

DNA 컴퓨팅 연산 과정의 통계 물리적 예측

김준식^{1,○}, 김종찬¹, 노영균², 이동윤³, 장병탁³

¹서울대 물리학과

²서울대 인지과학 협동과정

³서울대 컴퓨터 공학부

[○]shick@phy.snu.ac.kr

KISS Korea Computer Congress 2005

J.S. Kim, J.C. Kim, Y.-K. Noh, D.-Y. Lee, B.-T. Zhang

School of Physics, Seoul National University

Interdisciplinary Program in Cognitive Science

School of Computer Science and Engineering, Seoul National University

요 약

DNA 연산 과정의 열역학적 통계 물리학적 예측 모델을 기술한다. 온도를 천천히 내리는 시험관에서 DNA string 들의 결합은 Metropolis 알고리즘과 진화 연산의 일종인 simulated annealing 알고리즘으로 설명될 수 있다. 본 논문에서는 정리 증명 문제를 통해 위의 통계 물리학적 모델이 DNA 연산에 적용될 수 있음을 보인다. 여섯 종류의 DNA 가닥들의 시뮬레이션 결과와 온도에 대한 실험적인 fluorescence intensity의 비교를 통해 이 모델이 유효함을 보인다. 또한 목표 DNA 개수를 시뮬레이션으로 예측하고 그 결과를 electrophoresis gel image 와 비교하였다.

1. 서 론

DNA는 생명 현상의 정보 전달 물질이며 강한 수소 결합의 Watson-Crick 쌍으로 존재할 수 있다. 근래에 들어 DNA string 의 nano-science 의 탐구와 [1-3] 생물학적 기능을 이용한 암 치료제로서의 가능성이 [4] 시도 되었다. 또한 수학적 문제의 (예를 들자면 SAT 문제) 풀이 도구로서 DNA 컴퓨팅에 관한 연구도 [5] 활발히 이루어 지고 있다.

본 논문에서는 이렇게 주목 받는 DNA 의 반응을 물리학의 열역학과 통계 물리학적 모델을 [6] 사용하여 시뮬레이션 함으로써 DNA 컴퓨팅의 결과를 예측하고자 한다. dsDNA (double stranded DNA) 와 ssDNA (single stranded DNA) 사이에는 수소 결합의 변화에 따른 엔탈피(enthalpy) 변화와 자유도의 변화에 따른 엔트로피(entropy) 변화가 있는데 이는 깁스 자유 에너지(Gibbs free energy)의 변화로 나타난다. 이 깁스 자유 에너지 변화를 이용해 임의의 결합의 허락 여부를 Metropolis 방법으로 [7] 결정하며 온도를 95도에서 10도까지 서서히 내리면 이것이 곧 simulated annealing 이 [8] 된다.

우리는 이 방법을 실제 DNA 시험관의 느린 온도 내림의 시뮬레이션 이론으로 사용한다. C++ 언어로 여섯 가지의 DNA sequence 들에 대한 simulated annealing 을 구현하였고 [8] 실험과 비교하기 위해 이중 결합 쌍의 개수를 온도에 따라 구하였다. 이 결과와 SYBR Green I 을 [9] 이용한 real test tube 의 fluorescence 세기를 측정, 비교함으로써 모델이 유효함을 보인다. 마지막으로 관심이 있는 양인 문제의 해에 해당하는 목표 DNA 분자의 개수를 시뮬레이션으로 예측한 결과로 보이며 이의 간접적인 증명으로 실험으로 구한 electrophoresis gel image의 band를 제시한다.

2. 열역학적, 통계 물리학적 모델

시험관에서 결합한 두 가닥의 DNA single strand 를 우리가 관심 있어 하는 계로 보고 buffer solution 을 environment 로 볼 때 dsDNA 의 확률은 통계 물리학적 분포인 Boltzmann 분포를 따른다. 한 온도에서 Boltzmann 분포를 만드는 알고리즘이 Metropolis algorithm [7] 이며 partition function 을 구하지 않아도 되는 장점이 있다. 임의의 DNA

두 가닥을 잡아 임의의 결합을 상정하고 수소 결합에 의한 enthalpy 변화와 stacking energy 와 관련이 있는 entropy 변화를 구하여 이로부터 Gibbs free energy 변화를 구한다. Boltzmann 분포를 만들기 위해 만일 자유 에너지가 감소하면 그 반응은 무조건 허락하고 만일 자유 에너지가 증가한다면 그 차이의 Boltzmann factor로써 그 반응을 허락한다. 이것이 바로 Metropolis algorithm 이다. 이를 수식을 이용해 기술하면 아래와 같다.

