

Three Dimensional Visualization of Contact Region for a Protein Complex

Beom Sik Kang[†] · Ku-Jin Kim^{**} · Yukyeong Kim^{***}

ABSTRACT

In this paper, we present a method to visualize the contact region between two molecules in a protein complex in a three dimensional space. The contact region of two molecules shows compatibility in geometric aspects. Usually, the computation of the area of contact region has been used to show the strength of compatibility. The numerical value and simple drawing of contact region would be useful for comparing the relative strength of different contacts, but it is not appropriate for analysing the geometric characteristics of the contact region. In this paper, we present a method to show the compatibility between two molecules by visualizing the distance information between them.

Keywords : Protein Molecule, Protein Complex, Contact Region, Contact Area, Compatibility, Visualization, Depth Buffer

단백질 복합체를 위한 접촉 영역의 3차원 가시화

강 범 식[†] · 김 구 진^{**} · 김 유 경^{***}

요 약

본 논문에서는 두 개의 단백질 분자가 결합하여 복합체를 구성할 때, 두 분자 간의 접촉 영역을 발견하여 3차원 공간에서 가시화하는 방법을 제시한다. 두 분자 간의 접촉 영역은 기하학적인 측면에서 서로 상보성을 보이며, 상보적인 결합의 크기를 나타내기 위한 방법으로 접촉 영역의 면적을 계산하는 방법이 주로 사용되어 왔다. 접촉 영역의 면적을 수치화한 결과와 접촉 영역의 단순 표시는 서로 다른 접촉 영역에 대해 상대적인 결합 강도를 비교하기에는 적합하지만, 접촉 영역이 가진 기하학적인 특성을 분석하기에는 부적합하다. 본 논문에서는 접촉 영역에서 상보성을 표시하기 위해 상대 분자와의 거리 정보를 가시화하는 방법을 제시한다.

키워드 : 단백질 분자, 단백질 복합체, 접촉 영역, 접촉 면적, 상보성, 가시화, 깊이 버퍼

1. 서 론

단백질 분자 간의 도킹 (docking) 및 바인딩 (binding) 은 단백질의 기능 분석 및 새로운 단백질 복합체의 합성 또는 신약 개발 등의 분야에서 매우 중요한 역할을 하며, 이때 분자 간의 접촉 영역에 대한 발견 및 특성의 분석은 매우 중요하다[1-8]. 수많은 연구자들이 단백질 분자에서 상호작용하는 접촉 영역을 발견하거나, 접촉 영역의 면적을 구하기 위한 연구를 수행하여 왔다.

Gong 등 [9]은 PSIMAP 알고리즘 [10-12] 이 제안하는 5-5 rule을 이용하여 두 단백질 분자가 서로 상호작용을 한

다고 판단하였다. 5-5 rule은 최소한 5개의 잔기 쌍이 상호 간 5Å 이내에 존재할 때 두 분자가 서로 상호작용을 한다고 결정하는 방법이다.

Nye 등 [13]은 다른 분자와의 거리가 5Å 이내인 원자의 개수를 이용하여 접촉 면적을 계산하였다. Steinkellner 등 [14]은 결정 구조에서 단백질 분자 간의 contact patch를 발견하고, 서로 다른 곡면의 성질에 따라 patch들을 annotating하였으며, RNA, DNA, 또는 리간드 (ligand)와 접촉하는 단백질의 곡면 patch를 분석하였다.

Arab 등 [15]은 잔기(residue)에 속한 원자가 다른 분자와 접하는 구 표면적을 접촉 면적으로 정의하였다. 이들은 원자에 대응하는 구 표면의 표본 점 (sample point)에 프로브 (probe)들을 위치시키고 다른 원자와 겹치는 점의 개수를 이용하여 접촉 면적을 계산하였다.

단백질 분자 간의 결합에서 두 단백질의 결합 강도는 세포 내에서 혹은 용액 안에서 실제로 그 단백질의 기능이 수행될 수 있는 가능성을 말해 주는 중요한 요소이다. 단백질에 있어 파트너 단백질과의 결합의 강도는 접촉 영역에서 상호작용을

* 본 논문은 2012학년도 경북대학교 학술 연구비에 의하여 연구되었으며 2013년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (과제번호: NRF-2013R1A1A2A10004391).

