

열역학적 데이터에 기반한 DNA/DNA 연쇄 결합 반응 시뮬레이션

장하영^o 신수용 장병탁

서울대학교 공과대학 컴퓨터공학부 바이오지능 연구실
{hyjang^o, syshin, btzhang}@bi.snu.ac.kr

Simulation of DNA/DNA Hybridization Chain Reaction Using Thermodynamic Data

Ha-Young Jang^o Soo-Yong Shin Byoung-Tak Zhang

BioIntelligence Lab. School of Computer Science and Engineering, Seoul National University

요 약

DNA/DNA의 연쇄 결합 반응에 대한 시뮬레이션을 열역학적 데이터를 이용하여 구현하였다. 1-Base의 non Watson-Crick 결합과, dangling end(결합이 이루어진 두개의 DNA strand 중 한쪽 끝이 다른 쪽 끝보다 길거나 짧은 경우)를 허용하는 nearest-neighbor model을 사용하여 구현된 이 모델에서는 한번의 hybridization만을 예측하는 것이 아니라 연속적인 결합 반응의 시뮬레이션이 가능하다. 이를 통해서 분자 알고리즘의 설계와 검증이 가능할 뿐만 아니라, cross-homology의 검사를 통한 시퀀스의 검증까지도 가능하다. 이러한 *in silico*에서의 접근 방식은 효율적인 분자 알고리즘의 개발과 신뢰성 있는 시퀀스의 설계에 도움이 될 수 있다.

1. 서 론

DNA 컴퓨팅에 대한 연구가 진행되어 갈수록 새로운 어려움들이 나타나고 있다. 일반적으로 분자 생물학적 가작들은 매우 신뢰성 있는 구조를 가지고 있지만, 전통적인 컴퓨터에서 요구하는 수준의 무결성을 만족시키기에는 충분하지 않다. 또한 대부분의 실험들이 숙련된 전문가의 손에 의해서 직접 이루어져야 하기 때문에 규모가 큰 문제로의 확장이 힘들고 계산 과정이 비효율적이다. 이러한 문제 이외에도 대부분의 DNA 컴퓨팅 알고리즘이 인공적으로 합성된 DNA 시퀀스를 사용하고 있기 때문에 이에 따른 추가 비용이 감당하기 어려울 정도로 커지게 된다.

이러한 문제점들을 극복하기 위한 시도들이 계속해서 이루어지고 있는데, 그러한 시도중의 하나가 DNA에서 일어나는 분자 생물학 적인 반응들을 기존의 컴퓨터로 시뮬레이션 하는 것이다 [2, 3, 4]. 이를 통해 특정 문제를 위해 설계된 DNA 컴퓨팅 알고리즘의 문제점을 미리 알아볼 수 있을 뿐만 아니라, 이를 기반으로 하여 실험실에서 이루어지고 있는 일반적인 생물학 실험 과정들의 시뮬레이션이 가능하다. 이러한 시뮬레이션은 생화학이나 분자생물학의 연구에 있어서도 매우 큰 도움이 될 것이다.

본 논문에서는 현재 개발하고 있는 DNA 컴퓨팅 시뮬레이터인 NACST/Sim의 여러 구성 요소 중 하나인 DNA/DNA 연쇄 결합 반응 시뮬레이터를 소개하겠다. 연쇄 결합 반응의 시뮬레이션은 열역학적인 데이터와 artificial chemistry를 기반으로 하여 이루어진다. 제안된 시뮬레이터는 실제 실험을 통해서 얻어진 열역학적 데이터를 이용한 nearest neighbor(NN) 모델을 이용하여 구현되어 있다 [6].

본 논문의 구조는 다음과 같다. 2장에서는 DNA/DNA 연쇄 결합 반응의 시뮬레이션을 위해 사용된 모델들에

대해서 설명하고, 3장에서는 구현된 시뮬레이터를 이용한 실험 결과들을 제시하겠다. 이 후 4장에서 제시된 실험 결과와 시뮬레이션 모델에 대한 결론을 내리겠다.

2. DNA/DNA 연쇄 결합 반응 시뮬레이션 모델

DNA/DNA 연쇄 결합 반응 시뮬레이션 시스템은 collision 모델과 결합 모델로 구성되어 있다. collision 모델은 실험관 내에 있는 DNA 시퀀스들의 분포와 두개의 DNA 시퀀스가 서로 충돌할 확률을 계산한다. 결합 모델은 충돌이 일어난 두개의 DNA가 실제로 결합할 가능성을 계산한다. 이러한 hybridization 과정이 그림 1에 설명되어 있다.

