

PSAML를 기반으로 한 단백질 구조 비교 (Protein Structure Comparison using PSAML)

김진홍*, 안건태*, 조민수**, 이수현***, 이명준*

*울산대학교 컴퓨터정보통신공학부, ** ***창원대학교 컴퓨터공학과
*{avenue, java2u, mjlee}@mail.ulsan.ac.kr,
oops@pl.changwon.ac.kr, *suhyun@sarim.changwon.ac.kr

요 약

PSAML은 단백질의 2차구조 구성요소와 그들 사이의 관계를 이용하여 단백질 구조를 기술하는 PSA라는 단백질 구조 모델과 표준화된 문서 표현 양식인 XML을 이용하여 단백질 구조를 표현하는 언어이다. 대부분의 단백질 구조 비교하는 방법은 단백질 구조의 분류 및 공통 부분 구조를 파악하는데 효과적으로 활용되고 있으나, 단백질 구조를 비교하기 위하여 자신의 알고리즘에 맞는 고유한 형식의 데이터를 사용하고 있어 상호 호환성이 부족하다.

본 논문에서는 단백질 구조 비교 도구를 효율적으로 개발하기 위한 단백질 구조 표현 언어인 PSAML과 이를 활용하여 단백질 구조를 효과적으로 비교하는 방법에 대하여 기술한다. PSAML 기반의 단백질 구조 비교 방법은 두 단백질 구조 사이의 유사성 그래프를 생성한 후, 최대 유사 부분 그래프를 찾는 방법을 이용하여 유사한 부분 구조를 찾는다.

단백질 구조를 표현하는 방법에 따라 다양하게 존재하고 있다[1].

현재 개발된 단백질 구조 비교 방법은 자신들의 알고리즘에 맞는 고유한 형식의 단백질 구조 표현 데이터를 사용하고 있다. 일반적으로 단백질 구조를 표현하기 위하여 사용되는 정보에는 원자($C\alpha$), 잔기, 단백질 2차

1. 서론

단백질 기능을 파악하기 위하여 단백질의 물리적 및 구조적인 특징에 따라 단백질 구조를 분류하고 공통의 부분 구조를 찾아내는데 활용되고 있는 단백질 구조 비교 방법은

구조 등의 구성요소의 3차원 공간상의 위치, 화학적 특성, 아미노산의 종류, 2차구조의 종류 등이 있다.

대표적인 단백질 구조 알고리즘은 단백질 구조의 내부 분자들 사이의 거리 정보를 동적 프로그래밍 기법을 이용하는 방법[2], 원자들 사이에 RMSD가 최소가 되는 부분을 찾는 방법[3], 단백질 2차구조의 3차원 위치 정보를 유사 부분을 찾기 위하여 기하학적 해싱(geometric hashing)기법 사용하는 방법[4], 그리고 단백질 2차구조 사이의 거리 및 각도 관계를 이용한 방법[5] 등이 있다.

기존 단백질 구조 비교 방법에서 사용되는 데이터는 PDB 데이터베이스[6]에서 제공하는 단백질 구조 정보를 기반으로 하고 있다. 그러나 PDB에서 제공하는 데이터 형식은 텍스트 기반이고 정형화된 문법 명세가 부족하여 단백질 구조 비교를 위한 데이터 변형 과정에서 파싱(parsing) 에러의 가능성을 내포하고 있으며, 단백질 구조 비교를 위하여 부가적인 데이터 변환 과정과 계산 시간이 요구된다. 따라서 단백질 구조 비교에 필요한 구조적 정보를 미리 가공하여 표준화된 방법을 이용하여 보다 효과적으로 단백질 구조 비교에 적용될 수 있는 단백질 구조 표현 방법이 필요하다.

본 논문에서는 단백질 구조에 대한 표현을 2차구조를 이용하는 PSA[7,8] 및 이를 표준화된 XML 형식을 이용하여 기술하는 PSAML[7,8]을 소개한다. 또한, PSAML을 기반으로 표현된 단백질 간의 구조를 비교하는 방법에 대하여 기술한다. 제안된 단백질 구조 비교 방법은 두 단백질 구조를 2차구조와 그들 사이의 관계를 이용하여 유사성 그

래프[9]를 생성한 후, 최대 유사 부분 그래프[10]를 찾는 방법을 이용하여 단백질 간의 유사한 부분구조를 찾을 수 있다.

