



특집/Bioinformatics 연례간 미려

분야별 Bioinformatics 소개

Proteomics에서 계산화학의 역할

강 영 기 | E-mail: ykkang@cbacc.chungbuk.ac.kr
충북대학교 화학과 교수

		1. 서 론
	II. 계산화학과 생물정보학	
	III. 맺가키 예	
	IV. 결 론	

생물정보학은 유전체의 발견과 기능 및 신약설계의 개발 등에 관련된 정보들을 찾아 분석하고 관리하여 유용한 정보를 찾아내자는 학문이다. 특히 3-D 구조에 근거한 신약과 치료법의 개발은 생물정보학에서 아주 중요한 부분이다.

I. 서 론

얼마 전 인간유전체사업(Human Genome Project)의 결과가 발표된 직후, 미국화학회에서 주간으로 발간하는 "Chemical & Engineering News"의 Madeleine Jacobs 편집장은 다음과 같은 글을 썼다: "인간유전체를 해석한 것은 단지 생명이란 책의 제1장에 불과한 것이고, 나머지 장에서는 단백질에 대하여 쓰여질 것이다."^{1]}

인체유전체사업이 완결되어감에 따라서 많은 연구자들은 이제 DNA 서열에서 얻어진 정보를 인간의 질병 치료와 치료제 개발에 유용하게 사용하고자 시도하고 있다. 그런데 유전체에서 발견되고 전사된 단백질의 종류가 많고, 이러한 단백질-단백질 상호작용이 유전체보다는 훨씬 다양하다. 그러므로 여러 조건에서 여러 단백질들을 분리하고 정리하여, 특정 단백질-단백질의 상호작용을 밝혀서 그 생리활성 기능을 규명하려는 새로운 연구분야가 탄생한 것이 'proteomics'다.^{2]}

Proteomics는 단지 개개의 단백질을 연구하는 것보다 동시에 많은 단백질을 함께 다루어야 하기 때문에, 단백질의 3-D 구조를 효율적으로 예측할 수 있는 이론 및 알고리즘의 개발과 단백질들에 대한 많은 정보를

검색하고 그 특성을 규명할 수 있는 생물정보학(bioinformatics)의 여러 지식과 기술을 필요로 한다. 이 층에서는 계산화학적 지식과 연구방법이 단백질의 구조와 기능을 이해하는 데 어떻게 사용될 수 있는지에 대하여 몇 가지 예를 통하여 살펴보고자 한다.

II. 계산화학과 생물정보학

계산화학(computational chemistry)이란 화학분야에서 발생하는 여러 가지 현상인 분자의 구조와 성질, 분자상호 간의 반응 등을 물리화학과 전산학적인 지식과 기술을 사용하여 규명하는 학문인데, 최근 30여 년 전부터는 생명과학과 약의학과 관련되어 많은 연구가 진행되어 왔다. 그리고 무엇보다도 컴퓨터의 현저한 발전이 이 분야의 연구 수준을 급상승시키고 있고, 특히 최근들어 신약개발 분야에서는 계산화학의 유용성이 보다 더 확인되고 있다.

생물정보학은 유전체의 발견과 기능 및 신약설계의 개발 등에 관련된 정보들을 찾아 분석하고 관리하여 유용한 정보를 찾아내자는 학문이다.^{3]} 특히 3-D 구조에 근거한 신약과 치료법의 개발은 생물정보학에서 아주 중요한 부분이다. 그러므로 계산화학적 연구방법을

사용하여 생체분자 중에서도 특히 단백질이나 모형폴타이드에 대하여 구조분석(conformational analysis) 연구를 수행하는 것은 생물정보학에서 주요한 역할을 할 것으로 기대된다.

단백질과 펩타이드의 구조분석 연구에는 크게 두가지 계산방법이 사용되는데, 첫째는 양자화학적 계산방법인 분자궤도함수법(molecular orbital 또는 MO method)이고, 둘째는 고전역학적인 함수를 사용하는 힘장법(force field method)이다.^[4] 분자궤도함수법에는 크게 ab initio와 semiempirical MO 방법이 있는데, 비교적 큰 분자에 대하여는 semiempirical 방법이 사용되고 비교적 작은 분자에 대하여 ab initio 방법이 사용되어 왔으나, 컴퓨터와 관련 S/W의 발전으로 현재 비교적 큰 모형 생체분자의 구조분석 연구와 신약설계에도 그 결과가 더 정확한 ab initio 방법이 점점 더 널리 사용되고 있다.