† 정 회 원: 경북대학교 생명과학부 교수

** 정 회 원: 경북대학교 컴퓨터학부 교수

*** 준 회 원: 경북대학교 대학원 컴퓨터학부 석사과정

논문접수: 2013년 8월 19일

수정일: 1차 2013년 10월 16일

심사완료: 2013년 10월 17일

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

하는 원자들 사이의 물리화학적 결합 특성과 더불어 얼마나 많은 원자가 상호작용에 참여하고 있는가를 나타내는 접촉영역의 면적에 일반적으로 비례한다. 그러나, 결합 강도는 접촉영역의 단순한 면적뿐만 아니라 형태와 단백질에서의 위치에 의해 영향을 받는다. 따라서 접촉 영역이 단백질에서 어디에 분포하는지에 대한 시각적인 표시는 수치화된 면적이 제시할 수 없는 정보를 연구자에게 제공할 수 있다.

기존의 관련 연구에서는 접촉 영역의 면적과 더불어 분자간 접촉영역의 형태적인 특징을 단백질의 입체구조 상에서 표현하려고 하였다. Fig. 1에서는 두 단백질 분자의 복합체 구조(PDB id. 2HSN)의 예를 보이며, 그림에서 붉은 색과 노란 색이 각각 서로 다른 분자를 나타낸다. 두 분자의 접촉 영역을 가시화하기 위해 기존에 주로 사용된 방법은 Fig. 2와 같이 한 개의 분자는 구의 집합 또는 분자 곡면으로 표현하고, 다른 분자는 원자 중심점 간의 연결선을 나타내는 리본과 구성 아미노산의 본드(bond)를 stick의 집합으로 표현하는 방법 등이 있다.

두 단백질의 결합 영역이 비슷한 접촉 면적을 갖고 전체적인 형태가 유사할지라도 접촉 영역에서 상대 분자와의 거리에 따라 결합 강도가 달라질 수 있다. 상대 분자와의 거리는 두 분자 간의 요철에 대한 정보를 제공하여 두 분자간의 초기 결합 가능성과 결합력을 암시해 줄 수 있다. 단백질 분자 표면의 요철을 표시하기 위해서는 상대 단백질 분자와의 거리 측정이 필요하고, 특히 두 분자 간의 결합 방향에서의 거리가 중요하다. 두 단백질 간의 결합은 이들 분자의 용액 내에서의 운동과정에서 충돌에 의하여 발생하는 것이고, 접촉면에 서로 일치하게 충돌이 일어나야 한다. 따라서 두 분자가 서로 마주 보는 방향을 결정하고, 이 방향을 따라 분자 간의 거리 정보를 계산하여야 결합 방향에서의 요철의 정도를 보일 수 있다.

기존에 PyMOL (<http://www.pymol.org>)과 같은 단백질의 구조를 가시화하는 프로그램은 다수 존재하여, 이들 프로그램들은 단백질의 표면을 가시화한다. PISA (<http://www.ebi.ac.uk/pdbe>)와 같은 단백질 간의 상호결합을 분석하는 프로그램들은 접촉 면적과 참여 아미노산 등을 가시화한다. 실제 단백질과 단백질의 결합에 대한 연구에는 두 가지 종류의 프로그램에서 얻는 정보가 모두 필요하며, VASCo (<http://genome.tugraz.at/VASCo>)와 같이 접촉면을 계산하여 단백질 표면에 표시해주는 도구가 개발되고 있으나, 결합 파트너 단백질과의 상보적인 거리를 표시해주는 프로그램은 저자들이 아는 바로는 현재까지 개발되지 않았다.

본 연구에서는 복합체를 구성하는 단백질 분자 간의 접촉 영역에서 상대 단백질 분자까지의 거리에 따라 색상을 달리하여 3차원 분자 모델 위에 투영하여 가시화하는 방법을 제시한다. 제시된 방법은 그래픽 프로세서에서 제공하는 깊이 버퍼(depth buffer) 기능을 사용함으로써 효율적인 계산이 가능하며, 분자 간 상보적인 형태를 3차원적으로 가시화하는 독창적인 기법을 제안하였다. 본 논문의 구성은 다음과

같다. 2절에서는 접촉 거리 가시화 알고리즘을 제시한다. 3절에서는 실험 결과를 보인다. 4절에서 결론을 내린다.

2. 접촉 거리 가시화 알고리즘

단백질 분자를 기하학적인 측면에서 구성할 때, 분자 내의 각 원자를 반데르발스(van der Waals) 반경을 갖는 구로 표현하는 방법이 범용적으로 사용된다. 주어진 단백질 분자 복합체에 포함된 두 개의 분자는 각각 반경이 서로 다른 구의 집합으로 표현된다. 각 분자 M_k ($k = 0, 1$) 가 n_k 개의 원자로 구성될 때, M_k 를 아래와 같이 정의할 수 있다.

$$M_k = \{A_{k,i}(\mathbf{c}, r) \mid 0 \leq i < n_k\},$$

여기에서 $A_{k,i}(\mathbf{c}, r)$ 는 중심점 \mathbf{c} , 반경 r 를 갖는 구(sphere)를 나타내며, 한 개의 원자에 대응된다.

