백질과 상호작용하는 단백질의 상기와 같은 생물학적 정보를 검색 조건으로 하여 검색할 수 있다.

[0064] 상기 방법에 있어서, 단계 1)의 분석하고자 하는 단백질은 1가지의 특성을 지정할 수 있고, 또한 2가지 이상을 동시에 지정하여 입력된 단백질과 모두 상호작용하는 단백질 상호작용 정보를 검색함으로써 검색의 범위를 좁힐 수 있다.

[0065] 상기 방법에 있어서, 단계 2)의 유전자 온톨로지 데이터베이스는 GO(Gene Ontology)인 것이 바람직하다. 상기 생물학적 데이터베이스는 HPRD, BOND, DIP, MIPS, HPID, 및 BioGRID로 구성되는 군으로부터 선택되는 어느 하나인 것이 바람직하고 HPRD인 것이 더욱 바람직하다.

[0066] 상기 방법에 있어서, 단계 4)의 단백질 상호작용 정보를 검색하는 단계는 유전자 온톨로지 용어 간의 상호작용 관계에 지정된 변형된 괴델수와 분석하고자 하는 단백질에 지정된 변형된 괴델수의 정수 나눗셈 및 모듈로 연산을 통해 수행될 수 있다.

[0067] 상기 방법에 있어서, 단계 5)의 분석하고자 하는 단백질과 상호작용하는 단백질을 판정하는 단계는 검색 단계를 통해 수득한 연산값 중에서, 유전자 온톨로지 용어들 간의 상호작용 관계에 지정된 변형된 괴델수를, 분석하고자 하는 단백질에 지정된 변형된 괴델수로 나눈 나머지가 0일 경우, 상기 유전자 온톨로지 용어가 할당된 단백질과 상기 분석하고자 하는 단백질이 상호작용하는 것으로 판단할 수 있다.

[0068] 상기 입력 수단은 키보드, 스캐너, 바코드입력기, 마우스, 타블렛, 트랙볼, 전자펜 및 디지털카메라로 구성되는 군으로부터 선택되는 어느 하나일 수 있다.

[0069] 상기 출력 수단은 모니터, 프린터 및 플로터로 구성되는 군으로부터 선택되는 어느 하나일 수 있다.

[0070] 상기 단백질 상호작용 검색 방법은 마이크로소프트 계열 프로그래밍 언어, 자바 언어 또는 컴퓨터 운영체제에 따라 제공되는 기타 프로그래밍 언어로 입력된 데이터를 이용하여 계산이 수행될 수 있다.

[0071] 상기 검색 방법은 단백질 상호작용 검색 수단 및 판정 수단으로써, http://ontology.hpid.org/의 검색 엔진을 사용할 수 있고, 상기 검색 엔진은 자동완성기능을 제공하는 것이 바람직하다.

[0072] 변형된 괴델수에 의한 단백질 간의 생물학적 관계의 표현을 사용한 단백질 상호작용 검색 엔진을 통하여 특정 단백질과 상호작용하는 대부분의 단백질을 효율적으로 검색할 수 있고 상기 검색 결과는 유전자 온톨로지 용어 간의 계층구조를 반영하므로 더욱 특정한 용어를 사용하여 효율적으로 검색할 수 있으므로, 다양한 상호작용을 효율적으로 검색할 수 있는 온톨로지 기반의 단백질 상호작용 검색 방법으로서 유용하게 이용될 수 있다.

[0073] 이하, 본 발명을 하기 실시 예에 의해 상세히 설명한다.

[0074] 단, 하기 실시 예는 본 발명을 예시하는 것일 뿐, 본 발명의 내용이 하기 실시 예에 의해 한정되는 것은 아니다.

실 시 예 1

[0075] **<1-1> 변형된 괴델수 표현**

[0076] 괴델수(Godel number)는 하기 수학식 1에 의해 양의 정수(a_1, a_2, …, a_n)의 목록을 고유하게 암호화하기 위해 전형적으로 사용되고, 하기 P_k는 k번째 소수(prime number)를 나타낸다.

수학식 1

$$\varphi_{a_1 a_2 \ldots a_n} = \prod_{k=1}^{n} p_k^{a_k}$$

[0077]

[0078] 그러나, 상기 등식에 의해 정의된 상기 본래의 괴델수는 유전자 온톨로지의 방향성 비순환 그래프(directed acyclic graph) 구조를 표현할 수 없었다. 하기 표 1에 기재된 예시를 살펴보면, 고유의 자연수는 용어를 표현하고 괴델수는 상기 용어 간의 관계를 표현하였다. 용어4(Term4)는 관계3(R3)에 의해 용어2와 같은 종류이고 용어4는 관계 R1에 의해 용어1과 같은 종류이다. 그러나, 이러한 관계는 표 1의 괴델수 표현으로부터는 정확하게 추론될 수 없었다.

