

단백질 3차원 구조 비교 관찰 도구의 개발

오세종^{1*}

¹단국대학교 대학원 나노바이오횜과학과

A Development of Comparison and Observation Tools for Protein 3D Structure

Oh Se Jong^{1*}

¹Department of Nanobiomedical Science, The Graduate School of Dankook University

요 약 단백질은 생물의 주요 구성성분중의 하나로서 세포안에서 각종 화학반응의 촉매 역할을 담당하거나. 항체를 형성하여 면역을 담당하는 등 중요 역할을 수행한다. 단백질의 기능은 단백질 분자들의 사슬이 3차원 구조를 형성하면서 결정이 되기 때문에 단백질의 3차원 구조에 대한 연구는 매우 중요하다. 본 연구에서는 두 개의 단백질 3차원 구조를 하나의 화면에서 비교, 관찰할 수 있는 그래픽 도구를 개발하였다. 개발된 도구는 단백질의 3차원 구조를 보여줄뿐만 아니라 각종 비교 데이터를 제시하여 연구자들의 비교, 관찰이 용이하도록 지원한다.

Abstract Proteins are one of essential part of organisms; many proteins are enzymes that catalyze biochemical reactions and are vital to metabolism. Researching for 3D structure of proteins is important because functions of proteins are determined by 3D structure of them. In this study, we developed graphic tool that supports comparison and observation of the two proteins' 3D structure in the single screen. It also supports some comparison data to help researcher's easy comparison and observation works.

Key Words : protein, bioinformatics, BioJava

1. 서론

1990년 미국의 국립 보건원(NIH)과 에너지부(the U.S. Department of Energy)는 공동으로 인간유전체사업(HGP; Human Genome Project)을 시작하였으며, 2001년 2월 HGP와 셀레라제노믹스는 공동으로 인간의 전체 유전자에 대한 지도를 완성하였다고 공식 발표하였다. 이후 인간 게놈(genome)에 대한 연구는 유전자 서열을 알아내는 단계에서 알려진 서열로부터 유용한 정보를 얻어 내리는 포스트 게놈(post genome) 시대에 들어서게 되었다.

DNA는 유전과 관련된 정보를 가지고 있는데 이는 RNA로 전사(transcription)되고 다시 번역(translation)과정을 거쳐 단백질을 생성하게 된다. 단백질은 주로 아미노산(Amino Acid)이라고 불리는 분자들의 결합으로 구성이 되며, 아미노산은 표 1과 같이 20종이 존재한다.

단백질은 그 역할에 따라 구조 단백질(structural protein), 효소(enzyme), 조절 단백질(regulatory protein), 수송단백질(transport protein)로 분류되며 생물체 내의 구성 성분, 세포 안의 각종 화학반응의 촉매 역할(효소), 항체를 형성하여 면역을 담당하는 등 여러 가지 형태로 중요한 역할을 수행한다[1].

단백질이 만들어지는 과정은 다음과 같다. RNA의 정보에 따라 아미노산이 만들어지고, 만들어진 아미노산들은 정해진 순서대로 폴리펩티드 사슬이라고 불리는 1개의 줄로 연결된다. 각 종류의 단백질은 독특한 아미노산 서열을 가지며 이런 순서를 단백질의 1차 구조라 한다. 단백질의 2차 구조는 1차구조의 사슬이 나선 또는 병풍 구조로 접히거나 꼬이면서 만들어진다. 3차구조는 2차구조의 단백질 사슬이 다시 구부러지고 접혀짐으로 형성되며, 정도의 차이는 있지만 대부분 구형(球形)을 이루게

이 연구는 2008학년도 단국대학교 대학연구비의 지원으로 연구되었음

*교신저자 : 오세종(sejongoh@dankook.ac.kr)

접수일 09년 04월 07일

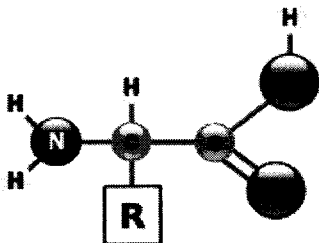
수정일 (1차 09년 05월 06일, 2차 09년 06월 10일)

게재확정일 09년 06월 17일

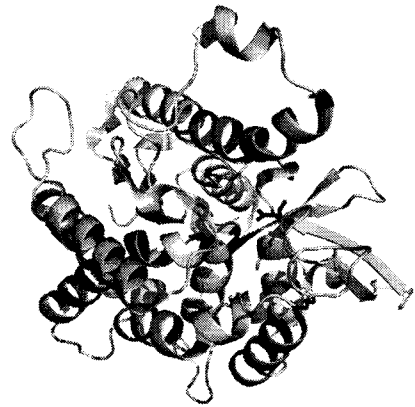
된다. 헤모글로빈과 같은 일부 단백질은 하나이상의 단백질 소단위가 모여 형성되는데, 이러한 공간적 구조를 4차 구조라 한다. 단백질의 3차 구조부터 단백질 고유의 특성 및 기능을 나타내게 된다.

