

# PDB 한글 뷰어

구형서<sup>o</sup>, 주영준, 박양수, 이명준  
울산대학교 컴퓨터정보통신공학부  
{masker, scjp, yspk, mjlee}@mail.ulsan.ac.kr

## PDB Korean Viewer

HyeongSeo Koo<sup>o</sup>, YoungJun Joo, YangSu Park, MyungJoon Lee  
School of Computer Engineering and Information Technology, University of Ulsan

### 요 약

단백질의 기능은 3차 구조(tertiary structure)에 의해 결정된다는 것이 밝혀짐으로써 단백질의 3차 구조에 대한 연구가 보다 활발하게 진행되고 있다. 단백질의 기능을 효과적으로 파악하기 위해서는 단백질의 3차 구조를 보여주는 도구가 필수적이며, 현재 단백질의 3차원 구조정보를 바탕으로 단백질의 3차 구조를 보여주는 다양한 뷰어가 존재한다.

본 논문에서는 PDB(Protein Data Bank) 파일을 분석하여 단백질의 3차 구조를 보여주는 PDB 한글 뷰어에 대해 기술한다. PDB 한글 뷰어는 기존 뷰어의 스테레오뷰(stereo view) 기능을 보완하여 한 단백질의 2가지 구조를 동시에 보여주는 기능을 제공한다. 또한, 아미노산 원자들에 대한 분석 기능과 PDB 데이터의 이해를 돕는 도움말을 제공함으로써 효과적으로 단백질의 3차 구조를 파악할 수 있다.

### 1. 서론

단백질은 생명현상 유지 및 생명체 내에서 발생하는 화학반응에 있어 필수적인 역할을 담당한다. 따라서 단백질의 기능을 정확하게 파악하는 것은 생명현상을 이해하는 데 매우 중요하다.

단백질의 기능은 단백질의 3차 구조에 의해 결정되는 것으로 밝혀졌다.[1] 즉 단백질의 내부 구성의 정확한 이해를 통해 단백질의 기능파악이 가능하다. 이를 위해 최근 단백질의 3차 구조를 파악하기 위한 연구가 활발하게 진행되고 있다. PDB(Protein Data Bank)는 X-ray 결정법(X-ray crystallography), NMR(nuclear magnetic response)과 같은 실험을 통해 단백질의 3차원 구조정보를 제공하고 있다.[2]

단백질의 기능을 효과적으로 파악하기 위해서는 단백질의 3차 구조를 보여주는 도구가 필수적이다. 현재 자바를 이용하여 개발된 단백질 3차 구조 뷰어로 QuickPDB[3], WebMol[4], JMV[5] 등이 있다.

이들 뷰어는 실험을 통해 밝혀진 단백질의 3차원 구조정보 데이터를 읽어 웹상에서 단백질의 3차 구조를 다양한 모양으로 보여준다. 그러나 기존의 뷰어들은 PDB가 제공하는 단백질의 3차원 구조정보의 내용 중 일부만을 사용하기 때문에 단백질 3차 구조에 대한 정보가 충분하지 않으며, 단백질 3차 구조를 파악하기도 힘들었다.

†본 연구는 정보통신부의 기초기술연구 지원사업의 지원으로 이루어졌음.

본 논문에서는 기존 뷰어들의 기능을 보완하여 PDB 파일에서 읽은 정보를 바탕으로 아미노산 원자간의 거리와 각도를 계산하고, 각 아미노산을 특성별로 분류할 수 있는 PDB 한글 뷰어에 대해 기술한다. PDB 한글 뷰어는 한 단백질의 두 가지 구조를 동시에 보여주는 기능을 제공함으로써 단백질의 구조를 보다 효과적으로 파악할 수 있다. 또한 개발된 뷰어는 일반 사용자가 PDB 파일을 이해하고 많은 정보를 얻을 수 있도록 PDB 파일의 요약 정보와, PDB 파일의 레코드 정보와 각 레코드별 의미정보를 도움말 기능을 통해 제공한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구로서 현재 널리 사용중인 단백질 3차 구조 뷰어에 대해 알아본 후, 3장에서는 PDB 한글 뷰어의 구조와 기능에 대해 기술한다. 마지막으로 4장에서는 결론 및 향후 연구과제를 기술한다.

