Reconstruction of a-helices in a Protein Molecule

Beom Sik Kang^{*} · Ku-Jin Kim^{**} · U Deok Seo^{***}

ABSTRACT

In a protein molecule, a-helices are important for protein structure, function, and binding to other proteins, so the analysis on the structure of helices has been researched. Since an interaction between two helices is evaluated based on their axes, massive errors in protein structure analysis would be caused if a curved or kinked long a-helix is considered as a linear one. In this paper, we present an algorithm to reconstruct a-helices in a protein molecule as a sequence of straight helices under given threshold.

Keywords : Protein Molecule, Protein Three-dimensional Structure, a-helix



단백질 분자 내에서 α-헬릭스는 단백질의 구조나 기능, 그리고 다른 단백질과의 결합, 활성 조절 등에 있어 중요한 역할을 하며, 이에 따라 헬릭스에 대한 구조적인 분석이 연구되어 왔다. α-헬릭스는 그 중심축을 기준으로 다른 α-헬릭스와의 상호위치를 평가하기 때문에 길 게 휘어지거나 꺾인 a-헬릭스들을 한 개의 헬릭스로 해석할 경우에는 단백질의 구조 분석에 있어서 상당한 오차가 발생할 수 있다. 본 논 문에서는 PDB 파일 내에 표시된 단백질 분자의 α-헬릭스를 주어진 오차 범위 내에서 여러 개의 곧은 형태의 헬릭스로 재구성하는 알고리 즘을 제안한다.

키워드 : 단백질 분자, 단백질 삼차원 구조, α-헬릭스

1. 서 론

단백질은 자신을 구성하는 폴리펩티드 사슬(polypeptide chain)이 접혀져 고유한 입체구조를 형성함으로써 제 기능을 가지게 된다. 폴리펩티드가 구성하는 국부적인 모양은 나선형 의 α-헬릭스(alpha-helix)와 길게 펼쳐진 β-가닥(beta-strand) 그리고 부정형의 루프(loop)로 나눌 수 있다. α-헬릭스는 다 른 헬릭스와 상호작용하여 coiled-coil 구조와 four-helix

bundle 그리고 globin fold 등 특징적인 형태를 만들어 전체 단백질의 일부 구조를 이루거나 다른 단백질과 결합 및 활 성 조절 등의 기능을 수행하기에 헬릭스는 구조적 분석의 대상이 되어 왔다[1, 2].

한 개의 단백질 분자 또는 단백질 복합체를 구성하는 아미 노산의 서열과 각 아미노산을 구성하는 원자들의 기하학적인 정보는 PDB 파일의 형태로 주어진다(http://www.pdb.org). PDB 파일에서는 각각의 a-helix를 구성하는 첫 번째 아미 노산부터 마지막 아미노산까지의 범위를 지정하여 알려 준다. 거의 모든 a-helix는 직선 형태, 휘어진 형태 그리고 꺾인 형태로 구분된다[2]. a-helix는 N번째 아미노산이 N+4번째 의 아미노산과 연속적으로 수소결합을 하여 곧은 원통형의 구조를 이루게 된다. 이때, 수소결합이 억제되는 proline[3]이

[※] 이 논문은 2011년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지 원을 받아 수행된 기초연구사업(No. 2011-0015987)이며, 2013년도 정부 (교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업

논문접수:2013년 12월 30일 수 정 일:1차 2014년 3월 5일

심사완료 : 2014년 3월 6일

^{*} Corresponding Author: Ku-Jin Kim(kujinkim@gmail.com)

164 정보처리학회논문지/소프트웨어 및 데이터 공학 제3권 제4호(2014.4)

나 주위의 다른 구조적 영향으로 헬릭스는 꺾임이 발생하거 나 휘게 된다[4, 5]. 헬릭스 간의 상호작용에 있어 한 헬릭스 가 다른 헬릭스와 배열된 각도는 헬릭스 간의 결합 패턴을 결정하며[6] 이는 단백질의 기능을 분석하는 데 중요한 도움 을 제공한다[7]. 이에 Prokink[8]나 HELANAL[9]같은 헬릭 스의 꺾인 정보에 대한 여러 분석용 프로그램이 개발되어 왔다. 이들을 이용한 대부분의 분석 대상은 수용성인 구형 의 단백질 내의 헬릭스[2, 10] 또는 막에 존재하는 단백질의 헬릭스[11]이다.

