

# Java3D를 이용한 PSAML 시각화 도구

류기현<sup>0\*</sup> 이명준<sup>\*\*</sup> 이수현<sup>\*</sup>

<sup>\*</sup>창원대학교 컴퓨터·정보통신공학부 <sup>\*\*</sup>울산대학교 컴퓨터·정보통신공학부  
rkh\_lhr@hanmail.net<sup>0</sup>

## A Visualization of PSAML Data using Java3D

Ki-Hyun Ryu<sup>0</sup> Myung-Joon Lee<sup>\*\*</sup> Su-Hyun Lee<sup>\*</sup>

<sup>\*</sup>School of Computer & Information Technology, Changwon National University

<sup>\*\*</sup>School of Computer Engineering & Information Technology, University of Ulsan

### 요 약

단백질은 생명현상 유지에 필수기능을 담당하며 이러한 기능이 단백질의 3차 구조에 의해 결정되므로 단백질 3차 구조에 대한 연구가 활발히 진행되고 있다. 본 논문에서는 단백질의 3차구조를 파악할 수 있는 Java3D 기반의 단백질 구조 뷰어인 PSAML Viewer에 관해서 기술한다. PSAML은 단백질의 2차구조와 2차구조 사이에서 발견되는 상호적인 관계를 이용하여 단백질 구조를 표현하는 방법이다. PSAML에 정의되어 있는 단백질 2차구조  $\alpha$ -나선과  $\beta$ -판상구조의 정보(서열, 길이, 공간상의 좌표)를 분석하여, 단백질 구조를 시각화한다. 이는 단백질 구조 정보를 보다 쉽게 이해하는데 도움을 줄 수 있다.

### 1. 서 론

단백질은 생명현상 유지 및 생명체 내에서 발생하는 화학반응에 있어 필수적인 역할을 담당하며, 20개의 아미노산의 배열에 의한 독특한 3차 구조를 형성한다. 이러한 3차 구조는 생명현상의 중요한 기능을 결정한다. 따라서 단백질의 3차 구조에 대한 활발한 연구와 단백질의 기능을 효과적으로 분석하기 위해서는 이러한 3차 구조 정보를 보여주는 도구 개발이 필요하다.

최근 컴퓨터 기술의 발전과 3차원 그래픽분야의 발전으로 단백질 3차 구조를 보다 용이하게 모델링 할 수 있게 되었다. Java를 이용하여 단백질 3차 구조를 모델링하고 있는 대표적인 뷰어는 QuickPDB[1], WebMol[2], KDBViewer[3] 등이 있다. 이들 뷰어는 실험을 통해 밝혀진 3차 구조 데이터를 기반으로 웹 상에서 단백질의 3차 구조를 다양한 모양으로 보여준다. 그러나 기존의 뷰어들은 PDB(Protein Data Bank)[4]가 제공하는 단백질의 3차 구조정보의 일부 내용만을 이용하기 때문에 거리, 각도와 같은 자세한 정보를 파악하기가 쉽지 않다.

또한, PDB에서 제공하는 데이터 형식은 텍스트 기반이고 정형화된 문법 명세가 부족하여 단백질 구조 비교를 위한 데이터 변형과정에서 파싱(parsing) 에러의 가능성을 내포하고 있으며, 단백질 3차구조 모델링을 위하여 부가적인 데이터 변환 과정과 계산 시간이 요구된다. 따라서 단백질 3차구조 모델링에 필요한 구조적 정보를 미리 가공하여 표준화된 방법을 이용하여 보다 효과적으로 단백질 3차구조 모델링에 적용될 수 있는 단백질 구조 표현 방법이 필요하다.

본 논문에서는 단백질 2차 구조 요소를 기반으로 표현[5,6]된 단백질간 구조를 Java3D[7]를 이용한 단백질 3차 구조 뷰어인 PSAML Viewer의 개발에 대하여 기술하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 단백질 구조에 대한 표현을 2차 구조를 이용하는 PSA 및 이를 표준화된 XML 형식을 이용하여 기술하는 PSAML을 소개하고, 3장에서는 프로그램 설계, 구현 및 사용법에 대해서 기술하고, 4장에서는 결론 및 향후 연구방향에 대해서 기술한다.

