

물질의 약리기전 및 독성기전관리를 위한 데이터베이스

이국경, 양원형, 박원근, 송태웅, 허창복, 박전홍, 홍용원*, 김송학*, 오찬우*,
제주대학교 농업생명과학 수의학과, D&A*

Development of Database for the Management of Pharmacological and Toxicological Mechanisms of Chemicals

Kookkyung Lee, Wonhyung Yang, Wonkeun Park, Taewoong Song, Changbok Hur,
Jun Hong Park, Yong Won Hong, Song Hak Kim, Chan Woo Oh

Department of Veterinary Medicine, College of Agriculture & Life Science, Cheju National
University, D&A

Abstract : Managing, processing, and expressing the data, the database system makes users save the time and labors required to look up and organize the useful data as well as helps to create some new information and ideas. However in Korean situation, our the history of research is not too long and the resources including database systems is not enough. Because other previous database systems manage the data in a form of text file. It is not easy to obtain a useful analyzed results at once. The goals of this project are to extract experimental data from the research on endocrine disruptors, to manage the data as a code form, to process the data, to create the useful analyzed results, to help toxicological researchers, and eventually to enhance the availability of raw data.

Before the systems design, we constructed the code tables from the reports and executed a pilot test. Depending upon the result of pilot test, we determined the factors to be extracted as data set from reports, constructed database

tables, and designed the software to manage the data-input and to search and analyze data. Data-input module help users easily to analyze the report in a paper form and to input the data. Developed in the web-based form, data-search module is able to give an access to public as well as researchers. Also depending on the data saved in a code form, data-search module can provide much useful information and have a potentiation to be developed as more powerful data-analyzing system.

Key words : database, pharmacological mechanisms, toxicological mechanism

서론

데이터베이스는 자료를 체계적으로 관리하고 입력된 자료를 여러 형태로 표현함으로써 사용자로 하여금 자료의 검색과 정렬함에 소모되는 시간과 노력을 덜어 주고 새로운 정보 및 아이디어의 창출을 도와준다. 그러나 국내에서의 본격적인 자연과학에서의 IT(Intellectual technology) 분야는 그 역사가 길지 않으며 아직까지 만족할 만한 정보를 제공해주는 데이터베이스는 부재한

Corresponding author : Kookkyung Lee, 제주도 제주시 아라1동 제주대학교 수의학과
Tel:064-754-3378, Fax:0505-754-3378, E-mail:syeon@cheju.ac.kr
이 연구는 2003년 식품의약품안전청 내분비계 장애물질 평가사업에 의해 연구된 결과임

실정이다. 또한 기존의 데이터베이스는 텍스트 구조로 되어 있으며 이들을 위한 독성정보의 조회나 비교가 용이하지 않다. 이러한 구조는 단지 텍스트 문서를 관리수준 정도에 지나지 않아 전문가적 활용도면에서 볼 때 기존의 종이로 보관하고 있는 자료의 활용도와 크게 다르지 않다.

최근에 다양한 분야에서의 IT 기술산업이 발전하면서 기존자료에 대한 분석을 통한 분석결과의 유용성과 그 가치가 인정되고 있다. 예를 들면 toxicogenomics 연구는 물질의 의해 변화되는 유전자 발현결과를 분석하여 세포에서의 물질의 작용기전을 이해하고자 한다. Toxicogenomics 연구는 분석을 위하여 수천 또는 수만 개의 유전자발현 조절 결과를 얻기 위하여 막대한 비용이 요구되며 물질의 기전 및 독성작용에 대한 뚜렷한 결과물을 아직까지 제시하지는 못하고 있으나 곧 유용한 자료가 산출될 전망이다. 데이터베이스를 이용하여 위와 같은 기능이 가능하려면 입력된 자료가 텍스트 형태로 구성되어 있으면 불가능하며 각각의 관찰항목 및 결과수치 등의 자료가 분리된 tabulation된 자료로 구축되어야 한다. 이에 물질에 대해 이미 보고된 실험결과의 tabulation은 물질의 작용기전을 이해함은 물론 작용기전을 증명할 실험의 필요성 여부를 결정해주는 유용한 도구가 될 것이다. 또한 물질별 세부적인 관찰결과의 검색이 가능함으로써 물질에 대한 반복시험을 피할 수 있음은 물론 데이터를 별도로 가공하지 않아도 다양한 응용분석에 활용될 수 있을 것이다. 따라서 본 연구에서는 기존의 연구결과를 분석·코드화된 데이터베이스를 구축함으로써 생약추출물을 비롯한 약물, 독성물질의 연구현황과 분야, 정보의 조회 및 물질별 비교를 한 개의 표로 보여지게 하여 물질의 약리학적 및 독성학적 이해를 증대시키고자 하며 향후 새로운 IT 분야의 한 영역으로서의 분석연구에 유용한 도구로 활용될 기반을 마련하고자 하였다.