### 2. 관련연구

#### 2.1 QuickPDB[3]

QuickPDB는 SDSC(San Diego Supercomputer Center)에서 단백질 3차 구조를 보여주기 위해 개발한 자바 애플릿으로 현재 PDB(RCSB:Research Collaboratory for Structural Bioinformatics) 웹 사이트에서 제공하고 있다.

QuickPDB는 단백질내 아미노산 서열과 3차 구조를 보여준다. 서열패널에서 선택된 아미노산이 구조패널에서 구분되므로 단백질 3차 구조에서 아미노산의 위치를 파악하기에 유용하다. 단백질

질 3차 구조의 회전, 이동, 확대/축소가 가능하다. 단백질의 2차 구조(7가지)와 아미노산 특성(11가지)을 색으로 구분하여 표시함으로써 단백질 3차 구조에서 단백질의 2차 구조와 아미노산의 특성을 효과적으로 파악할 수 있다. QuickPDB는 CGI 방식을 통해 PDB에서 제공하는 단백질 3차 구조에 대한 간략한 정보를 읽어 가장 신속하게 단백질의 구조를 보여준다.

2.2 WebMol[4]

WebMol은 단백질의 3차원 구조정보를 분석하여 보여주기 위해 UCSF(University of California, San Francisco)의 Cohen Group에서 개발된 자바 기반의 단백질 3차구조 뷰어이다.

WebMol은 PDB 파일 전체를 읽어 단백질의 3차 구조를 아미노산 원자단위까지 세부적으로 표현하고, 간단한 마우스 조작으로 3차 구조의 회전, 이동, 확대/축소가 가능하다. 또한 단백질의 3차 구조를 철망구조(wire-frame), 체인구조(chain), 골격구조(back-bone) 등 다양한 모양으로 보여주고, 아미노산 원자들간의 거리와 각도 등도 계산한다.

2.3 JMV(Java Molecular Viewer)[5]

JMV는 플랫폼에 상관없이 사용하기 쉽도록 자바 3D[7]를 이용하여 개발된 분자 구조 뷰어이다. JMV는 단백질을 비롯하여 여러 가지 분자를 표현할 수 있다. 또한 WEB, RCSB, BioCore 파일 시스템, 지역파일 시스템 등 다양한 위치에서, PDB, VRML 등과 같은 다양한 형식의 3차원 구조정보 파일을 읽을 수 있다.

3. PDB 한글 뷰어

PDB 한글 뷰어는 단백질 3차 구조의 효과적인 관찰을 위해 자바 3D 기술을 이용하여 구현되었다. 원격 혹은 지역파일시스템에서 PDB 파일을 읽어 각 객체에 저장함으로써 효과적으로 데이터를 관리할 수 있다. PDB 한글 뷰어는 단백질 3차 구조를 파악하기 위해 다양한 형태의 3차 구조와, 스테레오뷰, 원자간 거리, 각도를 보여주기 위한 다양한 도구, 그리고 PDB 파일의 이해를 돕기 위한 편리한 도움말 기능을 제공한다.

3.1 구조

PDB 한글 뷰어는 입력, 저장, 서열, 구조, 도구 등 기능별로 구분하여 모듈을 구성하였다.

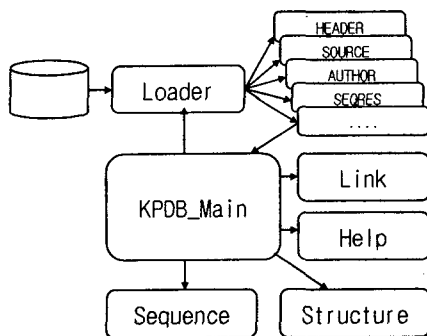


그림 1 PDB 한글 뷰어의 구조

PDB 한글 뷰어는 PDB 데이터를 읽어오는 입력모듈, 입력받은 데이터를 각 레코드별로 저장하고 관리하는 레코드모듈, 서열 관련 데이터를 처리하는 서열모듈, 구조 관련 데이터를 처리하는 구조 모듈, 도구의 기능을 가진 도구모듈, 그리고 시스템을 관리하는 메인모듈 등으로 구성된다. 그림 1은 PDB 한글 뷰어의 구조를 보여준다.

