

# 1-Base non Watson-Crick 결합을 허용하는 DNA Hybridization Simulation

장하영<sup>0</sup> 신수용 장병탁  
서울대학교 공과대학 컴퓨터공학부 바이오지능 연구실  
{hyjang<sup>0</sup>, syshin, btzhang}@bi.snu.ac.kr

## DNA Hybridization Simulation with Single Base Mismatches for DNA Computing

Ha-Young Jang<sup>0</sup> Soo-Yong Shin Byoung-Tak Zhang  
BioIntelligence Lab. School of Computer Science and Engineering, Seoul National University

### 요 약

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 DNA/DNA Hybridization 예측 시스템을 구현하였다. DNA 컴퓨팅을 기존의 실리콘 컴퓨터를 이용하여 접근하는 이러한 방법은 좀 더 효율적인 분자 알고리즘의 개발과 DNA 컴퓨팅에 사용될 수 있는 더욱 신뢰성 있는 DNA 시퀀스의 설계에 도움을 줄 수 있을 것이다.

### 1. 서 론

DNA 컴퓨팅에 대한 연구가 진행되어 갈수록 분자 컴퓨팅의 한계점들이 점점 밝혀지고 있다. 첫째로, 실험실에서 이루어지는 생물학적인 분자 반응들은 전통적인 컴퓨터에 비해서 신뢰성이 떨어진다. 일반적으로 분자 생물학적인 기작들은 매우 신뢰성 있는 구조를 가지고 있지만, 전통적인 컴퓨터에서 요구되어지는 수준의 무결성을 만족시키기에는 충분하지 않다. 두 번째로 현재의 실험 기술로는 대부분의 실험들이 숙련된 전문가의 손에 의해서 직접 이루어져야 하기 때문에 규모가 큰 문제로의 확장이 힘들고 계산 과정이 비효율적이다. 또한 이러한 전문가의 필요성 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 시험관이 아닌 생체 내에서 DNA 컴퓨팅이 이루어 질 수 있도록 하는 시도이다. 이러한 시도는 시험관에서 수행되는 DNA 컴퓨팅에서 발생하는 문제점들을 어느 정도는 극복하게 해 주고 있다. 또 다른 연구는 DNA 시퀀스의 설계에 관한 것이다. 잘 설계된 DNA 시퀀스는 DNA 컴퓨팅 알고리즘을 더욱 신뢰 있고 효율적으로 만들기 때문에 많은 연구자들이 에러를 최소화 시킬 수 있는 DNA 시퀀스 설계를 연구하고 있다. 또 다른 시도는 DNA에서 일어나는 분자 생물학 적인 반응들을 기존의 컴퓨터로 시뮬레이션 하는 것이다 [6, 7, 8]. 이러한 시뮬레이션을 통해 특정 문제를 위해 설계된 DNA 컴퓨팅 알고리즘의 문제점을 미리 알아보는 것이 가능하다.

본 논문에서는 현재 개발하고 있는 DNA 컴퓨팅 시뮬레이터인 NACST/Sim의 여러 구성 요소 중 하나인 DNA/DNA Hybridization 예측 시스템을 소개하겠다. hybridization은 PCR(Polymerase Chain Reaction)이나

bead separation등의 많은 실험 기술의 기초가 되는 과정이기 때문에, hybridization의 예측은 다른 실험 과정을 시뮬레이션하기 위해 반드시 필요한 일이다. 본 논문은 다음과 같이 구성되어 있다. 2장에서는 열역학적인 실험 데이터에 기반을 두어 만들어진 Hybridization 시뮬레이션 모델을 설명하고, 3장에서는 실제 실험결과 시뮬레이션 결과의 비교를 통해서 구현된 모델의 성능을 평가하겠다. 이후 4장에서 실험 결과와 시뮬레이션 모델에 대한 결론을 내리겠다.