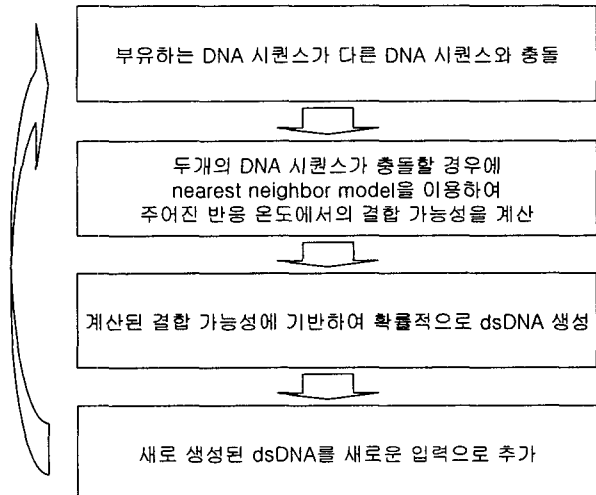


그림 1. DNA/DNA 연쇄 결합 반응 과정

이러한 과정을 반복함으로써 연쇄적인 DNA/DNA hybridization의 시뮬레이션이 이루어진다.

2.1 Collision 모델

시험관 내부의 버퍼에서 부유하는 DNA 시퀀스의 분포와 운동을 모델링 하기 위해서는 브라운 운동을 이용하는 것이 가장 적합하다. 그러나 브라운 운동을 이용한 모델링을 위해서는 추가적인 계산량의 증가가 너무 크기 때문에 제한한 모델에서는 artificial chemistry에서 사용되는 구조를 이용하였다.

artificial chemistry는 주로 오브젝트의 집합 S , 오브젝트들간의 충돌에 관해 정의하는 규칙의 집합 R , 그리고 이러한 규칙들이 오브젝트들에 어떻게 적용되는지 설명하는 알고리즘에 의해 정의되는 리액터 이렇게 3개의 구성 요소로 이루어진다. 모든 오브젝트들은 리액터 내에 존재하고 그 내부에서만 모든 반응들이 발생하기 때문에, 리액터 모델이 전체적인 구조에서 가장 중요한 역할을 담당하게 된다.

대부분의 DNA 컴퓨팅 알고리즘들은 미리 디자인된 DNA 시퀀스들을 초기 입력으로 하여 이 시퀀스들이 시험관 내에서 서로 결합한 결과를 분석하여 결과를 찾아내기 때문에, 모든 입력이 초기에 주어지고 이후에는 추가적인 입력이나 출력이 존재하지 않는다고 생각할 수 있다. 이러한 가정은 모든 오브젝트들이 리액터 내에 존재하고 그 내부에서만 모든 반응들이 발생한다는 리액터 모델에서의 가정과 동일하기 때문에, 리액터 알고리즘에서 사용되는 확률적인 분자 충돌에 기반하여 각각의 시퀀스들이 균일 분포를 이루고 있다는 가정에 충돌을 일으키는 시퀀스들을 선택할 수 있다 [1].

이러한 방식을 이용해서 브라운 운동 모델의 사용에 따른 추가적인 계산 부담이 없이 브라운 운동과 동일한 신뢰도로 충돌이 일어나는 시퀀스들을 선택할 수 있다.

2.2 결합 모델

결합 모델에서는 1-base non Watson-Crick base pair와 1-base dangling end를 고려하여 모든 가능한 경우의 hybridization 가능성을 계산하게 된다. hybridization이 일어날 가능성을 계산하기 위해서는 먼저 melting temperature(T_m)를 계산해야 하는데 이를 위해서 nearest neighbor(NN) 모델이 사용되었다 [6].

NN 모델에 의해서 T_m 이 계산된 후에는 사용자가 입력한 반응 온도에 따라서 아래와 같이 확률이 계산되고, 이를 이용하여 결합 여부가 결정이 된다. T_m 은 그림 2에서 보이는바와 같이, 버퍼 내에 존재하는 dsDNA의 절반이 ssDNA로 용해되는 온도이기 때문에 이 때의 ssDNA와 dsDNA간의 평형 상태는 다음과 같이 나타낼 수 있다.

$$W + C \leftrightarrow W \cdot C \tag{1}$$

전체 시퀀스의 농도는 다음과 같이 정의될 수 있다.

$$C_T = [W] + [C] + 2[W \cdot C] \tag{2}$$

이 때, 전체 dsDNA의 농도 f 는 다음과 같다.