단백질에 대한 연구가 중요한 이유는 단백질이 생명활동과 관련된 핵심적인 역할을 수행하기 때문이며, 각종 질병과도 관련이 있고 신약의 개발과도 깊은 관련성을 가지고 있다. 생물정보학(bioinformatics) 분야에서는 1차원의 아미노산 서열 정보로부터 3차원 구조를 예측하는 연구를 활발히 진행하고 있다. 그 이유는 아직도 기능이 알려지지 않은 단백질의 종류가 매우 많고, 단백질의 3차원 구조가 기능과 밀접한 연관을 가지고 있기 때문이다. 그러나 단백질의 1차원 구조가 3차원구조로 되는 과정에는 매우 많은 변수들이 관여하기 때문에 예측이 쉽지 않다.

본 연구는 아직 국내에 활성화 되지 않은 3D 그래픽 기반의 단백질 구조 비교 연구에 기여하는 것과 그래픽 기반과 텍스트 기반으로 양분되어 있는 분석도구들에 대해 통합 도구를 제시하는 것을 목적으로 시작하였다. 본 논문에서는 이미 알려진 단백질의 3차원 구조 데이터를 활용하여 두 개의 단백질간의 3차원 구조를 비교하고 관찰할 수 있는 도구의 개발내용을 제시한다. 단백질의 3차원 구조를 관찰할 수 있는 도구로는 Jmol, WebMol 등 다수의 알려진 3차원 그래픽 도구들[12]이 있으나 단일 단백질에 대한 구조만을 보여주기 때문에 단백질간 구조 비교를 위해서는 단백질별로 도구를 각각 실행시켜서 비교를 해야하는 등 불편이 있었다. 또한 텍스트에 기반한 비교분석 도구들[12,13]도 있으나 입력 및 출력 결과가 텍스트로 표시되기 때문에 3차원 구조와 연관지어 결과를 해석하는데 어려움이 있었다. 본 연구에서는 하나의 화면에 두 개의 단백질 구조를 표시하고, 두 단백질의 구조 비교를 위한 추가 데이터를 제공함으로써 연구자가 편리하게 단백질간 구조를 비교 관찰할 수 있도록 하였다.



[그림 1] 아미노산의 일반적인 구조



[그림 2] 단백질의 3차원 구조

[표 1] 아미노산의 종류와 이름

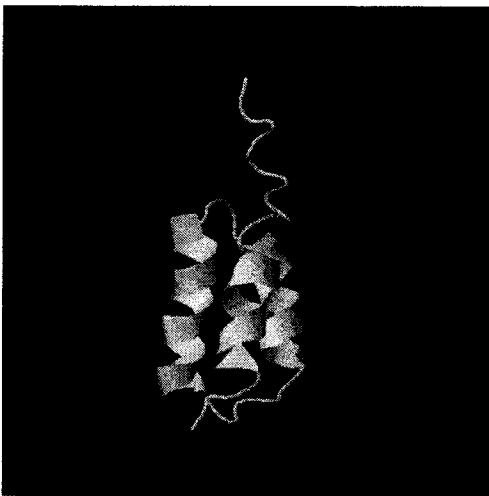
| 아미노산 이름 | 약자 | 기호 |
|---------|-----|----|
| 알라닌 | Ala | A |
| 시스테인 | Cys | C |
| 아스파르트산 | Asp | D |
| 글루탐산 | Glu | E |
| 페닐알라닌 | Phe | F |
| 글리신 | Gly | G |
| 히스티딘 | His | H |
| 이소루신 | Ile | I |
| 리신 | Lys | K |
| 류신 | Leu | L |
| 메티오닌 | Met | M |
| 아스파라긴 | Asn | N |
| 프롤린 | Pro | P |
| 글루타민 | Gln | Q |
| 아르기닌 | Arg | R |
| 세린 | Ser | S |
| 트레오닌 | Thr | T |
| 셀레노시스테인 | Sec | U |
| 발린 | Val | V |
| 트립토판 | Trp | W |

2. 개발의 필요성 및 관련 연구

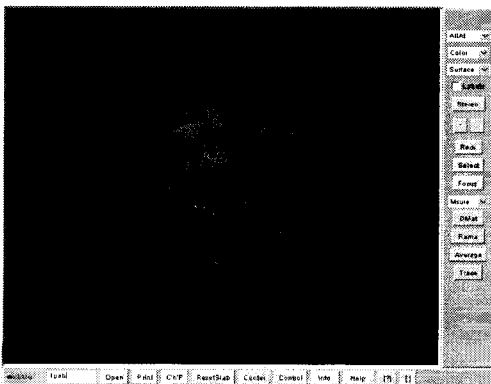
서론에서 설명한 바와 같이 단백질의 3차원 구조는 단백질의 기능과 밀접한 관계가 있기 때문에, 단백질의 구조를 비교 분석함으로써 구조의 차이와 단백질의 기능의 차이와의 관계를 알아낼수도 있고, 이를 바탕으로 구조가 알려지지 않은 새로운 단백질에 대해 1차 서열 데이터로부터 구조 및 기능을 예측하는데에도 사용할 수 있다. 단백질에 관한 데이터 및 3차원 구조에 대한 정보를 얻기

위해서는 일반적으로 'RCSB Protein Data Bank' 웹사이트[2]를 이용한다. 2009년 4월 현재 이곳에서는 약 5만 7천건의 단백질 구조 정보가 등록되어 있고, 각종 검색 도구 및 3차원 구조를 관찰할 수 있는 도구를 지원하고 있다.

