

RNA 구조 예측을 위한 컴퓨터 모델링

인하대학교 한경숙*
충남대학교 김흥진

1. 서론

전산화자의 입장에서 보았을 때, RNA, DNA, protein(단백질) 등의 분자 구조는, Bioinformatics의 여러 분야 중에서도 흥미롭고 복잡한 문제를 많이 제시하는 분야이다. 본고는 컴퓨터 모델링을 이용한 RNA 분자 구조 연구에 관한 것이다. 생물학적으로 RNA 분자는 단백질 합성의 모든 단계에서 필수적인 역할을 담당한다. messenger RNA(mRNA)는 단백질의 아미노산 서열 결정에 필요한 코드를 운반하고, transfer RNA(tRNA)는 그 코드를 단백질로 번역하며, ribosomal RNA(rRNA)는 단백질 합성에 필요한 구성 장치의 일부를 제공한다. 인간의 건강과 경제에 막대한 피해를 입히는 동·식물 바이러스가 RNA 바이러스일 수도 있는데, AIDS, 간염, 유행성 독감, 감기, 몇 종류의 암이 RNA 바이러스에 의한 질병이다. 이러한 RNA 바이러스를 연구하는 사람들은 한결같이 그 RNA의 구조를 알고자 하는데, 그 이유는 세포 내에서 RNA 분자의 생물학적 기능은 그 분자의 구조에 의해 결정되기 때문이다.

RNA 분자는 ribonucleotide가 연속적으로 연결된 사슬이다. 각 ribonucleotide는 네 가지 종류의 염기(adenine, cytosine, guanine, uracil) 중 하나를 포함한다. 이 염기 서열을 RNA 분자의 일차 구조(primary structure)라고 부르며, 각 RNA 분자는 고유의 염기 서열을 갖고 있다. 한 가닥으로 구성된 RNA는 adenine(A)과 uracil

(U) 사이, guanine(G)과 cytosine(C) 사이에서 수소 결합을 함으로써, 안정된 구조를 취하려고 한다. A-U 염기 쌍과 G-C 염기 쌍을 canonical pair라고 한다. G-C canonical pair만큼 강력한 결합은 아니지만, G-U 염기 쌍도 세포 내에서 자주 발견되며 이를 wobble pair라고 부

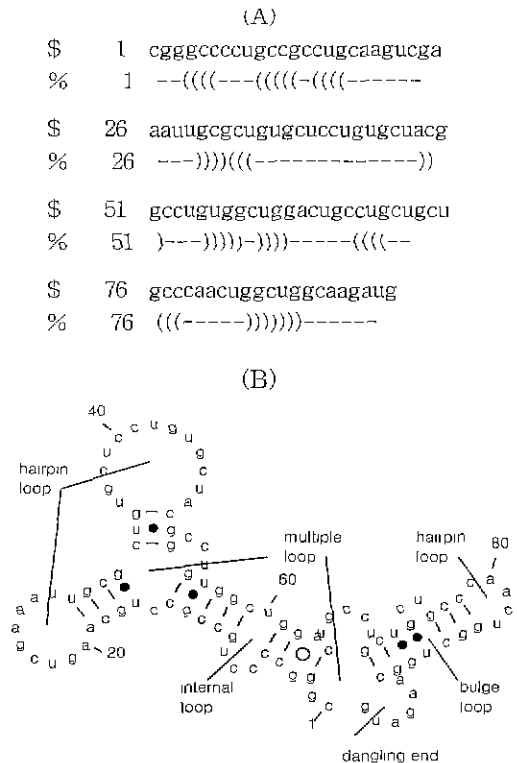


그림 1 (A) text form으로 표현된 mRNA 5' NTR 분자의 이차 구조 (B) graphic form의 이차 구조

*정희원

른다. 이러한 수소 결합에 의해 형성되는 이차원적 형태를 이차구조(secondary structure)라 하고(그림 1 참조), 삼차원적 형태를 삼차 구조(tertiary structure)라 한다. RNA의 생물학적 기능(function)을 완전히 결정하는 것은 삼차 구조이다. 단백질 구조와 비교하였을 때, RNA의 이차 구조 예측 기법은 단백질 이차 구조 기법과 거의 같은 수준이지만, RNA 삼차 구조는 단백질 삼차 구조 예측보다 훨씬 뒤떨어진 수준이다. DNA의 염기 서열을 밝히는 기술의 발전으로 인하여, genomic DNA의 분석으로부터 RNA의 일차 구조를 결정하는 것은 어렵지 않다. 최근 RNA 일차 구조에 대한 정보는 급격히 증가한 것에 반하여, 이차 구조 또는 삼차 구조를 결정하는 기술은 이를 따라가지 못하고 훨씬 느린 속도로 발전하고 있다. 일반적으로 RNA 분자 구조에 대한 컴퓨터 모델링 연구는 (1) 이차 또는 삼차 구조의 예측, (2) 구조 형성 과정(folding process)의 시뮬레이션, (2) 구조의 시각화의 세 분야로 나눌 수 있다. 본 고에서는 이 세 분야에서의 컴퓨터 모델링 연구를 소개한다.

