

# 항체 데이터베이스 구축에 대한 연구

이동준<sup>0</sup>, 심찬섭, 박형선  
 (주)엘피스바이오텍, 한국과학기술정보연구원  
 gump@elpis-biotech.co.kr, cshim@elpis-biotech.co.kr, seonpark@kisti.re.kr

## Construction of Antibody Database

Dong-Jun Lee<sup>0</sup>, Chan-Seob Shim, Hyung-Seon Park  
 ELPIS-BIOTECH, KISTI

### 요 약

게놈 프로젝트가 완성된 후 2년여가 지난 지금, 단백질체학(Proteomics)의 분야로서 기능분석위주로 연구가 진행되고 있다. 그 중에서도 동물계에서 면역체계를 구성하는 가장 핵심적인 요소로서 항체를 뽑고 있고 가장 활발하게 연구되고 있다. 현재, 세계적으로 구입이 가능한 연구용 항체 (Antibody)의 수는 약 70만종 정도로서 그 수는 생물학의 발전속도에 비례하여 급격히 증가하고 있는 추세이다. 항체는 생물학의 기초 연구 도구로서뿐만 아니라 질환의 진단 및 치료제는 물론 산업적으로도 그 활용가치가 큰 생물자원으로서 그 가치가 점차 중요하게 부각되고 있다. 향후 전개될 단백질체학의 주요 연구도구로서 항체가 맞춤형학을 위한 신약개발이나 신규단백질 동정에 필요한 항체칩(Immuno-chip)등을 생산할 목적으로 항체데이터베이스를 구축하는 것이며 항체의 기본적인 정보들은 앞으로 연구자의 연구방향에 지대한 영향으로 미칠 것으로 사료된다.

### 1. 서론

항체는 생물학의 기초 연구 도구로서뿐만 아니라 질환의 진단 및 치료제는 물론 산업적으로도 그 활용가치가 큰 생물자원으로서 그 가치가 점차 중요하게 부각되고 있다[1]. 특히 게놈프로젝트 (Genome project)의 연구결과로서 축적된 유전자 염기서열 분석을 통해 사람의 유전자 지도가 완성된 현시점에 있어서 생물학의 기본 방향이 유전자의 최종산물인 단백질의 기능성 연구로 전환되는 경향에 비추어 볼 때 단백질의 기능성 연구를 위한 항체의 사용이 급증할 것으로 전망된다. 단백질체학 (Proteomics)은 게놈프로젝트를 통해 얻은 유전자 정보로부터 유전자의 기능과 각 유전자간의 복잡한 상호 네트워크 (Intra- 또는 Inter-cellular network)을 규명하는 것을 최종 목적으로 하고 있으며 그 결과는 유용 유전자 산물의 산업적 응용이나 인간의 생명현상과 직결된 기초의학, 약학, 그리고 의공학 등에 직접적으로 이용될 수 있다. 항체의 높은 친화력과 다양한 응용성 등으로 인하여 향후 전개될 단백질체학의 주요 연구도구로서 항체가 깊은 관심을 모으고 있다. 더욱이 최근 맞춤형학을 위한 신약개발이나 신규 단백질의 분리 동정에 항체칩 (Immuno-chip) 등이 본격적으로 이용될 전망이어서 항체의 기본적인 정보의 취득은 앞으로의 연구방향에 지대한 영향을 미칠 것이다.

### 2. 항체 데이터 자료축적과 입력필드

#### 2.1. 판매용 • 비매용 항체 데이터 자료축적

약 160여개 항체제조회사의 on-line 카탈로그들을 검색하여 그 중에 유용한 데이터 정보가 있는 100개의 회사를 선별하여 판매용 항체 데이터 자료와 논문검색의 결과 발굴한 신규 비매용 항체를 계속적으로 수집·데이터화·입력작업을 하였다.

