단백질 구조 비교를 위한 이차구조의 상관관계 계산

조민수*, 안건태**, 이범준**, 이수현**
*창원대학교 컴퓨터정보통신공학부 **울산대학교 컴퓨터정보통신공학부
*oops@pl.changwon.ac.kr

Calculation of Relation between Secondary Structures for Protein Structure Comparison

*School of Computer & Information Technology, Changwon National University **School of Computer Engineering & Information Technology, University of Ulsan

요 약
단백질 구조의 표현 방식을 정의하고 호환성 및 상호작용성을 탐지하기 위하여 단백질의 이차구조 구성요소와 그들 사이의 관계를 이용하여 단백질 구조를 기술하는 PSA가 제안되었다. 본 논문에서는 PSA에서 정의된 단백질의 이차구조 사이에 존재하는 상관관계에 대한 계산 방법에 대해 기술하였으며, 이를 바탕으로 구현하여 그 결과를 확인하였다. 본 논문에서 제안한 방법은 단백질의 이차구조 사이의 상관관계를 포함하는 PSAML 데이터로부터 단백질의 구조 및 유사성을 비교하기 위한 단백질 구조 비교 시스템에서 사용할 수 있다.

1. 서론
단백질의 정형(folding)과 구조를 이해하고 분석하는데 있어 단백질 데이터의 전체를 이용하는 것보다 단백질 구조의 특성을 나타내는 대표적인 정보를 이용하는 것이 효과적이다. 단백질의 2차구조는 단백질 구조의 핵심적인 부분이기 때문에 많은 연구자들이 이에 의존하고 있다.


2. PSA
PSA는 단백질 구조를 구성하는 2차구조와 그들 사이의 관계를 이용하여 단백질 구조를 관찰하기에 표현할 수 있는 방법을 제공한다. 2차구조물 사이의 상관관계는 각도, 거리, 간격이 포함되며, 이들 상관관계는 2차구조의 3차원적 좌표 데이터로 계산되기 때문에 관계에 따라 다양한 계산방법이 존재한다. 본 논문에서는 PDB에서 제공하는 데이터 형식을 PSAML 형식으로 변환하여 2차구조물 사이의 상관관계를 계산하는 방법을 기술한다. 본 논문에서 제안하는 방법은 PDB로부터 PSAML로의 변환도구[3]의 구현에 적용될 수 있다.

본 논문은 다음과 같이 구성되어 있다. 1장에서는 단백질 구조를 표현하는 형식인 PSA에 대해 설명하고자 한다. 2장에서는 2차구조 사이의 상관관계 계산의 방법에 대해 설명하고, 3장에서는 단백질 구조 비교 시스템에서의 관계를 기술하고 있다. 본 연구는 한국과학재단 의학기술기초연구(R01-2001-000-00535-0) 지원으로 수행되었다.
두 2차구구인 $E_i$와 $E_j$의 각각의 값이 나타내며, $h$는 두 2차구
조인 $E_i$와 $E_j$ 사이의 구조적인 유무를 나타내고 있다. 그리고,
$d$는 두 2차구조 요소인 $E_i$와 $E_j$ 사이에 나타나는 방향성의 나타
낸다.
두 이차구조인 $E_i$와 $E_j$ 사이의 각도관계는 다음과 같이 기
술된다.

\[
\theta(E_i, E_j) = \angle(\theta_1, \theta_2, \theta_3, \theta_4)
\]

\(\theta\)는 두 이차구조 사이에 다음과 같은 데 가지의 각도를 나
타내고 있다. $\theta_1$과 $\theta_2$는 두 이차구조 ($E_i$와 $E_j$)에 평행한
평면에 평행한 두 벡터 사이에서 정의되는 각도로서, 푸링된 두 벡터에 평행한 중심선을 $L$이라고 할 때, $\theta_1$과 $\theta_2$ 각각은
$E_i$와 $L$사이의 각도와 $E_j$와 $L$사이의 각도를 말한다. 그리고,
$E_i$의 길장을 시작점으로 하고, $E_j$의 시작점을 길점을 함으로 하는 벡터를 $V$라고 할 때, $\theta_3$은 $E_i$와 $V$가 이루는 각도이며, $\theta_4$는
$E_j$와 $V$가 이루는 각도이다.
두 이차구조인 $E_i$와 $E_j$ 사이의 거리관계는 다음과 같이 기
술된다.