$$f = \frac{2[W \cdot C]}{[W] + [C] + 2[W \cdot C]}, \therefore [W \cdot C] = \frac{f}{2} C_T \tag{3}$$

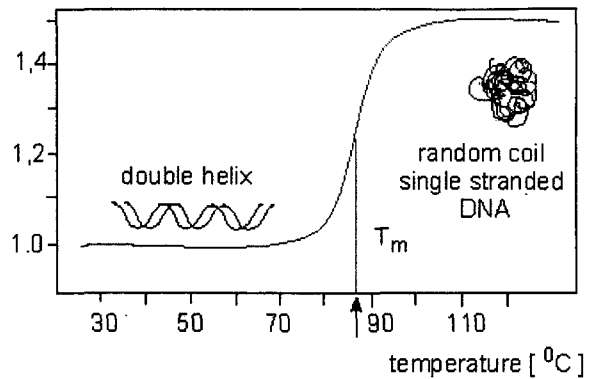


그림 2. melting temperature의 정의

평형상수 K_{eq} 은 다음과 같은 식으로 정의할 수 있다.

$$K_{eq} = \frac{[W \cdot C]}{[W][C]} \tag{4}$$

각각의 시퀀스 농도는 다음과 같이 나타낼 수 있다.

$$[W] = [W]^0 - [W \cdot C], [C] = [C]^0 - [W \cdot C] \tag{5}$$

이를 이용하여 평형상수 K_{eq} 를 다시 정의하면,

$$K_{eq} = \frac{\frac{f}{2} C_T}{[W]^0 [C]^0 - \frac{f}{2} C_T^2 + \frac{f}{4} C_T^2} \tag{6}$$

식 6을 f 에 관해서 풀면,

$$\therefore f = \frac{1 + C_T K_{eq} - \sqrt{C_T^2 K_{eq}^2 + 2C_T K_{eq} + 1 - 4K_{eq}^2 [W]^0 [C]^0}}{C_T K_{eq}} \tag{7}$$

van't Hoff relationship에 의해서,

$$K_{eq} = \exp\left(-\frac{\Delta G}{RT}\right) = \exp\left(-\frac{\Delta H}{RT} + \frac{\Delta S}{R}\right) \tag{8}$$

위의 식에 의해서 DNA 시퀀스가 결합할 확률인 f 가 계산되면, 0과 1 사이에서 랜덤하게 p 를 선택하여 $p \leq f$ 일 경우에 결합이 이루어지게 된다. 따라서 결합 반응은 반응 온도와 T_m 에 기반하여 확률적으로 일어나게 된다.

3. 실험 결과

연쇄 결합 반응 과정을 분석하기 위해서는 시간에 따른 각각의 시퀀스의 농도 변화의 관찰이 필요하다. 이를 위해서 그림 3과 같이 두개의 노드와 에지를 가진 그래프를 사용하였다.



그림 3. 2노드 그래프

각각의 노드를 표현하기 위해서 5'-AGGCGAGTATGGG GTATATC-3', 5'-CCTGTCAACATTGACGCTCA-3'의 조성을 가진 시퀀스를 사용하였다. 두개의 노드를 연결하는

에지는 노드 시퀀스와 상보적으로 디자인되었다.

노드와 에지를 의미하는 20mer 길이의 ssDNA 네종류를 초기 입력으로 사용한 실험 결과가 그림 4에 나와 있다.

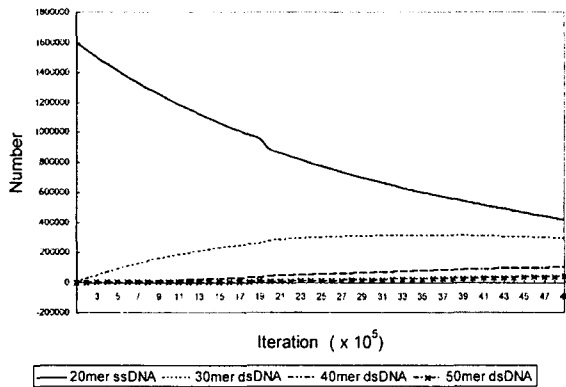


그림 4. 시간에 따른 길이별 시퀀스의 농도 변화

20mer ssDNA는 초기 입력을 의미하고, 30mer dsDNA는 노드와 에지의 결합, 40mer dsDNA는 2개의 노드와 1개의 에지 또는 1개의 노드와 2개의 에지가 결합한 결과를 보여 주고 있다. 시간이 흐름에 따라 초기 입력의 개수가 감소하고, 30mer dsDNA의 개수 변화를 살펴 보면 일정 횟수의 반복 이후에는 그 개수가 어느 정도 일정 수량으로 수렴함을 알 수가 있다. 그래프에는 표시되지 않았지만 이런 현상은 다른 길이의 dsDNA에 있어서도 비슷하게 관찰되었다.