#### 2.2. 항체 데이터 자료 구축현황

현재 판매용 항체 데이터 수는 3,400여건 수이고, 비매용 항체 데이터 수는 150여건 수로 전체 약 3,550의 항체 데이터가 한국과학기술정보연구원(이하 KISTI)의 SGI(Ruby)서버에 축적되어 있다.

#### 2.3. 항체 데이터 입력필드

각각의 항체 제조회사들의 항체 정보들을 검색한 결과 24개의 공통적인 데이터를 입력필드로 결정하였다(표 1)[2].

표 1. 24개의 항체데이터 입력필드

No	입력필드명	설명
1	Antigen name	항원이름
2	Antibody name	항체이름
3	Catalog name	카탈로그 이름
4	Description	상세설명
5	MW	분자량
6	Target tissue	적용조직
7	Antibody species	항체 종(Host의 단일클론과 다클론)
8	Isotype	면역글로불린의 종
9	Clone number	클론수
10	Procedures/Suppliers	항체제조회사
11	Amount	항체의 양
12	Catalog number	카탈로그의 고유번호
13	Application	용도범
14	Price	가격
15	Specificity	항체의 특이적인 성질
16	Presentation & Formulation	항체의 보관 및 보존
17	Storage	저장법
18	Stability	항체의 상태
19	Comments	항체 사용 시 주의할 점과 알아 두어야할 점
20	Image data	항체를 이용한 분석 사진
21	References of production	참고문헌
22	Related antibodies	관련된 항체들
23	DNA sequence for variable regions	항체의 가변부의 DNA의 염기서열
24	Epitope Sequence	에피토프 염기서열

본 연구는 KISTI 바이오인포매틱스센터 생물자원정보실의 지원에 의하여 수행되었습니다.

### 3. 항체 데이터베이스의 스키마

3.1. 항체 데이터 입력필드 형식

항체 데이터 입력필드는 모두 24개의 항목으로 되어 있으며 필드형식은 Int(\*\*), Varchar(\*\*), Text로 이루어져 있다. Int(\*\*)는 4byte의 정수 \*\*개까지 사용할 수 있으며, Varchar(\*\*)는 문자의 수가 최대 \*\*개까지 사용할 수 있는 문자열이다. 그리고 Text는 문자의 수가 최대 65535개까지 사용할 수 있는 문자열로 되어 있으며, Null은 중복되는 데이터의 유(Yes)/무(No)를 결정짓는 값이다[4].

3.2. 항체 데이터 입력필드 구조

KISTI에 구축하고 있는 항체 데이터베이스의 기본 모식도이다. 항체 데이터베이스는 6개의 카테고리가 있다. 항체의 전반적인 정보를 가지고 있는 "Antibody", 판매용 카다로그나 비매용 항체를 가지고 있는 항체제조회사나 항체보유자의 정보를 가지고 있는 "Catalog", 특정항체의 특이성이나 응용성 등의 정보를 가지고 있는 "Application", 특정항체의 사진이나 임상실험자료의 이미지 등의 정보를 가지고 있는 "Image", DNA 염기서열이나 에피토프 염기서열의 정보를 가지고 있는 "Sequence", 마지막으로 특정항체를 이용하여 실험하거나 임상에 사용한 논문이나 참고자료 등의 정보들이 있는 "Reference" 등이 있다. 각각 카테고리 안에는 서브 카테고리 가 있다(그림 1).