2. PSAML

PSAML(Protein Structure Abstraction Markup Language)은 단백질의 2차구조와 2차구조 사이에서 발견되는 상호 관계를 이용하여 단백질 구조를 표현하는 방법을 제공하는 PSA(Protein Structure Abstraction)를 기준으로 단백질 구조를 표준화된 문서 표현 양식인 XML로 기술할 수 있는 언어이다. 개발된 PSA 및 PSAML 기반으로 표현된 단백질 구조 비교 방법은 2차구조의 특징과 2차구조 사이의 관계를 비교하여 효과적으로 두 단백질 간의 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 이용한다.

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

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

S는 하나의 단백질을 구성하고 있는 2차 구조 요소의 집합이며, T는 2차구조의 종류에 대한 정보를 나타내며, α -나선(α -helix) 또는 β -판상조각(β -strand)에 대한 정보를 제공한다.

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

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

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

θ 는 두 2차구조인 E_i 와 E_j 사이의 각도 관계를 나타내고 있으며, γ 는 두 2차구조인 E_i 와 E_j 의 거리 관계를 나타낸다. γ 는 두 2차구조인 E_i 와 E_j 사이의 상대적인 거리에 대한 관계로써 3차원 공간에서 두 2차구조의 중점 간의 거리를 기술하고 있다.

ν 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(Protein Structure

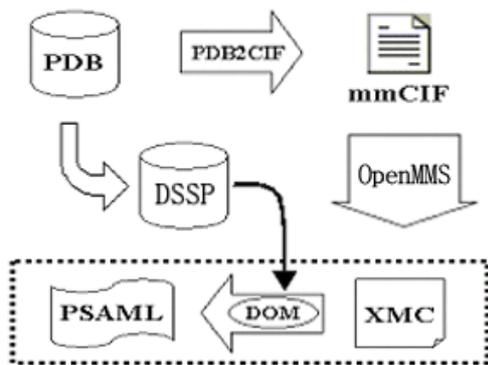
Abstraction Markup Language)

PSAML은 PSA를 이용하여 표현된 단백질 구조 정보를 XML 스키마(XML schema) [11]을 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 2차구조 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 2차구조 요소에 대한 정보를 포함한다. <R>요소는 단백질을 구성하고 있는 모든 구성요소의 각각의 쌍에 대하여 각도, 거리, 방향성과 같은 관계들을 표현한다.

PDB 데이터에서 자동적인 방법으로 XML 문서를 만들기 위하여 mmCIF[12]와 관련된 도구를 사용하여 변환 도구를 작성하였다. PDB 데이터를 PSAML 형태의 문서로 변환하는 전반적인 단계는 [그림 1]과 같다.

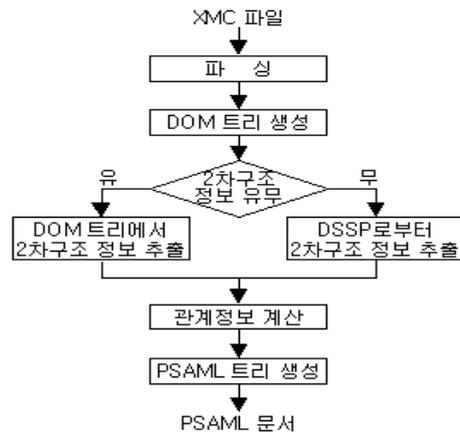


(그림 1) PSAML 문서로의 변환 과정

mmCIF 데이터는 고분자 화합물에 대한 구조 데이터를 표현하는데 있어서 결정학 정보를 가지고 있다.

