

# 단백질 가시화 형태에 따른 정보표현적합도 평가<sup>☆</sup>

## Evaluation of Information Representation Goodness-of-fit According to Protein Visualization Pattern

변재희<sup>1</sup>      최유주<sup>1</sup>      서정근\*  
Jaehye Byeon      Yoo-Joo Choi      Jung-Keun Suh

### 요약

단백질 기능을 규명하는 단백질 구조 정보는 단백질 의약품의 약효를 증진시키고, 개발을 단축시키는데 큰 영향을 미친다. 따라서 단백질의 구조를 효과적으로 분석하기 위한 단백질 가시화에 대한 연구가 증가하고 있다. 하지만 단백질 가시화에 대한 연구가 단백질의 구조를 예측하거나 렌더링의 속도를 향상시키는 것을 중심으로 이뤄지고 있으며, 단백질 가시화 형태에 따른 정보 전달 효율성에 대한 연구는 미비한 실정이다. 본 연구는 단백질 의약품에 대한 효율적인 정보 서비스의 사전 연구로써 단백질 1, 2차구조 혼합가시화 형태별 정보표현적합도를 분석하였다. 단백질 1, 2차구조 혼합가시화 형태는 대표적 가시화 서비스인 Chimera, PDB, Cn3D와 기존 가시화 서비스의 문제점을 개선한 단백질 1, 2차구조 혼합가시화 형태를 대상으로 하였다. 정보표현적합도를 구하기 위한 정보요인은 피험자 분석 결과를 바탕으로 단백질 1차구조, 아미노산 위치, 단백질 2차구조, 단백질 2차구조 비율정보로 구분하였으며, 피험자는 단백질 의약품 업계종사기간이 5년 이상인 전문가 집단을 대상으로 하였다. 그 결과 단백질 1, 2차구조 혼합가시화형태별 정보표현적합도에는 유의미한 차이가 있었으며, 가시화 형태별 정보 전달 효율성에 차이가 있음을 입증할 수 있었다.

☞ 주제어 : 바이오인포매틱스, 단백질 가시화, 정보표현

### ABSTRACT

The information about protein structure gives the clues for the function of protein. It is needed for the improvement for the efficacy and fast development of protein drugs. So, the studies visualizing the structure of protein effectively increase. Most studies of visualization focus on the structural prediction for protein or the improvement on the rendering speed. However, studies of information delivery depending on the form of protein visualization are very limited. The major objective of this study is to analyze the information representation goodness-of-fit for the patterns of the hybrid visualization with primary and secondary structures of protein. Those hybrid visualizations included the patterns which updated current representative visualization services, Chimera, PDB and Cn3D. Information factor to analyze information representation goodness-of-fit is assorted by protein primary structure, secondary protein structure, the location of amino acid and ratio information about protein secondary structure, based on the result of subject-analysis. Subject is the group of experts who are involved in protein drug development over 5 years. The result of this study shows the meaningful difference in the information representation goodness-of-fit by the patterns of hybrid visualization and proves the difference in the information by the pattern of visualization.

☞ keyword : Bioinformatics, Protein Visualization, Information Representation

## 1. 서론

단백질은 생체 내에서 특정한 3차원 구조 형태를 가지며, 생명현상 유지에 중요한 기능을 한다. 따라서 단백질의 구조를 밝히는 것은 단백질의 기능 및 작동 원리를 규명하는데 매우 중요한 정보이며, 단백질 의약품 개발을

가속화하는데 많은 영향을 끼친다[1-2].

단백질의 기능을 규명하기 위해 단백질의 구성 요소를 텍스트나 그래픽 등을 이용하여 구조적으로 가시화하는 것에 대한 요구가 증가하고 있다. 대표적인 단백질 가시화 서비스로는 바이오데이터베이스 센터인 NCBI(The National Center for Biotechnology Information), PDB (Protein Data Bank) 등에서 자체적으로 가시화 서비스를 제공하는 것과 Chimera, Jmol 등의 단백질 가시화 서비스가 있다. 최근 단백질 가시화 서비스는 단백질 구조 가시화 외에 수많은 단백질 정보 중 사용자가 원하는 정보만을 제공하거나 분석, 측정 등이 가능하도록 기능이 확장되고 있으며, 각 서비스 별로 특화된 기능을 제공한다[3-4].

<sup>1</sup> Dept. of Stereoscopic Media, Korean German Institute of Technology, Seoul, 157-930, Korea.

\* Corresponding author (jksuh@kgit.ac.kr)

[Received 10 February 2015, Reviewed 11 February 2015, Accepted 20 March 2015]

☆ 본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2014년도 산업계 맞춤형 인력양성 지원사업의 연구결과로 수행되었음

단백질 의약품 개발 시 단백질-단백질 상호작용 정보는 단백질의 효과를 예측하는데 필수적인 정보로 사용된다. 따라서 단백질 가시화에 대한 연구는 단백질-단백질 상호작용 추론이나 렌더링 속도를 향상하는 연구를 중심으로 이뤄지고 있다[4-5].