표 1

[0079] 괴델수 표현

용어	자연수	관계	괴델수
is a	1		
part of	2	R1: 용어2는 용어1이다.	$2^4 3^1 5^3$ = 6,000
용어1	3	R2: 용어3은 용어2의 부분이다.	$2^5 3^2 5^4$ = 180,000
용어2	4	R3: 용어4는 용어2이다.	$2^6 3^1 5^4$ = 120,000
용어3	5		
용어4	6		

[0080] 따라서, 상기 괴델수를 하기와 같이 변형하였다: 1) 하기 알고리즘 1 및 2를 사용하여 자연수 대신에 소수를 각 용어에 지정하였고, 2) 하기 알고리즘 3을 사용하여 변형된 괴델수에 의해 상기 용어 간의 각각의 관계를 표현하였다.

[0081] **알고리즘 1: 변형된 괴델수의 생성**

[0082] 알고리즘 1은 소수를 유전자 온톨로지(gene ontology, GO) 용어에 지정하여, 알고리즘 2를 호출함으로써 로컬 DB에 상기 GO 용어의 관계를 저장하고, 소수를 곱하여 변형된 괴델수를 생성하였다.

[0083] 1: $T = \{t_1, t_2, \cdots, t_G\}$ {G는 GO 용어의 개수}

[0084] 2: P = $\{p_1, p_2, \cdots, p_G\}$ {P는 소수의 순서집합}

[0085] 3: 모든 i는 i $\in \{1, 2, \cdots, G\}$

[0086] 4: $t_i \leftarrow p_i$ {소수를 GO 용어에 지정}

[0087] 5: end for

[0088] 6: 모든 t는 $t \in \{t_1, t_2, \cdots, t_G\}$

[0089] 7: 알고리즘 2(t. key, t) {로컬 DB에 소수 지정의 저장}

[0090] 8: end for

[0091] 9: 모든 t는 $t \in \{t_1, t_2, \cdots, t_G\}$

[0092] 10: 변형된 괴델수 ← 알고리즘 3(t. key)

[0093] 11: end for

[0094] **알고리즘 2: 표현저장(*Term.key, Term*)**

[0095] 알고리즘 2는 매개변수 용어가 계층의 루트 용어일 때까지 재귀호출에 의해 상기 GO 용어 및 그것의 모체의 소수 지정을 저장하였다.

[0096] 1: 만약 용어.루트이다 == 거짓이면

[0097] 2: 모든 t는 $t \in$ 용어.모체

[0098] 3: SetRelation(*Term.key*, *t.key*, *t.prime*) {용어 및 그것의 모체에 지정된 소수를 저장}

[0099] 4: 알고리즘 2(*Term.key*, *t*) {용어의 모체에 대한 재귀호출}

[0100] 5: end for

[0101] 6: end if

[0102] **알고리즘 3: 변형된 괴델수(*Term.key*)**

[0103] 알고리즘 3은 온톨로지 계층에서 매개변수 Term.key 및 그것의 조상 용어를 표현하는 소수의 곱셈으로써 변형된 괴델수를 계산하였다.

[0104] 1: ArrayList 리스트 = GetRelation(*Term.key*) {용어 및 그것의 조상에 대한 소수를 검색}

[0105] 2: *var* = 1

[0106] 3: 모든 t는 $t \in$ 리스트

[0107] 4: $var = var \times t.prime$

[0108] 5: end for

[0109] 6: 리턴(return) *var*

[0110] 상기 변형된 괴델수는 유전자 온톨로지 계층에 있어서 루트 용어를 포함하는 관계 및 그들의 조상에서의 용어를 표현하는 소수의 산물이었다. 예를 들어, 하기 표 2의 관계 R3은 42에 의해 표현되고, 이는 2{용어1, 계층에서 루트 노드(root node)}, 3(용어2), 및 7(용어3)의 소인수를 갖는다. 이러한 표현은 용어1, 용어2, 및 용어3을 표현하는 소수로 R3의 고유 인수분해에 의해 R3의 의미를 추론할 수 있도록 하였다.

표 2

[0111] 변형된 괴델수 표현

용어	소수	관계	변형된 괴델수
용어1	2		
용어2	3	R1: 용어2는 용어1이다.	관계(is a) = 3×2 = 6
용어3	5	R2: 용어3은 용어 2의 부분이다.	관계(part of) = 5×3×2 = 30
용어4	7	R3: 용어4는 용어2이다.	관계(is a) = 7×3×2 = 42

[0112] **<1-2> 도메인 간의 상호작용으로부터 예측되는 단백질 간의 상호작용**

[0113] 상기 실시예 <1-1>의 변형된 괴델수 표현을 사용하여 도메인 간의 상호작용으로부터 예측되는 단백질 간의 상호

작용의 예시를 하기 표 3에 나타냈다. 도메인A가 도메인B와 상호작용한다고 가정했을 때(표 3의 R4), 단백질A는 도메인A를 갖고(R5) 단백질B는 도메인B를 갖는다(R6). 이러한 관계는 표 3에 나타낸 바와 같이, 변형된 괴델수로 표현되었다. 이러한 표현을 사용함으로써, 단백질A가 단백질B와 상호작용한다는 가설은 정수 나눗셈 및 모듈로 연산(modulo operation)과 같은 간단한 산술연산으로서 확인할 수 있었다.