PDB 데이터에서 PSAML 문서를 생성하는 변환도구는 mmCIF 형태의 파일에 기술된 단백질과 핵산으로 기술되는 고분자 화합물에 대한 데이터를 분석할 수 있는 프로그램들을 제공하고 있는 OpenMMS 툴킷[13]과 응용 프로그램과 스크립트에서 문서의 내용, 구조 및 형식을 동적으로 변경할 수 있는 인터페이스를 제공하는 DOM(Document Object Model)[14]을 이용하여 개발되었다.

변환도구의 동작은 [그림 2]와 같다.



(그림 2) 변환도구의 동작

먼저 DOM 파서를 이용하여 XMC 파일을 파싱한 후 DOM 트리를 생성한다. 그 다음 단계로 생성된 DOM 트리를 재귀적인 방법으로 각 노드를 탐색하면서 원하는 정보가 있는 노드의 텍스트 데이터를 배열과 변수에 저장한다. 2차구조의 정보와 2차구조 사이의 관계 정보는 DOM 트리와 DSSP 파일의 정보를 바탕으로 계산되며, DOM 트리는 mmCIF 정보를 그대로 가지고 있다.

만약 XMC 파일에 PSAML를 생성하는데 필요한 정보, 즉 단백질의 2차구조에 관련된 태그가 존재하지 않을 수도 있는데 그러면 DSSP 파일에서 2차구조 정보를 추출한다.

얻어진 정보를 이용하여 필요한 데이터를 산출하고 새로 생성된 PSA 트리에 각 노드 및 텍스트 데이터를 추가한다. 마지막으로 PSA 트리를 재귀적으로 탐색하면서 각 노드와 데이터를 출력하게 된다.

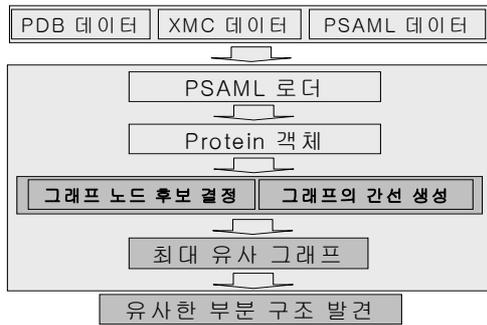
변환도구는 JAVA로 구현되었으며, Apache XML 프로젝트에서 제공하는 XML 파서 Xerces[15]를 이용하여 XMC 파일을 파싱했다.

3. 단백질 구조 비교 방법

단백질 2차구조 기반의 단백질간 구조 비교 방법은 두 단백질의 PSAML 정보를 바탕으로 유사한 부분구조를 내포하는 유사성 그래프를 생성한 후, 모든 노드 사이에 간선이 존재하는 부분 그래프[16]를 찾는 기존의 알고리즘[10]을 이용하여 최대 유사 부분 구조를 파악할 수 있다.

3.1 단백질 간 구조 비교 과정

[그림 3]은 PDB 형태로 입력되는 두 단백질 간의 구조 비교 방법의 과정을 보여주고 있다.



(그림 3) 단백질간 구조 비교 방법

입력되는 PDB 형태의 단백질 구조 데이터는 구현된 변환 도구를 통하여 PSAML 문서로 변환된다. 단백질 구조 비교를 수행하는 방법은 변환된 PSAML을 읽어 Protein 객체를 생성한다. 생성된 Protein 객체는 PSAML에서 정의하는 2차구조의 특징(아이디, 타입, 3차원 좌표)과 2차구조 사이의 관계(각도, 거리, 길이 차이, 수소결합 유무, 방향성)에 대한 정보를 가진다. 생성된 두

Protein 객체로부터 단백질 간 유사성을 내포하는 유사성 그래프를 생성하여 모든 노드들 사이의 간선이 존재하는 부분 그래프를 찾는 알고리즘을 이용하여 최대 유사한 부분 구조를 파악할 수 있다.

3.2 단백질 간 유사성 그래프

PSAML 데이터를 기반으로 단백질 구조 간의 유사성을 내포하는 유사성 그래프 G 는 <표 1>와 같이 정의된다.