분자 M_0 에 속한 한 개의 원자 A_0 와 다른 분자 M_1 간의 거리는 주어진 방향 d 에 대해 정의된다. 필요한 경우 M_0 와 M_1 에 대해 회전변환을 적용함으로써, d 가 z 축의 방향과 일치한다고 가정한다. 방향 d 는 graphic program 내에서 viewing direction으로 사용된다.

Fig. 3은 방향 d 를 viewing direction으로 적용한 결과를 제시한다. 그림에서 가로 방향, 세로 방향이 각각 x , y 축에 대응되며, z 축은 x 축과 y 축 방향 벡터에 대한 외적 벡터(cross vector) 방향에 대응된다.

가시화 알고리즘은 다음과 같이 요약된다.

단계 1. 두 개의 분자 M_0, M_1 를 각각 drawing 할 윈도우 W_0, W_1 를 설정하고, 이들의 width, height가 각각 w, h 값을 갖도록 한다. 각 윈도우의 깊이 버퍼 값을 저장하기 위하여 float type의 2차원 배열 $d_0[w][h], d_1[w][h]$ 를 생성한다.

단계 2. 결합 방향 d 에 대해 M_0 의 깊이 버퍼 값을 계산하여 $d_0[i][j]$ ($0 \leq i < w, 0 \leq j < h$)에 저장한다. $d_0[i][j]$ 는 최소 0부터 최대 1 사이의 값을 가지며, 주어진 view volume 내에서 시점으로부터 object의 거리값이 작을수록 0에 가까운 값을, 클수록 1에 가까운 값을 갖는다. Fig. 4A에서는 M_0 의 깊이 버퍼 정보를 표시한다. 이때 분자와 배경 부분을 구분하기 위해 분자가 그려지는 부분은 푸른 색상에 명암을 주어 표시하였고, 분자가 그려지는 부분에서 깊이 버퍼 값이 클수록(즉 시점에서 거리가 멀수록) 더 밝은 색상으로 표시를 하였다.

단계 3. 결합 방향의 반대 방향인 $-d$ 방향으로 M_1 의 깊이 버퍼 값을 계산하여 $d_1[i][j]$ 에 저장한 뒤, $d_1[i][j] = 1 - d_1[i][j]$ 를 수행한다. 이때 $d_1[i][j]$ 에 저장된 값은 M_1 을 $+d$ 방향으로 깊이값을 구할 때, 시점으로부터 가장 멀리 위치한 분자표면의 깊이 값을 저장하게 되며, Fig. 4B에서 깊이 버퍼를 가시화한 결과를 보인다.

단계 4. Depth map d_0, d_1 간의 차(difference)는 d 방향에 대해 각 픽셀 별로 ray를 발사하여 구한 M_0 와 M_1 사이의 거리에 대응하는 값이다. 이 값을 $diff[w][h]$ 에 저장한다.

$$diff[i][j] = d_1[i][j] - d_0[i][j]$$

단계 5. $diff[i][j]$ 를 주어진 threshold 및 양자화 단계에 따라 양자화한다(Fig. 5).

단계 6. 양자화된 $diff[i][j]$ 에 따라, $d_0[i][j]$ 와 $d_1[i][j]$ 값을 이용하여 3차원 M_0, M_1 에서 대응되는 point 위치를 계산하여 서로 다른 색상으로 point를 표시한다. 두 분자가 서로 마주 보는 면을 view direction d 에 수직인 면으로 설정하여, 이 면을 기준으로 mirror 상태로 두 분자를 그린다(Fig. 6A, 6B).