검색한 단백질 구조 데이터에 대해 3차원 구조를 보기 위해서는 Jmol, WebMol 등의 도구가 사용된다. Jmol[3]은 어떤 물질의 3차원 화학구조를 볼 수 있도록 해주는 도구로서 Java로 작성되어 웹페이지 내에서, 혹은 독립된 프로그램으로 다운로드 받아 사용할 수 있다. WebMol [4]은 Drick Walther 에 의해 개발된 프로그램으로 단백질의 3차원 구조를 전문적으로 보기 위한 도구로서 Jmol에서 제공하지 않는 상세한 부분까지도 보여 준다. 그림 3과 그림 4는 동일한 단백질 데이터를 가지고 Jmol 과 WebMol 도구로 관찰한 결과이다. Jmol 과 WebMol 외에도 KiNG Viewer, Rasmol Viewer, Swiss-PDB Viewer 등 다양한 형태의 도구를 이용할 수 있다[12].

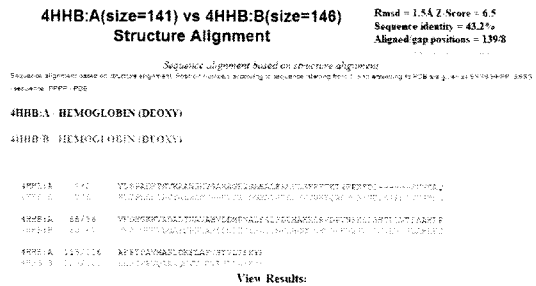


[그림 3] Jmol 화면의 예



[그림 4] WebMol 화면의 예

앞에서 설명한 도구들은 3D 그래픽에 기초하여 단일 단백질 구조를 관찰하기 위한 목적으로 개발되었다. 이와는 달리 두 개의 단백질 구조에 대해 텍스트 기반의 정보를 제공하는 도구들이 있다. C-alpha Match, FlexProt, MultiPro 등[13]이 그것인데 주로 두 개의 단백질 구조에 대한 비교 데이터와 아미노산 서열에 대한 정렬 (alignment) 작업을 지원한다. 그림 5는 [14]에서 제공하는 도구의 정렬 결과 화면이다.



[그림 5] 두 개의 단백질 서열에 대한 정렬 결과

3D 기반 도구들은 하나의 단백질 구조만을 볼 수 있도록 하기 때문에 두 개 이상의 단백질 구조를 비교하고자 하는 경우에는 불편함이 따른다. 연구자는 비교하고자 하는 단백질에 대해 각각 프로그램을 실행하여 비교 관찰을 해야 한다. 또한 제시된 도구들은 하나의 단백질에 대해서만 상세 정보를 제공하기 때문에 두 개 이상의 단백질을 비교하는 경우, 비교 데이터를 찾거나 만드는 일은 연구자의 몫이 된다. 또한 텍스트 기반 도구들은 텍스트 데이터만을 제공하기 때문에 3차원 구조를 알기 어렵다.

본 연구에서는 새로운 도구의 개발을 통해 두 개의 단백질에 대해 3차원 구조를 하나의 화면에 보여주는 것은 물론, 두 단백질에 대한 비교 데이터를 보여줌으로써 관찰자가 두 단백질의 구조상 차이를 쉽게 알 수 있도록 하였다. 특별히 3차원 구조간의 유사도를 유추할 수 있는 RMSD 지수를 계산하여 보여줌으로써 구조간 유사도를 수치화된 데이터로 알 수 있도록 하였다. 구체적인 개발 내용에 대해서는 3절에서 서술한다.

3. 비교 관찰 도구의 개발

3.1 개발 환경

3D 그래픽 도구들은 대체로 Java 또는 윈도우 기반의 C 언어로 개발이 되나, Java 환경이 우세한 편이다. 그 이유는 Java 로 개발된 도구들은 쉽게 애플릿 형태로 변환

되어 Web 브라우저로 실행할 수 있기 때문이다. 개발도구를 다수가 공유하도록 하기 위해서는 Web에서 실행되는 형태가 유리하다. 또한 Java에는 BioJava라고 하는 API 라이브러리가 제공되어 생명정보학 관련 애플리케이션을 쉽게 작성할 수 있도록 지원한다.