개발환경 및 방법

1. 프로그램의 기본구조 및 개발환경

본 프로그램은 크게 두 개의 모듈로 설계하였다. 입력 모듈은 로컬 클라이언트 프로그램으로 관리자가 서버컴퓨터에 접속하여 연구문헌의 결과를 입력할 수 있도록

하기 위한 모듈이며, 조회모듈은 웹 기반으로 작성하여 일반 사용자가 인터넷을 통해 접속하여 데이터를 조회할 수 있도록 하기 위한 모듈이다. 따라서 입력모듈은 다양한 코드의 관리와 코드를 이용한 연구결과의 입력이 용이하도록 하도록 델파이로 작성하였으며 조회모듈은 active server program(ASP)으로 작성하였다.

2. 데이터베이스 및 입력모듈의 설계

(1) 기초정보 및 기초코드의 수집 및 작성

프로그램에 활용되는 기초정보는 주로 물질과 문헌에 대한 기초정보이며 이외 다수의 기타 정보를 포함한다. 물질에 대한 기초정보는 물질명, 동의어, CAS No., 분자량, 끓는점, 녹는점, ACX No., EPA No., 화학구조식, 증발율, 비중, 조성으로 구성되며 현재 약 5,000여종의 물질정보를 수집하였다. 수집은 별도의 자동수집 프로그램을 이용하였다. 비주얼 베이직으로 작성된 자동수집 프로그램은 인터넷으로부터 데이터를 수집한 후 microsoft사의 access 프로그램의 데이터베이스 테이블에 자동저장되도록 하였다. 엑세스 테이블의 저장된 데이터는 개발된 프로그램의 데이터베이스 테이블에 일괄 삽입하였다.

문헌에 대한 기초정보는 고유번호, 용역연구과제번호, Pubmed ID, 과제명, 연도, 시작페이지, 끝페이지, 불림번호, 연구책임자, 연구자명, 논문게재여부로 구성하였으며 크게 용역연구과제, 연구문헌(Pubmed ID 포함), 기관보고서로 나누었으나 입력의 용이성과 메모리의 불필요한 사용을 줄이기 위하여 별도의 테이블이 아닌 하나의 테이블로 구성하였으며 입력시 한 화면에서 구분하도록 하였다. 문헌은 Pubmed ID를 통해 초록의 조회가 가능하도록 하였다.

기초코드 중 가장 기본적인 관찰항목코드는 유전자 발현관련 항목, 독성기전항목, 일반실험에서의 관찰항목, 병리조직학적 진단코드로 나누었다. 유전자 발현관련 항목(gene expression)은 물질에 의해 발현이 변하는 유전자명을 관찰항목명으로 설정하였으며, 유전자명은 문헌에서 일반적으로 인용되는 유전자명을 입력하였다. 작용기전(mechanism)은 물질의 특징적인 작용기전만을 모아 관찰항목으로 형성하였다. 일반 관찰항목(general observation)은 기존의 실험결과로부터 추출한 실험항목으로 표 1은 현재 입력된 유전자 발현관련 항목 및 독성기전, 일반관찰항목, 병리조직학적 진단코드의 예이다.

Table 1. Classification and partial lists of the parameters to be extracted from experimental results. Each code can be edited, deleted and added. Even though observational parameters are modified, No. of code remain unchanged permanently

Category	No. of Code	Observational parameters
Gene Expression	1001	Acyl CoA synthase 4
	1002	Calbindin d9k
	1003	Calbindin d9k, induced by estrogen
	1004	cDK inhibitor
	1005	cdk1
	1006	cdk2
	1007	cdk6
	1008	c-myc
	1009	cyclin A
	1010	cyclin B
	
Transcriptional regulations	1001	Transactivation by NFkB
	1002	Transactivation by PPARalpha
	1003	Transactivation by estrogen receptor
	1004	Transactivation by PAH receptor
	1005	Transactivation by CAR receptor
	1006	Transactivation by PXR
.....		
Toxic mechanism	1001	Protein kinase C activity
	1002	Alteration of calcium flux
	1003	Aromatase inhibition
	1004	Gonadotropin glycosylation
	1005	Alteration of norepinephrine or epinephrine synthesis
	1006	Blocking of LH receptor
	1007	Alteration of storage mechanism
	1016	Androgen receptor synthesis
.....		
General Observation	1001	11-hydroxyprogesterone induced by ACTH
	1002	Accessory sex organ weight
	1003	Acyl CoA synthase 4(ACS4) protein induced by ACTH
	1004	Adrenal gland weight
	1005	Adrenal gland weight, relative
	1006	Alanin aminotransferase(ALT) of blood
	1007	Albumin
	1008	Androgen receptor synthesis
	1009	Anogenital distance of offspring
	1010	Antioxidant effect
	1011	Aspartate aminotransferase(AST) of blood
.....		
Histopathological findings	1001	Abscess
	1002	Acanthocytosis
	1003	Accessory adrenocortical tissue
	1004	Accumulation
	1005	Acne
	1006	Adenomyosis
	1007	Adenoma
	1008	Adhesion
.....		