2. RNA 이차 구조 및 삼차 구조의 예측을 위한 컴퓨터 모델링

생화학적 실험에 의하여 RNA의 구조를 결정하는 기법은, 그 실험 방법에 따라 여러 단계의 해상도를 갖는 RNA 구조를 결정한다. 첫째, 가장 고해상도의 구조를 결정할 수 있는 방법은, X-ray crystallography에 의하여 RNA를 구성하는 원자의 3차원 좌표 값을 측정하는 것이다. 그러나 현재의 기술로는 tRNA를 제외한 대부분의 RNA 분자를 결정 형태로 할 수 없기 때문에, X-ray 분석을 적용할 수 있는 RNA는 제한된다. 둘째, X-ray 분석 다음으로 고해상도의 구조를 결정할 수 있는 방법은, NMR(nuclear magnetic resonance) spectroscopy에 의하여 가까운 양자 사이의 거리와 backbone torsion angle을 측정하는 것이다. 셋째, X-ray crystallography나 NMR spectroscopy보다 저해상도의 RNA 구조 데이터를 제공하는 방법은 화학적 또는 효소적 변이 실험을 하는 것으로서, 염기의 accessibility를 결정하거나 RNA의 변이가 생물학적 활동에

미치는 영향을 분석하여 RNA 구조를 밝히는 것이다. 생화학적 실험에 의하여 분자 구조를 결정하기 위해서는, 적은 개수의 염기로 구성된 RNA 분자에 대해서도 많은 시간과 시행 착오를 거쳐 실험이 설계되고 진행되는 것이 보통이다.

한편, 컴퓨터 모델링을 이용한 방법은 생화학적 실험과는 달리 이론적으로 RNA 분자의 구조를 예측하는 기법이다. 실험적 방법이 많은 시간과 노력을 요하는데 비하여, 컴퓨터 모델링에 의한 방법은 빠르고 수월하게 구조를 예측하기 때문에, 특히 새로 발견된 RNA의 기능을 빨리 이해하는데 유용하다. 컴퓨터 모델링은 실험 데이터를 해석하고, 구조에 대한 가설을 세우고, 삼차 구조 모델을 만들 수 있다. 따라서 컴퓨터 모델링을 이용하여, 화학적 또는 효소적 변이 실험이 보다 정확하고 효율적으로 수행될 수 있도록 설계할 수 있고, 역으로 생화학적 실험 결과는 보다 정확한 컴퓨터 모델링을 위한 데이터로 사용될 수 있다.

RNA 이차 구조를 이론적으로 예측하는 방법들은 열역학적 방법(thermodynamic method)과 계통 발생학적 비교법(phylogenetic comparative method)의 두 가지 유형으로 분류될 수 있다. 열역학적 방법은 energy model을 이용하여, 최소 또는 최소에 가까운 자유 에너지 값을 갖는 이차 구조를 추정하여 낸다[1, 2, 3]. 이 방법은 사용되는 energy model 자체가 부정확함으로 인하여 추정되는 구조가 불확실하다는 단점이 있다. 본 저자들의 연구[4]에서 밝혔듯이, 이 방법은 또한 RNA 염기서열의 국부적 변화나 범위에 상당히 민감하다는 단점도 갖고 있다.

계통 발생학적 비교법에서는 에너지 값 계산은 하지 않고, 동류성 염기서열을 비교함으로써 보존되는 stem을 찾아 구조를 추정한다[5, 6]. 동류성 염기서열은 같은 조상으로부터 유래한 염기서열이기 때문에 비슷한 구조를 가질 것이 예상된다. 동류성 염기서열간에 염기배열이 조금씩 다름에도 불구하고 어느 위치에 stem이 보존되는 것을 그 stem이 존재하는 것에 대한 증거로 간주한다. 16S rRNA와 23S rRNA의 현재의 이차 구조 모델은 계통 발생학적 비교법과 생화학적 실험 결과를 바탕으로 추정된 것이다[5, 6]. 계통 발생학적 비교법은 잠재적 stem들을 찾고