기존의 헬릭스 분석용 프로그램들이 상대적으로 길이가 짧은 헬릭스들이 꺾인 상태로 주어진 경우를 다루는 것과 달리, 본 연구에서는 특징적인 단백질에서 발견되는 아주 긴 헬릭스를 대상으로 한다. 이런 헬릭스를 하나의 긴 원 통모양의 구조로 해석할 경우 다른 헬릭스와의 상호작용 에 대한 분석에서 상당한 오차를 수반하게 된다. 긴 헬릭 스는 짧은 헬릭스에 비해 그 휘어진 정도가 약한 경향이 있는데[2], 헬릭스가 proline 등으로 인해 확실한 꺾임이 나타난 경우는 여러 개의 분할된 직선 형태의 헬릭스로 나누어 분석하지만[8, 9] 확실한 꺾임이 나타나지 않은 낮 은 정도로 휘어진 긴 헬릭스의 경우 그 기준점을 확실히 정할 수 있는 방법을 찾을 수 없었다. 길게 휘어진 헬릭스 를 분석하기 위해서는 여러 개의 곧은 형태의 헬릭스로 재구성하는 과정이 필요하며, 이를 수행하기 위한 알고리 즘을 제안한다.

본 논문의 구성은 다음과 같다. 2절에서는 사용자가 정한 오차 범위에 따라 α-헬릭스를 분할하여 재구성하는 알고리 즘을 제시한다. 3절에서는 실험 결과를 보인다. 4절에서는 결론을 내린다.

2. PDB a-헬릭스의 재구성 알고리즘

본 논문에서는 PDB 내의 단백질 2차구조를 나타내는 α-헬릭스와 구별하기 위하여 기하학적으로 정의된 곡선인 헬 릭스 (helix) (즉 중심축이 직선 형태인 헬릭스)를 나선구조 라고 표기한다. Fig. 1에서는 일반적인 나선구조를 도식화하 였다. 나선 구조는 파라미터 θ에 대한 3차원의 곡선식으로 나타낼 수 있으며, 곡선 식은 다음과 같다.

 $C(\Theta) = \mathbf{q} + \mathbf{b}_1 r \cos \Theta + \mathbf{b}_2 r \sin \Theta + \mathbf{N} k \Theta$





여기에서, **q**, **b**₁, **b**₂, **N**은 나선 구조가 구성되는 local frame을 나타내며, **b**₁, **b**₂, **N**은 서로 수직인 단위벡터들이다. r은 나선 구조의 반경이며, k는 **N** 방향에 대한 증가율을 나 타낸다(Fig. 1A). C(θ)를 **b**₁**b**₂-평면에 투영한 결과는 Fig. 1B와 같은 반경 r인 원이 된다. 나선구조의 수식 C(θ)는 다 음의 성질을 갖는다.

1) C(θ)의 N 방향에 대한 높이 변화는 일정하다.

- 2) C(θ)를 **b**1**b**2-평면에 투영하면 원으로 표현된다.
- 3) 나선구조의 중심축은 일직선 상에 존재한다.

단백질의 a-헬릭스는 단백질의 backbone 부분에 대해 정 의되는 구조이다. 모든 아미노산은 backbone과 side chain 부분으로 나누어진다. Backbone 부분은 네 개의 연속적인 원 자 배열을 가지며, 질소(nitrogen, N)-알파탄소(alpha carbon, CA)-탄소(carbon, C)-산소(O)로 구성된다. Side chain 부분 은 베타탄소(beta carbon) 및 이와 연속적으로 결합된 원자 들로 구성되며, backbone 내의 알파탄소와 베타탄소가 결합 하여 아미노산을 구성한다.