(2) 기타코드 테이블의 작성

기타코드로는 물질의 투여대상이 되는 동물종 및 세포주, 투여대상의 암수와 상태코드, 투여물질의 단위코드 등을 포함한다.

(3) 데이터베이스 설계를 위한 문헌분석

본 데이터베이스의 개발은 기존에 존재하지 않는 새로운 형태의 데이터베이스를 개발하는 것으로 다양한 문헌의 서로 다른 형태의 데이터 입력 및 처리가 가능해야 한다. 따라서 데이터베이스의 설계에 앞서 이들 다양한 형태의 데이터를 처리할 수 있는 적합한 입력모듈을 설계하기 위하여 문헌분석을 실시하였다. 문헌분석은 물질명과 문헌제목 등의 기초정보를 활용하여 데이터추출 및 데이터세트 작성의 적합성을 확인하였으며 결과는 microsoft사의 excel 프로그램으로 작성하였다.

문헌분석 결과, 실험결과로 구성된 데이터세트를 입력함에 있어서 먼저 문헌제목을 선택한 후 물질명과 연구분야를 선택해야 했다. 이후 물질의 투여대상이 되는 동물종 또는 세포주를 선택하였다. 투여대상은 모두 암수를 구분하도록 하였으며 표적장기 또는 세포주를 선택하도록 하고 관찰결과는 증가, 감소, 변화없음의 3단계로 분류하였다. 투여물질에 대해서는 투여량 또는 투여농도와 단위를 추출하였다. 추출에 필요한 인자의 수가 늘어날수록 데이터베이스의 세밀성은 향상되나 문헌으로부터의 데이터추출작업과 시스템메모리 활용에 있어서의 효율성을 저하시킴으로 적절한 수준의 추출인자를 설정해야 했다. 이상의 문헌분석결과를 정리하면, 연구문헌으로부터 문헌제목, 물질, 연구분야, 관찰항목, 투여대상동물 또는 세포주, 관찰결과, 투여량 또는 투여농도, 단위, 유의성 수준을 추출함이 바람직할 것으로 판단되었다.

(4) 입력모듈의 설계 및 개발

입력모듈은 관리자가 로컬 클라이언트 프로그램으로 서버컴퓨터에 접속하여 연구문헌의 결과를 입력할 수 있도록 하기 위한 모듈이다. 연구분야별로 입력해야 할 인자는 문헌분석결과에 기초하였다. 데이터의 입력순서는 기초코드를 먼저 입력하고 이를 이용한 연구문헌의 실험결과를 입력할 수 있도록 설계하였다. 따라서 입력모듈은 크게 기초코드 관리모듈과 데이터세트 입력모듈로 구분된다. 기초코드 관리모듈에서는 문헌분석에서 추출한 코드의 추가 및 수정, 삭제가 가능하도록 설계하였으며,

데이터세트 입력모듈에서는 문헌분석에서 이용된 방법에 따라 문헌제목 및 물질명을 필수적으로 선택하고 기타 데이터를 코드로부터 선택하여 입력하도록 하였다. 문헌명, 물질명, 투여대상이 동일한 상태에서 관찰항목과, 투여농도, 투여단위만을 달리 반복적으로 입력해야 하는 경우에는 사용자의 오류로 인하여 입력데이터의 정확성을 저하시키는 결과를 초래하였다. 따라서 반복적인 데이터의 정확성과 입력의 효율성을 향상시키기 위하여 일괄입력기능을 설계하였다. 또한 입력모듈에서도 입력된 데이터세트를 검색·확인하여 수정 및 삭제가 가능하도록 하였다.

3. 조회모듈의 설계 및 개발

조회모듈은 웹을 통해 일반인들의 조회가 가능하도록 하였으며 크게 물질별정보, 관찰항목별정보, 연구분야별정보로 나누었다.

물질별정보는 사용자의 편의성을 제공하기 위하여 물질명 또는 동의명의 일부분만으로도 검색되도록 하였다. 물질을 검색하면 물질에 해당하는 연구분야별 결과가 조회되도록 하였다. 이는 물질별로 수행된 연구분야 현황을 쉽게 알 수 있게 하기 위한 조회결과이며, 여기서 해당 연구분야를 선택하면 연구분야별 관찰항목의 입력된 결과들을 조회할 수 있게 하였다. 관찰항목 마다 증가, 변화없음, 감소의 관찰결과를 조회할 수 있으며 각 관찰결과를 선택하면 입력된 결과의 세부사항인 투여대상, 관찰대상, 연구문헌 정보를 볼 수 있게 하였다. 예비 문헌분석결과, 투여대상과 관찰대상이 다른 경우의 시험 또는 실험이 존재하며 이는 관찰항목명 및 관찰결과에 혼동을 야기시킬 우려가 발생하였다. 따라서 본 데이터베이스에서는 연구결과에 대한 대상을 투여대상으로 하고 관찰대상이 투여대상의 태아 또는 차세대의 경우에는 관찰항목명으로부터 투여대상과 관찰대상을 구분할 수 있도록 하였다. 예를 들면 관찰항목이 body weight 일지라도 차산자의 체중은 body weight of offspring으로 구분되어 조회가 되도록 하였다.