이에 반해 가시화 형태에 따른 정보 전달의 효율성 및 적합성에 관한 연구는 미비한 실정이다. 따라서 효율적인 단백질 구조 분석을 위해서는 단백질 가시화 형태에 따른 정보표현적합도 평가에 대한 연구가 필요하다.

본 논문에서는 효율적인 단백질 의약품 전용 서비스를 위한 사전 연구로써 대표적인 단백질 가시화 서비스의 가시화 유형을 비교 분석하였으며, 이를 기반으로 단백질 1, 2차구조 정보 제공을 위한 효율적인 단백질 가시화 형태를 제안하였다. 제안하는 단백질 가시화 형태의 정보표현적합도를 검증하기 위해 기존의 대표적인 단백질 가시화 유형의 정보표현적합도를 사용자 평가 방법으로 비교분석하였다. 평가 결과를 바탕으로 본 논문에서 제시한 단백질 가시화 형태가 사용자 입장에서 유용한 형태임을 입증하였다.

## 2. 단백질 가시화 서비스

단백질 가시화 서비스는 유의미한 데이터를 수집하고 분석할 수 있도록 하는 분석도구 기술과 단백질의 구조를 함축적으로 표시하는 그래픽 기술로 구성된다[6]. 단백질 정보는 구조에 따라 1차~4차로 나눌 수 있으며, 그 중에서도 단백질 3차구조는 단백질의 기능을 결정하는 역할을 한다. 복잡한 단백질 3차구조에서 단백질의 기능을 규명하기 위해서는 단백질 1, 2차구조의 정보가 필수적이다.

Chimera는 대표적 가시화 툴로써 다양한 단백질 분석 기능을 제공하는 장점이 있으며, 단백질에 대한 부가적인 정보를 제공하기 위해 점차 기능을 확장하고 있다[6].

PDB는 대표적인 단백질 데이터베이스 센터로 미국과 유럽, 일본이 협력하여 단백질 정보를 수집, 분석, 관리하고 있다[9]. PDB에서는 오픈소스인 Jmol을 이용해 단백질 3차구조 가시화 서비스를 하고 있으며, 단백질 1, 2차구조 가시화에 대해서는 웹을 통해 자체적으로 서비스하고 있다[10].

NCBI에서는 Cn3D 가시화 툴을 서비스하고 있다. Cn3D는 NCBI의 데이터 서비스와 연계하여 단백질 1, 2차구조 서비스 및 단백질 3차구조 가시화 서비스를 제공하고 있다[8-9].

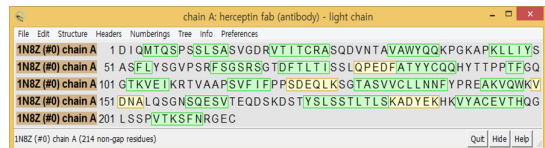
## 3. 단백질 1,2차구조 혼합 가시화

### 3.1 단백질 1, 2차구조 혼합가시화 형태 분석

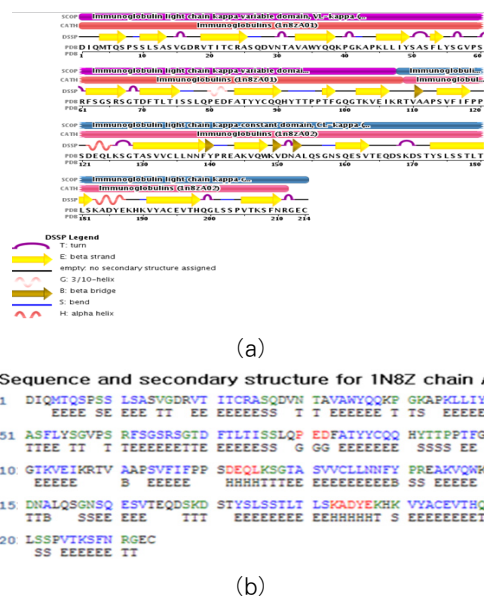
단백질 구조 가시화 서비스를 제공하는 Chimera, PDB, Cn3D는 단백질의 기능 분석을 위해 단백질 1~3차구조 정보를 제공하고 있으며, 단백질 1, 2차구조 정보는 혼합 가시화 형태로 제공하고 있다.

단백질 1, 2차구조 혼합가시화 형태를 분석하기 위해 Chimera, PDB, Cn3D에서 PDB ID 1N8Z를 검색하고 가시화 결과를 분석하였다. 그 결과 가시화 서비스에서 제공하는 단백질 정보요인에는 차이가 있었으며, 가시화 형태에도 큰 차이가 있었다[9].

Chimera는 (그림 1)과 같이 단백질 1, 2차구조 혼합가시화 서비스를 제공하였다.