\[
\gamma(E_i, E_j) = \text{distance}(D_{\text{max}}, D_{\text{min}}, D_{\text{max}}') D_{\text{min}}'
\]

\(\gamma\)는 두 이차구조인 $E_i$와 $E_j$ 사이의 상대적인 거리에 대한
관계로써 시베타인 값을 가지는다. $D_{\text{max}}$은 3차원 공간에서 두
이차구조의 중점간의 거리를 기술하고 있다. 반면에, 나머지
거리관계는 두 이차구조에 평행한 평면에 평행한 두 벡터 사이
에서 정의되는 거리관계이다. 푸링된 두 벡터에 평행한 중심선
을 $L$이라고 할 때, $D_{\text{max}}$, $D_{\text{min}}$, 각각 $E_i$와 $L$사이의 최대
거리 및 최소거리 값을 가지고, $D_{\text{max}}'$, $D_{\text{min}}'$은 각각 $E_j$와 $L$
사이의 최대거리 및 최소거리 값을 가진다.

3. 이차구조 사이의 상관관계 계산

3.1 두 이차구조에 평행한 평면
각각의 거리를 계산하기 위해서는 먼저 두 이차구조 ($E_i$와 $E_j$)
에 평행한 평면을 결정해야 한다. 공간상의 두 벡터에 평행한 평면은
수없이 많게 존재한다. 여기서는 수많은 평면 중에서 벡터 $E_i$를 포함
하는 평면을 기준으로 벡터 $E_j$와 평행한 평면을 선별하여 그 평면
$P$로 벡터 $E_j$를 푸링시킨다. 평면 $P$에 벡터 $E_j$를 푸링시키기 위
해서는 다음과 같은 과정을 거쳐진다.

첫 번째로 두 벡터 $E_i$와 $E_j$를 모두 원점으로 평행이동 시킨다.
이후 벡터의 시작점을 원점과 일치시킨다. 원점을 $P_i$라고 하고 이동
한 두 벡터 $E_i$와 $E_j$의 길점을 각각 $P_2$, $P_3$라고 하면, 두 벡터
$P_1P_2$와 $P_1P_3$는 두 벡터 $E_i$와 $E_j$를 원점으로 평행이동된 후
의 벡터이다. 그러므로 두 벡터 $P_1P_2$와 $P_1P_3$으로 공간상의 평면이
만들어 지는데 이 평면 $P_4$는 평면 $P$과 평행하다.

다음 단계로 평면에 수직인 렌선 벡터를 구해야 하는데 벡선 벡터
$N$은 \(\text{<그림 1>}\)과 같이 평면 위에 있는 두 벡터의 외적(cross
product)으로 쉽게 구할 수 있다.

\[
N = P_1P_2 \times P_1P_3
\]

\(\text{<그림 1> 벡선 벡터 } N\)을 구하는 방법

이제 구해진 벡선 벡터 $N$을 이용하여 벡터 $E_j$를 평면 $P$로 푸링
시키게 되는데 이때 다음과 같은 벡터의 분할이 이루어진다.

면에서 두 벡tor $u$와 $a$가 시점이 어떤 점 $Q$에서 일치하도록 위
치하였다면 $u$를 \(\text{<그림 2>}\)와 같이 분할할 수 있다. 그렇다면 벡터
$w_2$를 $u$에서 $a$로의 직교사영(orthogonal projection of $u$ on $a$)
이라 하고 $\text{proj}_u a$로 표기한다. 또한 벡터 $w_2$를 $a$에 직교하는 $u$
의 벡터상분(vector component of $u$ orthogonal to $a$)이라 한다.

\[
w_2 = u - \text{proj}_u a = u - \frac{u \cdot a}{\|a\|^2} a
\]

\(\text{<그림 2> 벡터의 분할}\)

여기서 벡터 $E_j$의 푸링에 필요한 벡터는 $w_2$인데 다음과 같은 공
식을 이용하여 구할 수 있다\([4]\).