[0114] 143 = 11(도메인A) × 13(도메인B) (1)

[0115] 46,189 = 187(단백질A) × 247(단백질B) (2)

[0116] 46,189 ≡ 0(mod 143) (3)

[0117] 상기 예시에 있어서, 도메인A가 도메인B와 상호작용하는 관계는 변형된 괴델수 143으로 표현되었고, 이는 11(표현된 도메인A)과 13(도메인B)(1)을 곱한 값이다. 단백질A가 단백질B와 상호작용한다는 가설은 247(단백질B)(2)을 187(단백질A)에 곱함으로써 변형된 괴델수 46,189로 표현하였다. 46,189(단백질A가 단백질B와 상호작용한다는 가설을 표현)를 143(도메인A가 도메인B와 상호작용한다는 관계를 표현)으로 나눈 후의 나머지가 0이기 때문에, 단백질A가 단백질B와 상호작용한다는 상기 가설은 참인 것으로 확인되었다.

표 3

[0118] 단백질 간의 상호작용 추론

용어	소수	관계	변형된 괴델수
도메인A	11		
도메인B	13	R4: 도메인A는 도메인B와 상호작용한다.	관계(상호작용한다) = 11×13 = 143
단백질A	17	R5: 단백질A는 도메인A를 가진다.	관계(가진다) = 17×11 = 187
단백질B	19	R6: 단백질B는 도메인B를 가진다.	관계(가진다) = 19×13 = 247

실 시 예 2

[0119] **검색 엔진의 사용자 인터페이스**

[0120] 온톨로지 기반의 검색 엔진의 시제품(http://ontology.hpid.org/)은 C# 프로그래밍 언어로 실시되었다. 에라토스테네스의 체(Sieve of Eratosthenes)(http://mathworld.wolfram.com/SieveofEratosthenes)를 사용하여 26,000개 이상의 소수를 생성하였고, 상기 숫자를 저장하고 그것의 곱셈 및 모듈로 연산(modulo operation)을 실시하기 위하여 java Biginteger class를 사용하였다. 사용자가 쿼리 단백질에 대한 GO 용어 또는 단백질 슈퍼패밀리(superfamily)를 지정할 때, 검색 엔진은 상기 지정된 GO 용어보다 더욱 특정한 용어로 주석이 달린 단백질뿐만 아니라 상기 GO 용어 또는 수퍼패밀리로 주석이 달린 단백질과 관련된 모든 상호작용을 가져온다. 상기 검색 엔진을 사용자가 사용하기 쉽도록 만들기 위하여, 상기 GO 용어 또는 단백질 수퍼패밀리에 대한 자동완성(autocomplete) 기능을 제공하였다(도 1). 따라서, 사용자에 의해 입력된 불완전한 용어는 상기 불완전한 용어와 일치하는 하나 또는 그 이상의 완전한 GO 용어 또는 수퍼패밀리로 확장되었다. 상기 검색 엔진에 있어서 GO 용어에 대한 자동완성기능을 사용한 예는 도 1에 나타냈다. 또한, 사용자는 Web Services Description Language(WSDL)를 사용한 검색 엔진의 웹 서비스로부터 단백질 간의 상호작용의 결과를 얻을 수도 있다. 상기 WSDL은 문서지향(document-oriented) 또는 절차지향(procedure-oriented) 정보를 포함하는 메시지에서 종점 운영(endpoints operating)의 세트로서 네트워크 서비스를 설명하기 위한 XML 포맷이다(http://www.w3.org/TR/wsdl). 상기 검색 엔진에 의해 나타난 상호작용 데이터가 표시되고 PSI-MI 포맷(Ceol A, et al., Nature Biotechnology 2004, 22(2):177-183) 또는 XML 스타일 시트를 가진 PSI-MI 포맷에 저장된다.

실 시 예 3

[0121] **온톨로지 기반의 검색과 ID-매칭 검색의 비교**

[0122] **<3-1> 온톨로지 기반의 검색 엔진을 사용한 방법과 ID-매칭 검색 방법에 의한 결과 확인**

[0123] HPRD(Prasad TSK, et al., Nucleic Acids Research 2009, 37:D767-D772)의 상호작용 데이터에서 온톨로지 기반의 검색 엔진을 시험하였고 그것을 ID-매칭 검색 방법과 비교하였다. 상기 두 가지 방법에 의해 HPRD에서 발견된 단백질 간의 상호작용의 개수를 하기 표 4에 나타냈다.