본 연구에서는 Java를 이용하여 단백질 3차원 구조 비교 관찰 도구를 개발하였다. 기본적인 Java 프로그래밍 환경은 JDK 1.6을 이용하였으며, 3차원 그래픽을 처리하기 위해서 Java 3D 라이브러리[11]를 이용하였다. 또한 단백질 구조 데이터로부터 자료를 읽어오고 처리하기 위해서 BioJava 라이브러리를 이용하였다.

- JDK 1.6
- Java 3D
- BioJava 1.6.1

본 논문에서 PBD 파일의 3차원 좌표를 읽어들이 3D 그래픽으로 표현하는 부분은 참고문헌 [6]의 PDBViewer 소스 코드를 참고 하였다. 참고한 소스 코드는 Java 및 Java 3D를 이용하여 단일 단백질 구조를 3차원으로 관찰 및 분석을 할 수 있도록 개발되었으며 10개의 클래스로 구성되어 있는데, 주요 클래스는 표 2와 같다.

[표 2] PDBViewer 의 구성 클래스

| class명 | 설명 |
|--------------|--|
| AtomObject | PBD 파일에서 단백질을 구성하는 각 분자들에 대한 정보인 ATOM, HETATM 항목 내용을 저장 |
| HelixObject | PBD파일에서 HELIXdp 대한 정보를 저장 |
| ThreeD Model | AtomObject 및 HelixObject 객체를 필드로 가지고 있는 Canvas3D를 통해 3차원 그래픽을 표현 |
| PDBViewer | 프로그램의 주화면 구성, 각종 이벤트 처리 |
| Protocol | 단백질 구조 표현방법, Cain의 색상, 확대/축소에 대한 속성 지정 |
| PickMouse | 마우스 클릭을 통해 3차원 공간에 표시된 객체를 선택 |
| TransModel | 3차원 공간에 표시된 객체를 이동하는 메소드 구현 |

Java3D는 3D 애플리케이션 개발을 위한 Java API 클래스들의 집합으로서 3차원 객체의 표현 및 세련된 렌더링이 핵심기능이다. 본 논문에서 참조한 PDBViewer도 Java3D API를 이용하여 3차원 그래픽을 구현하였다. 표

3은 개발에 이용한 주요 Java3D 클래스에 대한 설명이다.

[표 3] 개발에 이용된 주요 Java3D 클래스

| class명 | 설명 |
|----------------|--|
| Canvas3D | 렌더링을 위한 3D 캔버스를 제공 |
| SimpleUniverse | Java3D는 3차원 객체를 표현하기 위해 가상우주(virtual universe)라는 개념을 도입하였는데, 이를 사용하기 쉽게 단순화한 것이 SimpleUniverse 이다. |
| TransformGroup | 3차원 객체의 이동이나 회전과 같은 기하학적 변형을 처리 |

BioJava 는 바이오 데이터 및 프로그램을 편리하게 다루기 위한 Java API 클래스의 모음으로 약 80여개의 패키지과 1000개 이상의 클래스로 구성되어 있다. 본 연구에서는 PDB 파일을 읽고 파싱(parsing)하여 필요한 정보를 얻는데 BioJava를 이용하였으며 참조한 클래스는 표 4와 같다.

[표 4] 개발에 이용된 BioJava 클래스

| class명 | 설명 |
|---------------|------------------------------------|
| PDBFileReader | PDB 파일을 읽고 파싱한다 |
| Structure | 파싱한 객체로부터 파일 구조 정보를 얻는다 |
| Chain | Structure 객체부터 Chain 정보를 추출하고 다룬다 |
| Group | Structure 객체로부터 Group 정보를 추출하고 다룬다 |
| Atom | Structure 객체로부터 Atom 정보를 추출하고 다룬다 |

3.2 단백질 구조 데이터

RCSB Protein Data Bank 는 연구자들이 단백질 3차원 구조 데이터를 활용할 수 있도록 하기 위해서 표준화된 형태의 데이터 파일(PDB 파일)을 다운받아 사용할 수 있도록 지원하고 있다. 이 파일의 포맷에 대한 해설은 참고문헌 [5]에서 찾아볼 수 있다.

PDB 파일은 다양한 형태의 레코드들로 구성되어 있다. 표 5는 주요레코드에 대한 설명이다. 그림 6은 PDB 파일의 내용 일부를 보여준다. 본 연구에서는 연구자가 RCSB Protein Data Bank 로부터 비교를 원하는 PDB 파

일을 내려받아 저장한 것으로 가정하고, 프로그램에서 그러한 파일을 읽어들이어 3차원 구조 정보 및 비교 정보를 보여줄 수 있도록 하였다.