관찰항목별정보는 물질별 정보와 같이 관찰항목명의 전체 또는 일부 키워드로 검색한 후 관찰항목별로 입력된 물질의 연구결과를 조회하도록 하였다. 따라서 관찰항목별로 입력되지 않거나 아직 연구되지 않는 물질을 확인할 수 있으며 관찰항목에 대한 물질들간의 연구결과를 비교할 수 있다. 연구분야별정보는 연구분야별로 선택하여 최종적으로는 물질과 관찰항목명 또는 모니터링

분야별로 입력된 결과를 볼 수 있도록 하였다.

Ⅲ. 결과 및 고찰

1. 입력모듈

입력모듈은 연구문헌을 분석함과 동시에 입력하는 작업을 용이하게 할 수 있도록 제작한 프로그램이다. 입력모듈은 일종의 클라이언트 프로그램으로 현재 인터넷망을 통해 어디서든 접속하여 입력할 수 있으나 데이터베이스의 입력 및 수정의 권한을 갖는 관리자들만을 위한 모듈이다.

연구결과를 입력하기 전에 필요한 물질, 문헌 등의 기

초코드를 입력모듈 메뉴의 기초자료에 모았으며 연구문헌의 결과인 데이터세트는 입력관리를 통해 입력하였다. 5,000 여종 물질의 기본정보를 입력하였으며 fig. 1은 일례로 DEHP를 입력한 화면이다. 화면의 상단에서는 물질명, CAS No., 분자량 등의 기초정보를 입력하며 화면의 하단에서는 물질의 동명을 입력할 수 있게 하였다. 물질의 동명은 한 줄씩 추가 입력이 가능하도록 하였으며 화면 상단 오른쪽에는 저장된 물질 구조식을 볼 수 있다. Fig. 2는 문헌관리화면의 예로 문헌의 기초정보를 관리하는 화면이다. 화면의 상단에는 문헌제목은 비롯하여 문헌이 실린 보고서명 또는 저널명, 연도, 페이지 등을 입력할 수 있으며 하단에는 저자명 입력이 가능함은 물론 원문화일을 서버에 업로드할 수 있어 조회모듈에서 원문조회가 가능하다.

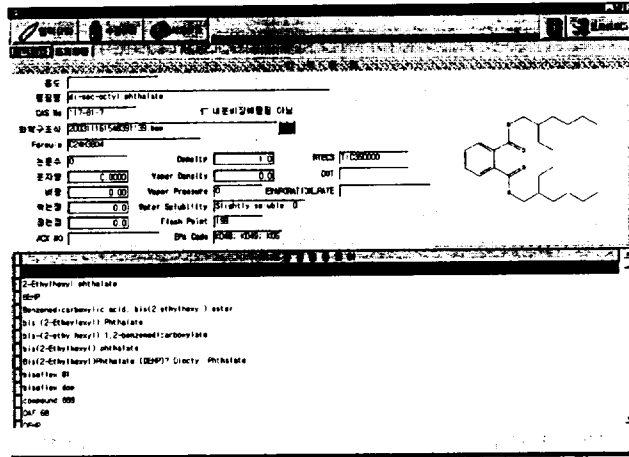


Fig. 1. Management of chemical information. The set of each information is composed of chemical name, CAS No., structure, molecular weight, boiling point, synonym, and et. al. The information set can be added, edited, and deleted.

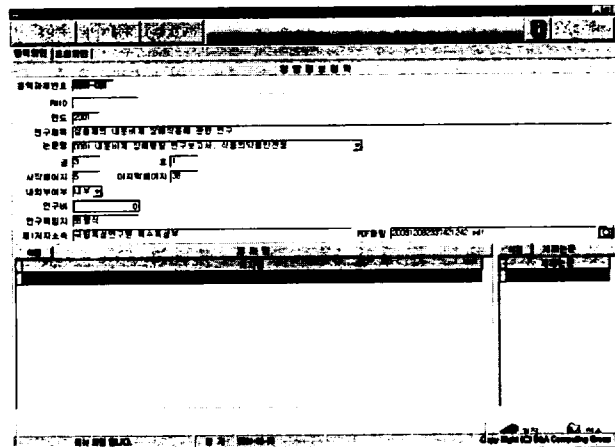


Fig. 2. Management of general information on research reports or articles. The set of each information is composed of title, journal name, volume, pages, publication year, authors, and et. al. The information set can be added, edited, and deleted.

