

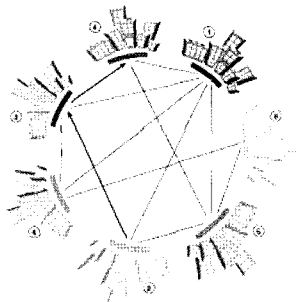
## DNA 컴퓨팅 기술의 응용

장병탁, 양경애(서울대학교 바이오지능연구실 및 바이오정보기술연구센터)

2007년 여름, 다양한 모양으로 변신할 수 있는 오토봇(autobot) 군단을 소재로 한 '트랜스포머(Transformers)' 라는 영화가 개봉되었다. 자유자재로 변화하는 오토봇만 보아도 흥미진진한 영화였지만, 영화 내용 중 미국 국방성에 원인을 해킹을 당하고 나서 등장인물들이 원인에 대해 추측하는 대화를 나누는 장면에서는 DNA 컴퓨팅 연구자로서 짜릿한 흥미를 느끼기도 하였다. 대화 내용은 이렇다. 현재 존재하는 슈퍼컴퓨터라도 국방 시스템의 방화벽을 뚫기 위해 죽히 20년은 걸림직한 일을 순식간에 해내고, 그 신호가 스스로 학습하며 진화하는 모습에 놀라면서 금발의 여배우는 "This is.....It's some kind of..... DNA-based computer! I know that sounds crazy, but..."라고 말한다. 'DNA-based computer' 정도가 되면 기계스스로 진화하면서 뭔가 월등한 능력을 가질 수 있는 듯 한 뉘앙스를 풍겼다. 이 영화에서 언급하는 수준의 DNA 컴퓨팅 연구가 현재까지는 이루어지지 않았지만, 언젠가 이처럼 진화하는 컴퓨터가 나타난다면 그것은 DNA 컴퓨터가 될 것이라고 생각하며, 지난 15년간 이루어진 DNA 컴퓨팅 연구 분야의 업적과 필자의 연구그룹에서 수행하고 있는 연구를 소개하면서 그 응용 가능성에 대해서 논해보고자 한다.

### DNA 컴퓨팅의 시작

아래에서 보이는 <그림 1>처럼 7개 도시를 일방통행으로만 이동한다고 했을 때 최단의 거리를 알아보려고 한다. 도시



【그림 1】 DNA 컴퓨팅에 의한 해밀턴 경로 문제 해결 (Adleman의 실험)[1]

의 개수가 7개 이므로 출발점에서 선택할 수 있는 길은 모두  $N-1$ 개로 6개이다. 0번 도시에서 출발한다고 하면 0번 도시에서 선택할 수 있는 가능성은 1, 3, 6의 세 가지이다. 이들 중 하나를 선택했다면, 두 번째 도시에서 선택할 수 있는 길은 자기 자신과 출발점을 제외한  $N-2$ 가 된다. 이와 같은 방식으로 한 걸음씩 전진해 나가면 선택할 수 있는 길의 개수가 하나씩 줄어들게 되므로  $N-1$  번째 도시에 도착했을 때 선택할 수 있는 길은 출발점으로 향하는 길 한 가지만 남게 된다. 즉  $(N-1)!!$  이고, 7도시를 다루는 경우라면 360가지의 해가 생긴다. 이러한 계산은 현재 실리콘 컴퓨터상에서 힘겹지 않게 수행할 수 있다. 하지만, 도시의 숫자가 하나씩 늘어날 때마다 그 복잡성은 기하급수적으로 증가하며 15도시 정도만 되어도 계산량이 4백억( $(15-1)!! = 43,589,145,600$ )을 넘어서게 되어 계산을 위해서 엄청난 시간이 필요하게 된다. 즉, 도시의 수와 연결 경로의 수가 늘어남에 따라 가능한 해의 수가 폭발적으로 늘어나는 NP (non-deterministic polynomial time)-완전문제에 해당한다. 이런 문제를 해결하는데 있어서, 만일 대규모의 병렬성과 데이터 집적도를 추가시킬 수 있다면 시간의 문제에 있어서 상황은 크게 달라질 수 있다. DNA 컴퓨팅에 관한 연구는 약 15년 전에 이처럼 현재의 실리콘 컴퓨터의 한계를 극복하는 목적을 가지고 시작되었고, 그간 진화하면서 여러 가지 다른 연구 분야를 파생시키게 되었다. 각각의 파생 분야들은 DNA 분자에 정보를 코딩하고 분자 수준에서 인식하여 물리화학적 반응을 통해 자기조직화 함으로써 정보를 처리한다는 면에서는 공통점이 있지만, 크게 세 가지 영역으로 구분할 수 있다. 한 가지는 새로운 계산 패러다임을 실현하는 신기술로써 DNA를 이용하는 것이고, 다른 한 가지는 나노특성을 이용하여 분자수준에서의 나노구조물을 자기 조립하는 것이며, 나머지는 바이오특성을 이용하여 실리콘 인터페이스 없이 바이오정보를 처리하기 위한 소자로 DNA 컴퓨팅을 응용하는 것이다.