Backbone에 속한 원자들은 어떤 아미노산에서도 거의 동 일한 상대적인 위치와 거리로 배열되어 있다. 또한, backbone이 a-헬릭스 형태일 때, a-헬릭스의 기하학적인 정보는 각 아 미노산에 속한 알파탄소 중심점의 sequence를 이용하여 추 출하는 것이 일반적이다[8, 9].

주어진 a-헬릭스의 알파탄소 중심점의 sequence가 p, p+1, p+2, ... 라 할 때, a-헬릭스가 나선구조와 정확히 일치 할 경우에는 Fig. 1A과 같이 점들이 배치될 것이다. a-헬릭 스의 성질에 따라 투영한 원 (즉, 나선구조의 한 개의 턴)에 는 3.6개의 아미노산이 포함된다. 즉, 나선구조가 투영된 한 개의 원에는 3.6개의 아미노산이 포함된다고 가정할 수 있 으며, 이에 따라 각 아미노산에 포함된 알파탄소 중심점은 투영 원에서 100° 간격으로 배치된다. Fig. 1B와 같이 점들 이 bib2-평면에 투영될 경우, 연속된 두 개의 점 p; p+1은 q 를 중심으로 100° 간격으로 배치된다.

주어진 a-헬릭스가 정확히 나선구조를 가질 경우에는 알 파탄소 중심점의 sequence를 이용하여 나선구조의 식을 유 도할 수 있다. 그러나, a-헬릭스는 형태가 나선구조와 유사 하지만, 중심축이 휘거나 꺾어진 형태로 주어지는 경우가 대다수이다.

본 논문에서는 PDB 파일에서 주어진 a-헬릭스를 분할하여 사용자가 지정한 오차 범위 내에서 나선 구조의 sequence로 재구성하는 알고리즘을 제안한다. 이는 a-헬릭스가 가지는 중심축이 휘어지거나 꺾인 상태일 때, 중심축을 piecewise linear approximation하여 수행되며, 이때 오차는 a-헬릭스 중심축과 중심축에 대한 근사 선분 간의 거리로 계산된다. 또 한, 분할되는 지점에서 아미노산 간의 수소결합이 깨어지는 부분을 두 개의 연속된 아미노산 진의 수소결합이 깨어지는 부분을 두 개의 연속된 아미노산 A; Ai+1에서 backbone에 포 함된 A,의 산소(O)와 Ai+1의 질소(N) 원자 간의 거리로 표시 하여, 3차원 공간 상에서 선분으로 가시화한다.

a-헬릭스에 속한 4개의 연속적인 알파탄소의 중심점이 po, pı, p₂, p₃라 할 때, 나선 구조의 반경 r과 k 값은 po.p₃ 값에 의하여 상수로 얻어진다. 이후, Eq. 1을 이용하여 나선 구조의 local frame **q**, bı, b₂, N을 계산할 수 있다.

$$\begin{bmatrix} \operatorname{rcos} \omega_0 \ \operatorname{rsin} \omega_0 \ \operatorname{k} \omega_0 \ 1 \\ \operatorname{rcos} \omega_1 \ \operatorname{rsin} \omega_1 \ \operatorname{k} \omega_1 \ 1 \\ \operatorname{rcos} \omega_2 \ \operatorname{rsin} \omega_2 \ \operatorname{k} \omega_2 \ 1 \\ \operatorname{rcos} \omega_3 \ \operatorname{rsin} \omega_3 \ \operatorname{k} \omega_3 \ 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ N \\ q \end{bmatrix} = \begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{bmatrix} \ (Eq.1)$$

여기에서 wo, w1, w2, w3은 각각 0°, 100°, 200°, 300°에 대응하는 radian 값이다.