이 stem들 중에서 이차 구조에 포함시킬 것들을 가려내는 과정 중 많은 부분이 수작업에 의존하는 경우가 많다. 수작업 자체가 쉽지 않을 뿐 아니라 작업 도중 실수로 간과되는 stem이나 구조도 생길 수 있다. 이 절차 중에서 어떤 부분은 컴퓨터 알고리즘으로 변환되었으나[7, 8], 모든 가능한 페턴을 저장해야 하기 때문에 페턴의 수가 많은 염기서열에 적용하는데 어려움이 있다. 각각 N 개의 염기로 구성된 M 개의 염기서열을 동시에 배열 (align)하고 공통된 구조를 추정하는 Sankoff의 알고리즘은 $O(N^{3M})$ 시간과 $O(N^{2M})$ 공간이 요구되어[9], 실용성이 떨어진다.

RNA의 삼차 구조를 형성하게 하는 상호 작용은 여러 가지가 있는데, 이 중에 대표적인 것이 pseudoknot 구조의 형성이다. pseudoknot 구조는 이차 구조의 loop에 있는 염기와 이 loop 외부에 있는 염기와의 결합으로 생성되는 삼차 구조 요소이다. 가장 전형적인 pseudoknot 구조는 H-type(hairpin-type)의 pseudoknot으로서, hairpin loop에 있는 염기가 hairpin loop-stem 외부에 있는 염기와 결합하여 형성된다. 그림 2 (B)-(D)에서 보듯이, topology가 서로 다른 세 종류의 H-type pseudoknot 구조가 있다.

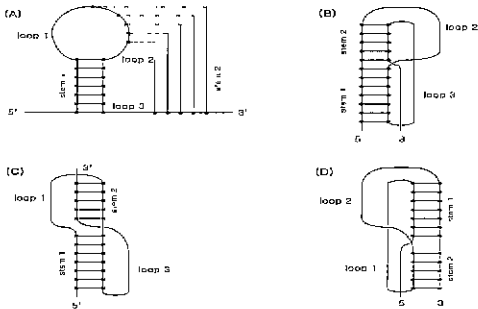


그림 2 H-type pseudoknot 구조의 종류

(A) 가장 일반적인 pseudoknot 구조로서, 점선은 hairpin loop에 있는 염기와 3' 방향에 있는 염기와의 결합을 나타낸다 (B)-(D)는 (A)에 있는 3개의 loop 중에서 하나가 생략되면서 stem 1과 stem 2가 일렬로 정렬되어 연속한 하나의 stem과 비슷한 구조를 취한다. (B)-(D)는 각각 (A)에서의 loop 1, loop 2, loop 3가 제외된 pseudoknot 구조인데, 이 중에서 (C)의 구조가 가장 흔한 RNA pseudoknot 구조이다.

이 pseudoknot 구조는 1980년대 초기에 turnip yellow mosaic virus (TYMV)를 비롯한 몇몇 plant RNA에서 최초로 발견되어 tRNA 유사 구조로 소개된 이후[10, 11], 거의 모든 종류의 RNA 분자, 특히 모든 종류의 RNA 바이러스 분자에서 존재하는 것으로 알려져 있다. TYMV RNA의 이차 구조는 tRNA의 이차 구조와는 다르지만, pseudoknot 구조의 형성으로 인하여 tRNA와 비슷한 삼차 구조를 이루고 이에 따라 tRNA와 비슷한 기능을 갖는 것으로 밝혀졌다.

topology 측면에서 보았을 때, RNA pseudoknot을 비롯한 삼차 구조가 이차 구조와 다른 점은 염기 쌍을 나타내는 선분이 서로 교차할 수 있다는 것이다. 그림 3에서 보듯이 이차 구조에서는 염기 쌍을 나타내는 선분이 서로 교차하는 일이 없지만, pseudoknot 구조에서는 염기 쌍을 나타내는 선분이 서로 교차한다. 그 이유는 「염기 서열에서 염기 i, j, k, l이 이러한 순서대로 존재하면, 이차 구조는 염기쌍 (i, k)와 (j, l)을 동시에 포함할 수 없다」는 제약조건을 만족하기 위하여, 이차 구조는 서로 교차하는 염기 쌍을 동시에 포함하지 않기 때문이다.

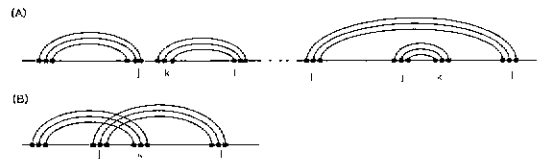


그림 3 (A) 이차 구조는 염기 쌍 (i, j)와 (k, l)을 포함하거나 (i, l)과 (j, k)를 포함할 수는 있지만, 염기 쌍 (i, k)와 (j, l)을 동시에 포함할 수 없다. (B) pseudoknot 구조는 염기 쌍 (i, l)과 (j, k)을 동시에 포함하는 구조이다.