위에서 구해진 벡선 벡터 $N$을 $a$라고 두고, 벡터 $E_j$의 시작점을
시작점, 벡터 $E_i$의 시작점을 끝점으로 하는 벡터를 $u$라고 두면 앞
의 공식에 의해서 벡터 $w_2$를 구할 수 있다. 이 벡터 $w_2$는 벡터 $E_j$
의 시작점에서 평면 $P$에 푸링된 벡터 $E_i$의 시작점까지의 벤위
(displacement)를 나타낸다. 또한 벡터 $P_1P_2$는 벡터 $E_j$의 시작점에
서 끝점까지의 벤위를 나타내므로 평면 $P$에 푸링된 벡터 $E_i$의 시작
점과 끝점은 다음과 같이 구할 수 있다.

\[
E_i의 시작점 + w_2 = 평면에 푸링된 E_i의 시작점
E_i의 시작점 + P_1P_2 = 평면에 푸링된 E_i의 끝점
\]

위와 같은 방법으로 벡터 $E_j$를 두 벡터 $E_i$와 $E_j$에 평행한 평면
$P$에 푸링하였다.

3.2 각도 상관관계의 계산

루핑된 두 벡터의 중점을 연결한 벡터를 $M$이라 하면 평행선과
직각각선각의 성질을 이용하여 가상의 중심선 $L$과 푸링된 두 벡터
사이의 두 각 $\theta_1$과 $\theta_2$는 \(\text{<그림 3>}\)과 같은 방법으로 구할 수 있는
데 두 벡터가 이루는 각은 아래에 나오는 벡터의 내적적이 이용된다. 그리고 나머지 두 각 $\theta_3$과 $\theta_4$는 평행선의 성질을 이용하여 쉽게 구할 수 있다.

\[
\cos \theta = \frac{a \cdot b}{||a|| ||b||} = \theta_1, \quad \cos(\theta_1) = \theta
\]

3.3 거리 상관관계의 계산
두 영역 두 벡터의 중심을 연결한 벡터 $M$의 길이를 $m$이라 하면 가상의 중심선 $L$과 두 영역 두 벡터 사이의 값가 거리의 계산은 $\langle \text{그림 4} \rangle$에 나타나는 식으로 그 값을 구할 수 있다. 그러나 여기서 벡터의 방향과 위치에 따라 최대거리와 최소거리가 되어질 수도 있으며 벡터 $E_i$나 벡터 $E_j$가 중심선 $L$과 만날 경우 최소거리로 0으로 처리한다.

\[
D_{\text{med}} = ||T||
D_{\text{max}} = m^2 - a \cos A
D_{\text{min}} = m^2 - b \cos B
D_{\text{max}} = m - D_{\text{min}}
D_{\text{max}} = m - D_{\text{min}}
\]

3.4 실험 및 결과
앞 장에서 기술한 방법을 JAVA 프로그램으로 구현하여 그 결과를 살펴보았다.

1B01이라는 단백질은 내 개의 a-나선과 두 개의 $\beta$-판상과로 이루어져 있는데, 그 구조는 $\langle \text{그림 5} \rangle$와 같다. 여기서 a-나선은 원통형으로 표시되었으며, 원통 속의 원심표는 두 개의 a-나선 A2와 A3의 방향을 나타내고 있다. 이 두 이차구조 요소 사이의 각도 및 거리 관계를 생성된 1B01의 PSAML 문서(\langle \text{그림 6} \rangle)에서 보여준다. 이 문서에서, $\langle \text{그림 5} \rangle$에서 보면 바라 보면 두 일차구조 요소는 서로 교차하게 되는데 가상의 중심선 $L$을 생각해 보았을 때, 두 일차구조 요소 사이의 관계 중 최소거리가 있는 것을 확인할 수 있으며 나머지 각도 및 거리 관계에 대한 수치를 얻을 수 있다.

4. 결론
본 논문에서는 PSA에 의한 단백질의 이차구조 간의 각도와 거리 관계를 계산하는 방법에 대해서 검증하고 그 결과를 살펴보았다. PSA에서 제공되는 이차구조의 상관관계들은 단백질의 구조에 관한 공간상의 여러 가지 각도 및 거리 정보를 제공함으로써 단백질 구조를 비교하는 여러 형태의 시스템을 개발하는데 유용하게 이용될 수 있다.

 앞으로는 생성된 PSAML 파일을 이용하여 단백질 구조비교와 유사도 측정을 위한 시스템을 개발할 예정이다. 또한 PSAML 형태로 표현된 단백질 구조를 논리적 표현으로 변환하는 방법과 재현 프로그램 개발을 이용하는 방법을 개발할 예정이다.

참고문헌