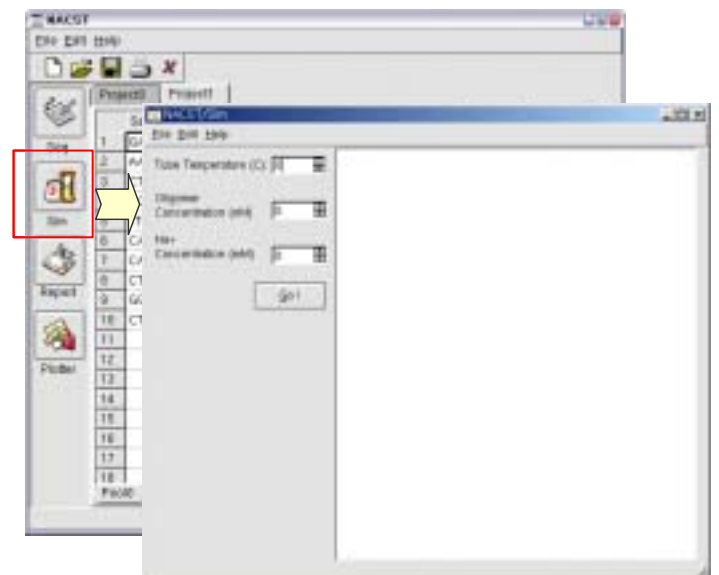


그림 1 DNA hybridization 예측 시스템의 구현.

### 2. 열역학 데이터를 이용한 Hybridization 시뮬레이션

DNA/DNA Hybridization 시뮬레이션 모델에서는 두 개의 ssDNA가 서로 hybridize될 수 있는 확률을 계산하게 된다. 우리가 구현한 모델에서는 1-base non

Watson-Crick base pair와 1-base dangling end를 고려하여 모든 가능한 경우의 hybridization 확률을 검사하게 된다. 이 모델은 PCR이나 전기영동등과 같은 모든 DNA 컴퓨팅의 연산자를 구현할 수 있는 첫 단계가 될 수 있다.

Hybridization이 일어날 가능성을 검사하기 위해서 nearest neighbor method가 사용되었다. 완전한 Watson-Crick pair에 대해 적용될 수 있는 nearest neighbor(NN) 파라미터들은 이미 많은 연구가 진행되고 있다 [9, 12, 14]. 이 외에도 최근에는 1-base non Watson-Crick pair에 적용될 수 있는 NN 파라미터들을 찾고자 하는 노력이 계속되고 있다. GT 결합 [1], GA 결합 [2], CT 결합 [4], AC 결합 [3], AA, CC, GG, TT 결합 [13], dangling end 파라미터 [5].

이러한 파라미터들과 함께 시험관 내의 염분의 농도, DNA의 농도, 그리고 실험 온도 등을 고려하여 계산된 결합 가능성은 사용자가 직접 입력한  $T_m$ (melting temperature)값에 따라서 최종적으로 결합 여부를 결정하게 된다. 이러한 열역학적 데이터를 사용하는 방법의 경우에는 자유 에너지를 이용한 계산이 좀 더 일반적이지만, 실제 실험에서 사용되는 것은 자유 에너지가 아니라  $T_m$ 이기 때문에 우리가 구현한 모델에서는  $T_m$ 을 이용하여 최종적인 결정을 내리게 하였다.

nearest neighbor model에 의해 계산되는  $T_m$ 의 계산식은 다음과 같다.

$$T_m = \frac{\Delta H^\circ}{\Delta S^\circ + R \ln([C_T]/4)} - 273.15 \quad (1)$$

R: 볼츠만 상수(1.987 cal/(K Mol)),  $[C_T]$ : DNA 농도.

엔탈피( $H^\circ$ )와 엔트로피( $S^\circ$ )는 다음의 식에 의해 계산된다.

$$\Delta H^\circ = \Delta H^\circ_{init} + \sum_{k \in \{stacks\}} \Delta H^\circ_k + \Delta H^\circ_{end} \quad (2)$$

$$\Delta S^\circ = \Delta S^\circ_{init} + \sum_{k \in \{stacks\}} \Delta S^\circ_k + \Delta S^\circ_{end} \quad (3)$$

식 (1)에서 소금의 농도는 1M로 가정되어 있다. 만약 소금의 농도가 1M이 아닐 경우에는 다음의 식을 사용한다.

$$\Delta S^\circ = \Delta S^\circ(1M NaCl) + 0.368 \times n \times \ln[Na^+] \quad (4)$$

n: 결합이 일어난 염기쌍의 개수 [9].

### 3. 실험 결과

구현된 시뮬레이터의 성능을 평가하기 위하여 시뮬레이션 결과를 실제 DNA를 가지고 실험한 결과와 비교하였다. 이를 위해서 Traveling Salesman Problem문제의 해결을 위해서 사용하였던 DNA 시퀀스를 이용하였다 [10]. 시퀀스는 표 1에 나온 것과 같다.