[0124] 그 결과, HPRD release 7은 8,800개의 인간 단백질 사이에서 38,190개의 상호작용을 포함하였고, 이는 470개의 GO 용어로 주석이 달렸다. GO 용어의 총합은 26,000개 이상이었으나, 상기 26,000개의 GO 용어 중 1.8%가 HPRD에서 인간 단백질에 주석을 달기 위해 사용되었다. 본 발명의 검색 엔진은 인간 단백질에 주석을 달기위해 사용된 상기 470개의 GO 용어로부터 인간 단백질에 대한 698개까지의 GO 용어를 추론할 수 있다.

[0125] 하기 표 4 및 도 2에 나타낸 바와 같이, GO 용어, 아이디 GO:0008150은 생물학적 과정(biological process)에 대한 GO 계층의 루트 노드(Root Node)이다. GO:0008150 쿼리에 의해, ID-매칭 검색은 오직 5개의 상호작용(총 38,190개의 HPRD 상호작용의 0.01%)을 검색하였으나, 온톨로지 기반의 검색 엔진은 36,523개의 상호작용(총 38,190개의 HPRD 상호작용의 95.6%)을 검색하였다. GO 계층에서 상기 GO:0008150의 자손 노드(descendent node)인, 대사 과정에 대한 GO:0008152 쿼리에 의해, 상기 ID-매칭 검색은 2,862개의 상호작용(총 38,190개의 HPRD 상호작용의 7.5%)을 검색하였으나, 상기 온톨로지 기반의 검색 엔진은 19,828개의 상호작용(총 38,190개의 HPRD 상호작용의 51.9%)을 검색하였다. 상기 ID-매칭 검색은 덜 특정적인 것보다 더 특정적인 용어를 사용할 때 더욱 많은 검색 결과를 나타냈다. 또한, 상기 ID-매칭 검색에서, GO:0019538은 GO:0044238 또는 GO:0043170 보다 더 하위 수준임에도 불구하고, 상기 GO:0044238 또는 GO:0043170의 쿼리와 상호작용하는 것은 검색되지 않았으나, GO:0019538의 쿼리에 대해서는 5,324개의 상호작용이 검색되었다. 반면에, 상기 온톨로지 기반의 검색은 쿼리에 의한 특정 GO 용어에 대한 상호작용뿐만 아니라 상기 GO 용어에 대하여 더욱 특정화된 용어에 의한 상호작용도 검색하였다.

표 4

[0126] **온톨로지 기반의 검색과 ID-매칭 검색 방법에 의한 검색 결과의 비교**

GO 용어 ID	GO 용어 이름	ID-매칭 검색	온톨로지 기반의 검색
생물학적 과정			
GO:0008150	생물학적 과정	5 (0.01%)	36,523 (95.63%)
GO:0008152	대사과정	2,862 (7.49%)	19,828 (51.92%)
GO:0044238	1차 대사과정	0 (0.00%)	17,434 (45.65%)
GO:0043170	고분자 대사과정	0 (0.00%)	7,211 (18.88%)
GO:0019538	단백질 대사과정	5,324 (13.94%)	5,659 (14.82%)
분자 기능			
GO:0003676	핵산 결합	10 (0.03%)	8,733 (22.87%)
GO:0003677	DNA 결합	1,944 (5.09%)	6,935 (18.16%)
GO:0003700	전사인자 활성	5,164 (13.52%)	5,164 (13.52%)
세포성분			
GO:0005622	세포 내	0 (0.00%)	31,694 (82.99%)
GO:0005737	세포질	17,312 (45.33%)	20,990 (54.96%)
GO:0005829	사이토졸(cytosol)	452 (1.18%)	452 (1.18%)

[0127] **<3-2> 단백질 상호작용 네트워크**

[0128] 인간 단백질의 상호작용을 상기 두 방법으로 검색하여 cytoscape(Shannon P, et al., Genome Research 2003, 13:2498-2504)를 사용하여 상호작용 네트워크를 나타냈다.

[0129] 그 결과, 표 5 및 도 3에 나타낸 바와 같이, 도 3의 네트워크 1 내지 11은 HPRD 데이터로부터 "뉴클레오티드 결합(GO 용어 ID GO:0000166)"의 쿼리에 의한 상기 온톨로지 기반의 검색을 통해 검색된 70개의 단백질 간 상호작용을 나타내고, 표 5에서는, 상기 70개의 상호작용 중 오직 5개만이 "뉴클레오티드 결합"의 분명한 주석을 갖는 단백질과 관련된 것으로 나타났다. 나머지 65개의 상호작용은 "뉴클레오티드 결합" 보다는 "ATP 결합", "FAD 결합" 또는 "GTP 결합"과 같은 더욱 특정화된 용어로 주석이 달린 단백질 검색에 의해 유전자 온톨로지로부터 추론하였다. 도 3의 네트워크 7 내지 11은 "ATP 결합"(GO:0005524)의 쿼리에 의한 온톨로지 기반의 검색을 통해 검색된 31개의 상호작용을 나타내고, 온톨로지 계층에 있어서 상기 "ATP 결합"은 "뉴클레오티드 결합"보다

하위 수준이다. 따라서, "ATP 결합"에 의한 검색 결과가 "뉴클레오티드 결합"에 의한 검색 결과 내에 포함되는 것은 당연한 것이었다. 반면에, 도 3의 네트워크 5 내지 6과 같이, ID-매칭 검색에서는 "뉴클레오티드 결합" 쿼리에 의해 오직 5개의 상호작용만이 검색되었고 나머지 65개의 상호작용은 누락되었다. ID-매칭 검색을 통한 "ATP 결합"의 쿼리에 있어서는 상기 온톨로지 기반의 검색 방법과 같은 31개의 상호작용이 검색되었다.