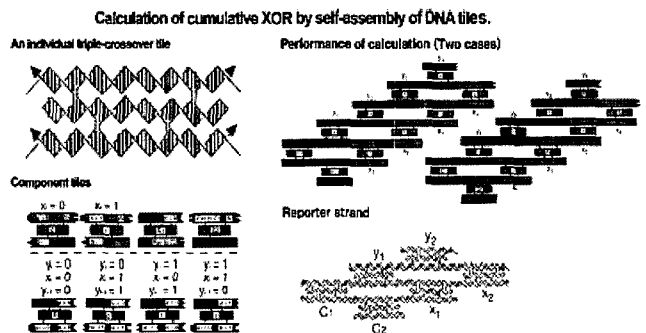
## 새로운 계산 패러다임을 실현하는 신기술로써의 DNA 컴퓨팅

1994년에 미국 University of Southern California의 Adleman 박사는 합성 DNA 분자에 정보를 코딩하고 분자 수준에서 물리화학적으로 조작함으로써 <그림 1>의 경로탐색에 관한 조합 문제를 해결할 수 있음을 실험적으로 보여 주었다[1]. 문제 풀이 과정은 다음과 같다. <그림 1>의 7개 도시를 연결하는 경로를 나타내는 방향성 그래프가 있다고 하자. 이를 위해 도시 이름(그래프의 노드)을 DNA 분자 서열로 코딩하였고, 도시간의 연결 경로(그래프의 에지)에 대해서는 해당 도시 이름에 대한 상보적인 서열로 DNA 코드를 구성하였다. 많은 수의 도시와 연결선에 대한 DNA 분자들을 합성한 후 이를 하나의 시험관에 넣어 DNA의 상보적인 혼성화 반응에 의해 가능한 모든 경로에 해당하는 긴 이중 가닥의 DNA 분자들이 자기조립 되도록 하였다. 예를 들어, 도시0을 AAAAAAAAA로 코딩하고, 도시1을 CCCCCCCC로 코딩한다면, 도시0으로부터 도시1로 가는 경로는 TTTTGGGG로 코딩하여 도시0에 해당하는 코드의 후반부 AAAA와 도시1에 해당하는 코드의 전반부 CCCC가 TTTTGGGG와 혼성화 반응하여(AAAA는 TTTT와 CCCC는 GGGG와 상보 결합) 이중가닥을 형성한다. Adleman은 이와 같이 모든 도시들과 항로들을 DNA 분자를 합성하여 만들어 두고 이를 하나의 시험관 내에서 모두 넣어 반응하게 함으로써 DNA 반응의 병렬적 조합 가능성을 이용하여 모든 가능한 경로들을 생성시킨 후, 이 중에서 해밀턴 경로 문제의 해의 조건을 만족하는 경우들을 순차적으로 여과해 냄으로써 답을 찾도록 하였다. 답이 되기 위한 한 가지 조건은 모든 도시가 방문되어야 한다는 것이다. 각 도시를 20개의 염기로 표시할 경우 7개의 도시를 모두 방문해야 하므로 140개의 염기로 구성된 DNA 분자만이 답의 가능성이 있고 더 짧거나 더 긴 DNA 분자는 답이 될 가능성이 없다. 또 다른 조건은 각각의 도시가 적어도 한번은 나타나야 한다. 이를 검사하기 위해서 에이들만은 자화된 비드에 도시 이름에 상보적인 코드를 가진 DNA 분자를 부착한 후 이를 시험관 내의 DNA 분자들과 결합시킴으로써 후보를 골라내었다. 마지막으로 해는 반드시 특정 시작 도시에서 출발하여 특정 종착 도시로 끝나야 한다. 시험관 내에 남은 DNA 분자들 중에서 이 조건을 만족하는 것이 있으면 해밀턴 경로를 찾은 것이며 따라서 "예"라고 답하