관찰항목코드를 비롯한 연구분야, 동물종, 지역, 단위, 상태 등의 코드 또한 fig. 3에서와 같이 각 해당 코드관리 페이지에서 입력, 수정, 삭제가 가능하다. 단, 한번 입력한 코드명은 고유한 번호를 계속적으로 차지하게 되며, 고유 번호는 입력된 데이터세트와 연계된다. 따라서 코드명을 수정하게 되면 코드번호는 유지된 상태에서 코드명만 바뀌게 된다. 이미 이 코드를 이용하는 데이터세트가 저장되어 있다면 코드명의 변경 후에는 해당 데이터세트는 달라진 코드명을 갖게 되므로 코드의 변경은 매우 신중을 기해야

한다. 이처럼 코드의 변경에 있어서 주의해야 하는 단점이 있으나 코드의 추가나 정확한 코드명으로 수정해야 할 경우 매우 편리함을 제공한다. 특히 본 시스템은 많은 연구 문헌의 다양성을 수용해야 하는 어려움이 따르기 때문에 코드명의 변경이나 추가가 수시로 요구된다. 따라서 이처럼 코드명만 수정하면 코드번호를 사용하고 있는 모든 데이터세트는 변경된 코드명을 사용하게 된다. 향후 착오에 의한 코드의 삭제 또는 수정에 의한 데이터의 손상을 막기 위해서는 보완이 요구된다.

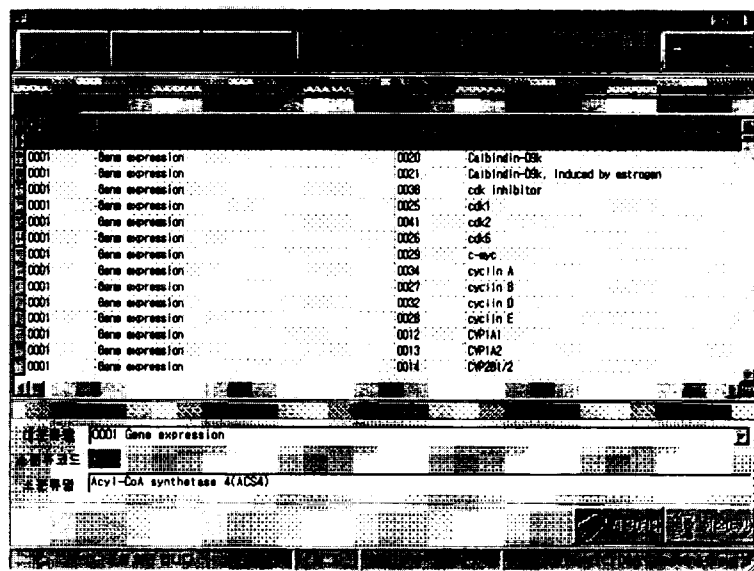


Fig. 3. Management of the observational parameters and its code. This parameter is composed of the 4 classes(gene expression, mechanism, general observation, and histopathological diagnostic lesions) and each class included detailed parameters(right column). The each codes can be added, selected, modified, and deleted.

2. 조회모듈

조회모듈은 입력된 연구결과를 사용자들이 쉽게 조회하여 유용한 정보를 활용하기 위한 프로그램이다. 따라서 접속면에서 가장 용이한 웹을 기반으로 제작하였으며 데이터베이스 테이블로부터 바로 입력한 또는 수정한 내용을 실시간으로 조회할 수 있다. 조회모듈은 물질별, 관찰항목별, 연구분야별정보의 조회기능을 주요기능으로 설정하였으며 fig. 5는 물질별정보의 조회화면으로 물질명을 검색하여 물질별로 해당하는 연구분야의 관찰항목 결과 및 모니터링 검출농도를 조회할 수 있다. 예를 들어 물질명검색란에 'glycyrrhetic'을 입력한 결과를 fig. 6에서 볼 수 있으며 검색결과로 glycyrrhetic

acid가 검색되었음을 확인할 수 있다. Fig. 6은 fig. 7에서 glycyrrhetic acid의 자료수를 선택한 결과화면으로 해당시험분야에서의 관찰항목과 관찰결과로 구성된 표를 볼 수 있다. 관찰결과와 자료수를 선택하면 입력된 실험결과에 대한 상세내용을 볼 수 있다(Fig. 7). 관찰항목별정보와 연구분야별정보의 조회 역시 물질별정보에서와 같이 검색 또는 트리(tree) 구조로 구성되어 있어 항목별로 조회할 수 있다(Fig. 9, 10). Fig. 9는 관찰항목별 정보화면으로 fig. 5에서 입력한 melanin content의 일부인 melanin만을 입력하여 검색한 화면으로 fig. 5에서 입력한 내용이 조회됨을 볼 수 있다. 물질별정보에서와 같이 문헌정보를 선택하면 실험 데이

터를 포함하고 있는 문헌의 정보를 볼 수 있다. Fig. 10은 로 조회되며, 각각의 연구분야별 자료수를 선택하면 세부 연구분야별 정보화면으로 입력된 모든 자료가 연구분야별 연구별, 관찰항목별, 물질별순으로 조회가 가능하다.