N 개의 염기로 구성된 RNA에서 가능한 염기 쌍의 종류는 $N(N-1)/2$ 이다. 각 염기 쌍은 RNA 이차 구조에 포함될 수도 있고 안될 수도 있다는 점을 고려하면, N 개의 염기로 구성된 RNA가 취할 수 있는 이차 구조의 종류는 $O(2^{N^2})$ 이다. pseudoknot과 같은 삼차 구조 요소를 포함하면 가능한 구조의 종류는 이보다 훨씬 많아지고, 이에 따라 삼차 구조 예측은 이차 구조 예측보다

계산학적으로 훨씬 복잡한 문제이다. 컴퓨터 모델링을 이용한 이론적 방법이 pseudoknot 구조의 예측에 그다지 사용되지 못했던 가장 큰 이유는, pseudoknot 구조에 대한 열역학적 파라미터가 아직 알려지지 않았기 때문이다. 따라서, 에너지 모델을 사용하는 기존의 이차 구조 예측 프로그램은, 예를 들면 MFOLD[12], pseudoknot 구조를 전혀 고려하지 않는다. loop과 stem의 에너지 값을 변형하여 pseudoknot 구조에 적용한 프로그램[13, 14]이 개발되었으나, 이미 밝혀진 작은 RNA 구조에 포함된 pseudoknot 구조를 예측하는 정도에 그쳤다. 비교적 최근에 개발된, 유전자 알고리즘을 이용한 pseudoknot 구조 예측 프로그램들[15, 16]은 에너지 값을 fitness criteria로 사용하는데, 프로그램 실행 시간이 지나치게 길거나 병렬처리를 하는 슈퍼 컴퓨터에서만 실행되는 단점이 있다.

pseudoknot 구조 예측에 현재 가장 많이 사용되고, 예측 결과도 다른 방법보다 상대적으로 좋은 것이 유전자 알고리즘을 이용한 방법이다. 네덜란드 Leiden 대학의 연구 그룹에서 개발한 유전자 알고리즘[15]은 열역학적 에너지 값을 fitness criteria로 사용하고, 에너지 값 계산은 MFOLD 프로그램[12]을 이용한다. 그러나, 500개의 염기 서열로 구성된 RNA의 구조를 예측하는데 20 시간이 소요되는 정도로 프로그램 실행 시간이 지나치게 긴 단점이 있다. 이 유전자 알고리즘에서 사용하는 mutation, crossover, selection operator는 Goldberg가 제안한 본래의 유전자 알고리즘[17]과 같이 binary string인 것에 반하여, 미국 NIH의 Shapiro와 Wu가 개발한 H-type pseudoknot 구조를 예측하는 유전자 알고리즘[16]에서는 본래의 유전자 알고리즘이 많이 변형된 것으로 병렬 처리를 하는 슈퍼 컴퓨터에서 수행된다. 그러나, 이 두 알고리즘은 공통적으로 열역학적 에너지 값을 fitness criteria로 사용하는 기법이다. 단, 이차 구조 예측에서 사용되는 열역학적 방법과 다른 점은, dynamic programming 대신 유전자 알고리즘을 사용하여 많은 수의 가능한 구조를 생성하여 최적화하는 과정을 통하여 구조를 예측한다는 것이다. pseudoknot 구조에 대한 연구는 네덜란드의 Leiden 대학에서 선도하고 있는데, 현재까지 알려진

pseudoknot 구조에 대한 정보는 이 대학에서 구축한 web site[40]에 정리되어 있다.

3. RNA 구조 형성 과정의 시뮬레이션

열역학적 방법과 계통 발생학적 비교법으로 대표되는, 기존의 RNA 구조 예측 방법들은 RNA의 구조 형성 과정을 고려하지 않고 최종 구조를 추정하는 것이다. 그러나, RNA 분자의 구조는 그 RNA의 염기 서열이 다 자라기를 기다렸다가 가장 안정된 형태를 취하는 것이 아니라, 염기 서열이 생성됨과 거의 동시에 구조가 형성된다[18, 19]. 따라서 folding의 동력학(kinetics) 측면에서 어려움이 있으면, 열역학적 또는 계통 발생학적으로 안정된 형태에 도달하지 못할 수도 있고, RNA 분자의 최종 구조는 일차 구조가 시간이 지남에 따라 자라면서 형성하게 되는 중간 단계의 구조들에 의해 어느 정도 영향을 받게 된다. 또한, 한 구조에서 다른 구조로의 동적인 이동이 어떤 RNA 분자에 있어서는, 기능의 중요한 부분일 수도 있다.