표 5

[0130]

쿼리 단백질 HPRD_ID	기능	파트너 단백질 HPRD_ID	기능
HPRD_02944	ATP 결합	HPRD_02431	아실트랜스퍼라제 활성
HPRD_01368	ATP 결합	HPRD_02147	ATP 결합
HPRD_01368	ATP 결합	HPRD_02300	ATP아제 활성
HPRD_02147	ATP 결합	HPRD_12171	촉매활성
HPRD_01368	ATP 결합	HPRD_02110	세포외기질 구조적 성분
HPRD_02944	ATP 결합	HPRD_02682	GTP아제 활성
HPRD_09468	ATP 결합	HPRD_08986	포스포리파제 활성
HPRD_01368	ATP 결합	HPRD_02610	단백질 결합
HPRD_02944	ATP 결합	HPRD_03913	단백질 결합
HPRD_09468	ATP 결합	HPRD_01496	단백질 세린/트레오닌 키나제 활성
HPRD_09468	ATP 결합	HPRD_02619	백질 세린/트레오닌 키나제 활성
HPRD_09468	ATP 결합	HPRD_03479	단백질 세린/트레오닌 키나제 활성
HPRD_05802	ATP 결합	HPRD_04066	단백질 세린/트레오닌 키나제 활성
HPRD_09468	ATP 결합	HPRD_05428	단백질 세린/트레오닌 키나제 활성
HPRD_05802	ATP 결합	HPRD_02963	수용체 활성
HPRD_09468	ATP 결합	HPRD_01158	수초(myelin sheath)의 구조적 성분
HPRD_02147	ATP 결합	HPRD_01235	전사조절자 활성
HPRD_09468	ATP 결합	HPRD_00591	번역조절자 활성
HPRD_09468	ATP 결합	HPRD_06774	번역조절자 활성
HPRD_09468	ATP 결합	HPRD_06802	번역조절자 활성
HPRD_09468	ATP 결합	HPRD_09084	번역조절자 활성
HPRD_01368	ATP 결합	HPRD_03051	수송 활성
HPRD_16742	FAD 결합	HPRD_11762	DNA 결합
HPRD_16742	FAD 결합	HPRD_02171	가수분해효소 활성
HPRD_04100	GTP 결합	HPRD_07135	아실트랜스퍼라제 활성
HPRD_04100	GTP 결합	HPRD_01721	보조 수송 단백질 활성
HPRD_04100	GTP 결합	HPRD_01722	보조 수송 단백질 활성
HPRD_04100	GTP 결합	HPRD_04738	GTP 결합
HPRD_04738	GTP 결합	HPRD_06716	GTP 결합
HPRD_11978	GTP 결합	HPRD_00743	GTP아제 활성
HPRD_11978	GTP 결합	HPRD_00766	GTP아제 활성
HPRD_04100	GTP 결합	HPRD_03297	GTP아제 활성
HPRD_10360	GTP 결합	HPRD_03297	GTP아제 활성
HPRD_04100	GTP 결합	HPRD_12228	GTP아제 활성
HPRD_04738	GTP 결합	HPRD_12228	GTP아제 활성
HPRD_10360	GTP 결합	HPRD_06692	GTP아제 활성
HPRD_10360	GTP 결합	HPRD_08555	GTP아제 활성
HPRD_11978	GTP 결합	HPRD_09191	GTP아제 활성
HPRD_11978	GTP 결합	HPRD_09973	GTP아제 활성
HPRD_11978	GTP 결합	HPRD_11820	GTP아제 활성
HPRD_10360	GTP 결합	HPRD_04398	단백질 결합
HPRD_11978	GTP 결합	HPRD_01265	단백질 세린/트레오닌 키나제 활성
HPRD_06419	GTP 결합	HPRD_03384	단백질 세린/트레오닌 포스포타제 활성
HPRD_04100	GTP 결합	HPRD_06288	RNA 결합
HPRD_04738	GTP 결합	HPRD_01853	세포골격(cytoskeleton)의 구조적 성분
HPRD_04100	GTP 결합	HPRD_01451	구조적 분자 활성
HPRD_06419	GTP 결합	HPRD_01859	전사인자 활성
HPRD_06419	GTP 결합	HPRD_16515	전자조절자 활성