먼저 평형 상태에서 detailed balance에 의해

$$W_{ij}P_i = W_{ji}P_j \quad (1)$$

이다. 위에서 P 는 상태의 확률, W 는 전이 확률이며 위의 식은 i 상태가 된 뒤 j 상태로 가는 확률과 j 상태가 된 뒤 i 상태로 가는 확률이 같음을 의미 한다. 상태의 확률이 Boltzmann 분포를 가진다고 가정하면 W 의 비는 아래와 같다.

$$\begin{aligned} \frac{W_{ij}}{W_{ji}} &= \frac{P_j}{P_i}, \\ &= \frac{g_j}{g_i} \exp(-\Delta H / kT), \\ &= \exp\{-(\Delta H - T\Delta S) / kT\}, \\ &= \exp(-\Delta G / kT). \end{aligned} \quad (2)$$

위에서 g 는 degeneracy, H 는 엔탈피, k 는 Boltzmann 상수, T 는 온도, S 는 엔트로피, G 는 깁스 자유 에너지 이다. Metropolis 방법에서는 아래와 같이 전이 확률을 잡아서 식 (2) 의 비를 유지한다.

$$W = \min\{1, \exp(-\Delta G / kT)\}. \quad (3)$$

우리는 시뮬레이션에서 다음의 값들을 사용하였다. $\Delta H_{G-C} = 9.0$ kcal/MBP(Mole Base Pair), $\Delta H_{A-T} = 7.2$ kcal/MBP, $\Delta H_{others} = 4.0$ kcal/MPB 그리고 $\Delta S_{base-pair} = 23$ cal/degMBP [10].

만일 높은 온도에서 결합을 다 떨어트린 뒤 서서히 온도를 낮추면서 Metropolis 알고리즘을 이용하면 바로 simulated annealing 이다. 즉 DNA 가닥들은 자유 에너지를 최소로 하는 결합들을 찾게 되며 이것은 DNA population 의 일종의 진화 연산이다.

3. 정리 증명 문제

우리는 DNA 컴퓨터의 적용 예로서 정리 증명(theorem proving) 문제를 [11] 선택했다. 6가지 문장이 주어질 때 이들이 과연 일관성이 있는가 하는 것을 알아 보는 방법이다. 여섯 문장은 아래와 같다.

- ㄱ. S 이고 T 이면 Q 이다.
- ㄴ. P 이고 Q 이면 R 이다.
- ㄷ. S 는 참이다.
- ㄹ. T 는 참이다.
- ㅁ. P 는 참이다.
- ㅂ. R 은 참이다.

위의 여섯 문장의 일관성을 묻는 것이다. 이 문제는 resolution refutation 이라는 방법으로 접근 할 수 있다. 즉 먼저 R 을 부정하고 ㄱ 문장과 ㄴ 문장을 or 들의 결합으로 바꾼 뒤 상보적 명제들을 차례로 지워 가서 결국 공집합이 됨을 보여 처음에 부정을 취했던 R 이 나머지 문장들과 일관성을 가짐을 증명한다. 우리는 여기서 상보적 명제들이 지워지는 것을 Watson-Crick pair 의 상보적 결합을 이용하여 구현하고 마지막으로 공집합이 생기는 것을 75 base pair 의 target dsDNA로써 표현을 한다. 결국 6 문장을 표현하는 6 종류의 DNA 가닥들을 섞어서 반응 시킨 뒤 공집합을 표현하는 target dsDNA 가 생성되면 정리는 증명이 되는 것이다.

4. 시뮬레이션 결과

시뮬레이션은 섭씨 95도에서 10도까지 온도를 내리면서 simulated annealing 을 구현했으며 1도 내리는 데 10^8 번 반응을 시도 했다. DNA string 의 개수는 6가지 종류가 각각 3000개, 6000개, 9000개 일 때로 했으며 첫째로 bond 의 수의 총합을 온도에 따라 구하였고 둘째로는 75 base pair 의 target dsDNA 의 개수를 온도에 따라 구하였다.

아래의 그림들은 시뮬레이션 결과들이다.

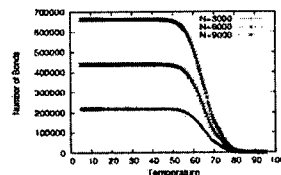


그림 1. 온도에 따른 bond 개수의 총 수

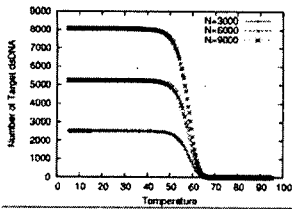


그림 2. 온도에 따른 목표 dsDNA의 개수.