[표 5] PDB 파일의 주요 레코드 타입

| Record Type | 설명 |
|-------------|---|
| ATOM | 단백질 분자를 구성하는 각 원자들의 이름과 3차원 위치를 나타내는 좌표값을 기술 |
| HETATM | 단백질을 구성하는 분자들중 비표준 분자(아미노산, 핵산을 제외한 분자)들에 대한 좌표를 기술 |
| TER | 잔기(residues)들의 체인(chain)이 끝났음을 나타냄. |
| HELIX | 3차원 그래픽에서 나선에 해당하는 부분에 대한 개수 및 위치 정보를 기술 |
| SEQRES | 3차원 구조를 구성하는 체인(chain)의 개수와 체인에 포함된 아미노산의 종류를 기술함 |
| HEADER | PDB 등록번호, 단백질이름, 등록일자 등을 기술 |

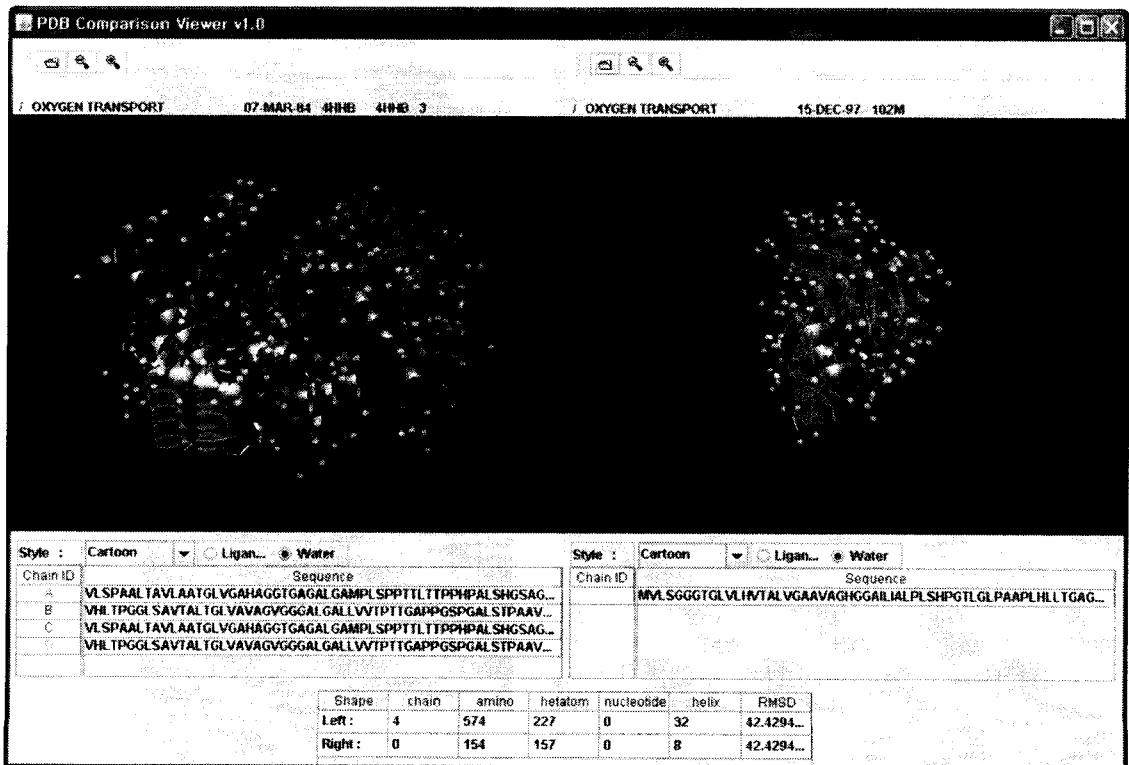
```

REMARK 999 ILE 32, DURING SEQUENTIAL ASSIGNMENT RESIDUE 244 IN THE
REMARK 999 PROTEIN PAD GENE (PDB RESIDUE 32, SEE DREF/SEQADV BELOW)
REMARK 999 WAS SEEN TO BE AN ILE INSTEAD OF A VAL. DETAILS ARE GIVEN
REMARK 999 IN THE PAPER CITED IN REMARK 1.
DREF 16AB ILE 1 53 GB 1333728 27864 213 265
SEQADV 16AB ILE 32 GB 1333728 VAL 244 CONFLICT
SEQRES 1 53 THR ILE ASP GLN TRP LEU LEU LYS ASN ALA LYS GLU ASP
SEQRES 2 53 ALA ILE ALA GLU LEU LYS LYS ALA GLY ILE THR SER ASP
SEQRES 3 53 PHE TRP PHE ASN ALA ILE ASN LYS ALA LYS THR VAL GLU
SEQRES 4 53 GLU VAL ASN ALA LEU LYS ASN GLU ILE LEU LYS ALA HIS
SEQRES 5 53 ALA
HELIX 1 HI ALA 10 LEU 18 1 9
HELIX 2 HI ASP 26 LYS 34 1 9
HELIX 3 HI VAL 38 ALA 51 1 14
CRYST1 1.000 1.000 1.000 90.00 90.00 90.00 P 1 1
ORIGX1 1.000000 0.000000 0.000000 0.00000 0.00000
ORIGX2 0.000000 1.000000 0.000000 0.00000 0.00000
ORIGX3 0.000000 0.000000 1.000000 0.00000 0.00000
SCALE1 1.000000 0.000000 0.000000 0.00000 0.00000
SCALE2 0.000000 1.000000 0.000000 0.00000 0.00000
SCALE3 0.000000 0.000000 1.000000 0.00000 0.00000
MODEL 1
ATOM 1 N THR 1 -1.820 24.919 -5.344 1.00 0.00 N
ATOM 2 CA THR 1 -1.256 24.379 -4.074 1.00 0.00 C
ATOM 3 C THR 1 -1.754 22.950 -3.864 1.00 0.00 C
ATOM 4 O THR 1 -2.804 25.725 -3.295 1.00 0.00 O
ATOM 5 CB THR 1 -1.707 25.254 -2.901 1.00 0.00 C
ATOM 6 OG1 THR 1 -1.518 24.547 -1.683 1.00 0.00 O
    
```