HPRD_04100	GTP 결합	HPRD_03967	유비퀴틴-특이적 프로테아제 활성
HPRD_04738	GTP 결합	HPRD_03967	유비퀴틴-특이적 프로테아제 활성
HPRD_09704	뉴클레오티드 결합	HPRD_03356	신호 트랜스듀서(Signal transducer) 활성
HPRD_09704	뉴클레오티드 결합	HPRD_16544	전사인자 활성
HPRD_09704	뉴클레오티드 결합	HPRD_03221	전사조절자 활성
HPRD_13115	뉴클레오티드 결합	HPRD_03015	알려지지 않음
HPRD_09704	뉴클레오티드 결합	HPRD_04484	알려지지 않음

[0131] **<3-3> 다중 조건에 의한 검색 결과의 비교**

[0132] 쿼리 단백질에 대해 다중 조건으로 검색한 결과를 비교하기 위하여, 두 개의 GO 용어인 "단백질 대사과정"(GO:0019538)과 "세포질"(GO:0005737)을 쿼리 단백질에 대한 조건으로서 지정하였다.

[0133] 그 결과, 하기 표 6에 나타낸 바와 같이, 온톨로지 기반의 검색이 ID-매칭 검색보다 모든 쿼리에 대하여 더욱 많은 상호작용을 검색하였다.

표 6

[0134] 쿼리 단백질에 대해 다중 GO 용어의 지정에 의한 단백질 간 상호작용

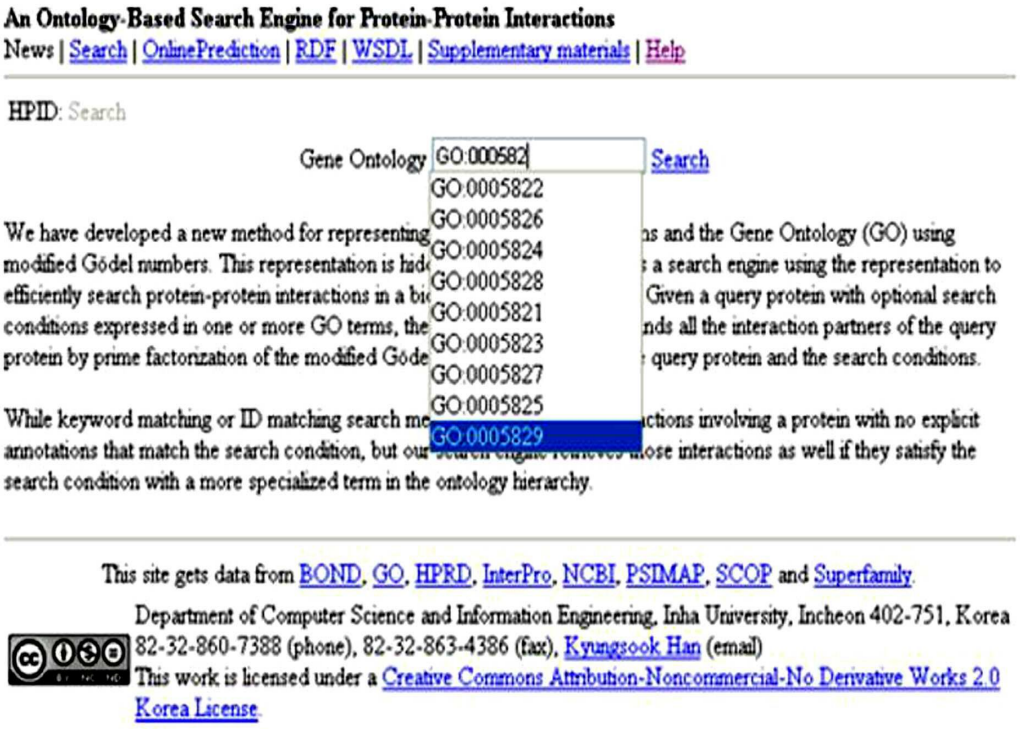
다중 GO 용어 쿼리 단백질		검색 방법	
		온톨로지 기반의 검색	ID-매칭 검색
생물학적 과정	세포성분		
GO:0019538	GO:0005737	3062 (8.02%)	1994 (5.22%)
GO:0019538	GO:0005576	769 (2.01%)	753 (1.97%)
분자 기능	세포 성분		
GO:0003700	GO:0005737	592 (1.55%)	576 (1.51%)
GO:0003700	GO:0005576	103 (0.27%)	6 (0.02%)