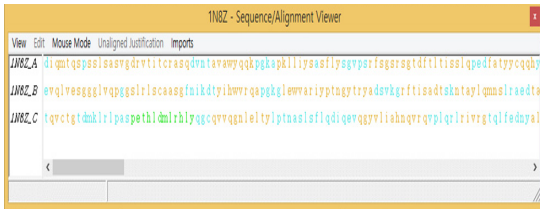
(그림 1) Chimera의 단백질 1, 2차구조 혼합가시화 형태 (Figure 1) Hybrid visualization of primary and secondary structures in Chimera



(그림 2) PDB의 단백질 1, 2차구조 혼합가시화 형태. (a) 혼합 가시화 형태, (b) 2차 구조 가시화 형태 (Figure 2) Hybrid visualization of primary and secondary structures in PDB

PDB는 단백질 1, 2차구조 혼합 가시화를 (그림 2)와 같이 두 가지 형태로 제공하였다. 두 형태 모두 텍스트와 그래픽 요소를 적용하여 가시화 하는 것에는 공통점이 있었으나 제공하는 단백질 정보요인에는 차이가 있었다[9].

Cn3D는 (그림 3)과 같이 단백질 1, 2차구조 혼합가시화 서비스를 제공하였다.



(그림 3) Cn3D의 단백질 1, 2차구조 혼합가시화 형태 (Figure 3) Color code for protein secondary structure

각 가시화 형태별 정보요인과 문제점을 분석한 결과는 (표 1)과 같다.

(표 1) 단백질 1, 2차구조 혼합가시화 형태 분석 (Table 1) Hybrid visualization pattern analysis for primary and secondary structures

	Chimera	PDB(a)	PDB(b)	Cn3D
정보요인	<ul style="list-style-type: none"> <li>단백질 1차구조</li> <li>아미노산 위치</li> <li>단백질 2차구조</li> </ul>	<ul style="list-style-type: none"> <li>단백질 1차구조</li> <li>아미노산 위치</li> <li>단백질 2차구조</li> <li>단백질 2차구조 비율정보</li> </ul>	<ul style="list-style-type: none"> <li>단백질 1차구조</li> <li>아미노산 위치</li> <li>단백질 2차구조</li> </ul>	<ul style="list-style-type: none"> <li>단백질 1차구조</li> <li>단백질 2차구조</li> </ul>
문제점	<ul style="list-style-type: none"> <li>단백질 2차구조 정의 정보가 없음</li> <li>단백질 2차구조의 비율 정보가 없음</li> <li>단백질 2차구조의 일부 정보만 제공함</li> <li>아미노산 서열의 시작 번호만 표시하여 서열이 길수록 서열의 번호를 알기 어려움</li> </ul>	<ul style="list-style-type: none"> <li>아미노산 서열과 맵핑되는 2차구조의 위치를 잘못 파악할 수 있음</li> <li>단백질 2차구조 중 일부에 대해서만 비율정보 제공</li> </ul>	<ul style="list-style-type: none"> <li>단백질 2차구조 정의 정보가 없음</li> <li>단백질 2차구조 비율 정보가 없음</li> <li>아미노산 서열의 시작 번호만 표시하여 서열이 길수록 서열의 번호를 알기 어려움</li> </ul>	<ul style="list-style-type: none"> <li>단백질 2차구조 정의 정보가 없음</li> <li>단백질 2차구조 비율 정보가 없음</li> <li>단백질 2차구조의 일부 정보만 제공함</li> <li>서열 번호를 표시하지 않아 아미노산의 위치를 알기 어려움</li> </ul>

### 3.2 단백질 1, 2차구조 혼합가시화 형태 제안

Chimera, PDB, Cn3D의 단백질 1, 2차구조 혼합가시화 형태의 문제점을 바탕으로 개선된 정보요인을 적용한 가시화 형태를 제안한다.

먼저 네 개의 가시화 형태 모두 단백질 2차구조에 적용하는 색상에 차이가 있어 (그림 4)와 같이 단백질 2차구조에 대한 표준화된 색상코드를 생성하였다. 표준화된 색상코드의 생성 기준은 다음과 같다[9].

1. Chimera, PDB, Cn3D의 사용 색상 중 중복되는 색상 선택
2. 중복되는 색상이 없을 경우 기준 색상 중 주목성이 높은 색 선택
3. 색상 중 주목성이 낮은 색상은 다른 색상으로 대체

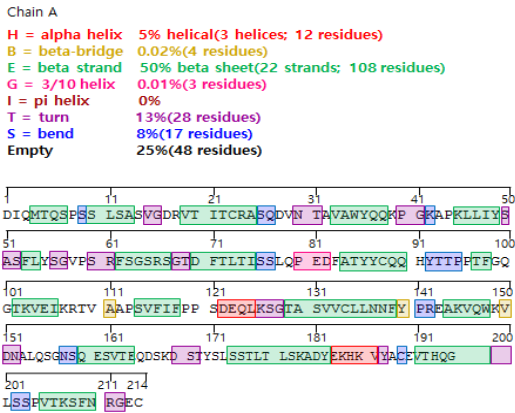
구분	PDB(mol)	Cn3D	Chimera	제안하는 색상
H = alpha helix	빨강	녹색		빨강
B = beta-bridge	황금			황금
E = beta strand	노랑	황금	연한 녹색	초록
G = 3/10 helix	다홍			분홍
I = pi helix	진한 빨강			진한 빨강
T = turn	진한 자주		연한 노랑	진한 자주
S = bend	파랑			파랑
Empty	검정	바다		검정

(그림 4) 단백질 2차구조에 대한 색상코드 (Figure 4) Color Code for Protein Secondary Structure

생성된 단백질 2차구조 색상 코드는 제안하는 가시화 형태에 적용하여 단백질 1차구조와 2차구조 간 정보의 혼동이 없도록 하였다.