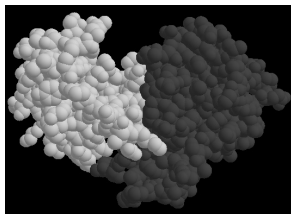


Fig. 1. Two proteins in a crystal (PDB id. 2HSN)

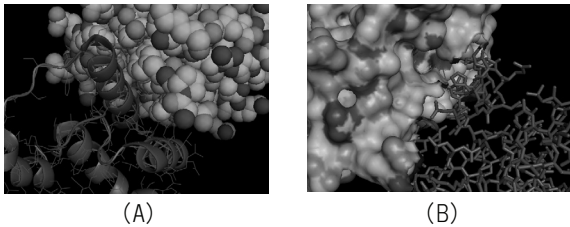


Fig. 2. Examples of visualizing the compatibility between two protein molecules

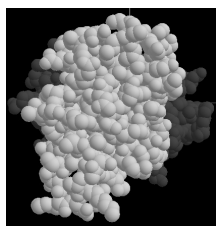


Fig. 3. Two molecules in 2HSN with a given viewing direction

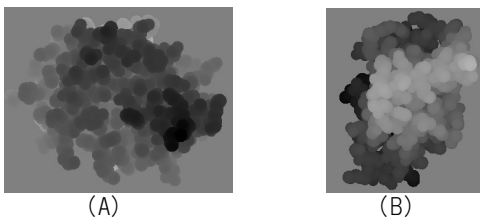


Fig. 4. Colored depth buffer image for M_0 and M_1

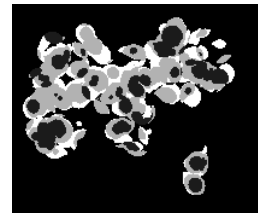


Fig. 5. Quantized distance map with different colors according to the distance

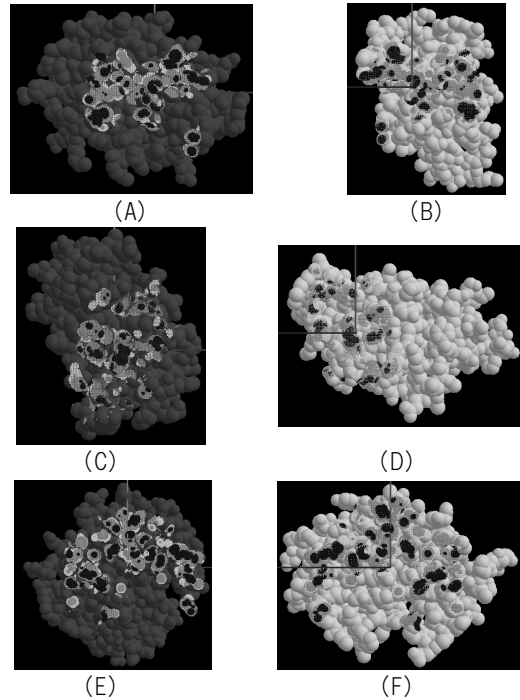


Fig. 6. Protein complexes (PDB id.: 2HSN, 2UZ8, 4BJW from top to bottom), where two proteins in a complex are shown in mirror image with 3-dimensional distance information

3. 실험 결과

실험은 nVidia GTX590 GPU가 장착된 컴퓨터에서 실행하였다. 제시된 알고리즘은 Window 환경에서 Visual C++와 OpenGL을 이용하여 구현되었다.

Fig. 6은 세 개의 PDB 파일 2HSN (각 분자의 원자개수 왼쪽: 1,239개, 오른쪽: 964개), 2UZ8 (각 분자의 원자 개수 왼쪽: 1,275개, 오른쪽: 1,281개), 4BJW (각 분자의 원자 개수 왼쪽: 1,636개, 오른쪽: 1,385개)에 제안된 알고리즘을 적용하여 상대 분자가 거리 5Å 이내에 있는 영역을 표시한 결과를 보인다. 각 단백질 복합체는 두 개의 분자로 구성되어 있으며, 두 분자를 mirror image로 제시하였다. 거리 영역은 3단계로 구분되어, 상대 분자가 1Å 이내인 경우 파란색, 1Å 부터 3Å 사이인 경우 시안(cyan)색, 3Å 부터 5Å 이내인 경우 흰색으로 표시되었다.

Fig. 6A와 6B의 경우는 다른 경우와 비교하여 접촉 영역이 밀집되어 있다는 특징을 볼 수 있다. Fig. 6C, 6D의 경우

는 분자의 전체 영역에 비해 접촉 영역의 면적이 좁다고 분석되며, Fig. 6E, 6F는 다른 분자들에 비해 접촉 영역이 넓게 분산되어 있다는 특징을 갖는다.

4. 결 론

본 논문에서는 단백질 복합체를 구성하는 두 개의 단백질 분자에 대해 접촉 영역을 발견하고 두 분자 간의 거리를 3차원 공간에서 가시화하는 방법을 제안하였다. 제안된 방법은 그래픽 프로세서에서 제공하는 깊이 버퍼 기능을 이용하여 효율적으로 수행된다. 두 분자가 서로 마주 보는 면을 설정하여 이 면을 기준으로 두 분자를 mirror 상태로 놓고 가시화하는 독창적인 방법을 제안하였다. 가시화된 결과는 두 단백질 간의 거리 정보를 보임으로써, 결합 방향에서의 오차의 정도를 분석하기 위해 유용하게 사용된다.