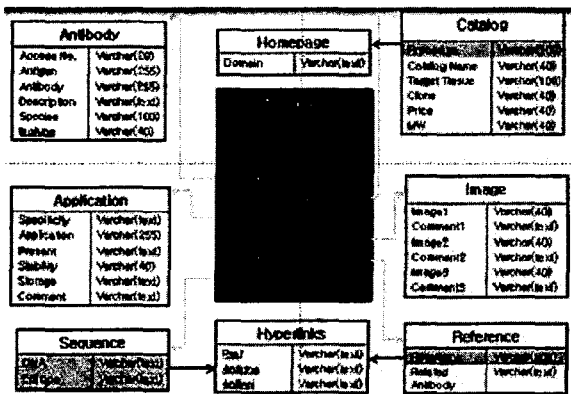


그림 1. 항체 데이터베이스 입력필드의 구조

4. 항체 검색엔진 툴 시스템 구현

4.1. 항체 검색엔진 툴 모듈

본 검색엔진 모듈은 현재 KISTI의 SGI(Ruby) 서버 SunOS 5.7에서 PHP언어로 시스템을 구현하였다. 그리고 Database Management System은 Mysql을 사용하였다. 현재 항체 검색엔진 필드는 주로 연구자가 많이 사용하는 항체명과 항원명으로 이루어지는 점에 착안하였으며 이 이외에도 항체사용방법, 면역체, 적용조직, 제조회사, 에피토프의 키워드로 검색이 가능하도록 시스템을 구현하였다[3][4].

4.2. 항체 검색엔진을 통한 검색 Processing 과정

7개의 검색필드에 특정 항체 정보를 검색하기 위하여 각 항목에 맞는 키워드를 입력하고 체크박스에 체크를 한 후, or나 and를 클릭하고 검색아이콘을 클릭하면 각 항목에 맞는 항체데이터 리스트가 검색결과 창에 출력된다. 검색조건에 맞는 항체데이터가 순차적으로 출력이 되고, 검색하고 싶은 항체 데이터의 번호를 클릭하면 그 특정 항체데이터의 전체 데이터 정보가 나타난다. 그리하여 특정 항체에 대한 사진 정보가 있으면 그 사진 필드에 있는 이미지를 클릭함으로써 큰 이미지 사진이 나타

난다(그림 2).

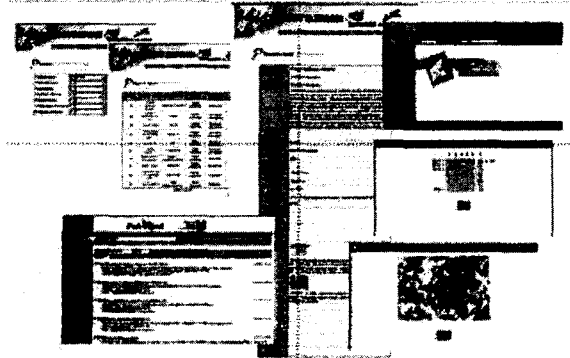


그림 2. 항체 데이터베이스의 검색 모식도

5. Bioinformatics를 이용한 항체분석 기반구축

5.1. 항체 가변부의 염기서열분석

항체 양쪽 가변부에는 특정 항원과 결합을 위한 특정 DNA 염기서열이 있다. 가변부에는 무거운 사슬(이하 VH)과 가벼운 사슬(이하 VL)이 함께 존재하고 있는데 약 110개의 아미노산이 내재되어 있다. 그 중에서 VH와 VL의 가변영역에는 3 ~ 10개의 아미노산이 특이성을 보여지고 있다. 이 특이성을 가진 아미노산들이 항원과 결합할 수 있는 열쇠가 된다. 그러므로 이들 아미노산을 코딩하는 DNA의 신규발굴 작업들이 진행중이다. 그러나 아직 데이터베이스로 구축하기엔 가변부 DNA 염기서열이 많이 밝혀진 않았다. 이들 DNA 염기서열들로 많은 분자적 차원에서 DNA Sequence Alignment, Motif Sequence Analysis, DNA Sequence Functional Phylogeny 등으로 항체의 분류(Phylogeny)[6][8]와 항원·항체의 특이성 분석하여 유사성을 나타내는 특징적인 부위들을 찾아 연관성과 진화적인 기능(Functional Analysis)을 탐지(예, ClustalW, BlockMaker, MEME, Match-Box 등의 프로그램 이용.)할 수 있는 기반을 구축하고 있다(그림 3).