Chimera의 단백질 1, 2차구조 혼합가시화 형태를 기반으로 제안하는 가시화 형태는 (그림 5)와 같으며, 개선점은 다음과 같다.

- 표준화 된 단백질 2차구조 색상코드 적용
- 단백질 2차구조 정의 정보 표시
- 단백질 2차구조의 비율 정보 표시
- 10 단위로 아미노산 서열 번호 표시

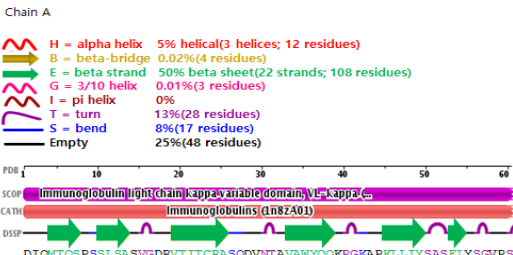


(그림 5) Chimera 기반 단백질 1, 2차구조 혼합가시화 형태 제안

(Figure 5) The modified hybrid visualization of primary and secondary structures based on Chimera pattern

PDB(a)의 단백질 1, 2차구조 혼합가시화 형태를 기반으로 제안하는 가시화 형태는 (그림 6)과 같으며, 개선점은 다음과 같다.

- 표준화 된 단백질 2차구조 색상코드 적용
- 단백질 2차구조의 모든 비율 정보 표시
- 단백질 2차구조와 맵핑되는 아미노산 서열에 색상 부여



(그림 6) PDB(a) 기반 단백질 1, 2차구조 혼합가시화 형태 제안

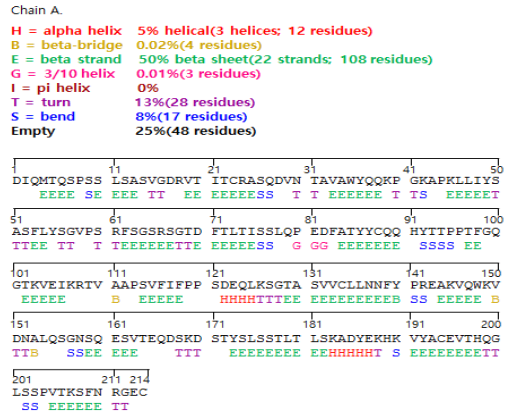
(Figure 6) The modified hybrid visualization of primary and secondary structures based on PDB(a) pattern

PDB(b)의 단백질 1, 2차구조 혼합가시화 형태를 기반으로 제안하는 가시화 형태는 (그림 7)과 같으며, 개선점은 다음과 같다.

- 표준화 된 단백질 2차구조 색상코드 적용
- 단백질 2차구조 정의 정보 표시
- 단백질 2차구조의 비율 정보 표시
- 10 단위로 아미노산 서열 번호 표시

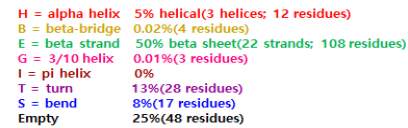
Cn3D의 단백질 1, 2차구조 혼합가시화 형태를 기반으로 제안하는 가시화 형태는 (그림 8)과 같으며, 개선점은 다음과 같다.

- 표준화 된 단백질 2차구조 색상코드 적용
- 단백질 2차구조 정의 정보 표시
- 단백질 2차구조의 비율 정보 표시
- 10 단위로 아미노산 서열 번호 표시
- 단백질 2차구조의 모든 정보를 제공함



(그림 7) PDB(b) 기반 단백질 1, 2차구조 혼합가시화 형태 제안

(Figure 7) The modified hybrid visualization of primary and secondary structures based on PDB(b) pattern



(그림 8) Cn3D 기반 단백질 1, 2차구조 혼합가시화 형태 제안

(Figure 8) The modified hybrid visualization of primary and secondary structures based on Cn3D pattern

각 가시화 형태별 개선점을 바탕으로 개선된 정보요인은 (표 2)와 같다.