본 논문에서 제안하는 알고리즘은 7개 이상의 아미노산이 포함된 a-헬릭스에 대해 적용된다고 가정한다. 최소 2개의 턴에 해당하는 알파탄소 중심점의 sequence가 주어질 때, 이 sequence가 나선구조에 가까운 지 더 정확히 판단할 수 있 으며, a-헬릭스가 2개의 턴을 가질 경우 최소 7개의 아미노 산이 포함되기 때문이다.

사용자가 정한 오차 범위가 ɛ이라 할 때, a-헬릭스를 나 선 구조의 sequence로 재구성하기 위한 기본 알고리즘은 다 음과 같다.

- 단계 1. a-헬릭스를 구성하는 아미노산 서열을 Aı.. An (n ≥ 7)이라 할 때, Ai (1 ≤ i ≤ n) 에 포함된 알파 탄소의 중심점을 pi라고 표시한다. 4개의 연속된 알파탄 소 pi..pi+3 (1 ≤ j ≤ n-3)이 나선 구조를 이룬다는 가 정 하에, Eq.1에 좌표값을 대입하여 나선 구조의 local frame qi, b1.j, b2.j, Nj을 계산한다. Fig. 2A의 그림에서 는 pi 점들을 빨간색 점으로, sequence는 빨간 선으로 표 시하였다. 각 pi..pi+3 에 대해 구한 qi 점을 파란 점으로 표시하였고, 파란 선분들로 점들을 연결하여 sequence를 보였다. Fig. 2B는 Fig. 2A에서 qi 점들에 대한 정보만 을 표시하였다.
- 단계 2. 계산된 각 나선구조의 local frame reference point 중 양 끝에 해당하는 **q**₁, **q**_{n-3}을 연결하는 선분 *L* 을 구한다. Fig. 2C와 Fig. 2D에서 예를 보인다.
- 단계 3. 선분 L과 각 q_j 에 대한 거리를 계산하여 d_j라 하 고, d_j가 최대 거리가 되는 j값을 선택하여 M이라 한 다. Fig. 2E와 Fig. 2F에서 선분 L과 각 q_j 의 거리를 표시하였다.

단계 4. dM ≤ ε이면 α-헬릭스를 분리하지 않고 현재



Fig. 2. Examples for each step in the reconstruction algorithm (an α -helix from PDB id. 3GIG)

function의 수행을 마친다. *dM* > ε 이면, **q***M*을 계산하는 데 사용된 네 개의 알파탄소 중심점 **p***M*, **p***M*¹, **p***M*², **p***M*⁴3 과 **q***M*간의 거리 중 최대가 되는 알파탄소에서 α-헬릭스를 두 개로 분리한다. 즉, **p**_{1..}**p**ⁿ으로 구성된 한 개의 α-헬릭스를 **p**_{1..}**p***M* 과 **p***M*⁴1..**p**ⁿ 으로 분리한다. 분리 된 α-헬릭스 각각에 대해 알고리즘을 재귀적으로 수행 한다. Fig. 2G와 Fig. 2H는 제안된 알고리즘을 재귀적으 로 수행하여 α-헬릭스가 3개로 분할된 예를 보인다.

3. 실험 결과

제안된 알고리즘은 Visual C++과 OpenGL을 이용하여 구 현하였다. PDB 파일에서 추출한 α-헬릭스 정보에 대해 ε을 1.5Å로 정하여 알고리즘을 적용한 결과로 Fig. 3과 같은 수 행 결과를 얻었다.