문제 해결을 위해 미리 설계된 시퀀스였기 때문에 대부분의 시퀀스들이 미리 예상했던 대로 정해진 위치에서만 붙는다는 결과가 나왔고, 1-base mismatch를 감안하더라도 최대 5개쌍 미만의 잘못된 결합만이 발견되었다. 그러한 결합의 예는 다음과 같다.

표 1 TSP에 사용된 시퀀스.

Vertex sequence	
No.	Sequence(5'→3')
0	AGGCGAGTATGGGGTATATC
1	CCTGTCAACATTGACGCTCA
2	TTATGATTCCACTGGCGCTC
3	ATCGTACTCATGGTCCCTAC
4	CGCTCCATCCTTGATCGTTT
5	CTTCGCTGCTGATAACCTCA
6	GAGTTAGATGTCACGTCACG
Weight sequence	
Edge cost	Sequence(3'→5')
3	ATGATAGATATGTAGATTCC
5	GGATGTGATATCGTTCTTGT
7	GGATTAGCAGTGCCTCAGTT
9	TGGCCACGAAGCCTTCCGTT
11	GAGCTGGCTCCTCATCGCGC

5'-CATCTAACTCGGAATCTACATATCTATCATGATATACCCC-3' (노드 0에서 노드 6으로 가는 웨이트가 3인 에지)와 5'-GGATGGAGCGGAATCTACATATCTATCATGTAGGGACCA-3' (노드 3에서 4로 가는 웨이트가 3인 에지)가 밀줄 친 부분에서 잘못된 GT 결합을 포함한 채로 hybridization이 일어나고 있다. 즉, 5'-CATCT-3' 와 3'-GTAGG-5' 사이에서 GT mismatch가 발생하고 있는 것이다. 이 결과를 실제 실험 결과와 비교하기 위한 전기영동 사진이 그림 2에 나와 있다. 이미 예상하였던 것처럼 미리 계획되지 않은 Hybridization을 거의 찾아보기 힘들다.



그림 2 20% polyacrylamide gel에서의 전기영동 결과. (A)레인M: 마커(50bp 래더), 레인1:노드 2와 3의 결합(각각 200pmoles), 레인2: 노드 1과 3의 결합, 레인3:노드3과 웨이트 5의 결합, 레인4:노드5와 웨이트7의 결합, 레인5:노드2와 에지13(노드1에서 3으로 가는 웨이트가 7인 에지)의 결합, 레인6:노드4와 에지56(웨이트3)의 결합, 레인7:에지12와 에지34(웨이트3)의 결합, 레인8:에지23과 에지56(웨이트3)의 결합, 레인9:에지12(웨이트3, ssDNA 200pmoles). (B)레인1:노드1(ssDNA 200pmoles), 레인2:노드0과 1의 혼합(각각 200pmoles), 레인3:노드0과 1의 결합, 레인5:노드3과 4의 결합, 레인6:노드5와 6의 결합, 레인7:노드0, 1, 2의 결합, 레인8:노드3, 4, 5, 의 결합.

이와 함께 버전 스페이스 러닝을 해결하기 위해 사용하였던 DNA 시퀀스에 대해서도 확인을 해 보았다 [11]. 이 경우에는 실험과정 상에서 문제점을 일으킬 수도 있는 시퀀스를 발견하였다. 시퀀스 5'-ACAAGGGCTCA GAACCAATG-3' 와 5'-TTCTTGTGTACAACCGCGGC-3' 가 밑줄 친 부분에서 잘못된 GT 결합을 만들어 낼 수도 있다는 것이 시뮬레이션 결과 밝혀졌고, 이러한 사실은 실험 결과에 큰 영향을 미칠수도 있는 것이다. 실제로도 이 시퀀스들은 실험 과정에서 문제점이 발생해서, 최종적인 실험은 위에서 지적한 시퀀스들을 수정해서 사용하게 되었다.

#### 4. 결론

열역학적인 데이터에 기반을 둔 nearest neighbor model을 이용하여 만들어진 Hybridization 예측 시스템을 이용하여, 1-Base non Watson-Crick 결합과 dangling end를 포함하는 Hybridization을 예측할 수 있었다. 또한 두개의 ssDNA 시퀀스가 결합할 가능성과 결합시의 안정성뿐만 아니라, 두개의 ssDNA 사이에서 발생할 수 있는 모든 가능한 경우를 모두 검사하기 때문에 실제적인 실험에서 발생할 수 있는 문제점들을 미리 확인해 볼 수 있다.