(표 2) 제안하는 단백질 1, 2차구조 혼합가시화 형태 정보요인  
(Table 2) Additional information conveyed by the modified hybrid visualization of primary and secondary structures

	Chimera	PDB(a)	PDB(b)	Cn3D
기존 정보요인	<ul style="list-style-type: none"> <li>단백질 1차구조</li> <li>아미노산 위치</li> <li>단백질 2차구조</li> </ul>	<ul style="list-style-type: none"> <li>단백질 1차구조</li> <li>아미노산 위치</li> <li>단백질 2차구조</li> <li>단백질 2차구조 비율정보</li> </ul>	<ul style="list-style-type: none"> <li>단백질 1차구조</li> <li>아미노산 위치</li> <li>단백질 2차구조</li> </ul>	<ul style="list-style-type: none"> <li>단백질 1차구조</li> <li>단백질 2차구조</li> </ul>
개선된 정보요인	<ul style="list-style-type: none"> <li>단백질 2차구조 정의</li> <li>단백질 2차구조 비율 정보</li> <li>아미노산 위치(10단위)</li> <li>단백질 2차구조 (모든 정보)</li> </ul>	<ul style="list-style-type: none"> <li>단백질 2차구조 비율 정보(모든 정보)</li> </ul>	<ul style="list-style-type: none"> <li>단백질 2차구조 정의</li> <li>단백질 2차구조 비율 정보</li> <li>아미노산 위치 (10단위)</li> </ul>	<ul style="list-style-type: none"> <li>단백질 2차구조 정의</li> <li>단백질 2차구조 비율 정보</li> <li>아미노산 위치(10단위)</li> <li>단백질 2차구조 (모든 정보)</li> </ul>

### 3. 단백질 1, 2차구조 혼합가시화 형태에 따른 정보표현적합도 실험 설계

#### 3.1 연구 목표와 가설 설정

각 가시화 형태별 정보표현적합도를 구하기 위한 정보요인은 단백질 1차구조, 아미노산 위치, 단백질 2차구조, 단백질 2차구조 비율정보 이상 4가지로 구분하였다. 각 정보요인 별 정보표현적합도를 평가하기 위한 세부 연구 가설은 다음과 같다.

가설 1. 각 가시화 형태별 단백질 1차구조 정보표현적합도에는 차이가 있을 것이다.

가설 2. 각 가시화 형태별 아미노산 위치 정보표현적합도에는 차이가 있을 것이다.

가설 3. 각 가시화 형태별 단백질 2차구조 정보표현적합도에는 차이가 있을 것이다.

가설 4. 각 가시화 형태별 단백질 2차구조 비율 정보표현적합도에는 차이가 있을 것이다.

#### 3.2 연구대상 및 자료 수집

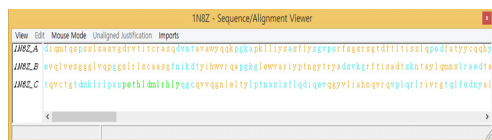
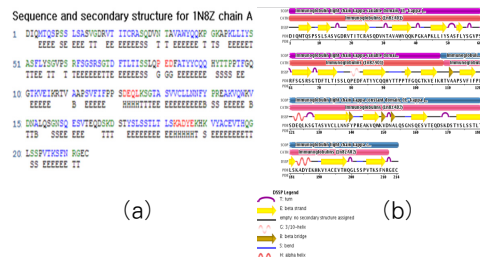
본 논문에서는 단백질 가시화 형태별 정보표현적합도 조사를 위해 해당 분야에 종사하고 있는 전문가 12명을 대상으로 설문을 진행하였다. 해당 분야 종사자 모두 석사학위 이상이며, 업계종사기간이 5년 이상이고, 단백질 가시화 서비스를 일주일에 1~2회 이상 사용하였다. 수집된 설문지 중 불성실한 응답이 없어 수집된 12부 모두 설문 분석에 사용하였다.

#### 3.3 조사방법 및 체계

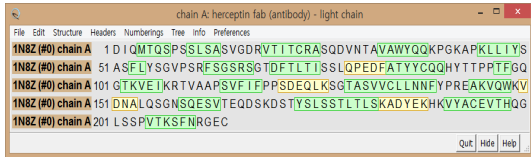
설문 항목 내용은 크게 2가지로 개인정보, 단백질 1, 2차구조 혼합가시화 형태별 정보표현 적합도에 대해 설문하였다. 개인정보는 피험자의 학력과 업계종사기간, 단백질 의약품 개발 시 필요한 단백질 정보 항목으로 구성하였으며, 명목척도로 정보를 수집하였다.

단백질 1, 2차구조 혼합가시화 형태별 정보표현적합도는 리커트 5점 척도로 하여 필드 별 적합도를 수집하였다. 단백질 1, 2차구조 혼합가시화 형태 별 정보표현적합도 설문을 위한 가시화 형태는 (그림 9)와 같다.

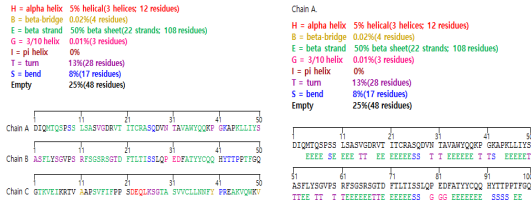
각 가시화 형태별 단백질 정보표현적합도에 차이가 있는지 검증하기 위하여 각 단백질 정보 필드를 독립변수로 하여 One-Way Repeated Measures ANOVA 분석을 하였다. Mauchly의 구형성 검정을 만족할 경우 Within-Subjects Effect의 Sphericity Assumed을, 만족하지 못할 경우 Greenhouse-Geisser 값을 참조하였다.



