

단백질 아미노산의 단순 곡면 표현 기법

김구진

경북대학교 컴퓨터학부

e-mail : kujinkim@gmail.com

Simple Surface Representation for Protein Amino Acids

Ku-Jin Kim

School of Computer Science & Engineering, Kyungpook National University

요 약

본 논문에서는 잔기 유연성(residue flexibility)을 가진 단백질 아미노산에 대한 기하학적인 계산 효율성을 높이기 위해 아미노산을 단순 곡면으로 표현하는 기법을 제안한다. 잔기 유연성을 가진 아미노산은 고정된 백본(backbone)에 대해 잔기가 회전하는 특성을 갖는다. 단순 곡면 표현 기법을 이용하여 아미노산의 충돌 감지 속도 효율성을 높일 수 있다.

1. 서론

단백질 분자에 대해 기하학적인 접근 방법을 사용하는 연구에서는, 아미노산의 각 원자 별로 반데르바스 반경(van der Waals radius)을 갖는 구(sphere)를 대응시켜 계산을 진행하는 것이 일반적이다 [1-3]. 이러한 경우, 단백질 분자는 적개는 수백에서 많게는 수십만 개의 구의 집합으로 표현된다. 정적인 상태의 단백질 분자에 대해서는 이러한 형태의 표현 방법으로도 적절한 시간 내에 기하학적인 계산을 수행할 수 있다. 그러나, 유연성을 가진 단백질 분자가 동적으로 형태가 변할 때, 단백질 분자의 움직임을 모든 원자들의 위치를 변경시켜 표현하는 것은 계산 시간 상 비효율적이며 단백질 분자의 연속적인 변화를 구현하기 어렵다.

본 논문에서는 각 아미노산 별로, 서로 공유결합된 원자들의 열(sequence)을 밀접하게 둘러싸는 단순 곡면들을 계산함으로써, 단백질 분자를 단순 곡면들의 집합으로 표현하고자 한다. 각 아미노산은 기하학적인 계층 구조를 가지며, 잔기가 회전할 경우 회전축의 위치 및 이에 대한 하위 원자들의 상대적인 거리가 거의 일정하게 유지된다.

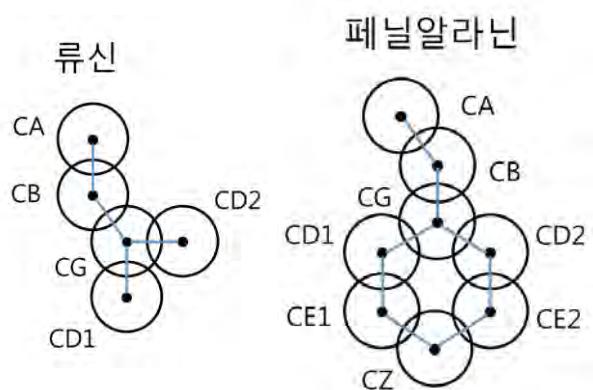
2. 아미노산 구조 분석

아미노산의 구조는 일반적으로 공유결합이 있는 원자의 중심점들을 서로 연결하여 표현한다. 그림 1에서 아미노산의 종류인 류신과 페닐알라닌의 구조를 예로 보인다. 기하학적인 측면에서 각 아미노산은 서로 교집합이 존재할 수 있는 구의 집합으로 표시되며, 공유결합이 있는 원자들은 원자에 대응하는 구들이 서로 교차한다.

그림 1에서 제시된 바와 같이, 아미노산 구조는 페닐알라닌에서 포함된 것과 같은 링(ring) 상태의 구

조와 그 외의 체인(chain) 상태의 구조로 구분할 수 있다.

아미노산 중에서 링 구조가 포함된 것은 히스티딘, 페닐알라닌, 트립토판, 티로신이고, 나머지 아미노산들은 링 구조를 포함하지 않고 체인 상태로 구성되며, 트리 형태로 표시할 수 있다.