고 그렇지 않으면 "아니오"라고 답한다. Adleman은 DNA 분자로 해를 표시하고 생화학적 실험 방법에 의한 DNA 분자 조작을 통하여 그 해를 찾는 방법을 제시하였다. 즉 DNA 분자의 집적도를 이용하여 가능한 모든 경로를 생성한 후 답을 검사함으로써 이 문제를 해결하였다. 이 실험은 개념적으로는 기존의 실리콘 기반의 연산 개념에 기반하고 있지만, DNA 분자의 물질적 특성을 활용하여 대규모 메모리 공간과 병렬 처리 특성을 사용함으로써 계산 시간을 줄일 수 있는 예를 보여주었다. 인공 DNA 합성 능력과 비용 및 다루는 기술의 제약을 받고 있어서 현재까지 20가지 변수를 갖는 3-SAT문제를 푼 것이 최대 문제를 푼 경우에 해당되지만, 바이오테크놀로지 발전과 더불어 발전할 수 있는 차세대 컴퓨팅 영역에 해당한다.

## DNA 나노조립에 의한 연산 및 DNA 나노조립을 위한 연산

DNA 자기 조립을 이용하여 계산 알고리즘을 수행할 수 있다는 "Algorithmic Self-Assembly"의 개념을 처음으로 제시한 사람은 캘리포니아 공대의 Winfree 이다. 이 개념에 기반을 두어 뉴욕대학의 Seeman 그룹은 순차적인 XOR 논리 연산을 수행할 수 있음을 실험적으로 데모하였다[2]. 이를 위해서 <그림 2>에 보인 바와 같이 세 개의 DNA 가닥으로 구성된 "DNA 타일" 구조를 설계하였다. 이를 이용하여 입력값(청색 DNA 타일)을 표현하고 이들이 계산과정에서 순차적으로 <그림 2>의 왼쪽아래에서 오른쪽 위로

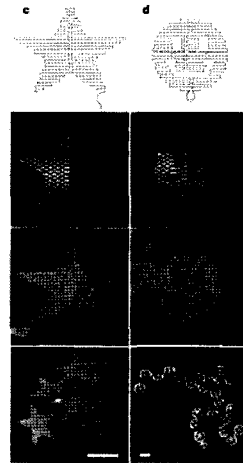


[그림 2] DNA 나노 조립에 의한 순차적 논리 연산

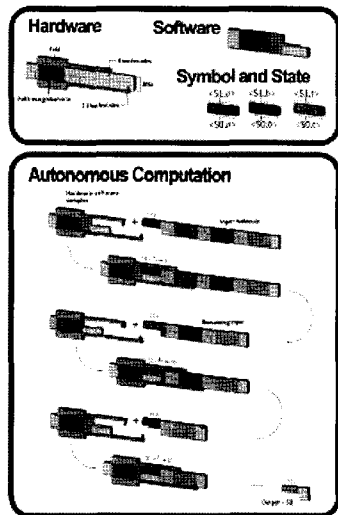
순차적으로 자기조립함으로써 오른쪽 끝에 그 결과가 출력 (적색 DNA 타일)된다. 타일들이 자기조립될 때 리포터 DNA 조각이 같이 연결되도록 설계함으로써 계산의 결과를 후에 원자 현미경(AFM)으로 관찰할 수 있다. 보다 구체적으로 <그림 2>에서는 입력 이진수 (X<sub>1</sub>, X<sub>2</sub>, X<sub>3</sub>, X<sub>4</sub>) = 1110와 초기치로 캐리 C<sub>1</sub> = 0 가 주어질 때, 그 결과로 출력 (Y<sub>1</sub>, Y<sub>2</sub>, Y<sub>3</sub>, Y<sub>4</sub>) = (Y<sub>0</sub> + X<sub>1</sub>, Y<sub>1</sub> + X<sub>2</sub>, Y<sub>2</sub> + X<sub>3</sub>, Y<sub>3</sub> + X<sub>4</sub> = 1011)가 계산되는 과정을 보여주고 있다. 이 과정에서 입력이 Y<sub>i-1</sub> = 0, X<sub>j</sub> = 0 이면 출력은 Y<sub>j</sub> = 0 즉 0 + 0 = 0가 계산되며, 비슷한 방식에 의해서 0 + 1 = 1, 1 + 0 = 1, 1 + 1 = 0의 Y<sub>j</sub> 값이 순차적으로 계산된다. 이러한 계산은 기존의 실리콘 컴퓨터에서도 잘 수행될 수 있으나, 이 연구에서는 DNA 타일에 입력을 기록하고 이들이 결합하여 출력 타일들을 순차적으로 만들고 그 결과를 리포트 함으로써 DNA의 자기조립을 통해서 입력, 연산, 출력의 하나의 완전한 정보처리 과정을 수행할 수 있음을 보인데 그 의의가 있다고 볼 수 있다.