RNA의 구조는 계층적 과정을 거쳐 형성되는 것으로 추정된다. RNA 분자가 자라면서 이차 구조가 먼저 형성되고, 삼차 상호 작용(tertiary interaction)에 의하여 간결한 삼차 구조의 형태를 취한다. 이차 구조를 형성할 때에도, 염기 서열에서 가까운 거리에 있는 염기들간의 상호 작용으로 인한 hairpin loop은, 먼 거리에 있는 염기들간의 상호 작용으로 인한 구조 요소들(internal loop, bulge loop, multiple loop 등)보다 먼저 형성된다. 이러한 계층적 구조 형성에 대한 가설은 구조 요소가 만들어지는 속도를 측정할 실험 결과[20]에 비추어 볼 때, 타당성이 있는 것으로 보인다.

이와 같은 RNA 분자의 구조 형성 과정을 시뮬레이트해 보려는 시도가 있었는데, 중간 구조의 적합성 여부를 결정하는 기준으로서 자유 에너지가 강조되었다. 그 중, additive approach라고 특징 지을 수 있는 방법들은 자유 에너지를 최소로 증가시키면서 형성되고 있는 구조에 이미 존재하는 stem과 공존할 수 있는(즉, 이들과 conflict가 없는) stem을 차례로 골라 현 구조에 덧붙이는 방법을 취한다[21, 22]. 이와 같이 구조의 재편성(structure reorganization)을 허용하지

않는 additive approach는 생화학 실험의 관찰과 상반되는 면이 있다; 이차 구조는 동적 평형 상태에서 존재하는데, 염기 서열이 성장하거나 화학적 처리로 잘라지게 되면, 보다 안정된 구조를 이루기 위해 기존의 구조는 분해된다[23]. 구조의 재편성을 허용하는 방법이 있기는 하나, 이들은 각 중간 단계에서 총 자유 에너지를 최소로 할 수 있는 구조를 구하기 때문에 역시 생물학적 실현 가능성이 미약하다[24, 25]. Monte Carlo method를 사용하여 각 단계에서 확률이 가장 높은 구조를 선택하는 방법도 있으나 매우 복잡한 계산을 요하고, 작은 RNA 분자에만 적용되었다[26, 27, 28].

유전자 알고리즘을 이용하여 RNA의 구조 형성 과정을 모의실험하는 시도가 있었다[15]. 유전자 알고리즘을 이용한 방법들은 유용한 면이 있으나, 많은 계산량을 필요로 하기 때문에 커다란 분자에 적용하기에는 아직 실용적이지 못한 측면이 있다. 최근, 본 저자가 개발한 QFolder 프로그램[29]은 구조 변이의 kinetics를 직접적으로 고려하여 구조 형성 과정을 시뮬레이트하는 그래픽 프로그램이다.

4. RNA 구조의 시각화

RNA 분자의 이차 구조는 본질적으로, 구성 염기간의 연결관계가 결정되면 유일하게 결정되는 topological structure이지, geometric structure가 아니다. 이것을 굳이 그래픽 형태로 나타내려는 목적 중의 하나는, 사람이 육안으로 보아서 쉽게 파악하고 비교할 수 있도록 표현하기 위한 것이다. 그림 1 (A)는 96개 염기를 포함하는 RNA 분자의 이차 구조를 text form으로 나타낸 것으로서, 비교적 작은 분자에 해당한다. 그러나 RNA 분자를 구성하는 염기의 개수가 몇 백개 이상으로 많아지면, 그 RNA의 이차 구조가 결정되었다 하더라도, 그림 1 (B)와 같이 시각화를 하지 않고서는 이차 구조를 파악하는 것은 거의 불가능하다. 이차 구조를 시각화할 때 가장 큰 어려움은, 구조의 구성 요소들이 겹치지 않도록 하면서 동시에 전체적인 구조가 간결하게 보이도록 그려야 한다는 것이다. 많은 염기를 포함하는 RNA 분자일수록, 이 상충되는 조건을 만족시키는 것이 쉽지 않다.