[그림 6] PDB 파일의 일부

3.3 비교분석 도구의 구현

본 연구에서 구현한 도구의 화면은 그림 7과 같다. 개발된 도구는 2개의 PDB 파일을 읽어들이어 화면 좌우에 3차원 그래픽 데이터를 보여주며, 마우스로 그래픽 데이터를 원하는 방향으로 회전시켜가며 관찰할 수 있다. 3차원 그래픽 부분은 Java 3D 라이브러리가 많이 사용되었다.



[그림 7] PDB구조 비교관찰 도구 화면

화면 하단에는 두 개의 단백질 구조에 대한 비교 데이터를 표시하였다(chain, amino acid, hetatom, nucleotide, RMSD, 6개 항목). RMSD를 제외한 나머지 5개 정보는 BioJava 라이브러리를 이용하여 PDB 파일로부터 추출하였다. 각 비교 항목의 의미는 다음과 같다.

- chain : 단백질의 단위구조를 형성하는 체인의 개수. 단백질의 단위구조는 하나 또는 그 이상의 체인으로 구성된다.
- amino : 체인을 형성하는 분자중 아미노산의 개수. 체인을 형성하는 분자는 아미노산(amino), 핵산(nucleotide), 기타(hetatom)로 구성되는데 일반적으로 아미노산이 가장 많다.
- hetatom : 체인을 형성하는 분자중 아미노산, 핵산을 제외한 분자의 수
- nucleotide : 체인을 형성하는 분자중 핵산의 갯수
- RMSD : 두 개의 3차원 구조의 유사도를 나타내는 지수

3.4 RMSD

본 연구에서는 RMSD 지수[7]를 이용하여 두 개의 단백질의 3차원 구조간의 유사성을 계산하였다.

$$RMSD = \sqrt{\frac{\sum_i (d_i)^2}{n}}$$

n = 원자들의 갯수

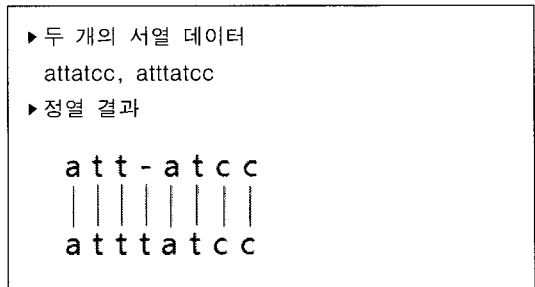
d_i = 두 개의 구조에서 i번째 대응하는 두 개의 원자간 거리

RMSD는 Root Mean Square Deviation 의 약자로서 기본 원리는 두 개의 단백질 3차원 구조에서 서로 대응하는 원자간의 거리를 구해 이들의 평균값을 구한 것으로 RMSD 지수가 0 에 가까울수록 구조적 유사도가 높다는 것을 의미한다.

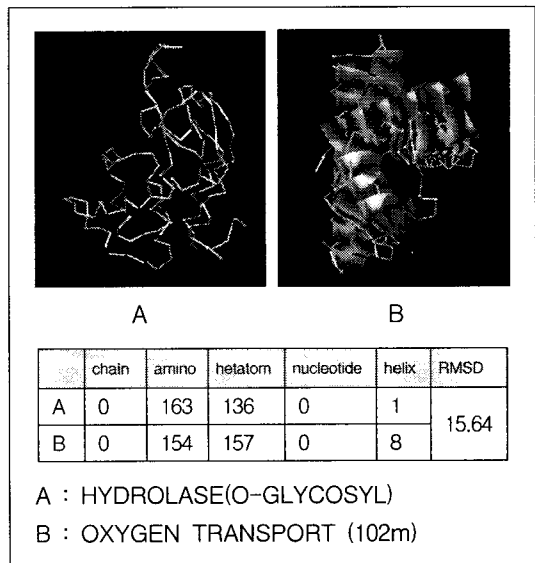
본 연구에서는 대응하는 원자들간의 거리를 계산하기에 앞서서 니들만-분쉬 알고리즘(Needelman-Wunsch algorithm)[8-10]을 사용하여 두 단백질의 아미노산 분자 서열을 정렬(alignment)함으로써 원자간 거리 계산시 보다 연관성이 있는 원자들이 대응될 수 있도록 하였다. 그림 8은 서열 정렬의 예이다.