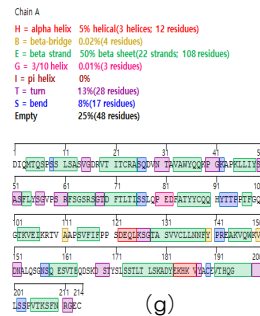
(c)



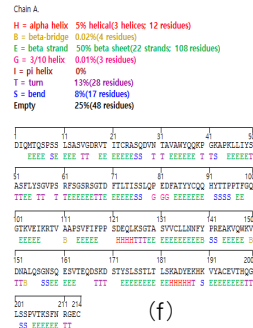
(d)



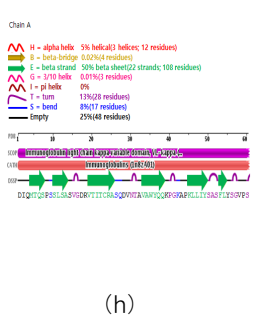
(e)



(g)



(f)



(h)

(그림 9) 정보표현적합도 설문을 위한 가시화 형태 (Figure 9) Protein information visualization patterns for information representation goodness-of-fit test

### 4. 단백질 의약품 정보 서비스를 위한 설문 결과 분석

#### 4.1 피험자 분석

설문에 응답한 피험자의 개인정보의 분포는 (표 3)과 같으며 12명 중 9명이 박사졸업 및 10년이상 업계에 종사하여 전문가 집단으로 분류할 수 있다.

(표 3) 피험자 분석 (Table 3) Subject of Survey

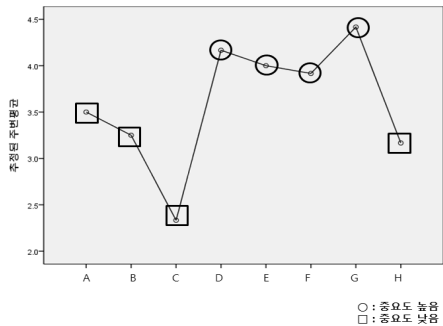
		서비스 이용횟수			
		1~2회	3~4회	4회이상	기타
학력	대학원졸(석사)	-	2	1	-
	대학원졸(박사)	1	2	6	-
업계 종사 기간	5년이상~10년미만	-	2	1	-
	10년이상	1	2	6	-

단백질 의약품 개발 시 반드시 필요한 단백질 정보에 대해서는 11명이 단백질 1차구조, 단백질 2차구조, 단백질 2차구조 비율정보라고 응답하였으며, 아미노산 위치 정보에 대해서는 3명이 응답하였다.

#### 4.2 단백질 1차구조에 대한 정보표현적합도분석

단백질 1차구조에 대한 가시화 형태 별 정보표현적합도를 분석한 결과 (Mauchly's  $W=0.034, df=27, p>.05$ ) 가설 1이 채택되었다.

단백질 1차구조에 대한 Pairwise Comparisons 결과 D, E, F, G의 정보표현적합도가 상대적으로 높음을 확인하였다 (그림 10).



(그림 10) 단백질 1차구조의 추정된 주변 평균 (Figure 10) Estimated marginal means of protein primary structure

#### 4.3 아미노산 위치에 대한 정보표현적합도분석

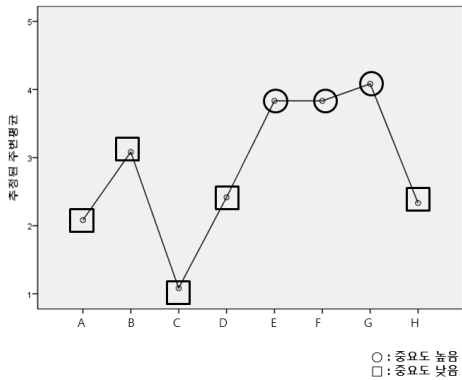
아미노산 위치에 대한 가시화 형태 별 정보표현적합도를 분석한 결과 ( $F(7,77)=42.109, p<.05$ ) 가시화 형태 별 정보표현적합도에 유의한 차이가 있었으므로, 가설 2가 채택되었다 (표 4).

(표 4) 아미노산 위치에 대한 가시화 형태 별 정보표현적합도 분석 결과

(Table 4) Analysis result of information representation goodness-of-fit for amino acid sequence

Source	Mauchly's test of sphericity (p-value)	Within subjects effect (df)	Within subject effects (mean square)	Within subjects effect (F-value)	Within subjects effect (p-value)
단백질 1 2차구조 혼합 가시화 형태	.000	7	13.153	42.109	.000
Error (단백질 1 2차구조 혼합 가시화 형태)		77	.312		

아미노산 위치에 대한 Pairwise Comparisons 결과 E, F, G의 정보표현적합도가 상대적으로 높음을 확인하였다 (그림 11).