PDB 파일로부터 입력된 모든 데이터는 PDB 데이터에서 지정된 레코드별로 구분되어 레코드마다 객체를 생성한 후, 해당 객체에 데이터가 저장된다.[6] PDB 한글 뷰어는 각 레코드마다 객체를 생성하여 관리함으로써 효과적으로 데이터를 저장하고 처리할 수 있다. 그림 2는 HEADER 레코드의 데이터가 관리되는 자료구조를 보여준다.

```

public class HEADER
{
    String classification;
    String depDate;
    String idCode;

    public HEADER() {
        classification = null;
        depDate = null;
        idCode = null;
    }
    protected void read(String s) {
        classification = s.substring(10, 50).trim();
        depDate = s.substring(50, 59).trim();
        idCode = s.substring(62, 66).trim();
    }
    .....
}
    
```

그림 2 HEADER 레코드의 구조

3.2 기능

·불러오기(Load)

PDB 한글 뷰어는 PDB 파일을 3가지 방식으로 입력받는다. 먼저 PDB ID 만을 입력했을 경우 RCSB(PDB) 웹사이트에 접속하여 CGI 인터페이스를 통해 제공되는 단백질의 3차원 구조정보를 불러온다. PDB 파일의 위치정보를 알고 있을 경우 입력 필드에 URL 정보를 입력하여 웹으로 불러오거나, 로컬 파일시스템에서 PDB 파일을 검색하여 단백질의 3차원 구조정보를 불러올 수 있다.

·보기(Graphical Representation)

PDB 한글 뷰어는 단백질의 3차 구조를 효과적으로 보여주기 위한 서열보기, 구조보기 및 색깔지정 기능을 제공한다.

PDB 한글 뷰어는 단백질이 어떤 아미노산으로 구성되어 있는지 쉽게 파악할 수 있도록 화면 상단의 서열패널에 아미노산의 전체 서열정보를 제공한다.

구조보기는 각 아미노산의 Ca(alpha carbon)를 연결한 골격구조(back-bone), 단백질의 모든 원자를 선으로 연결한 철망구조(wire-frame), 철망구조에서 원자들을 작은 구로 표현하여 원자와 원자를 연결한 구조(ball and stick), 골격구조를 확장하여 Ca

를 지나는 매끄러운 곡선을 표현한 튜브구조(tube), 그리고 리본구조(ribbon) 등을 제공한다.

한글 PDB 뷰어는 효과적인 구조파악을 위해 원자, 아미노산, 체인, 2차구조 등을 서로 다른 색깔로 구분하여 표현할 수 있다. 본 뷰어는 아미노산을 11가지 특성(hydrophobic, polar, small, proline, tiny, aliphatic, aromatic, positive, negative, charged, glycine)으로 분류하고 색깔을 구분하여 표현함으로써 단백질 3차구조 내에서의 아미노산의 위치 및 특성 파악이 용이하다.

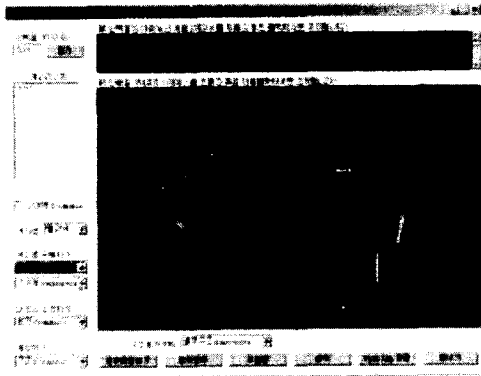


그림 3 단백질 구조 보기

·도구(Tool)

PDB 한글 뷰어는 단백질의 구조를 관찰 및 분석할 수 있는 도구를 제공한다. 마우스 조작을 통해 단백질 3차 구조가 회전, 이동, 확대/축소될 수 있으며, 간단한 수학적 공식을 이용하여 단백질 3차 구조 내에서의 원자들 사이의 거리와 각도정보를 계산하여 제공한다 또한, 기존 뷰어에서 제공하는 스테레오뷰(stereo view) 기능을 보완하여 그림 4와 같이 한 화면에 한 단백질의 2가지 구조를 한꺼번에 보여줌으로써 단백질의 구조를 보다 효과적으로 파악할 수 있다.