아미노산 분자 서열을 정렬한 다음의 절차는 상호 대

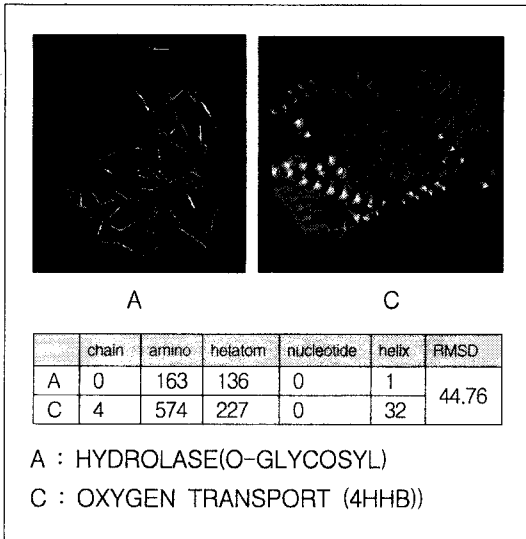
응하는 각 분자를 구성하는 원자들 사이의 거리를 구하는 것인데, 하나의 분자는 10~20 개의 원자로 구성이 되므로 각각의 원자들간 거리를 구하는 것은 번거로운 일이다. 본 연구에서는 계산의 단순성을 위해 분자를 구성하는 원자들중 3차원 구조의 뼈대를 형성하는 C_α 원자(그림 1에서 H와 R에 연결된 C 원자)들간의 거리만을 계산하였다. 그림 9와 그림 10은 세 개의 단백질에 대한 구조 비교 데이터의 산출 결과를 보여준다. 그림 9와 그림 10을 보면 직관적으로 A,B 가 A,C 에 비해 구조적으로 더 유사함을 알 수 있고 수치 데이터 및 RMSD 지수상으로도 확인됨을 알 수 있다.



[그림 8] 서열 정렬의 예



[그림 9] 구조 비교 결과의 예 1



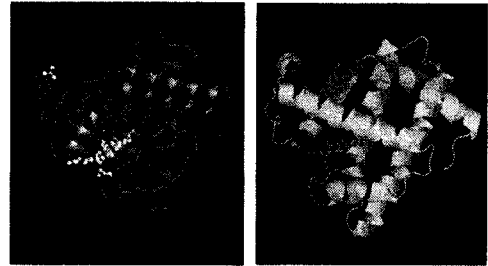
[그림 10] 구조 비교 결과의 예 2

3.5 기존 도구들과의 비교

표 6은 기존의 도구들과 본 연구를 통해 제안된 도구를 비교한 내용이다. 기존의 도구들은 그 수가 많고 각 도구별로 조금씩 특징이 있기 때문에 기존 도구들을 3D 그래픽 기반 도구와 텍스트 기반 도구로 분류하고 분류별 대표 특성을 가지고 비교하였다. 표에서 알 수 있는 바와 같이 본 논문에서 제안한 도구는 3D 기반 도구와 텍스트 기반 도구의 특성을 종합하였으며, 기존의 3D 기반 도구가 단일 단백질에 대한 구조만을 보여주는 단점을 극복 하였다.

그림 11은 동일한 PDB 파일에 대한 논문의 개발 도구와 Jmol 의 3D 이미지 비교이다. 전체 구조와 윤곽에 있

어서는 유사하나, 제안된 도구는 Jmol 에 비해 체인이 부드럽지 못하고, 체인과 나선의 접점 부분이 매끄럽지 못하다. 또한 나선의 회전 방향이 표시되지 못하고 있어 향후 개선 과제로 남아 있다.



(a) 제안된 도구 (b) Jmol

[그림 11] 제안된 도구의 3D 이미지 비교

본 논문에서는 두 개의 단백질 구조에 대한 비교 데이터를 표시하고 있으며 그중 중요한 것이 RMSD 이다. RMSD 는 계산 방식에 따라 절대값에 있어서 차이가 나기 때문에 절대값 자체는 큰 의미가 없고 절대값 간의 차이가 의미가 있다. 표 7은 그림 9 및 그림 10에 있는 단백질 구조에 대한 RMSD 값과 동일 구조에 대해 기존의 도구에서 계산한 값을 비교하여 보여준다. 기존 구조 파일은 PDB ID : 102L을 사용하였고 이에 대해 5개 구조 파일과 비교하였다. 비교 대상도구는 CE [14]를 사용하였다. 표에서 보는 바와 같이 동일 비교 대상에 대하여 절대값은 다르지만 동일한 증가 패턴을 보이고 있음을 알 수 있다. 이는 본 논문에서 RMSD 공식을 적절히 구현하였음을 의미한다.