다음으로는 그림 3의 그래프에서 '0→1→0→1→0'의 경로가 해답을 의미한다고 가정하고 해의 분포를 관찰해 보았다. 주어진 경로를 만족하는 후보해가 되기 위해서 dsDNA는 100mer의 길이를 가져야만 하고 그림 5에서 그 결과를 확인할 수 있다.

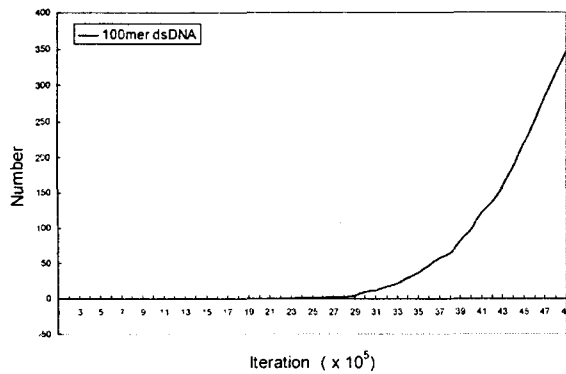


그림 5. 시간의 흐름에 따른 후보해의 분포

각각의 ssDNA의 개수는 4×10^5 개 이고, 전체 population의 크기는 1.6×10^6 이다. 소금의 농도는 1M, 각각의 시퀀스 농도는 100pmoles라고 가정하였을 때, 10^7 번의 반복을 통해 해의 분포를 관찰하였다. 2,600,000번 정도의 반복

이후에 최초의 후보해가 발견되었고 이후에는 그 수가 매우 급격히 증가해서 일정 수량에 수렴한 이후에 증가가 둔화된다.

4. 결론

본 논문에서는 1-Base의 non Watson-Crick 결합과, dangling end를 허용하는 nearest-neighbor model을 사용하여 DNA/DNA의 연쇄 결합 반응을 시뮬레이션 하고 그 결과를 제시하였다. 구현된 시뮬레이터는 NACST/seq[5]에서 사용된 multiobjective evolutionary algorithm에 대한 decision maker로 사용되는 것 뿐만 아니라 실제 실험에서 발생할 수 있는 모든 가능한 연쇄 결합 반응을 시뮬레이션 할 수 있기 때문에 실제 실험에 앞서 DNA 컴퓨팅에 있어 발생할 수 있는 문제점들을 미리 확인해 볼 수 있다. 이러한 과정을 통해서 좀 더 신뢰성 있는 DNA 시퀀스의 디자인과 분자 알고리즘의 설계가 가능해 질 것이다.

감사의 글: 본 연구는 산업자원부 차세대 신기술 과제 및 과학기술부 국가지정 연구실 과제에 의해 지원되었음. 이 연구를 위해 연구 장비를 지원하고 공간을 제공한 서울대학교 컴퓨터 연구소에 감사드립니다.

5. 참고문헌

- [1] P. Dittrich, J. Ziegler, and W. Banzhaf, "Artificial Chemistry - A Review," *Artificial Life*, vol. 7, pp. 225-275, 2001.
- [2] M. Garzon and C. Oehmen. Biomolecular computation in virtual test tubes. In *Proceedings of 7th International Workshop on DNA-Based Computer*, pages 117-128, 2001.
- [3] A. J. Hartemink and D. K. Gifford. Thermodynamic simulation of deoxyoligonucleotide hybridization of DNA computation. In *Proceedings of 4th DIMACS Workshop on DNA Based Computers*, pages 25-38, 1998.
- [4] T. Hinze, U. Hatnik, and M. Sturm. An object oriented simulation of real occurring molecular biological processes for DNA computing and its experimental verification. In *Proceedings of 7th DIMACS Workshop on DNA Based Computers*, pages 1-13, 2001.
- [5] D. Kim, S.-Y. Shin, I.-H. Lee, and B.-T. Zhang, NACST/Seq: A Sequence Design System with Multiobjective Optimization, *Lecture Notes in Computer Science*, vol. 2568, pp. 242-251, 2003.
- [6] 장하영, 신수용, 장병탁, 1-Base non Watson-Crick 결합을 허용하는 DNA Hybridization Simulation, *한국정보과학회 봄 학술발표 논문집 (B)*, 제30권 1호, pp. 476-478, 2003.