힘장계산에는 여러 함수 형태의 고전역학적인 함수가 사용되고, 각 함수의 변수들은 실 값이나 고급 수준에서 계산된 ab initio 결과들을 이용하여 유도한다.^[5]

그러나 각 함수의 변수들은 단백질이나 펩타이드의

모형분자들을 이용하여 유도하기 때문에 사용한 모형분자의 종류나 물리학적 성질에 따라서 서로 다른 힘장이 얻어진다. 그러므로 단백질이나 펩타이드의 구조분석을 수행할 때 사용하는 힘장에 대한 각별한 주의와 검증이 필요하다.

힘장법은 구조분석을 수행하려는 단백질이나 펩타이드의 결합길이와 결합각을 고정시키거나 또는 이완시키거나에 따라서 'fixed geometry' 힘장과 'rigid geometry' 힘장으로 크게 분류할 수 있는데, 전자에는 ECEPP과 OPLS 등의 힘장이 있고, 후자에는

힘장법은 구조분석을 수행하려는 단백질이나 펩타이드의 결합길이와 결합각을 고정시키거나 또는 이완시키거나에 따라서 'fixed geometry' 힘장과 'rigid geometry' 힘장으로 크게 분류할 수 있는데, 전자에는 ECEPP과 OPLS 등의 힘장이 있고, 후자에는 CHARMM과 AMBER 등의 힘장이 있다.^[6]

CHARMM

과 AMBER 등의

힘장이 있다.^[4] 특히 fixed

geometry 힘장은 rigid geometry 힘장에 비하여 구조에너지를 계산하는 데 기억용량과 계산속도를 훨씬 줄일 수 있기 때문에 단백질-단백질 상호작용에 대한 구조분석 연구에 유용하다.

III. 몇 가지 예

계산화학적인 연구방법이 단백질이나

펩타이드의 구조분석에 어떻게 사용될 수 있는지를 다음의 예로 설명하고자 한다.

1. Proline의 펩타이드 Cis-Trans 이성질화

Proline(Pro)는 곁가지(side chain)가 골격(backbone) 아미이드의 질소원자와 직접 공유결합되어 있는 유일한 필수 아미노산으로서, Pro의 backbone은 수소결합을 할 수 없다. 그리고 Pro는 오각형 고리화합물이기 때문에 구조가 비교적 정직되고 고리의 일부 원자가 평균평면에서 벗어난 구조를 가진다.



분야별 Bioinformatics 소개

이러한 계산결과는 분자 내 수소결합이 단백질에서 Pro의 알 펩타이드 결합의 trans-cis 이성질화를 촉진시키는 요소가 될 수 있다는 최근의 화학반응속도와 분광학적 실험결과를 입증해주고 있다. 또한 본인의 양자화학적 계산결과는 용매의 극성이 증가할수록 trans 구조에 대한 cis 구조의 상대적인 안정도가 증가하고, trans-to-cis 이성질화에 대한 자유에너지 장벽의 높이가 증가한다는 실험적인 사실을 확인하고 있다.

최근 들어서 Pro의 바로 앞에 위치하는 펩타이드 결합의 cis-trans 이성질화가 여러 단백질의 접힘(folding)을 결정하는 중요한 요소인 것이 밝혀졌고, 이에 관여하는 효소들에 대한 연구가 활발히 진행되고 있으며, 특히 몇몇 효소는 번역역제계의 결합단백질인 것이 알려져서 세포 신호전달과 cis-trans 이성질화의 관련에 대한 연구가 진행되어 오고 있다.

이에 본인은 ab initio 계산법인 HF, MP2와

B3LYP 수준에서 6-31+G* basis set을 사용하여 모형펩타이드인 Ac-Pro-NHMe에 대한 구조분석 연구를 수행하였다.⁶⁾ HF/6-31+G* 방법으로 최적화된 Ac-Pro-NHMe의 구조를 (그림 1)에 도시하였다. 그림에서도 알 수 있듯이, Pro의 알 펩타이드 결합이 trans와 cis인 구조는 서로 비슷한 정도로 오각형 고리의 구조가 유사하나 전이상태의 구조는 특히 질소 원자의 결합상태가 현저히 다른 것을 알 수 있고, 특

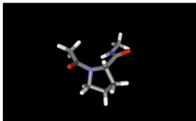
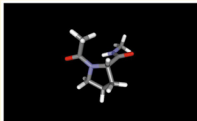
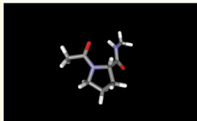


그림 1 양자화학적 계산방법인 HF/6-31+G* 수준에서 최적화된 Ac-Pro-NHMe의 구조들(상단 왼쪽: trans 구조, 상단 오른쪽: cis 구조, 하단: 전이상태 구조)

히 전이상태에서 Pro의 질소원자는 바로 다음에 위치하는 펩타이드의 수소원자와 분자 내 수소결합을 형성할 수 있다.