산업상 이용가능성

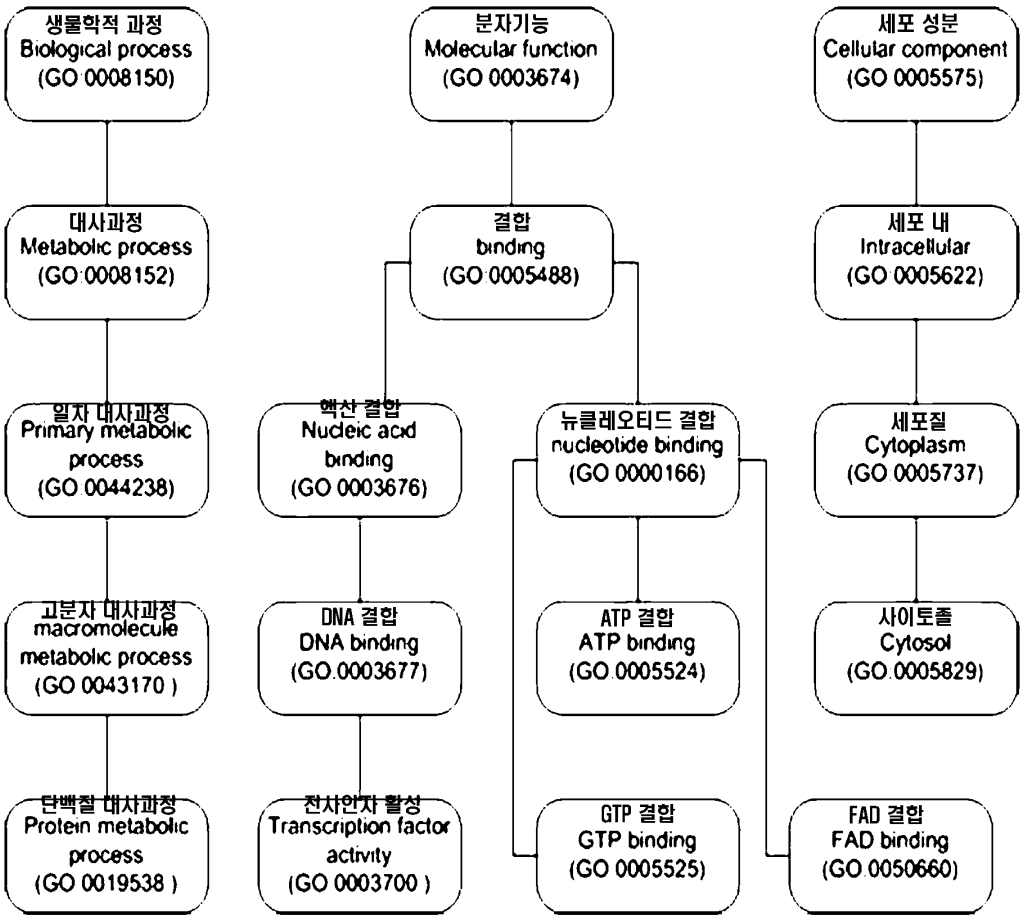
[0135] 상기와 같이, 온톨로지 기반의 단백질 상호작용 검색 엔진은 분석하고자 하는 단백질에 대한 대부분의 단백질 상호작용을 효율적으로 검색할 수 있으므로, 온톨로지를 기반으로 하는 단백질 상호작용 검색 및 검색 장치의 개발과 제조에 유용하게 사용될 수 있을 뿐만 아니라 단백질 상호작용의 효율적인 검색을 통하여 신약 개발 등에 유용하게 사용될 수 있다.