<표 1> 유사성 그래프 정의

$G(A, B) = \{V, E\}$, A, B 는 단백질 구조 $V = \{(ai, lj) \mid ai \in A, lj \in B\}$ $E = \{(ai, bk), (cj, bl) \mid 2차구조\ 유사성\ 비교\}$

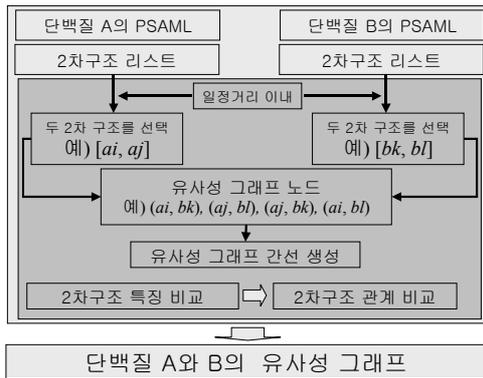
<표 1>에서, 유사성 그래프 G 는 단백질 A 와 단백질 B 의 2차구조의 특징 및 관계를 이용하여 유사성이 있는 2차구조를 표현하고 있다. V 와 E 는 각각 그래프 G 의 노드와 간선의 집합을 나타내고 있다. V 에 속한 각 노드는 단백질 A 의 한 2차구조와 단백질 B 의 한 2차구조의 쌍으로 이루어져 있다. E 에 속한 각 노드 사이의 간선은 노드에 포함된 단백질 2차구조 간의 관계가 유사하면 존재한다.

<표 2> 비교 인자 및 비교 식

관계	비교 식
타입	$T(ai) = T(bk), T(aj) = T(bl)$
각도	$ \theta(ai, aj) - \theta(bk, bl) < \theta d$
거리	$ \gamma(ai, aj) - \gamma(bk, bl) < \gamma d$
길이	$ \nu(ai, aj) - \nu(bk, bl) < \nu d$

노드 (ai, bk) 와 노드 (aj, bl) 사이의 간선은 2차구조 ai 와 aj 에 존재하는 관계와 bk 와 bl 에 존재하는 관계가 유사할 때 생성한다. 이 경우, 단백질 A 의 ai 와 단백질 B 의 bk 및 단백질 A 의 aj 와 단백질 B 의 bl 가 유사하다는 것을 의미한다.

두 단백질 간 유사성 그래프의 생성과정은 [그림 4]와 같다.



(그림 4) 유사성 그래프 생성 방법

[그림 4]에서, 단백질 A 와 B 각각에서 두 2차구조 요소를 선택한다. 이때, 2차구조 사이의 거리가 일정 거리(threshold distance) 이내에 있는 2차구조를 선택한다. 선택되는 2차구조의 쌍은 $[ai, aj]$ 와 $[bk, bl]$ 과 같은 형태이며, (ai, bk) , (aj, bl) , (aj, bk) , (ai, bl) 와 같은 유사성 그래프 노드가 생성된다. 유사성 그래프 노드 사이의 간선은 <표 2>에 기술된 각 관계의 인자 값(θd , γd , νd)보다 작으며 같은 타입일 때 생성된다.

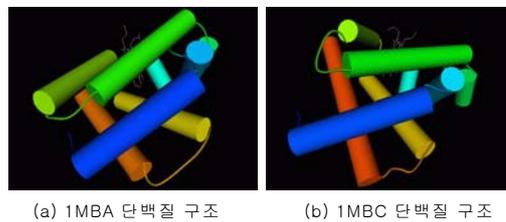
이와 같은 과정을 단백질 A 에서 형성되는 2차구조의 모든 쌍과 단백질 B 에서 형성되는 2차구조의 모든 쌍에 대하여 적용하여 유사성 그래프 G 를 형성한다.

3.3 최대 유사 부분 그래프

형성된 유사성 그래프 G 에서 최대로 유사한 부분 구조는 모든 노드 사이에 간선이 존재하는 부분 그래프(Clique)를 찾음으로써 발견된다. 현재 최대 유사 부분 그래프를 찾는 방법은 SICStus 프롤로그[17]에서 제공하는 API를 이용하여 작성되었다.