RNA 이차 구조의 시각화를 위해서 개발된 대부

분의 드로잉 프로그램은 일단 구조 요소들이 겹치는 구조를 생성하고, (1) 사용자가 수작업에 의하여 구조 요소들을 변형하는 작업(회전, 길이 줄임 또는 늘림 등) 수행하여 겹침 현상(overlap)을 제거하거나[30, 31], (2) 프로그램 스스로 backtrack을 하거나, 점진적으로 겹침이 줄어드는 구조를 반복적으로 생성함으로써 겹침을 제거하게 된다[32, 33, 34, 35, 36]. 프로그램에 의하여 겹침 현상을 제거하는 경우, 구조 요소의 겹침을 피하기 위하여 구조 요소를 변형하게 되는데, 이 때 도입되는 변형 규칙이 모든 구조 요소에 일률적으로 적용된다 보니 왜곡되게 보이는(예를 들면, 특정 구조 요소가 지나치게 휘거나 찌그러진 모양) 이차 구조를 생성할 때가 많고, backtrack이나 반복적인 과정으로 인하여 시간적 효율성도 낮다. 또한, 이차 구조 시각화 프로그램이 필요로 하는 계산량이 많아서, 그 운용 환경이 대부분 PC보다는 mainframe 컴퓨터나 workstation급 컴퓨터이어서, RNA를 연구하는 사람들에게 널리 보급되어 사용되기에 제약이 있다. 최근의 Nakaya의 알고리즘[37]은, N개 입자간의 인력과 척력을 계산하기 위하여 개발된 알고리즘[38]을 RNA 이차 구조 시각화에 적용함으로써, N개 염기로 구성된 RNA의 이차 구조를 $O(N \log N)$ 시간에 시각화하였다. 그러나, $O(n \log n)$ 의 time complexity는 병렬 처리를 하였기 때문에 가능했다. 본 저자가 개발한 VizQFolder 프로그램[39]의 시각화 알고리즘은 N개의 염기를 갖는 RNA의 이차 구조를 $O(N^2)$ 시간에 그리는데, 단일 프로세서에서 실행되는 시각화 알고리즘으로서는 가장 시간적 효율성이 높다.

미국 NCBI(National Center for Biotechnology Information)에서 관리하는 Structure 웹 서버[41], MMDB(Molecular Modelling Database)의 데이터에 대한 삼차 구조를 시각화하거나 구조를 비교하는 프로그램을 web browser에서 수행할 수 있도록 한다. MMDB에는 실험적 방법으로 구조가 밝혀진 RNA, DNA, 단백질에 대한 정보가 수록되어 있는데, 아직까지는 RNA보다는 단백질 위주의 정보가 많이 있다.

5. 결론

컴퓨터 모델링을 이용하여 분자 구조를 예측하고 시각화하는 연구는 Bioinformatics의 중요한

분야로서 컴퓨터공학과 생명과학의 공유 영역에 있는 학제간 학문이다. 외국에서는 컴퓨터의 보급이 확산되면서 프로그램을 이용하여 RNA 분자의 구조를 이론적으로 예측해 보려는 노력이 1980년대부터 활발히 시도된 반면, 국내에서는 아직 생소한 분야이다. 국내는 컴퓨터공학과 생명과학의 각 분야의 연구 인력이 상당히 풍부한 편이나, 이 두 분야의 공통 영역인 Bioinformatics 분야의 연구 인력은 취약한 편이다. 금년에 미국에서 완료될 Human Genome Project를 계기로, 인체의 DNA 정보를 이용한 연구가 국내에서도 많이 시작되리라 추측되는데, RNA 구조 예측을 위한 컴퓨터 모델링 연구도 Human Genome Project 결과물에 적용될 가능성이 있다. 가령, 인체의 messenger RNA가 형성하는 구조와 그와 관련되어 일어날 수 있는 기능을 추정함으로써, 관련 질병의 예측에 기여할 수 있다. 무엇보다도, 컴퓨터를 이용한 분자 구조의 연구는 생화학적 실험에 의존하는 연구보다 상대적으로 비용과 시간이 적게 드는 이점이 있다. 뿐만 아니라, 컴퓨터 프로그램이 예측한 구조를 바탕으로 이를 검증하고 보완하는 실험을 설계할 수 있기 때문에, 훨씬 효율적인 생화학 실험을 하도록 도와준다. 또한, 컴퓨터 공학을 생명과학 문제에 응용하는 것은, 현존하는 컴퓨터 방법을 단순히 적용하는 것에 그치지 않고 컴퓨터 공학에서 새로운 연구 분야(예를 들면, molecular graphics and modeling)를 창출하게 함으로써, 해당 분야들이 서로 기여하고 혜택을 주는 기회를 마련하여 준다고 하겠다.