Fig. 4. Example of data input. In this example, article title was selected from previous saved database and glycyrrhetic acid was entered as a chemical. In addition to article title and chemical name, the general observation parameter(melanin content), the result(decrease), cell lines(BL16), dosage, and dosage unit were input, respectively.

Fig. 5. Chemical search screen. The screen shows the result of searching 'glycyrrhetic'. It is possible to search for the synonyms using a partial name.

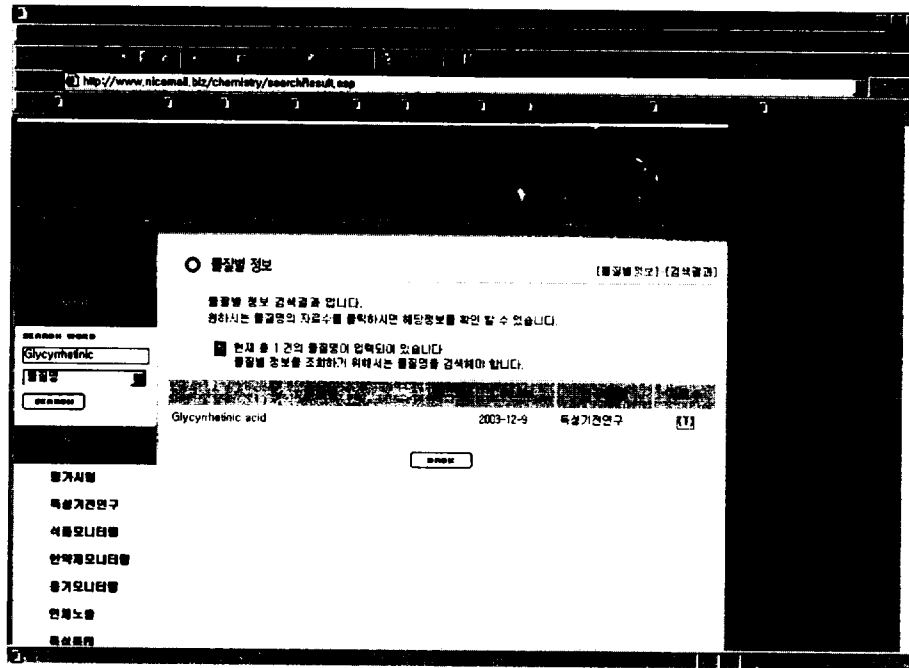


Fig. 6. This is a result screen of fig. 5. Glycyrrhetic acid is searched. The research data of each chemical are divided into research fields. The number in the rightmost column means the number of data about the chemical.

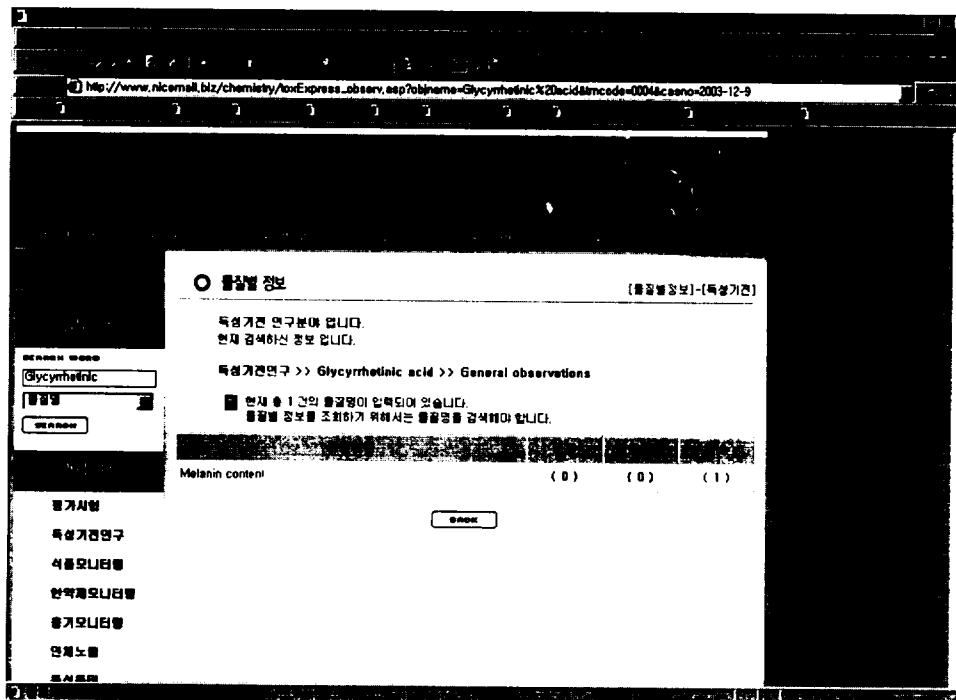


Fig. 7. The third result screen of chemical search. In Fig. 6, selecting the number, its contents are divided into three kinds of results, increase, no change, and decrease.