4. PSAML 기반의 단백질 구조 비교 결과

PSAML 기반으로 표현된 단백질 구조를 비교하는 방법을 이용하여 PDB ID $1MBA$ [그림 5-(a)]와 $1MBC$ [그림 5-(b)] 단백질 간 구조를 비교하였다. $1MBA$ 와 $1MBC$ 는 각각 8개의 α -나선으로 이루어진 단백질이다. 두 단백질은 Myoglobin 계열에 속하며, 산소를 저장하는 기능을 담당한다.



(그림 5) 비교 대상 단백질 구조

● PSAML 문서 생성

$1MBA$ 와 $1MBC$ 의 PDB 정보를 바탕으로 PSAML 형식으로 변환하는 변환도구를 이

용하여 PSAML 형식의 데이터를 생성한다. [그림 6]은 생성된 *IMBA*와 *IMBC*의 PSAML 문서를 보여주고 있다.

```

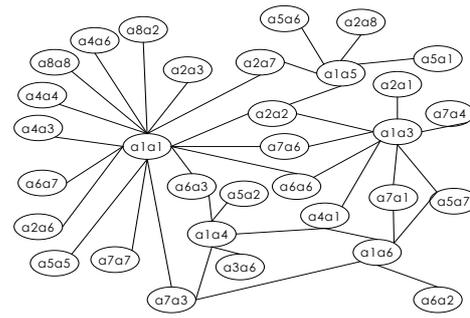
<?xml version="1.0" encoding="UTF-8" ?>
<PSA>
  <Desc>IMBA</Desc>
  <DATA>
    <SSE>
      <Alpha id="A1" n="1">
        <Alpha id="A2" n="2">
          <Alpha id="A3" n="3">
            <Alpha id="A4" n="4">
              <Alpha id="A5" n="5">
                <Alpha id="A6" n="6">
                  <Alpha id="A7" n="7">
                    <Alpha id="A8" n="8">
                      </SSE>
                    </Alpha>
                  </Alpha>
                </Alpha>
              </Alpha>
            </Alpha>
          </Alpha>
        </Alpha>
      </SSE>
    </DATA>
    <Relation s1="A1" s2="A2">
      <S1>A1</S1>
      <S2>A2</S2>
      <Angle>97.66143748273089</Angle>
      <Distance>17.853649192067795</Distance>
      <Length>1.1079969833610032</Length>
    </Relation>
    <Relation s1="A1" s2="A3">
      <S1>A1</S1>
      <S2>A3</S2>
      <Angle>106.48443308139</Angle>
      <Distance>18.173678564210306</Distance>
      <Length>0.11796603473912981</Length>
    </Relation>
  </PSA>
  <Desc>IMBC</Desc>
  <DATA>
    <SSE>
      <Alpha id="A1" n="1">
        <Alpha id="A2" n="2">
          <Alpha id="A3" n="3">
            <Alpha id="A4" n="4">
              <Alpha id="A5" n="5">
                <Alpha id="A6" n="6">
                  <Alpha id="A7" n="7">
                    <Alpha id="A8" n="8">
                      </SSE>
                    </Alpha>
                  </Alpha>
                </Alpha>
              </Alpha>
            </Alpha>
          </Alpha>
        </Alpha>
      </SSE>
    </DATA>
    <Relation s1="A1" s2="A2">
      <S1>A1</S1>
      <S2>A2</S2>
      <Angle>106.48443308139</Angle>
      <Distance>18.173678564210306</Distance>
      <Length>0.11796603473912981</Length>
    </Relation>
    <Relation s1="A1" s2="A3">
      <S1>A1</S1>
      <S2>A3</S2>
  </PSA>
  </Desc>
</PSA>
  