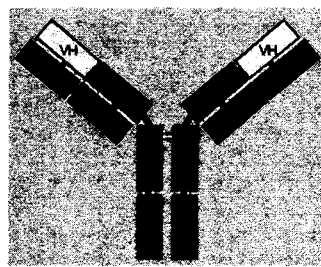


그림 3. 항체 가변부의 모식도

5.2. 에피토프의 염기서열과 기능분석

에피토프란 항원 분자상의 구조이며 항원항체 반응의 특이성을 결정하고 있는 분자구조이다. 그러므로 에피토프 염기서열을 이용한 Amino-acid Sequence Alignment, Motif Sequence Analysis, Amino-acid Sequence Functional Phylogeny(예, Blocks, Prints, ProDom, Pfam 등의 프로그램 이용.)의 방법 [6][8]들을 통하여 항원·항체간의 구조적 특이적 기질을 예측 분석하여 항원과 항체반응의 원인을 밝혀 질병의 근원을 찾을 수 있는 기반을 구축하고 있다(그림 4).

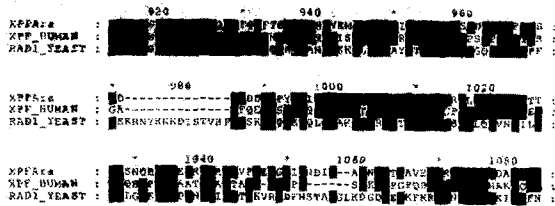


그림 4. Blocks 프로그램을 이용한 다중서열정렬로 유전자간 유사성을 비교분석

## 6. 항체 데이터베이스 웹 서비스

### 6.1. 웹 서비스의 구조

현재 항체 데이터베이스의 웹 서비스는 KISTI와 엘피스바이오텍 서버를 이용하고 있다. KISTI 서버에서 제공하고 있는 것은 항원-항체에 대한 정보와 그 특정항체의 실험한 데이터 사진이나 임상에서 사용중인 것들의 사진 정보들을 수록해 놓았다. 그리고 엘피스바이오텍 서버는 항원-항체를 분석할 수 있는 프로토타입과 연구자들과 공유할 수 있는 학술 자료들을 네트워크로 할 수 있도록 구축되어 있다(그림 5).

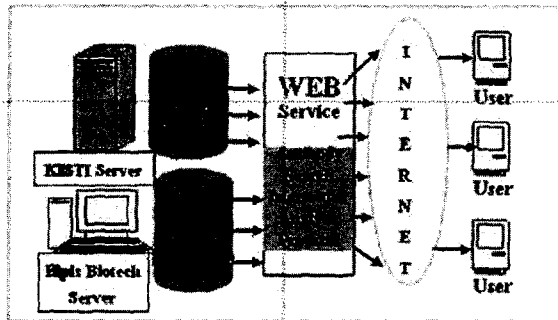


그림 5. 웹 서비스의 모식도

### 6.2. 항체 분석방법의 프로토콜

분석방법프로토콜사이트 내에는 항체와 관련된 면역실험기법, 분자생물실험기법, 생물정보분석기법등 실험기법들이 있다. 면역실험기법에는 Western 분석법, 면역학적 침전법, 면역표지법 등 면역학을 이용한 분석방법의 프로토콜을 제공하고, 분자생물학실험기법에는 cDNA library 제작법, RNA 분석방법, SDS-PAGE 등 분자생물학을 이용한 최근 분석방법의 프로토콜을 제공하고, 최근 생물정보학이 활발히 연구되고 있어 이를 이용한 많은 유전자분석기법들(유전자서열이 알려진 항체의 경우에는 유전자 서열분석, epitope sequence search, NCBI에서 이용한 유전자를 찾기, 1-D, 2-D 이미지 분석 툴 기법, Motif domain 찾기, 등등)을 제공한다.