### 2. 단백질 구조 표현

PSAML(Protein Structure Abstraction Markup Language)은 단백질의 2차구조와 2차구조 사이에서 발견되는 상호적인 관계를 이용하여 단백질 구조를 표현하는 방법을 제공하는 PSA(Protein Structure Abstraction)를 기준으로 단백질 구조를 표준화된 문서 표현 양식인 XML을 이용하여 기술 할 수 있는 언어이다. 개발된 PSA 및 PSAML 기반으로 표현된 단백질 구조 비교 방법은 2차구조의 특징과 2차구조 사이의 관계를 비교하여 효과적으로 두 단백질간의 유사한 부분 구조를 찾을 수 있다.

#### 2.1 PSA(Protein Structure Abstraction)

PSA는 단백질 구조를 구성하는 2차구조와 그들 사이의 관계를 이용하여 단백질 구조를 추상화하여 표현할 수 있는 방법을 제공한다.

한 단백질 구조를 표현하기 위해서, PSA는 구조를 결정하고 있는 2차구조를 3차원 공간상의 벡터(Vector)로 표현하여 공간적인 정보 및 임의의 두 2차구조 쌍에 대한 각도, 거리, 길이, 그리고 수소 결합 및 방향성 등의 관계에 대한 정보를 이용한 다.

하나의 단백질  $P$ 에 대하여, 추상화된 표현은 다음과 같이 기술될 수 있다.

$$PSA(P) = (S, T, C, A, R)$$

S는 하나의 단백질 구조를 구성하고 있는 2차구조 요소의 집합이며, T는 2차구조의 종류에 대한 정보를 나타내며,  $\alpha$ -나선( $\alpha$ -helix) 또는  $\beta$ -판상조각( $\beta$ -strand)에 대한 정보를 제공한다.

C는 2차구조를 3차원 공간에 위치한 벡터에 대한 정보(시작점 및 끝점)를 나타내고 있으며, A는 2차구조를 구성하고 있는 아미노산 서열과 아미노산 서열의 길이 정보를 가지고 있다.

R은 단백질 구조를 비교할 때 사용될 수 있는 한 단백질 내의 두 2차구조( $E_i$ 와  $E_j$ ) 사이에 정의되는 관계이다.

$$R = (\theta, \gamma, \nu, h, d) \text{ 단, } E_i, E_j \in S, \\ i \neq j, i, j \text{ 는 이차구조 인덱스}$$

$\theta$ 는 두 2차구조인  $E_i$ 와  $E_j$  사이의 각도 관계를 나타내고 있으며,  $\gamma$ 는 두 2차구조인  $E_i$ 와  $E_j$ 의 거리 관계를 나타낸다.  $\gamma$ 는 두 2차구조인  $E_i$ 와  $E_j$  사이의 상대적인 거리에 대한 관계로써 3차원 공간에서 두 2차구조의 중점들간의 거리를 기술하고 있다.  $\nu$ 는 두 2차구조인  $E_i$ 와  $E_j$ 의 각각의 길이를 나타내며,  $h$ 는 두 2차구조인  $E_i$ 와  $E_j$  사이의 수소결합의 유무를 나타내고 있다. 그리고,  $d$ 는 두 2차구조 요소인  $E_i$ 와  $E_j$  사이에 나타나는 방향성을 나타낸다.

### 2.2 PSAML(PSA Markup Language)

PSAML은 PSA를 이용하여 표현된 단백질 구조 정보를 XML 스키마[8]를 이용하여 XML로 기술하는 언어이다.

PSAML 문서는 식별(Identity) 부분과 데이터(Data) 부분으로 나누어 단백질 구조에 대한 정보를 표현한다. 식별 부분은 단백질의 주석을 나타내며, 데이터 부분은 단백질을 구성하고 있는 2차 구조 요소에 대한 기술과 더불어 그들 사이의 관계를 나타낸다.

데이터 부분은 <SSE>과 <R>의 두 요소(elements)를 가진다. <SSE> 요소는 단백질을 구성하고 있는 모든 2차 구조 요소에 대한 정보를 포함한다. <SSE> 요소의 하위요소로는 2차구조의 종류를 나타내는 <ID>, 서열 정보를 나타내는 <AA>와 <Len>, 3차원 공간상에 위치한 좌표 정보를 포함하는 <Vector>로 구성되어 있다. <R> 요소는 단백질을 구성하고 있는 모든 구성요소의 각각의 쌍에 대하여 각도, 거리, 방향성과 같은 관계들을 표현한다.

### 3. PSAML Viewer의 구현

PSAML Viewer는 PSAML파일에 포함된 단백질의 3차구조 정보를 분석하고 Java3D를 사용하여 단백질의 구조를 보여준다.