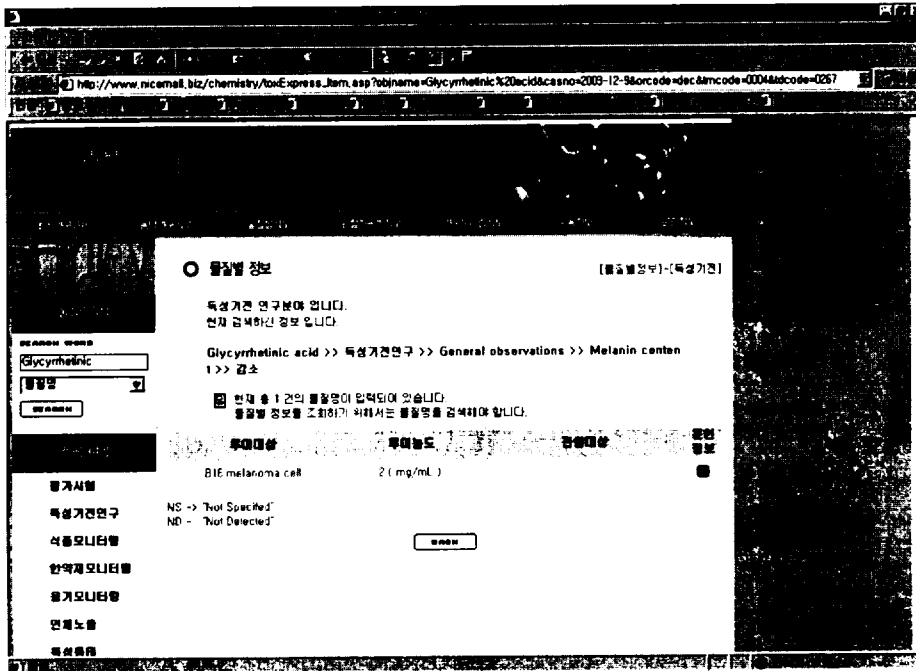


Fig. 8. The fourth step of chemical search. This result screen shows a detail information such as chemical, experimental animal(or cell line), dosage, and its reports.

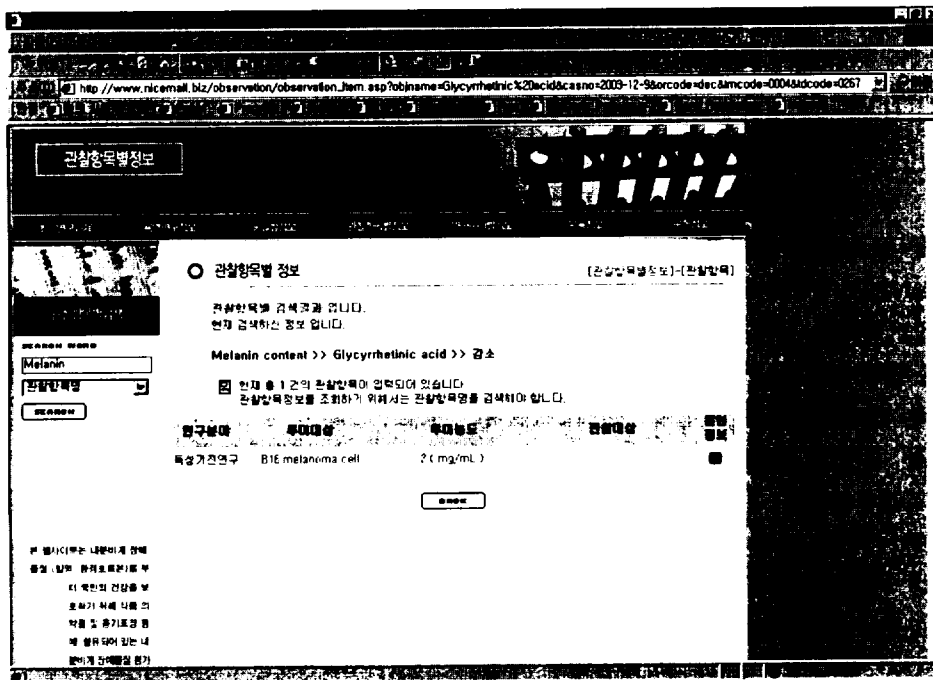


Fig. 9. Search result of observational parameters. This window shows the result when 'melanin', as a search term, was looked up.