참고문헌

- [1] Sankoff, D., Kruskal, J. B., Mainville, S., Cedergren, R. J In *Time warps, string edits, and macromolecules: the theory and practice of sequence comparison* (Sankoff, D. and Kruskal, J. B., eds.), pp.93-120, Addison-Wesley Publishing Company, 1983
- [2] Zuker, M. and Stiegler, P., "Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information," *Nucleic Acids Res.*, Vol.9, No.1, pp.133-148, 1981.
- [3] Zuker, M., "On Finding All Suboptimal Foldings of an RNA Molecule," *Science*, Vol.244, pp.48-52, 1989.
- [4] Han, K. and Kim, H-J., "Prediction of Common Folding Structures of Homologous RNAs," *Nucleic Acids Res.*, Vol.21, No.5, pp.1251-1257, 1993.
- [5] Noller, H. F. and Woese, C. R., "Secondary Structure of 16S Ribosomal RNA," *Science*, Vol.212, No.4493, pp.403-411, 1981.
- [6] Noller, H. F. "Structure of Ribosomal RNA," *Ann. Rev. Biochem.*, Vol.53, pp.119-162, 1984.
- [7] Waterman, M. S., Arratia, R., and Galas, D. J., "Pattern Recognition in Several Sequences: Consensus and Alignment," *Bull. Math. Biol.*, Vol.46, No.4, pp.515-527, 1984.
- [8] Waterman, M. S. "Computer Analysis of Nucleic Acid Sequences.," *Methods in Enzymology*, Vol.164, pp.765-793, 1988.
- [9] Sankoff, D., "Simultaneous Solution of the RNA Folding, Alignment and Protosequence problems," *SIAM J. Appl. Math.*, Vol.45, No.5, pp.810-825, 1985.
- [10] Rietveld, K., van Poelgeest, R., Pleij, C. W. A., van Boom, J. H. and Bosch, L., "The tRNA-like structure at the 3' terminus of turnip yellow virus RNA: Differences and similarities with canonical tRNA," *Nucleic Acids Res.*, Vol.10, pp.1929-1946, 1982.
- [11] Pleij, C. W. A., Rietveld, K., and Bosch, L., "A new principle of RNA folding based on pseudoknotting," *Nucleic Acids Res.*, Vol.13, pp.1717-1731, 1985.
- [12] Zuker, M. and Stiegler, P., "Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information," *Nucleic Acids Res.*, Vol.9, pp.133-148, 1981.

- [13] Abrahams, J. P., van den Berg, M., van Batenburg, E., and Pleij, C. W. A., "Prediction of RNA secondary structure, including pseudoknotting, by computer simulation," *Nucleic Acids Res.*, Vol.18, pp.3035-3044, 1990.
- [14] Chen, J.-H., Le, S.-Y., and Maizel, J. V., "A procedure for RNA pseudoknot prediction," *CABIOS*, Vol. 8, pp.243-248, 1992.
- [15] Gultyaev, A. P., van Batenburg, F. H. D., and Pleij, C. W. A., "The computer simulation of RNA folding pathways using a genetic algorithm," *J. Mol. Biol.*, Vol.250, pp.37-51, 1995.
- [16] Shapiro, B. A., and Wu, J. C., "Predicting RNA H-type pseudoknots with the massively parallel genetic algorithm," *CABIOS*, Vol.13, pp.459-471, 1997.
- [17] Goldberg, D. E. *Genetic algorithm in search, optimization and machine learning*, Addison-Wesley Publishing Co., Reading, 1989.
- [18] Gluick, T. C. and Draper, D. E., "Thermodynamics of folding a pseudoknotted mRNA fragment," *J. Mol. Biol.*, Vol.241, pp.246-262, 1994.
- [19] Zarrinkar, P. P. and Williamson, J. R., "Kinetic Intermediates in RNA Folding," *Science*, Vol.265, pp.918-923, 1994.
- [20] Zarrinkar, P. P. and Williamson, J. R., "Kinetic Intermediates in RNA Folding," *Science*, Vol.265, pp.918-923, 1994.
- [21] Kozlov, N. N. and Kugushev, E. I., "Computer simulation of tRNA secondary structure folding," *CABIOS*, Vol.9, No.3, pp.253-258, 1993.
- [22] Martinez, H. M., "An RNA folding rule," *Nucleic Acids Res.*, Vol.12, No.1, pp.323-334, 1984.
- [23] Kramer, F. R. and Mills, D. R., "Secondary structure formation during RNA synthesis," *Nucleic Acids Res.*, Vol.9, No.19, pp.5109-5124, 1981.
- [24] Modelevsky, J. L. and Akers, T. G., "Reiterative RNA folding and occupancy rate analysis for mRNA modeling," *CABIOS*, Vol.4, pp.161-166, 1988.
- [25] Nussinov, R. and Tinoco, I., "Sequential Folding of a Messenger RNA Molecule," *J. Mol. Biol.* Vol.151, pp.519-533, 1981.
- [26] Mironov, A., Dyakonova, L. P., and Kister, A., "A Kinetic Approach to the Prediction of RNA Secondary Structures." *J. Biomolecular Structure & Dynamics*, Vol.2, No.5, pp.953-962, 1985.
- [27] Mironov, A. and Kister, A., "RNA Secondary Structure Formation during Transcription," *J. Biomolecular Structure & Dynamics*, Vol.4, No.1, pp.1-9, 1986.
- [28] Kister, A., Magarshak, Y., and Malinsky, J., "The theoretical analysis of the process of RNA molecule self-assembly," *Biosystems*, Vol.30, pp.31-48, 1993.
- [29] Han, K., "A graphical tool for parametric simulation of the RNA structure formation," *Mol. Cells*, Vol.10, No.3, pp.348-355, 2000.
- [30] Devereux, J., Haerberli, P., and Smithies, O., "A comprehensive set of sequence analysis programs for the Vax," *Nucleic Acids Res.*, Vol.12, pp.387-395, 1984.
- [31] Shapiro, B. A., Maizel, J., Lipkin, L. E., Currey, K., and Whitney, C., "Generating non-overlapping displays of nucleic acid secondary structure," *Nucleic Acids Res.*, Vol.12, pp.75-88, 1984.
- [32] Brucoleri, R. E. and Heinrich, G., "An improved algorithm for nucleic acid secondary structure display," *CABIOS*, Vol.4, pp.167-173, 1988.
- [33] Lapalme, G., Cedergren, R. J., and