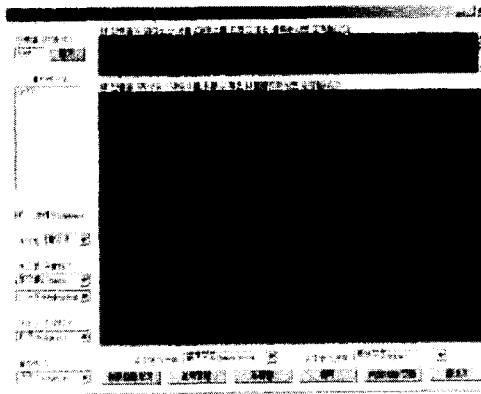


그림 4 스테레오뷰(stereo-view)

·기타

PDB 한글 뷰어는 해당 단백질 3차 구조에 대한 간략한 요약정보를 제공하고, 사용자의 PDB 파일에 대한 보다 폭넓은 이해를 돕기 위해 PDB 파일의 전체 내용과 함께, PDB 파일내의 각 레코드의 의미와 필드의 정보를 설명한 도움말을 제공한다.

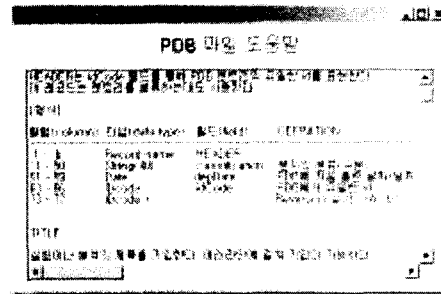


그림 5 PDB 파일 도움말

4. 결론 및 향후 연구과제

본 논문에서는 단백질의 3차원 구조정보를 담고 있는 PDB 파일을 분석하여 단백질의 3차 구조를 보다 효과적으로 파악할 수 있는 자바 기반의 PDB 한글 뷰어에 대해 기술하였다. PDB 한글 뷰어는 원격 또는 지역파일시스템에서 단백질의 3차원 구조정보를 읽을 수 있다. PDB 데이터의 효과적인 저장과 사용을 위해 각 레코드별로 객체를 생성하여 관리하며, 동일한 단백질의 2가지 구조를 동시에 보여주는 스테레오뷰 기능을 제공함으로써 단백질의 3차 구조를 보다 쉽게 파악할 수 있다. 또한 PDB 원본 데이터와 3차 구조에 대한 요약정보를 제공하고 기존의 PDB 뷰어에서 부족한 도움말 기능을 보완하여 PDB 데이터의 레코드별 의미정보를 제공하였다.

현재, 사용의 편의를 위해 보다 친숙한 사용자 인터페이스에 대해 연구중이며, 보다 다양한 형식의 3차원 단백질 구조정보를 읽을 수 있고, 단백질의 3차 구조를 그림파일로 저장할 수 있는 기능을 제공할 계획이다.

5. 참고문헌

[1] Carl Branden, John Tooze, "Introduction to Protein Structure 2nd ed.", Garland Publishing Inc., 1999.  
 [2] Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T.N.Bhat, Helge Weissig, Ilya N. Shindyalov and Philip E. Bourne, "Protein Data Bank", Nucleic Acids Research, Vol.28, No.1, pp.235-242, 2000.  
 [3] QuickPDB at "http://cl.sdsc.edu/QuickPDB.html"  
 [4] WebMol at "http://www.cmpchem.ucsf.edu/cgi-bin/webmol.pl"  
 [5] JMV at "http://www.ks.uiuc.edu/Development/jmv/"  
 [6] Protein Data Bank Contents Guide at "http://www.rcsb.org/pdb/docs/format/pdbguide2.2/guide2.2\_frame.html"  
 [7] 박동규, "자바 3D 프로그래밍", 도서출판 대림, 2001.5.