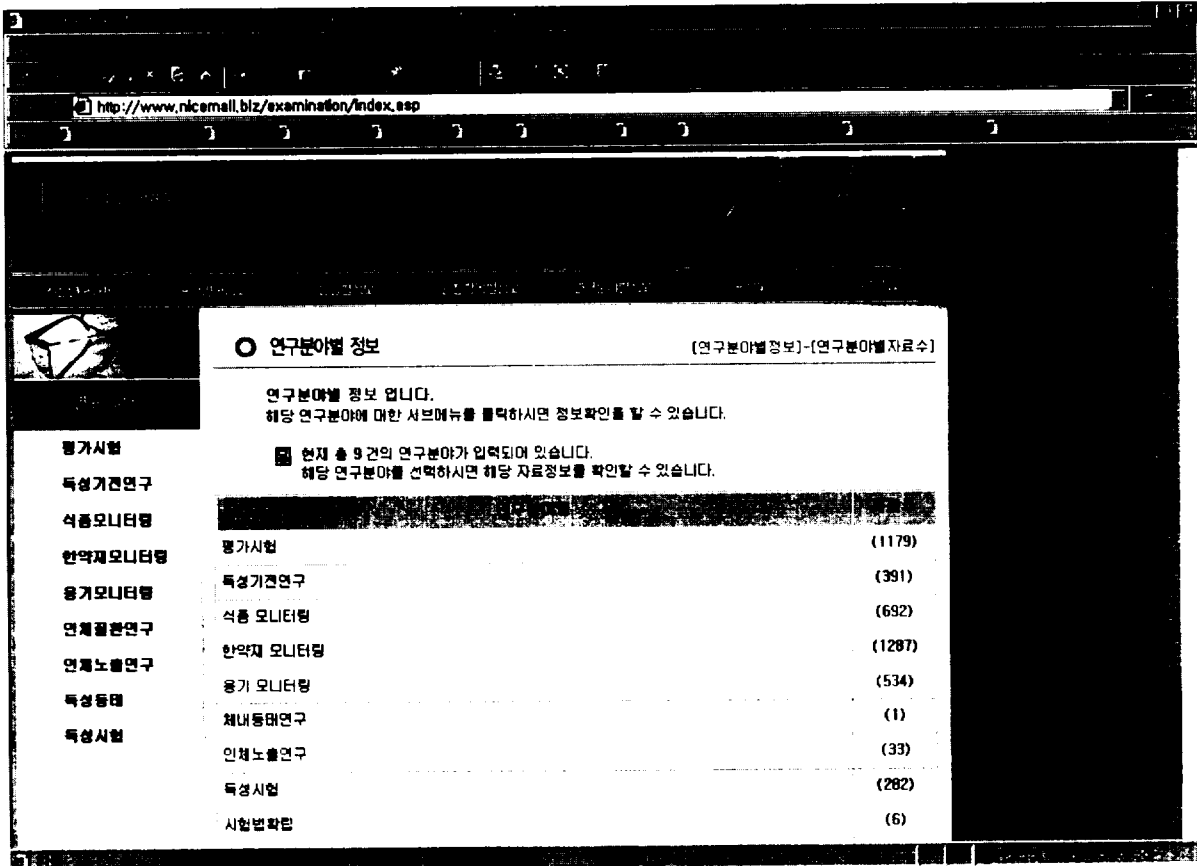


Fig. 10. Search result depending on research fields. This screen show that the input data are divided depending upon their research field. The numbers mean the total data sets input into the research field of database.

이상의 물질별, 연구분야별, 관찰항목별 정보의 조회는 SQL의 쿼리문(query)을 이용한 결과이다. 다소 복잡한 쿼리문(query)을 이용하여 유용한 결과를 보여주나 쿼리문 검색 그 자체는 데이터베이스를 이용하는 가장 기초적인 활용방법이다. 더욱 합리적이고 유용한 응용분석을 위해서는 물질들간의 상관관계, 유사물질들의 군 형성, 그래프를 이용한 결과의 표현, 특정 관찰항목들의 통합분석, 특정물질들의 통합분석 등과 같이 데이터베이스의 세밀한 가공과 이를 구현시킬 응용프로그램이 필요할 것이다. 현재 본 시스템은 응용분석을 위한 데이터베이스의 구조를 갖추고 있으나 이를 분석할 응용프로그램은 개발되지 않았다. 그러나 본 시스템의 데이터베이스는 응용분석 모듈확장을 위한 충분한 호환성을 갖추고으로써 향후 독성연구의 한 연구분야로서의 유용한 도구로 성장할 기반이 될 것이다.

또한 본 데이터베이스는 물질의 독성기전 및 식품 및 한약재에서의 검출농도 등을 위주로 개발하였으나 fig. 3에서와 같이 감초추출물과 그 작용 및 결과, 투여농도, 투여단위의 입력이 가능하였다. 또한 fig. 7에서는 세부 정보의 조회가 가능함으로써 본 시스템이 의약품 또는 일반추출물의 약리학적 기전 및 효과도 충분히 관리할 수 있음이 입증되었다. 따라서 데이터베이스 테이블의 분할과 모듈을 구분하여 활용한다면 생약 또는 한약, 해양산물 등과 같은 다양한 분야에서의 물질의 작용기전 및 유전자발현 효과를 별도 또는 종합적으로 구분할 수 있는 시스템으로 활용이 가능할 것으로 사료된다.