(그림 11) 아미노산 위치의 추정된 주변 평균

(Figure 11) Estimated marginal means of amino acid sequence

#### 4.4 단백질 2차구조에 대한 정보표현적합도분석

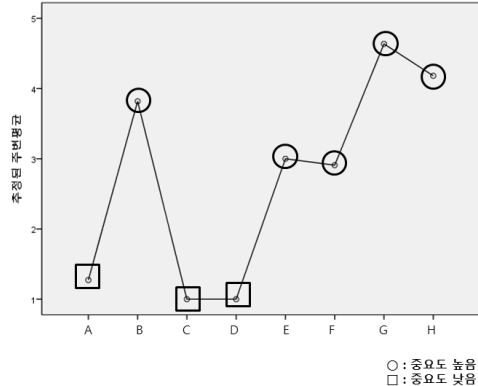
단백질 2차구조에 대한 가시화 형태 별 정보표현적합도를 분석한 결과(F(7,70)=66.078, p<.05) 가설 3이 채택되었다 (표 5).

(표 5) 단백질 2차구조에 대한 가시화 형태 별 정보표현적합도 분석 결과

(Table 5) Analysis result of information representation goodness-of-fit for protein secondary structure

Source	Mauchly's test of sphericity (p-value)	Within subjects effect (df)	Within subject effects (mean square)	Within subjects effect (F-value)	Within subjects effect (p-value)
단백질 1,2차 혼합 가시화 형태	.000	7	23.792	66.078	.000
Error (단백질 1,2차 혼합 가시화 형태)		70	.360		

단백질 2차구조에 대한 Pairwise Comparisons 결과 B, E, F, G, H의 정보표현적합도가 상대적으로 높음을 확인하였다 (그림 12).



(그림 12) 단백질 2차구조의 추정된 주변 평균

(Figure 12) Estimated marginal means of protein secondary structure

#### 4.5 단백질 2차구조 비율에 대한 정보표현적합도 분석

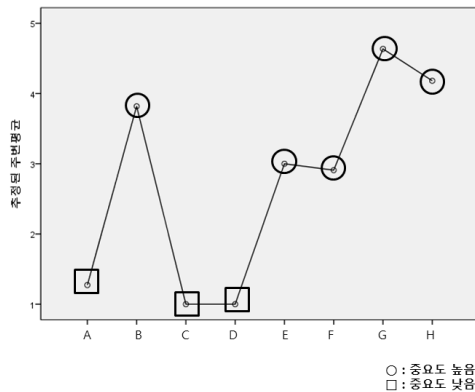
단백질 2차구조 비율에 대한 가시화 형태 별 정보표현적합도 분석 결과(F(7,70)=260.040, p<.05) 가설 4가 채택되었다 (표 6).

(표 6) 단백질 2차구조 비율에 대한 가시화 형태 별 정보표현적합도 분석 결과

(Table 6) Analysis result of information representation goodness-of-fit for protein secondary structure rate

Source	Mauchly's test of sphericity (p-value)	Within subjects effect (df)	Within subject effects (mean square)	Within subjects effect (F-value)	Within subjects effect (p-value)
단백질 1,2차 혼합 가시화 형태	.000	7	31.830	260.040	.000
Error (단백질 1,2차 혼합 가시화 형태)		70	.122		

단백질 2차구조 비율에 대한 Pairwise Comparisons 결과 B, E, F, G, H의 정보표현적합도가 상대적으로 높음을 확인하였다 (그림 13).

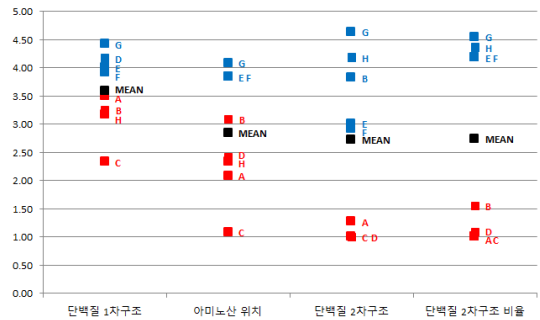


(그림 13) 단백질 2차구조 비율의 추정된 주변 평균 (Figure 13) Estimated marginal means of protein secondary structure rate

#### 4.6 결과

단백질 1, 2차구조 혼합가시화 형태 별 정보표현적합도의 설문 조사 결과 PDB(a)의 가시화 형태인 B와 Chimera 가시화 형태인 D, 본 연구에서 제안한 가시화 형태 중 E, F, G가 상대적으로 적합하다는 것을 확인하였다 (그림 14).

Chimera, PDB(b), Cn3D와 비교하여 PDB(a)는 단백질 2차구조 비율정보요인을 제공하는데 적합한 가시화 형태였다. 따라서 PDB(a)의 가시화 형태를 기반으로 제안한 H가 단백질 2차구조 비율정보에 대한 정보표현적합도에서 다른 가시화 형태와 비교했을 때 상대적으로 높은 것을 확인할 수 있었다[9].