```

(a) *IMBA* (b) *IMBC*

(그림 6) 두 단백질의 PSAML

● 유사성 그래프 생성

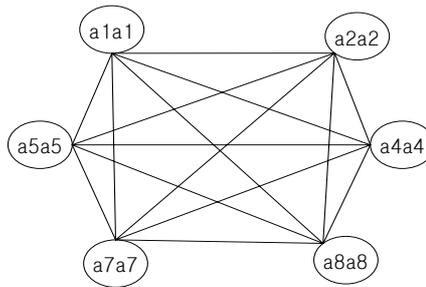
생성된 *IMBA*와 *IMBC*의 PSAML 문서를 읽어 Protein 객체를 생성한 후, [그림 7]와 같은 두 단백질 구조 사이에 유사한 2차구조 쌍에 대한 정보를 나타내는 유사성 그래프를 생성한다. [그림 7]은 노드 사이의 간선의 복잡성으로 노드 (*a1*, *a1*) 중심의 부분 그래프만을 보여주고 있다. 생성된 유사성 그래프에서 두 노드 사이에 간선이 존재할 경우, 이는 두 단백질 구조의 유사한 2차구조 쌍을 나타낸다. 예를 들어, 노드 (*a1*, *a1*)와 노드 (*a2*, *a3*)에 간선이 존재한다는 것은 *IMBA*에 속하는 *a1*과 *a2*의 쌍과 *IMBC*에 속하는 *a1*과 *a3*의 쌍 사이에 나타나는 2차구조의 특징과 관계 정보와 유사함을 보여주고 있다. [그림 7, 8]에서는 유사성 그래프 노드 표현을 *ala2*와 같이 기술한다.



(그림 7) 유사성 그래프 일부

● 최대 유사 부분 그래프

[그림 8]은 *IMBA*와 *IMBC*의 유사 그래프에서 최대 유사 부분 그래프를 보여주고 있다.

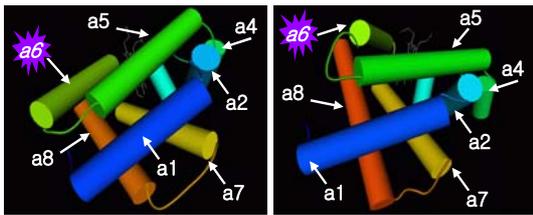


(그림 8) 최대 유사 그래프

*IMBA*와 *IMBC*의 유사 부분구조는 <표 3>과 [그림9]가 보여준다. [그림9]는 VRML[18]을 이용하여 *IMBA*와 *IMBC*의 유사한 부분구조를 보여주고 있다.

<표 3> 일치된 이차구조

ID	일치된 이차구조					
IMBA	a1	a2	a4	a5	a7	a8
IMBC	a1	a2	a4	a5	a7	a8



(a) 1MBA 단백질 구조 (b) 1MBC 단백질 구조

(그림 9) 두 단백질간 유사 부분 구조

● 고려 사항

개발된 PSAML 기반 단백질 구조 비교 방법은 한 단백질 구조에 속한 두 2차구조 사이의 관계와 다른 단백질 구조에 속한 두 2차구조 사이의 관계를 비교할 때 <표 2>에 기술된 관계의 인자 값의 크기에 따라 찾은 유사한 부분 구조의 정확도가 결정된다.

<표 3>에 나타난 결과보다 정확하고 유사한 부분 구조를 찾기 위하여 인자 값의 크기를 작게 하여 단백질 구조를 비교한 결과는 <표 4>와 같다. <표 4>는 각도와 거리에 대한 인자 값을 작게 하여 단백질 구조를 비교한 결과이다.

<표 4> 일치된 이차구조

ID	일치된 이차구조				
1MBA	a1	a2	a4	a5	a7
1MBC	a1	a2	a4	a5	a7

※ 인자 값이 작은 경우

● 타 단백질 구조 비교 방법과 비교

제안된 단백질 구조 비교 방법은 원자(C α)를 이용하여 단백질 구조를 비교하는 방법[2,3]에 비하여 빠르게 구조 비교 결과를

보여 주며, 타 단백질 2차구조 기반의 단백질 구조 비교 방법[4,5]과 비슷한 결과를 보인다.