Fig. 3. Examples of reconstruction of α -helices(PDB id. 3GIG)

Fig. 3A는 PDB id. 3GIG인 단백질복합체에 대해 한 개의 subunit에 해당하는 단백질을 가시화하였다. Fig. 3B는 단백 질의 α-헬릭스에 해당하는 알파카본 중심점들을 빨간 점으 로, 중심점의 sequence를 빨간 선으로 가시화하였다. PDB 파일에서는 6개의 α-헬릭스가 표시되어 있고, 이에 따라 그 림에 표시하였다. 이중 하나는 한 개의 아미노산이 한 개의 a-헬릭스로 표기되어 있다. 연속된 아미노산 backbone에 위 치한 수소결합에 참여하는 산소-질소 간의 거리가 3.2A 이 상일 경우에는 산소와 질소 간에 검은 선분으로 표시를 하 였다. Fig. 3C에서는 7개 이상의 아미노산으로 구성된 a-헬 릭스에 대해 양 끝 부분에서 나선 구조의 local frame들을 계산한 뒤, 참조점들을 연결한 선분을 축으로 원통을 그린 예제를 보인다. Fig. 3D에서는 재구성 알고리즘을 적용하여 a -헬릭스를 재구성한 결과를 보인다. 오른쪽의 최장 길이의 첫 번째 a-헬릭스가 3개의 나선구조로 분할되었다. Fig. 3E 는 분할 된 a-헬릭스에 대해 각각 원통을 추가하여 분할된 상태를 표시하였다.

분할된 세 개의 헬릭스를 Ha, Hb, Hc라 할 때, 헬릭스 의 분할이 실제로 단백질의 구조상에서 의미를 가지는 지 를 살펴보면 다음과 같다. 헬릭스 Ha는 이 분자의 두 번 째 헬릭스와 다른 분자의 두 헬릭스와 함께 four-helix bundle 구조에 참여하고 그 다음의 Hb와 Hc 부분은 다른 분자의 헬릭스와 다시 상호작용에 참여하고 있다. 또한 Hb와 Hc로 나뉘는 지점은 다른 분자의 헬릭스와 접촉하 는 지점이다.

기존의 Prokink와 같은 프로그램을 사용할 경우에는, 긴 첫 번째 헬릭스에 proline이 없어 꺾이는 지점을 선정하기가 어렵 지만 제안된 알고리즘을 이용하여 헬릭스의 분할이 가능하 였다. 또한 첫 번째 헬릭스를 가시화하였을 때 두드러지지 않는 두 번째 휘는 지점 (Hb와 Hc 사이)의 선정은 이 단백 질의 구조와 기능의 분석에 도움을 준다. 이는 선정된 지점 이 이 단백질이 dimer를 형성할 때 상대 헬릭스와의 특징적 인 교차가 있는 지점으로[12] 이 헬릭스의 다른 헬릭스와의 결합 기능을 제대로 평가할 수 있기 때문이다.

PDB id. 3GIG 이외에도 긴 a-헬릭스를 갖는 여러 개의 단백질에 대해 실험을 수행하였고, 지면이 제한된 관계로 이에 대한 실험결과는 http://bh.knu.ac.kr/~kujinkim/helix_sample.zip 로 제공한다.

4.결론

본 논문에서는 단백질 분자 내의 a-헬릭스를 분할하는 방법을 제시하였다. PDB id. 3GIG의 한 폴리펩티드 사슬의 첫 번째 헬릭스는 길이가 매우 긴 휘어진 형태로 PDB 파일 에서는 하나의 헬릭스로 표시되었지만 본 논문에서 제시한 알고리즘을 적용하여 세 개의 헬릭스로 분할할 수 있었다. 또한, 이외에 다수의 다른 단백질에 대해서 헬릭스 분할 결 과를 웹페이지로 제시하였다.

제안된 알고리즘을 이용하여 헬릭스가 분할되는 지점을 선정하고, 이를 이용하여 분할된 헬릭스를 단위로 단백질의 구조와 기능을 분석할 수 있었다. 향후에는 길게 휘어진 헬 릭스들을 거리 정보를 이용하여 분할할 뿐 아니라 휘어진 형태를 분류할 수 있는 기하학적인 정보를 추출하는 방법을 연구할 계획이다.