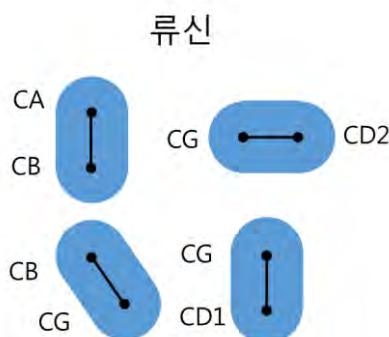
(그림 1) 아미노산 구조의 예

3. 단순 곡면 표현

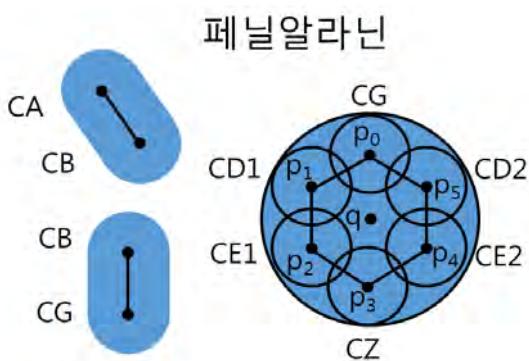
아미노산 내에서 함께 회전하는 원자들의 집합을 한 개 이상의 단순 곡면으로 표현하고자 하며, 잔기의 하위 구조의 형태에 맞게 단순 곡면을 적용한다.

공유결합된 원자 쌍 A_1, A_2 가 주어지고, 이들은 링에 포함되지 않는다고 가정하자. A_1, A_2 의 중심점이 각각 p_1, p_2 이고 반경이 r_1, r_2 ($r_1 \leq r_2$)라 하자. 이에 대한 단순 곡면은 중심점의 궤적인 양끝점 p_1, p_2 를 갖는 선분이고 반경이 r_2 인 구의 swept volume으로 표현한다.

원자 $A_0, A_1, A_2, \dots, A_{n-1}$ 이 링을 구성하고, 모든 A_i ($0 \leq i < n$)와 $A_{(i+1) \% n}$ 사이에 공유결합이 있다고 가정하자.



(그림 2) 류신의 단순 곡면 표현



(그림 3) 페닐알라닌의 단순 곡면 표현

링에 포함된 원자들은 토러스(torus)를 이용하여 표현하고자 한다. 토러스의 중심점이 q , major radius 가 R , minor radius 가 r , 토러스의 main circle 을 포함하는 평면의 normal vector 가 N 이라 할 때, q , N , R , r 는 다음과 같이 구해진다. 링에 포함된 모든 $A_i (0 \leq i < n)$ 의 중심점이 p_i 로 표시될 때, q 는 모든 p_i 의 평균으로 다음과 같이 계산한다.

$$q = (1/n) \sum_{0 \leq i < n} p_i$$

토러스의 main circle 은 q 를 지나며, normal vector N 은 q , p_i , $p_{(i+1)\%n}$ 삼각형의 normal vector 들의 평균으로 구한다.

$$N = (1/n) \sum_{0 \leq i < n} (p_i - q) \times (p_{(i+1)\%n} - q) / \| (p_i - q) \times (p_{(i+1)\%n} - q) \|$$

원자 A_i 의 반경이 r_i 라 할 때, 토러스의 major radius R 은 다음 식을 이용하여 구한다.

$$R = (1/n) \sum_{0 \leq i < n} \|p_i - q\|$$

각 점 p_i 와 토러스 main circle 간의 거리가 d_i 라 하자. 토러스의 minor radius r 은 다음과 같이 계산한다.

$$r = \text{MAX}_{0 \leq i < n} (d_i + r_i)$$

그림 2 와 그림 3 에서는 류신과 페닐알라린의 예를 들어 단순 곡면 표현을 제시한다.

4. 결론 및 향후 연구 방향

본 논문에서는 각 아미노산을 단순 곡면으로 표현하는 방법을 제안하였다. 제안된 단순 곡면 표현은 잔기 유연성을 가진 아미노산의 최소 거리 및 충돌 탐지 계산의 효율성을 높일 수 있다. 향후 이를 이용하여 잔기 유연성을 가진 단백질 분자를 표현하고, 단순 곡면들로 표현된 단백질 분자에 대한 기하학적인 계산을 효율적으로 수행하고자 한다.

감사의 글

이 논문은 2013 년도 정부(교육부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(NRF-2013R1A1A2A10004391).

참고문헌

- [1] T. Can, C. -I. Chen, Y. -F. Wang, "Efficient molecular surface generation using level-set methods," Journal of Molecular Graphics and Modelling 25 (2006) 442–454
- [2] W. Chen, J. Zheng, Y. Cai, "Kernel modeling for molecular surfaces using a uniform solution," Computer-Aided Design 42 (2010) 267-278.
- [3] J. Cortes, S. Barbe, M. Erard, T. Simeon, "Encoding molecular motions in voxel maps," IEEE/ACM Trans. On Computational Biology and Bioinformatics, 8(2) (2011) 557-563.