단백질 1MBA와 1MBC의 비교 결과

```

141 Ca-atoms ( 96%), rmsd = 1.91, 26% identical residues

alignment of 9 SS elements: 9 alpha and 0 beta
a01a02a03a04a05a06a07a08a09
a01a01a02a03a04a05a07a07a09
(a) 9개의 2차구조가 일치된 경우

42 Ca-atoms ( 28%), rmsd = 1.88, 11% identical residues

alignment of 3 SS elements: 3 alpha and 0 beta
a07a08a09
a07a07a09
(b) 3개의 2차구조가 일치된 경우
    
```

(그림 10) 타 단백질 구조 비교 실행결과 일부분

[그림 10]은 1MBA와 1MBC의 단백질 구조 비교를 타 2차구조 기반 도구[5]에서 실행한 결과의 일부분이다. 제안된 구조 비교 방법의 결과[그림 9]와 비교하여, [그림 10-(a)]는 발견되지 않은 2차구조를 포함하고, [그림 10-(b)]는 2차구조 사이에 유사한 관계가 있는 2차구조를 포함하지 않고 있다.

5. 결론

본 논문에서는 단백질 구조 비교를 위한 XML을 이용한 단백질 구조 표현 명세 언어인 PSAML와 이를 이용하여 단백질 구조를 비교하는 방법에 대하여 기술하였다.

PSAML은 단백질 2차구조와 이들 사이의 관계를 이용하여 단백질 구조를 정의하는 PSA를 기반으로 설계되었다. PSAML은 단백질 구조를 2차구조에 초점을 맞추고 있으므로 다른 원자(C α) 또는 잔기 등을 이용한

표현들보다 간결하다. PSAML 문서는 PDB 데이터로부터 개발된 변환기를 통하여 자동적으로 생성되며, 2차구조의 특징(공간적 정보, 타입)과 다른 2차구조와의 관계(각도, 거리, 길이)에 대한 정보를 표현하고 있다. 또한, PSAML은 단백질을 비교할 때 사용되는 데이터를 PDB 데이터에서 가공하여 제공함으로써 보다 효과적인 단백질 구조 비교 방법을 개발할 수 있도록 한다.

PSAML 기반 단백질 간 구조 비교 방법은 두 단백질 사이에 있는 유사 부분 구조를 찾기 위하여 단백질 2차구조의 정보(형태, 길이)와 그들 사이의 관계(각도, 거리, 길이)를 이용한다. 이 방법은 두 단백질 사이의 유사 부분 구조를 유사성 그래프로 표현하고, 이 그래프에서 최대 유사 부분 그래프를 찾는 기존의 방법을 이용하여 찾을 수 있다.

추후연구 과제로는 단백질 구조를 형성하는 2차구조에 대한 특징을 보다 자세히 기술하는 요소를 PSAML에 추가로 기술하여 더 정확한 단백질 구조 비교를 할 수 있도록 한다. 또한, 기존의 단백질 구조 비교 방법을 제한 프로그래밍 기법을 이용하여 더 효과적으로 비교할 수 있도록 할 예정이다.

감사의 글

본 연구는 한국과학재단 목적기초연구(R01-2001-000-00535-0) 지원으로 수행되었습니다.

참고문헌

- [1] I Eidhammer, I Jonassen, W R. "Structure Comparison and Structure Patterns", *Reports in Informatics*, 7, 1999.
- [2] L. Holm and C. Sander, "Protein structure comparison by alignment of distance matrices," *Journal of Molecular Biology*, Vol. 233, pp. 123-138, 1993.
- [3] A. P. Singh and D. L. Brutlag, "Hierarchical Protein Structure Superposition using both Secondary Structure and Atomic Representations," *Proc. Intelligent Systems for Molecular Biology* 97, 1997.
- [4] A. P. Singh and D. L. Brutlag, "Protein Structure Alignment: A Comparison of Methods", 1999.
- [5] N.Alexandrov and D Fischer, "Analysis of topological and nontopological structural similarities in the PDB: New examples with old structures," *Proteins, Structure, Function, and Genetics*, Vol 25, No. 3, pp.354-365, 1996.
- [6] H. M. Berman, J. D. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The Protein Data Bank," *Nucleic Acid Research*, Vol. 28, No. 1, pp. 235-242, 2000.
- [7] Su-Hyun Lee, Jin-Hong Kim, Geon-Tae Ahn, and Myung-Joon Lee, "An XML Representation of Protein Data for Efficient Structure Comparison", *Second ICIS*, No. 1, pp. 313, 2002.
- [8] 김진홍, 안건태, 이수현, 이명준, "구조비교를 위한 단백질 데이터의 XML 표현 기법", *한국정보과학회 프로그래밍언어연구회*, 16권, 2호, 15-16, 2002.
- [9] Hiroaki KATO and Yoshimasa TAKA-

HASHI, "Automated Identification of Three-Dimensional Common Structural Features of Proteins", *J. Chem. Software*, Vol. &, No. 4, p. 161-170, 2001.