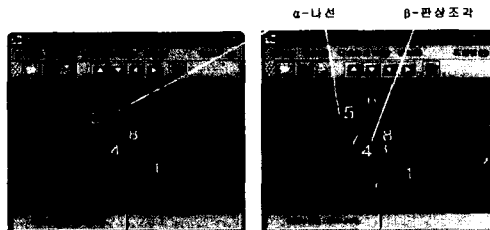
#### 3.1 프로그램 설계

PSAML의 데이터 부분인 <SSE>의 정보 중에서 <ID>와 <Vector>는 <표 1>과 같이 PSAML Viewer에서 표현된다.

<표 1> Viewer에서 사용된 <SSE> 정보

<SSE>요소	내용	PSAML Viewer
<ID>	$\alpha$ -나선	나선
	$\beta$ -판상조각	원통
<Vector>	시작점과 끝점	공간상의 좌표

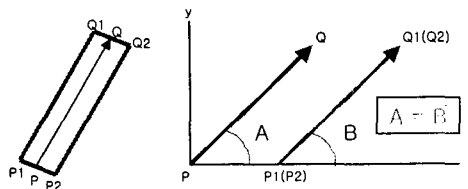
PDB데이터에는 단백질 2차 구조간의 시작과 끝을 연결하는 루프가 존재하지만, PSAML에서는 이를 포함하지 않는다. 이로 인해 본 Viewer에서 각 2차 구조 사이의 연결관계를 보는데 한계가 있다. 이에 본 Viewer에서는 이 루프를 대신하여 각 2차 구조의 시작부분에 번호를 붙여 2차 구조들간의 연결관계를 나타내었다. <그림 1>은 PSAML Viewer에서 표현된  $\alpha$ -나선,  $\beta$ -판상조각 그리고 연결관계를 나타내는 번호를 보여주고 있다.



<그림 1>  $\alpha$ -나선과  $\beta$ -판상조각의 표현

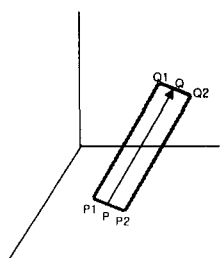
#### 3.2 프로그램 구현

PDB에서는 단백질 2차 구조의 3차원 공간상의 정보는 2차구조를 구성하고 있는 모든 원자들의 좌표 값이 나타나 있지만, PSAML에서 제공되는 3차원 공간상의 2차 구조 정보는 벡터정보(시작점과 끝점)만 제공되고 있다. 2차원 공간상에서 <그림 2>의 왼쪽과 같이  $\vec{PQ}$ 를 기반으로 P를 중점으로 하는  $\vec{P_1P_2}$ 와 Q를 중점으로 하는  $\vec{Q_1Q_2}$  그리고  $\vec{PQ}$ 와 평행한  $\vec{P_1Q_1}$ ,  $\vec{P_2Q_2}$  이 네 선분들로 이루어진 사각형을 그리고자 할 때, <그림 2>의 오른쪽과 같이  $\vec{PQ}$ 와 평행한 벡터들은  $\vec{PQ}$ 와 x축이 이루는 각도와 동일해야 한다.

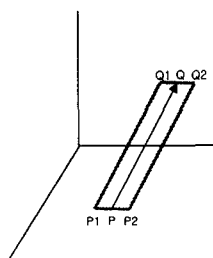


<그림 2> 2차원공간에서의 사각형 표현과 각도 계산

평행한 벡터들이 동일한 각 축에 대해서 동일한 각도를 가지고 있어야 한다는 것은, 3차원공간에서도 동일하게 적용된다. <그림 3>과 <그림 4>에서 보듯이 각도 계산이 없을 경우 3차원 공간상에서 원하는 도형을 제대로 표현하지 못한다.

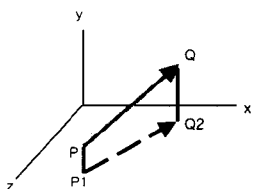


<그림 3> 각도 계산  
한 후 도형 표현

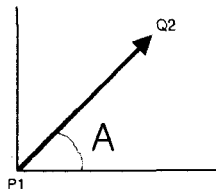


<그림 4> 각도 계산  
없이 도형 표현

3차원공간에서는 <그림 5>와 같이 한 벡터를 각 축의 평면에 투영을 하여, <그림 6>과 같이 2차원공간에서 각도를 계산하는 방법과 동일하다.