5. Wet 실험 결과

먼저 앞에 그림 1과 비교하기 위해 SYBR Green I 을 이용한 fluorescence intensity를 측정하였다. SYBR Green I 은 bond 의 수에 비례하여 fluorescence 내므로 이 intensity 는 수소 결합의 총 수에 비례한다. 1도에 10분씩 기다리며 온도를 내리면서 fluorescence를 측정한 그림이 그림 3 이다.

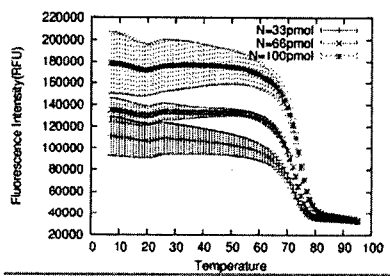


그림 3. fluorescence intensity.

마지막 실험 결과로써 세 가지 농도의 annealing 결과의 electrophoresis gel image 가 그림 4에 있다.

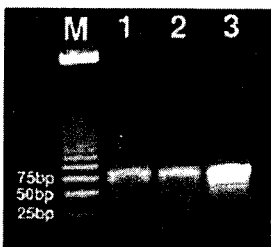


그림 4. 세 농도의 electrophoresis gel image.

6. 결 론

그림 1과 그림 3의 비교를 통하여 서서히 온도를 낮출 때 bond 의 총수의 온도에 따른 변화 형태가 시뮬레이션과 wet

실험이 유사함을 알 수 있다. 이는 우리의 접근 방식이 유효하다는 첫번째 간접적 증거가 된다.

그림 2에서 알 수 있듯이 시뮬레이션 결과에서는 75bp target dsDNA 의 수가 총 가능한 수의 약 85퍼센트 이상 이었다. 이는 온도를 천천히 내리며 반응시킬 때 나오는 많은 수의 결과물들이 우리가 의도하였던 target dsDNA 임을 의미한다. 실제로 그림 4에서 보듯이 세 농도 모두에서 75bp 길이의 깨끗한 band를 볼 수 있다. 이것 역시 우리의 접근 방식의 유효함의 간접적 증거이다.

DNA 연산은 NT (Nano Technology) 와 BT (Bio Technology) 그리고 IT (Information Technology) 의 접합점으로서 위상을 가지며 이 연산의 통계 물리학적 모델의 연구는 이 위상으로부터 그 중요성을 가진다고 생각한다.

7. 감사의 글

본 연구는 산자부 차세대 연구 개발 사업 분자 진화 컴퓨터 (MEC) 과제에 의하여 지원 되었음.

8. 참고 문헌

- [1] Winfree, E., Liu, F., Wenzler, L.A., Seeman, N.C., Design and self-assembly of two-dimensional DNA crystal. *Nature* 394, 539-544, 1998.
- [2] Elghanian, R., Storhoff, J.J., Mucic, R.C., Letsinger, R.L., Mirkin, C.A., Selective colorimetric detection of polynucleotides. *Science* 277, 1078-1081, 1997.
- [3] Yan, H., Zhang, X., Shen, Z., Seeman, N.C., A robust DNA mechanical device controlled by hybridization topology. *Nature* 415, 62-65, 2002.
- [4] Benenson, Y., Gil, B., Ben-Dor, U., Adar, R., Shapiro, E., An autonomous molecular computer for logical control of gene expression. *Nature* 429, 423-429, 2004.
- [5] Braich, R.S., Chelyapov, N., Johnson, C., Rothmund, P.W.K., Adleman, L., Solution of a 20-variable 3-SAT problem on a DNA computer. *Science* 296, 499-502, 2002.
- [6] Azbel, M.Y., Phase transition in DNA. *Phys. Rev. A* 20, 1671-1684, 1979.
- [7] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Equation of state calculations by fast computing machines. *J. Chem. Phys.* 21, 1087-1092, 1953.
- [8] Kirkpatrick, S., Gelatt Jr., C.D., Vecchi, M.P., Optimization by simulated annealing. *Science* 220, 671-680, 1983.
- [9] Skedsvoll, J., Ueland, P.M., Analysis of double-stranded DNA by capillary electrophoresis with laser-induced fluorescence detection using the monometric dye SYBR Green I. *Anal. Biochem.* 231, 359-365, 1995.
- [10] Klump, H., Ackermann, T., Experimental thermodynamics of the helix random coil transition. IV. Influence of the base composition of DNA on the transition enthalpy. *Biopolymers* 10, 513-522, 1971.
- [11] Lee, I.-H., Park, J.-Y., Jang, H.-M, Chai, Y.-G, Zhang, B.-T., DNA implementation of theorem proving with resolution refutation in prepositional logic. *Lecture Notes in Computer Science* 2568, 156-167, 2003.