### 6.3. 관련사이트

현재 150여 개의 항체에 관련된 사이트 링크를 하였고, 각각의 항체판매회사, 생물공학회사, 생물정보학, 분자생물학, 연구기관, 생물학저널 등의 카테고리들을 가지고 있으며, 각각의 사이트에 간단한 설명들이 있다. 앞으로도 생성되는 홈페이지를 계속적으로 링크할 예정이다.

## 7. 활용방안

연구자 입장에서 적합한 항체의 검색 툴을 제공하여 단백질

기능성 연구에 필요한 기초적 데이터를 제공한다. 또한 항체간 특이성 및 특성비교를 통한 질환진단 및 치료제 개발활성화를 도모 시키며 항체와 관련된 단백질체학에 대한 정보제공을 한다.

## 8. 고찰

전년도에는 최적의 검색환경 조성과 검색엔진 툴의 개발에 중점적인 목표를 두었기 때문에 실질적인 항체 자료는 아직 미흡하다고 할 수 있다. 따라서 올해에 이를 보완하는데 최대의 역점을 두었다. 대량의 항체 자료 축적을 위해 기존 자료를 바탕으로 항체의 기초자료는 물론 현재 사용되고 있는 임상자료의 예, 참고문헌, 항체 특성에 맞는 응용 가능성 등 자료를 추가 또는 수정하고 있다. 또한 부가적으로 외국의 항체판매회사와 기존의 데이터베이스를 링크시켜 관련항체의 부가자료를 직접 검색할 수 있는 기능 추가와 한국과학기술정보연구원 바이오인포매틱스 센터(CCBB)에 있는 국제적인 데이터베이스 (PDB, Swiss-plot, PIR)를 항체 데이터베이스 홈페이지에 포팅하여 직접적인 검색을 할 수 있도록 구축할 예정이며 이에 따른 부가적으로 추가할 수 있는 검색 시스템 개발도 할 예정이다. 생물정보학을 이용하여 기존의 유전자 데이터베이스로부터 항체가변부를 코딩하는 DNA의 염기서열을 발굴하고 항원 Epitope sequence 데이터베이스로 계속적으로 축적하면서 아미노산서열과 기능분석이 가능한 체계의 구축을 앞으로도 계속 시도할 것이다(그림 6).

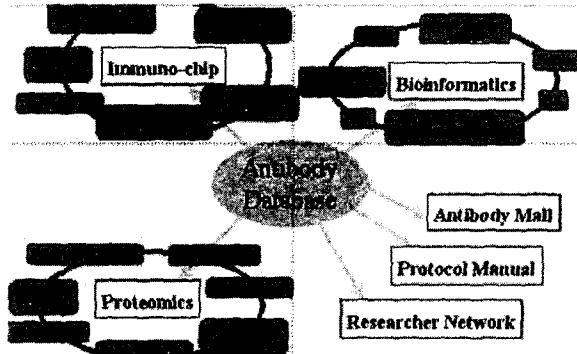


그림 6. 항체 데이터베이스 구축에 따른 파생되는 연구범위에 대한 기대효과의 개념도

## 9. 참고문헌

- [1] 김광택, 김영상 외, 세포분자면역학, 정문사, 1998.
- [2] 김재식, 면역학용어사전, 대학서림, 1988.
- [3] 김태훈 여인춘, 알기쉬운 PHP, 정보문화사, 2000.
- [4] 신상호 외, 데이터베이스 프로그램과 JDBC, 인포북, 1997.
- [5] Bioinformatics: Sequence, Structure and Databanks : A Practical Approach (The Practical Approach Series, 236), Des Higgins (Editor), Willie Taylor (Editor), 2000, Oxford Univ Press
- [6] Bruce Albert, Dennis Bray., et al, Molecular Biology of THE CELL -THIRD EDITION- p.1195-1254 Garland Publishing, Inc.1994.
- [7] Eli Benjamini., et al, IMMUNOLOGY A Short Course - THIRD EDITION- WILEY-LISS, 1996.
- [8] Bioinformatics: Methods and Protocols, Stephen Misener Krawetz (Editors),1999, Humana Press.