Reference

- D. Barlow and J. M. Thornton. "Helix geometry in proteins," J. Mol. Biol., Vol.201, pp.601–619, 1988.
- [2] S. Kumar S. and M. Bansal, "Geometrical and sequence characteristics of a-helices in globular proteins," Biophysical Journal, Vol.75, pp.1935–1944, 1998.
- [3] T. L. Blundell, D. Barlow, N. Barkakoti, and J. M. Thornton., "Solvent induces distortions and curvature of α-helices," Nature, Vol.306, pp.281-283, 1983.
- [4] P. Chakarabarti, M. Bernard, and D. C. Rees. "Peptide bond distortion and curvature of a helices," Biopolymers, Vol.25, pp.108–1093, 1986.
- [5] S. Kumar, and M. Bansal. "Structure and sequences characteristics of long a helices in globular proteins," Biophys. J., Vol.71, pp.1574–1586, 1996.
- [6] C. Chothia, M. Levitt, and D. Richardson, "Helix to helix packing in proteins," J. Mol. Biol., Vol.145, pp.215–250, 1981.
- [7] S. A. Tatulian, "Determination of helix orientations in proteins," Comp. Biol. Chem., Vol.32, pp.370–374, 2008.
- [8] B. Visiers, B. Braunheim, H. Weinstein, "Prokink: a protocol for numerical evaluation of helix distortions by proline," Protein Engineering, Vol.13, No.9, pp.603–606, 2000.
- [9] M. Bansal, S. Kumar, and R. Velavan, "HELANAL: A program to characterize helix geometry in protein," J. Biomol. Struc. Dynamics, Vol.17, pp.811–819, 2000.

- [10] J. Deville, J. Rey, and M. Chabbert, "Comprehensive analysis of the helix-X-helix motif in soluble proteins," Proteins: structure, function and bioinformatics, Vol.72, pp.115–135, 2008.
- [11] Y. H. Huang and C. M. Chen. "Statistical analyses and computational prediction of helical kinks in membrane proteins," J Compt Aided Mol Des, Vol.26, pp.1171–1185, 2012.
- [12] D. Albanesi, M. Martin, F. Trajtenberg, M. C. Mansilla, A. Haous, P. M. Alzari, D. de Mendoza, A. and Buschiazzo. "Structural plasticity and catalysis regulation of a thermosensor histidine kinase", Proc. Natl. Acad. Sci. USA, Vol.106, pp.16185–16190, 2009.



강 범 식

e-mail:bskang2@knu.ac.kr 1988년 연세대학교 생물학과(학사) 1997년 연세대학교 생물학과(박사) 2004년~현 재 경북대학교 생명과학부 교수 관심분야:단백질의 구조와 기능



김 구 진

e-mail:kujinkim@gmail.com 1992년 이화여자대학교 전자계산학과(학사) 1992년 한국과학기술원 전자계산학과(석사) 1998년 포항공과대학교 컴퓨터공학과(박사) 1998년~2000년 미국 Purdue University (PostDoc.) 2000년~2002년 아주대학교 정보통신전문

대학원 BK21조교수

2002년~2003년 Dept. of Mathematics and Computer Science, University of Missouri-St. Louis, Visiting Assistant Professor

현 재 경북대학교 컴퓨터학부 정교수

관심분야:계산생물학, 컴퓨터 그래픽스, 곡면 및 기하모델링, 병렬처리 등 168 정보처리학회논문지/소프트웨어 및 데이터 공학 제3권 제4호(2014.4)



e-mail:udseo@knu.ac.kr 2014년 경북대학교 컴퓨터학부(학사) 2014년~현 재 경북대학교 컴퓨터학부 석사과정 관심분야:컴퓨터 그래픽스, 계산생물학