참 고 문 헌

[1] D. Levine, M. Facello, P. Hallstrom, G. Reeder, B. Walenz, F. Stevens, "Stalk: an Interactive System for Virtual Molecular Docking," IEEE Computational Science and Engineering, Vol.4, No.2, pp. 55-65, 1997.

[2] T. J. A. Ewing, S. Makino, A. G. Skillman, I. D. Kuntz, "DOCK4.0: Search Strategies for Automated Molecular Docking of Flexible Molecule Databases," Journal of Computer Aided Molecular Design, Vol.15, No.5, pp.411-428, 2001.

[3] H. Nagata, H. Mizushima, H. Tanaka, "Concept and Prototype of Protein Ligand Docking Simulator with Force Feedback Technology," Bioinformatics, Vol.18, No.1, pp.140-146, 2002.

[4] R. D. Taylor, P. J. Jewsbury, J. W. Essex, "A Review of Protein Small Molecule Docking Methods," Journal of Computer Aided Molecular Design, Vol.16, No.3, pp.151-166, 2002.

[5] C. M. Venkatachalam, X. Jiang, T. Oldfield, M. Waldman, "LigandFit: a Novel Method for the Shape Directed Rapid Docking of Ligands to Protein Active Sites," Journal of Molecular Graphics and Modelling, Vol.21, No.4, pp.289-307, 2003.

[6] S. K. Lai Yuen, Y. -S. Lee, "Interactive Computer Aided Design for Molecular Docking and Assembly," Computer Aided Design and Applications, Vol.3, No.6, pp.701-709, 2006.

[7] Y. Zhao, M. F. Sanner, "Protein Ligand Docking with Multiple Flexible Side Chains," Journal of Computer Aided Molecular Design, Vol.22, No.9, pp.673-679, 2008.

[8] O. Trott, A. J. Olson, "AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading," Journal of Computational Chemistry, Vol.31, No.2, pp.455-461, 2010.

[9] S. Gong, G. Yoon, I. Jang, D. Bolser, P. Dafas, M. Schroeder, H. Choi, Y. Cho, K. Han, S. Lee, "PSIbase: a database of protein structural interactome map (PSIMAP)," Bioinformatics, Vol.21, No.10, pp.2541-2543, 2005.

[10] J. Park, M. Lappe, S. Teichmann, "Mapping Protein Family Interactions: Intramolecular and Intermolecular Protein

Family Interaction Repertoires in the PDB and Yeast," J. Mol. Biol., Vol.307, pp.929-938, 2001.

[11] M. Lappe, J. Park, O. Niggemann, L. Holm, "Generating protein interaction maps from incomplete data: application to Fold assignment," Bioinformatics, pp.149-156, 2001.

[12] K. S. Han, B. K. Park, H. G. Kim, J. S. Hong, J. Park, "HPID: The Human Protein Interaction Database," Bioinformatics, Vol.20, pp.2466-2470, 2004.

[13] T. M. W. Nye, C. Berzuini, W. R. Gilks, M. M. Babu, S. Teichmann, "Predicting the strongest domain domain contact in interacting protein pairs," Statistical Applications in Genetics and Molecular Biology, Vol.5, No.1, 2006.

[14] G. Steinkellner, R. Rader, G. G. Thallinger, C. Kratky, K. Gruber, "VASCO: computation and visualization of annotated protein surface contacts," BMC Bioinformatics, Vol.10, No.32, 2009.

[15] S. Arab, M. Sadeghi, C. Eslahchi, H. Pezeshk, A. Sheari, "A pairwise residue contact area based mean force potential for discrimination of native protein structure," BMC Bioinformatics, Vol.11, No.16, 2010.



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
 1990년 이화여자대학교 전자계산학과(학사)
 1992년 한국과학기술원 전자계산학과(석사)
 1998년 포항공과대학교 컴퓨터공학과(박사)
 1998년~2000년 미국 Purdue University
 (PostDoc.)

2000년~2002년 아주대학교 정보통신전문대학원 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



김 유 경

e-mail : morph84@naver.com
 2010년 계명대학교 공과대학 컴퓨터
 공학과(학사)
 2012년~현 재 경북대학교 대학원
 컴퓨터학부 석사과정
 관심분야: 컴퓨터 그래픽스, 계산생물학