(그림 14) 가시화 형태 별 정보표현적합도 평균 비교 (Figure 14) Comparison of information representation goodness-of-fit according to visualization pattern

## 5. 결론

본 연구에서는 단백질 가시화 서비스를 제공하는 Chimera, PDB, Cn3D의 가시화 1, 2차구조 혼합가시화 형태를 비교 분석하였다. 각 가시화 형태 별 문제점을 바탕으로 각 가시화 형태 기반의 단백질 1, 2차 혼합가시화를 제안하고 해당 분야 전문가를 대상으로 가시화 형태별 정보표현적합도를 평가하였다.

정보표현적합도를 평가하기 위한 정보요인은 단백질 의약품 개발 시 반드시 필요한 단백질 정보에 대한 피험자의 응답 결과를 바탕으로 단백질 1차구조, 아미노산 위치, 단백질 2차구조, 단백질 2차구조 비율정보로 구분하였다.

정보요인에 대한 단백질 1, 2차구조 혼합가시화 형태별 정보표현적합도 평가 결과 정보표현적합도에는 유의미한 차이가 있었으며 가시화 형태별 정보 전달 효용성에 차이가 있음을 입증할 수 있었다.

본 연구는 단백질 의약품 서비스를 위한 사전연구 진행되었으며, 본 연구의 결과는 단백질 의약품 정보제공을 효율적으로 제공하기 위한 서비스 개발에 활용될 수 있을 것이라 기대된다.



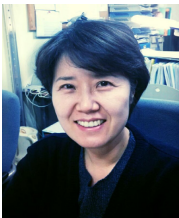
## 참 고 문 헌 (Reference)

- [1] H. S. Choi, "Visualization of protein domain interaction interfaces," KAIST, Doctoral thesis, 2005.
- [2] C. Y. Park, "Flexible Protein Structure Alignment and Visualization," Graduate School Chungnam National University, Doctoral thesis, 2007.
- [3] J. H. Byeon, Y. J. Choi, J. H. Lee, and J. K. Suh, "Evaluation of Usefulness of the Protein Drug Feature Information Filed," Journal of Korean Society for Internet Information, Vol. 15, No. 4, 2014, pp. 21-31. <http://dx.doi.org/10.7472/jksii.2014.15.4.21>
- [4] K. E. Song, "Visualization for Protein-Protein Interaction Data," KGIT, Master's thesis, 2013.
- [5] J. K. Seong, N. Baek and K. J. Kim, "Real-time approximation of molecular interaction interfaces based on hierarchical space decomposition," Computer-Aided Design, Vol. 43, Issue. 12, 2011, pp. 1598-1605. <http://dx.doi.org/10.1016/j.cad.2011.07.001>
- [6] Chimera, <https://www.cgl.ucsf.edu/chimera/>
- [7] Jmol, <http://jmol.sourceforge.net/>
- [8] Cn3D, <http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml>
- [9] J. H. Byeon, "A Study on the Website Design for Efficient Delivery of Protein Drug Information," KGIT, Master's thesis, 2014.
- [10] PDB, <http://www.pdb.org>

## ● 저 자 소 개 ●

**변 재 희 (Jaehee Byeon)**

2010년 덕성여자대학교 컴퓨터시스템학과(공학사)  
 2012년 호서대학교 벤처전문대학원 IT응용기술학과 졸업(공학석사)  
 2015년 한독미디어대학원대학교 입체영상미디어학과 졸업(미디어학석사)  
 현재 바이오엔시스시스템즈주식회사 연구원  
 관심분야 : 바이오인포매틱스, 컴퓨터그래픽스, etc.  
 E-mail : bjach9188@gmail.com

**최 유 주 (Yoo-Joo Choi)**

1989년 이화여자대학교 전자계산학과(이학사)  
 1991년 이화여자대학교 일반대학원 전자계산학과(이학석사)  
 2005년 이화여자대학교 과학기술대학원 컴퓨터공학학과(공학박사)  
 1991년~1993년 한국컴퓨터주식회사 기술연구소 주임연구원  
 1994년~1999년 포스데이타주식회사 기술연구소 주임연구원  
 2005년~2010년 서울벤처대학원대학교 컴퓨터응용기술학과 조교수  
 2010년~현재 한독미디어대학원대학교 뉴미디어콘텐츠학과 부교수  
 관심분야 : 컴퓨터그래픽스, 가상현실, HCI, 컴퓨터비전, etc.  
 E-mail : yjchoi@kgit.ac.kr

**서 정 근 (Jung-Keun Suh)**

1987년 서울대학교 식물학과(이학사)  
 1989년 서울대학교 대학원 식물학과(이학석사)  
 1996년 University of Texas at Austin Department of Biochemistry(이학박사)  
 2000년~2007년 LG생명과학 기술연구원 부장  
 2007년~2009년 한독산학협동단지 BT연구센터 센터장  
 2009년~현재 한독미디어대학원대학교 뉴미디어콘텐츠학과 교수  
 관심분야 : 데이터베이스, 바이오인포매틱스, etc.  
 E-mail : jksuh@kgit.ac.kr