이러한 계산결과는 분자 내 수소결합이 단백질에서 Pro의 알 펩타이드 결합의 trans-cis 이성질화를 촉진시키는 요소가 될 수 있다는 최근의 화학반응속도와 분광학적 실험결과를 입증해주고 있다.

또한 본인의 양자화학적인 계산결과는 용매의 극성이 증가할수록 trans 구조에 대한 cis 구조의 상대적인 안정도가 증가하고, trans-to-cis 이성질화에 대한 자유에너지 장벽의 높이가 증가된다는 실험적인 사실을 확인하고 있다. 그러므로 고급 수준의 양자화학적인 계산방법은 실험으로 확인할 수 없는 단백질의 이

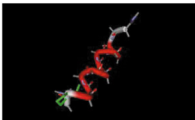
러 가지 흥미로운 구조적 특성과 물성을 자세히 확인할 수 있게 해준다.

2. 단백질의 α -Helix에서 Proline의 구조적 역할

Pro는 바로 앞절에서도 설명하였듯이, 그 구조적인 특성으로 인하여 수소결합에 의한 단백질의 3-D 구조를 형성하는 중요한 요소 중의 하나인 helix의 형성을 방해하는 요인이라는 것이 오래 전부터 잘 알려져 왔다.

그럼에도 불구하고 Pro는 다른 아미노산에 비하여 helix가 시작되는 첫 번째 위치에서 가장 큰 분포율을 가지는 것이 단백질의 X-ray 구조분석으로 알려졌다.

이에 본인은 힘장 계산법 중에서 그 역사가 가장 오래된 ECEPP 힘장과 본인의 수화각이론(hydration



□그림 2) Proline이 단백질 α -helix의 중간에 위치하는 경우(왼쪽)와 첫 번째에 위치하는 경우(오른쪽)의 힘장과 수화각이론을 사용하여 수화상태에서 최적화된 모형 단백질의 3-D 구조들

shell model)을 사용하여 모형단백질인 Ac-(Ala) $_n$ -Pro-(Ala) $_m$ -NHMe(이때 $n + m = 14$)에 대하여 구조분석 연구를 수행하였다.²⁷ 수화상태에서 최적화된 Pro가 helix의 중간에 위치하는 경우와 첫 번째에 위치하는 경우의 3-D 구조를 (그림 2)에 도시하였다. 그림에서 쉽게 알 수 있듯이, Pro는 helix의 중간에서 수소결합의 형성을 방해하여 소위 'kinked helix'를 형성하지만, helix가 시작되는 처음의 위치에서는 α -helix의 형성을 방해하지 않는다.

그리고 구조에너지의 각 성분을 비교분석하여, Pro 위의 정전기적 에너지와 비결합 에너지가 각 모형단백질의 구조를 결정하는 가장 중요한 요인임을 확인할 수 있었고, 수화는 오히려 'kinked helix'인 경우에 더 유리한 것을 증명화할 수 있었다. 이와 관련하여 RNA의 가수분해 효소 중의 하나인 RNase T1의 첫 번째 helix에서 Ala21을 다른 아미노산으로 치환했을 때, helix 구조와 자유에너지 변화를 연구 수행중이다.



proteomics는 단지 개개의 단백질을 연구하는 것보다 동시에 많은 단백질들을 함께 다루어야 하기 때문에 단백질의 3-D 구조를 효율적으로 예측할 수 있는 이론 및 알고리즘의 개발과 단백질들에 대한 많은 정보를 검색하고 그 특성을 규명할 수 있는 생물정보학의 여러 지식과 기술을 필요로 한다.

3. Arg-Gly-Asp를 포함하는 펩타이드의 구조와 활성

Arg-Gly-Asp (RGD)는 integrin이라고 불리는 여러 세포의 단백질의 인식 아미노산 서열이라는 것이 잘 알려져 왔다. 특히 RGD를 포함하는 단백질이나 펩타이드는 혈소관의 당단백질인 GPIIb/IIIa에 결합하여 혈액의 응고를 저해하거나, v3에 결합하여 중앙의 전이를 저해하는 것으로 알려져서 이에 대한 연구가 활발히 진행되어 왔다.