[표 6] 기존 도구들과의 비교

| 항목 | 3D 그래픽 기반 도구 | 텍스트 기반 도구 | 본 논문에서 제안한 도구 |
|-------------------------|---|---|---|
| 주요 도구 | Jmol, WebMol, RASMOL, NOC, mmPDBViewer, VMD, WPDB | C-alpha Match, FlexProt, MultiProt, MASS, STACCATO | PDB Comparison Viewer 1.0 |
| 주 개발 목적 | 단일 단백질 구조에 대한 3차원 구조의 분석과 관찰 | 두 개의 단백질 구조에 대한 아미노산 서열 비교, 정렬(alignment), 유사도 데이터 제시 | 두 개의 단백질 구조에 대한 3차원 구조 동시 비교 관찰, 아미노산 서열 비교, 유사도 데이터 제시 |
| 두 개의 단백질 구조에 대한 비교가능 항목 | 단백질 1차 구조 및 2차 구조, 3차원 구조 ¹⁾ | 아미노산 서열의 정렬 결과, 서열 일치도(sequence identity), RMSD 지수 | 단백질 1차 구조 및 2차 구조, 3차원 구조, chain, amino 등 주요 구성 요소 데이터, RMSD 지수 |

1) 3D기반 도구는 단일 단백질 구조에 대한 정보만 제공하기 때문에 도구를 두 번 실행하여 비교한 것으로 가정

[표 7] RMSD 값의 비교

| | 1 | 2 | 3 | 4 | 5 |
|--------|------|------|-----|-----|-----|
| 제안된 도구 | 0.19 | 0.28 | 43 | 44 | 51 |
| CE | 0.2 | 0.3 | 3.9 | 4.3 | 9.1 |

(1-239L, 2-200L, 3-1PRB, 4-1GAB, 5-12CA)

4. 결론

단백질의 3차원 구조를 분석하고 관찰하는 일은 생물정보학 분야에서 매우 중요한 작업중의 하나이다. 이러한 작업을 위해 많은 도구들이 개발 되었으나 국내에서는 아직 이렇다 할 결과를 내놓지 못하고 있는 실정이다. 또한 기존에 개발된 도구들을 그래픽 기반과 텍스트 기반으로 양분되어 비교 결과를 종합하여 검토하는데 어려움이 있었다. 본 연구에서는 두 개의 단백질 3차원 구조를 비교하고 관찰할 수 있는 그래픽 도구를 개발하였다. 기존 3D 도구들의 문제점을 보완하여 하나의 화면에서 두 개의 3D 이미지를 비교할 수 있도록 하였으며, 텍스트 도구들에서 제공하는 비교 데이터를 함께 제공하여 비교 분석 작업의 효율성을 높였다. 아직은 기초적인 수준의 비교 데이터만을 제공하지만 향후 연구개발을 통해 보다 유용한 정보를 제공할 예정이다.

참고문헌

[1] 이명석, “알기쉽고 재미있는 분자생물학”, 라이프사이언스, 2008.
 [2] RCSB protein Data Bank, <http://www.rcsb.org/>
 [3] Jmol Viewer, <http://www.jmol.org>
 [4] webMol Viewer, <http://www.cmpfarm.ucsf.edu/cgi-bin/webmol.pl>
 [5] Protein Data Bank Contents Guide, http://mmcif.pdb.org/dictionaries/mmcif_pdbx.dic/Index/index.html.
 [6] 박현석, “자바로 배우는 바이오인포매틱스”, 사이텍미디어, 2006
 [7] Anderson, M.P., Woessner, W.W., “Applied Groundwater Modeling: Simulation of Flow and Advective Transport”, (2nd Edition ed.). Academic Press, 1992
 [8] Needleman SB, Wunsch CD., "A general method

applicable to the search for similarities in the amino acid sequence of two proteins". Journal of Molecular Biology, vol.48 No.3, pp.443-53, 1970.

[9] Hans-Joachim B., Dirk B., "Algorithmic Aspects of Bioinformatics", Springer, 2007.
 [10] 김진, 최홍식, 류웅진, “분자생물학과 알고리즘”, 정보과학회지 제18권 제8호, pp29-34, 2000.
 [11] 박동규, “자바 3D 프로그래밍”, 도서출판 대림, 2001
 [12] Bio-Soft Net, <http://en.bio-soft.net/3d.html>
 [13] Bioinformatics Links Directory, http://bioinformatics.ca/links_directory/?subcategory_id=136
 [14] Databases and Tools for 3-D Protein Structure Comparison and Alignment, <http://cl.sdsc.edu/ce.html>

오 세 중 (Se-Jong Oh)

[정회원]



- 1989년 2월 : 서강대학교 컴퓨터학과 (공학사)
- 1991년 2월 : 서강대학교 컴퓨터학과 (공학석사)
- 2001년 8월 : 서강대학교 컴퓨터학과 (공학박사)
- 2003년 9월 ~ 현재 : 단국대학교 대학원 나노바이오의과학과 조교수

<관심분야>

바이오인포매틱스, 임베디드시스템