이를 이용하면 높은 비용이 드는 실제 실험을 수행하기 전에 실험에서 발생할 수도 있는 문제점을 검증할 수 있을 뿐만 아니라, DNA 컴퓨팅으로의 적용 면에서는 보다 효율적인 분자 컴퓨팅 알고리즘의 설계나 DNA 시퀀스의 설계 등에도 도움을 줄 수 있다.

현재는 사용자가 입력한 Tm이 시뮬레이터가 계산해 낸 값보다 낮을 경우에는 Hybridization이 발생하는 것으로 가정하고 있다. 그러나 melting temperature의 정의가 50%의 DNA strand는 서로 결합이 되어 있고, 나머지 50%는 결합이 풀려있는 상태라는 것을 감안할 때 현재의 가정으로는 보다 정확한 시뮬레이션 결과를 얻는데 한계가 있다. 따라서 Tm 곡선이 시그모이드 함수 형태를 이루고 있다는 점을 감안하여, 이를 이용한 통계적인 선택의 과정을 도입하는 것이 반드시 필요할 것이다. 즉, 이러한 경우에는 단순히 hybridization이 일어났다, 또는 일어나지 않았다가 아니라 사용자가 낮은 Tm을 입력할 경우 보다 높은 확률로 hybridization이 발생하고, 높은 Tm을 입력하게 될 경우 보다 낮은 확률로 hybridization이 발생한다는 결과가 나오게 될 것이다.

**감사의 글:** 본 연구는 산업자원부 차세대 신기술 과제 및 과학기술부 국가지정 연구실 과제에 의해 지원되었음.

#### 5. 참고문헌

[1] H. T. Allawi and J. SantaLucia Jr. Thermodynamics and NMR of internal GT mismatches in DNA. *Biochemistry*, 36:10581 - 10594, 1997.  
 [2] H. T. Allawi and J. SantaLucia Jr. Nearest neighbor thermodynamic parameters for internal GA mismatches in

DNA. *Biochemistry*, 37:2170 - 2179, 1998.

[3] H. T. Allawi and J. SantaLucia Jr. Nearest neighbor thermodynamics of internal AC mismatches in DNA: Sequence dependence and pH effects. *Biochemistry*, 37:9435 - 9444, 1998.

[4] H. T. Allawi and J. SantaLucia Jr. Thermodynamics of internal CT mismatches in DNA. *Nucleic Acids Research*, 26(11):2694 - 2701, 1998.

[5] S. Bommarito, N. Peyret, and J. SantaLucia Jr. Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Research*, 28(9):1929 - 1934, 2000.

[6] 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.

[7] 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.

[8] 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.

[9] J. SantaLucia Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. USA*, 95:1460 - 1465, 1998.

[10] J. Y. Lee, S.-Y. Shin, S. J. Augh, T. H. Park, and B.-T. Zhang. Temperature gradient-based DNA computing for graph problems with weighted edges. In *Preliminary Proceedings of 8th International Meeting on DNA Based Computer*, pages 41 - 50, 2002.

[11] H.-W. Lim, J.-E. Yun, H.-M. Jang, Y.-G. Chai, S.-I. Yoo, and B.-T. Zhang. Version space learning with dna molecules. In *Preliminary Proceedings of 8th International Meeting on DNA Based Computer*, pages 261 - 270, 2002.

[12] R. Owczarzy, P. M. Vallone, F. J. Gallo, T. M. Paner, M. J. Lane, and A. S. Benight. Predicting sequence-dependent melting stability of short duplex DNA oligomers. *Biopolymers*, 44:217 - 239, 1998.

[13] N. Peyret, A. Senevirante, H. T. Allawi, and J. SantaLucia Jr. Nearest neighbor thermodynamics and NMR of DNA sequences with internal AA, CC, GG, and TT mismatches. *Biochemistry*, 38:3468 - 3477, 1999.

[14] J. G. Wetmur. DNA probes: applications of the principles of nucleic acid hybridization. *Crit. Rev. Biochem. Mol. Biol.*, 26:227 - 259, 1991.