도면

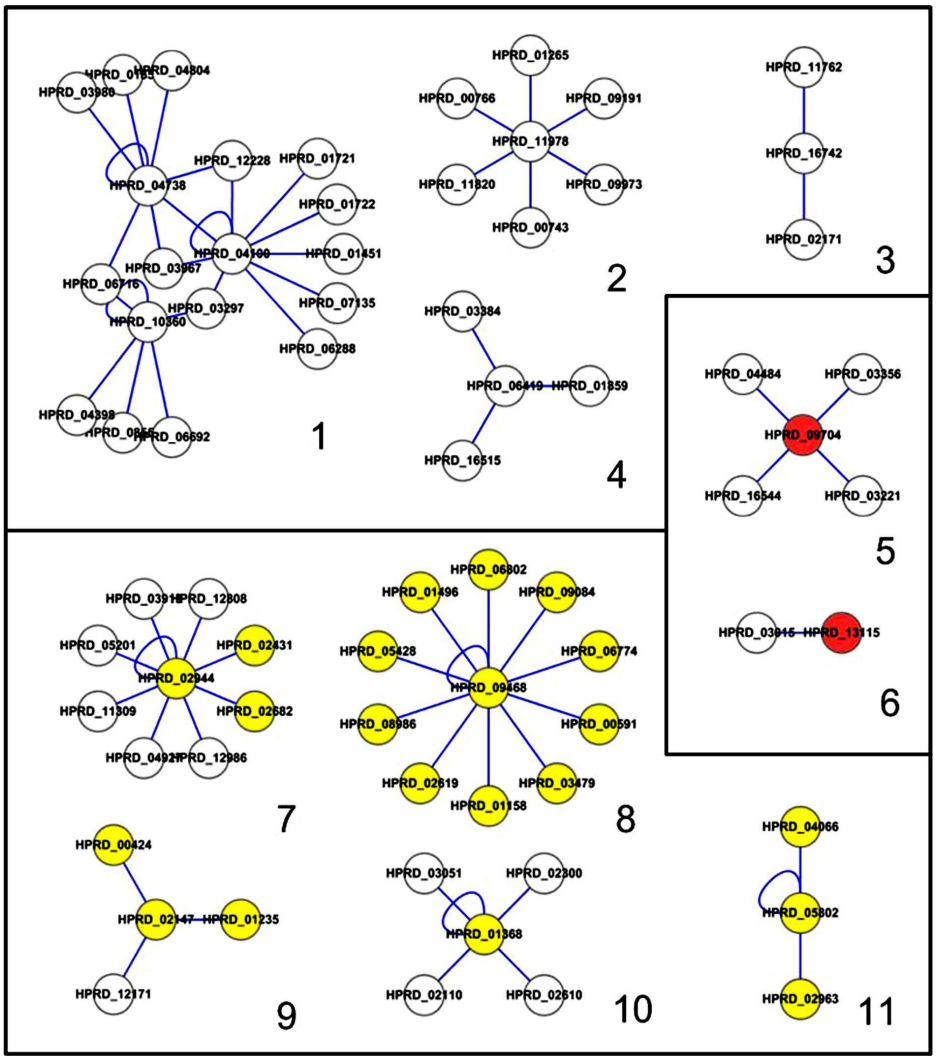
도면1


An Ontology-Based Search Engine for Protein-Protein Interactions
News | Search | OnlinePrediction | RDF | WSDL | Supplementary materials | Help
HPID: Search
Gene Ontology GO:000582 Search
GO:0005822
GO:0005826
GO:0005824
GO:0005828
GO:0005821
GO:0005823
GO:0005827
GO:0005825
GO:0005829
We have developed a new method for representing
ns and the Gene Ontology (GO) using
modified Gödel numbers. This representation is hid
a search engine using the representation to
efficiently search protein-protein interactions in a bi
Given a query protein with optional search
conditions expressed in one or more GO terms, the
nds all the interaction partners of the query
protein by prime factorization of the modified Göde
query protein and the search conditions.
While keyword matching or ID matching search me
ctions involving a protein with no explicit
annotations that match the search condition, but our
ose interactions as well if they satisfy the
search condition with a more specialized term in the ontology hierarchy.
This site gets data from BOND, GO, HPRD, InterPro, NCBI, PSIMAP, SCOP and Superfamily.
Department of Computer Science and Information Engineering, Inha University, Incheon 402-751, Korea
82-32-860-7388 (phone), 82-32-863-4386 (fax), Kyungsook Han (email)
This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 2.0 Korea License.

도면2


생물학적 과정
Biological process
(GO 0008150)
분자기능
Molecular function
(GO 0003674)
세포 성분
Cellular component
(GO 0005575)
대사과정
Metabolic process
(GO 0008152)
결합
binding
(GO 0005488)
세포 내
Intracellular
(GO 0005622)
일차 대사과정
Primary metabolic process
(GO 0044238)
핵산 결합
Nucleic acid binding
(GO 0003676)
뉴클레오티드 결합
nucleotide binding
(GO 0000166)
세포질
Cytoplasm
(GO 0005737)
고분자 대사과정
macromolecule metabolic process
(GO 0043170)
DNA 결합
DNA binding
(GO 0003677)
ATP 결합
ATP binding
(GO 0005524)
사이토졸
Cytosol
(GO 0005829)
단백질 대사과정
Protein metabolic process
(GO 0019538)
전사인자 활성
Transcription factor activity
(GO 0003700)
GTP 결합
GTP binding
(GO 0005525)
FAD 결합
FAD binding
(GO 0050660)

도면3


HPRD_04804
HPRD_03989
HPRD_12228
HPRD_01721
HPRD_04738
HPRD_01722
HPRD_04100
HPRD_01451
HPRD_03967
HPRD_06716
HPRD_03297
HPRD_07135
HPRD_10360
HPRD_06288
HPRD_04398
HPRD_06692
1
HPRD_01265
HPRD_00766
HPRD_09191
HPRD_11978
HPRD_11820
HPRD_09973
HPRD_00743
2
HPRD_11762
HPRD_16742
HPRD_02171
3
HPRD_03384
HPRD_06449
HPRD_01359
HPRD_16515
4
HPRD_04484
HPRD_03356
HPRD_09704
HPRD_16544
HPRD_03221
5
HPRD_03845
HPRD_13115
6
HPRD_03912
HPRD_12808
HPRD_05201
HPRD_02431
HPRD_02944
HPRD_11309
HPRD_02582
HPRD_04927
HPRD_12986
7
HPRD_06802
HPRD_01496
HPRD_09084
HPRD_05428
HPRD_06774
HPRD_09468
HPRD_08986
HPRD_00591
HPRD_02619
HPRD_03479
HPRD_01158
8
HPRD_00424
HPRD_02147
HPRD_01235
HPRD_12171
9
HPRD_03051
HPRD_02300
HPRD_01368
HPRD_02110
HPRD_02610
10
HPRD_04066
HPRD_05802
HPRD_02963
11