Sankoff, D., "An algorithm for the display of nucleic acid secondary structure," *Nucleic Acids Res.*, Vol.10, pp.8351-8356, 1982.

[34] Muller, G., Gaspin, Ch., Etienne, A., and Westhof, E., "Automatic display of RNA secondary structures," *CABIOS*, Vol.9, pp.551-561, 1993.

[35] Perochon-Dorisse, J., Chetouani, F., Aurel, S., and Iscolo, N., "RNA_d2: a computer program for editing and display of RNA secondary structures," *CABIOS*, Vol.11, pp.101-109, 1995.

[36] Stüber, K., "Visualization of nucleic acid sequence structural information," *CABIOS*, Vol.1. pp.35-42, 1985.

[37] Nakaya, A., Taura, K., Yamamoto, K., and Yonezawa, A., "Visualization of RNA secondary structures using highly parallel computers," *CABIOS*, Vol.12, pp.205-211, 1996.

[38] Barnes, J. and Hut, P., "A hierarchical O(NlogN) force-calculation algorithm," *Nature*. Vol.324, pp 446-449, 1986.

[39] Han, K., Kim, D and Kim. H-J., "A vector-based method for drawing RNA secondary structure," *Bioinformatics*, Vol.15, pp.286-297, 1999.

[40] <http://www.bio.leidenuniv.nl/~Batenburg/PKB.html>

[41] <http://www.ncbi.nlm.nih.gov/Structure/>

한 경 속



1983 서울대학교 자연대학 지질학과 학사 (컴퓨터 공학 부전공)
 1985 한국과학기술원 전산학과 석사
 1989 미국 University of Minnesota 전산학과 석사
 1994 미국 Rutgers University 전산학과 박사
 1995~현재 인하대학교 공과대학 조교수

관심분야 분자구조 예측과 시각화, 모델링 및 시뮬레이션. 컴퓨터 그래픽스, 인공지능
 E-mail khan@inha.ac.kr

김 흥 진



1988 미국 Fairleigh Dickinson University 분자생물학과 석사
 1992 미국 Rutgers University 미생물학과 박사
 1992~1995 미국 American Cyanamid Co 의학연구소 연구원
 1996~현재 중앙대학교 약학대학 부교수

관심분야 RNA 바이러스의 구조와 기능, HCMV 분자생물학적 연구
 E-mail hongjunk@cau.ac.kr

• 제27회 정기총회 및 추계학술발표회 •

- 일 자 : 2000년 10월 27 ~ 28일
- 장 소 : 숙명여자대학교
- 발표논문 접수마감 : 2000년 9월 2일(토)
- 문 의 처 : 한국정보과학회 사무국
 Tel. 02-588-9246/7, 4001/2
<http://www.kiss.or.kr>