[10] C. Bron, J. Kerbosch. Algorithm 457: Finding All Cliques of an Undirected Graph. *CACM*, 16(9):575-577, 1973.

[11] D. C. Fallside, "XML Schema Part 0: Primer", *W3C*, May 2001.

[12] P. Bourne, H. Berman, B. McMahon, K. Watenpaugh, J. Westbrook, and P. Fitzgerald, "The Macromolecular Crystallographic Information File (mmCIF)," *Methods in Enzymology*, Vol. 277, pp. 571-590, 1997,

[13] D. S. Greer, J. D. Westbrook, and P. E. Bourne, "OpenMMS: An Ontology Driven Architecture for Macromolecular Structure," *Objects in Bio and Cheminformatics*, 2001.

[14] W3C, "Document Object Model (DOM)," *WWW document* (<http://www.w3.org/DOM/>).

[15] Xerces: XML parsers in Java, Apache XML Project, *The Apache Software Foundation*. (<http://xml.apache.org/>)

[16] Edith Cohen, Lecture Note in Combinatorial Algorithms & Data Structures, *University of California*, 1997. (www.cs.berkeley.edu/~edith/cs270/Lectures/lecture21.ps)

[17] SICStus Documentation and Manuals, *SICStus Prolog Leading Prolog Technology*, 2002. (<http://www.sics.se/~sicstus/>)

[18] VRML Plug-in, 2002. (<http://www.rcsb.org/pdb/he-lp/graphics.htm>)

김진홍



1999년 2월 울산대학교 전자계산학과 졸업(학사)
2001년 2월 울산대학교 컴퓨터정보통신 공학부 졸업(석사)

2001년 3월 ~ 현재 울산대학교 컴퓨터정보통신 공학부 박사과정
관심분야 : 생물정보학, 제한프로그래밍, 협업지원 시스템, 이동에이전트 시스템 등.

안건태



1999년 2월 울산대학교 전자계산학과 졸업(학사)
2001년 2월 울산대학교 컴퓨터·정보통신공학부 졸업(석사)

2001년 3월 ~ 현재 울산대학교 컴퓨터·정보통신공학부 공학박사과정
관심분야 : 생물정보학, 협업지원 시스템, 분산시스템, 이동에이전트 시스템 등.

조민수



2002년 2월 창원대학교 컴퓨터공학과 졸업(학사)
2002년 3월 ~ 현재 창원대학교 컴퓨터공학과 석사과정

관심분야 : 생명정보학, 리눅스 시스템 등.

이수현



1987년 2월 광운대학교 전
자계산학과 졸업(학사)

1989년 2월 한국과학기술원
전산학과 졸업(석사)

1994년 8월 한국과학기술원
전산학과 졸업(박사)

1994년 9월 ~ 1996년 2월 한국전자통신연구
원 선임연구원

1996년 3월 ~ 현재 창원대학교 컴퓨터공학
과 부교수

관심분야 : 프로그래밍언어, 제한프로그래밍,
생명정보학 등.

이명준



1980년 2월 서울대학교 수
학과 졸업(학사)

1982년 2월 한국과학기술원
전산학과 졸업(석사)

1991년 8월 한국과학기술원
전산학과 졸업(박사)

1982년 3월 ~ 현재 울산대학교 컴퓨터정보
통신 공학부(교수)

1993년 8월 ~ 1994년 7월 미국 버지니아대
학 교환교수

관심분야 : 프로그래밍언어, 분산 객체 프로
그래밍 시스템, 병행 실시간 컴퓨팅, 인터넷
프로그래밍시스템, 생물정보학 등.