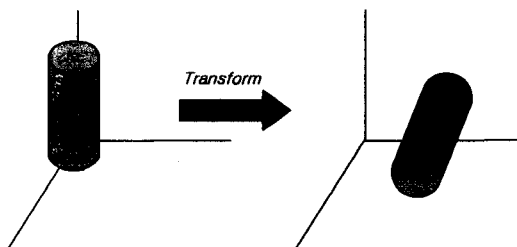


<그림 5> 벡터 투영



<그림 6> 각도 계산

3차원공간에서 단백질 2차 구조(나선, 원통, 화살표)를 표현하고자 할 때, 위와 같은 과정을 반복하는데 있어 많은 계산을 필요로 한다. 중복을 피하고 효율적인 작업을 위해 이 프로그램에서는 원점에서 단백질 2차 구조를 표현한 뒤에, 3차원공간에서 벡터가 가지는 각도를 계산한 뒤에, Transform을 적용시켜 원하는 공간에 위치하도록 구현했다. <그림 7>은 위 과정을 보여주는 예이다.



<그림 7> PSAML Viewer에서의 도형 표현법

### 3.3 프로그램 메뉴

프로그램의 메뉴는 크게 파일, 편집, 디스플레이, 정보로 구성되어 있으며, 파일 메뉴는 PSAML을 열고, 화면상의 내용을 저장하거나 프린트한다. 편집 메뉴는 프로그램 환경을 설정하며, 디스플레이 메뉴는 단백질 2차 구조 표현의 종류를 선택한다. 정보 메뉴는 3차원 공간상에 표현된 단백질 2차 구조의들 사이의 각도, 거리등과 같은 관계 정보를 표시한다. PSAML

Viewer의 주요 메뉴는 <표 2>와 같다.

<표 2> PSAML Viewer 주요 메뉴

주메뉴	부메뉴	설명
파일	이미지로..	단백질 3차 구조를 JPEG으로 저장
	프린트	단백질 3차 구조를 프린트
편집	Arrows	$\beta$ -판상조각의 높이와 크기를 조절
	Cylinders	원통의 반지름을 조절
	Helixs	나선의 반지름을 조절
	Colors	배경화면, 화살표, 나선 그리고 원통의 색깔을 조절
디스플레이	Arrow & Helix	화살표와 나선을 표현한다.
	Arrow & Cylinder	화살표와 원통을 표현한다.
정보	All list	모든 단백질의 구조정보를 트리로 표시
	Relation	입력의 두 단백질 구조간의 정보를 트리로 표시

## 4. 결 론

본 논문에서는 단백질 3차 구조를 보다 효과적으로 파악할 수 있는 Java3D 기반의 단백질 구조 뷰어에 대해서 기술하였다. PSAML Viewer는 사용자 편의를 위해 한글화된 인터페이스를 제공하고 단백질의 3차 구조를 특성별로 파악할 수 있는 여러 가지 구조 보기를 제공함으로써 단백질 3차 구조 파악이 보다 쉬워졌다.

향후 연구 과제로는 단백질의 3차 구조 뷰어의 속도를 좀더 개선하고, 2차 구조의 순서관계를 시각적으로 보이기 위하여 각 2차 구조에 대한 컬러링 기능을 제공할 예정이다.

## 참고문헌

- [1] Applet QuickPDB v1.1 © 1996-1998 SDSC, by Ilya Shindyalov and Phil Bourne
- [2] Waither D. "WebMol - a Java based PDB viewer." Trends Biochem Sci, 22: 274-275, 1997
- [3] 변상희, 김진홍, 문남두, 이명준, "KPDBViewer : Java3D를 이용한 PDB 뷰어 개발", "한국정보과학회, '2003년도 가을 학술발표논문집 Vol. 30 No. 2'
- [4] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne: The Protein Data Bank. Nucleic Acids Research, 28 pp. 235-242, 2000
- [5] 김진홍, 안건태, 변경익, 윤형식, 이수현, 이명준, "단백질 3차 구조의 추상적인 표현기법", 한국정보과학회, '2001가을 학술발표논문집(B) 제 28권 2호, 595-597, 2001.
- [6] Su-Hyun Lee, Jin-Hong Kim, Geon-Tae Ahn, Myung-Joon Lee, "An XML Representation of Protein Datafor Efficient Structure Comparison", Second ICIS, No. 1, pp. 313, 2002
- [7] Java3D at "http://java.sun.com/products/java-media/3D/"
- [8] D. C. Fallside, "XML Schema Part 0: Primer", W3C, May 2001.