그러나 단백질이나 펩타이드의 3-D 구조에 근거하여 그 활성기구를 규명하려는 연구는 아직까지 제한되어 있다.

이에 본인은 ECEPP 힘장과 수화각이론을 사용하여 모형 펩타이드인 NH₂-Arg-Gly-Asp-Xaa-OH(또는 RGD_X, 이때 Xaa = Phe(F), Trp(W), Val(V), Cys(C), Gln(Q), Ser(S), Tyr(Y)과 Leu(L))에 대하여 구조분석 연구를 수행하였다.³⁾ 여러 실험으로 얻어진 혈액응고 저해활성은 RGDF > RGDV > RGDC > RGDQ > RGDS > RGDY > RGD_L 순이고, 어떠한 turn 구조를 형성하는 선호도가 활성을 좌우할 수 있는 요소인지를 파악하였다(여기에서 turn 구조는 펩타이드 서열 중에서 i번째의 C=O와 (i+3)번째의 H-N 사이에 수소결합을 형성



그림 3) 수화상태에서 구조최적화를 최적화하여 얻은 혈액응고 저해활성이 가장 큰 펩타이드인 H-Arg-Gly-Asp-Phe-OH의 3-D구조

하여 C_αi와 C_αi+3 원자들의 거리가 7Å 이하인 2차구조를 의미한다).

이것은 최근에 수행한 본인의 Ac-RGD-NHMe에 대한 구조분석 연구결과³⁾와 관련된 펩타이드의 분광학적인 실험결과를 기초한 것이다.


(그림 3)에 수화상태에서 최적화하여 얻은 저해활성이 가장 큰 펩타이드 RGDF의 최소 자유에너지 구조를 도시하였다.

계산결과로부터 혈액응고 저해활성이 큰 펩타이드일수록 Gly-Asp와 Asp-Xaa 위치에서 V와 IV 형태의 turn 구조를 잘 형성하고, 동일한 turn 구조를 형성하더라도 Xaa 아미노산 곁가지의 크기와 그 배향이 활성을 가지게 하는 데 중요할 것이라는 것을 알았다.

IV. 결 론

앞서 언급하였듯이, proteomics는 단지 개개의 단백질을 연구하는 것보다 동시에 많은 단백질들을 함께 다루어야 하기 때문에 단백질의 3-D 구조를 효율적으로 예측할 수 있는 이론 및 알고리즘의 개발과 단백질들에 대한 많은 정보를 검색하고 그 특성을 규명할 수 있는 생물정보학의 여러 지식과 기술을 필요로 한다.

이 총설에서는 한정된 예들이지만, 계산화학적인 연구방법이 단백질이나 핵타이드의 구조와 기능의 관계를 규명하는 데 어떻게 사용될 수 있는가를 살펴보았다. 보다 효율적으로 단백질의 3-D 구조와 단백질-단백질 상호작용을 규명하기 위해서는 단백질의 구조분

석에 대한 기초이론과 알고리즘을 개발하고, 이러한 일련의 연구를 지원할 수 있는 국가 차원의 집약적인 센터의 설립이 절실히 요구되고, 무엇보다도 이 분야에 종사할 수 있는 전문가를 양성하고 지원하는 것이 시급하다. 

■ 참고문헌

- (1) Jacobs, M. Chem. Eng. News 2000, July 31, 5.
- (2) Borman, S. Chem. Eng. News 2000, July 31, 31.
- (3) Trauer, A. M. Chem. Eng. News 2000, February 7, 19.
- (4) *강원대 | 생화학뉴스*, 1990, 10(1), 17.
- (5) Kang, Y. K.; No, K. T.; Scheraga, H. A. J. Phys. Chem. 1996, 100, 15588.
- (6) Jhon, J. S.; Kang, Y. K. J. Phys. Chem. A 1999, 103, 5436.
- (7) Kim, M. K.; Kang, Y. K. Protein Sci. 1999, 8, 1492.
- (8) Pak, H. S.; Kim, O.; Kang, Y. K. *in manuscript preparation*.
- (9) Kang, Y. K.; Jhon, J. S. J. Peptide Res., 2000, *in press*.

Genome

Bioinformatics