또 한 가지 흥미로운 연구는 나노조립에 의한 연산이 아니라, DNA 나노조립을 위한 연산 연구이다. 2006년 캘리포니아 공과대학 원프리 연구실의 연구원이던 Rothemund는 자연에서 존재하는 바이러스 M13mp18의 genomic DNA (4247 nucleotides)에 staple 가닥들을 첨

가함으로써 원하는 나노구조를 만들어냈다. Staple 가닥들은 기다란 DNA의 일정 부분에 결합하여 가닥을 접히고 할 수 있다. 종이접기 하듯이 DNA를 자유재로 다뤘다고 해서 이런<그림 3> DNA 오리гами[3] 방식으로 생성된 나노구조를 '오리가미(origami)'라는 표현을 사용하기도 한다[3]. <그림 3>에서는 Rothemund가 디자인 한 DNA들이 staple DNA 가닥에 의해서 자기조립화한 문양을 보여주고 있다. 이러한 연구결과는 DNA 자기조립성을 이용하여 나노머신에 사용되는 컴포넌트들을 만들어낼 수 있으며, 자연에서 존재하는 DNA를 이용하여 정보전달을 위한 나노구조 제작 및 고밀도의 정보의 저장가능성을 보여주고 있다. 이미 고가의 명화나 물품에 진품 여부를 알게 하기 위해 DNA 바코드를 사용하게 될 가능성이 제기되고 있기도 하다. DNA 화가가 등장하기도 하고, 나노부품을 찍어내는 미생물 공장들이 가동되기도 하고, 생체 내 부품화된 DNA 기반의 회로를 삽입하여 정보처리를 하게 되는 것이 머지않은 미래에 가능하게 될지도 모른다.



[그림 3] DNA 오리гами [3]



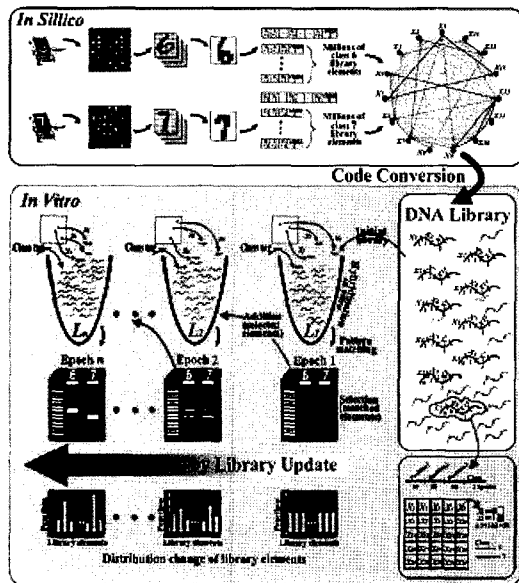
[그림 4] DNA 오토마타 기반 분자 진단

### DNA Computing for Biotechnology: DNA 프로그램 기반 분자 진단

이스라엘의 와이즈만 공대에서는 자기조립의 반대 과정 즉 자기해체 과정을 통해서 정보처리가 가능하고 특히 이를 질병진단에 응용할 수 있음을 보였다[4, 5]. 이 방법에서는 합성 DNA 분자를 이용하여 유전자 진단 규칙을 코딩한 선형 구조의 프로그램으로 만들어 둔다. 이 프로그램 DNA는 끝부분이 일부 단일 가닥으로 존재하는(sticky end) 이중 가닥 구조를 가지고 있으며, 환자로부터 추출한 DNA 데이터

가 이 단일 가닥 부분과 일치하면 완전한 이중 가닥 DNA를 형성하고, 그렇게 되면 이중 가닥을 인식하는 특수 효소에 의하여 잘려나간다. 이는 프로그램 DNA의 다음 명령으로 전이하고 다시 여기에 데이터 DNA가 결합하여 이중 가닥 DNA를 형성하면 또 효소에 의하여 잘려나간다. 이와 같은 자기해체 과정을 반복함으로써 마지막에 남는 프로그램 DNA는 진단 결과를 표시한다. 최종 진단 결과에 약물에 해당하는 서열을 저장하고 있으면 약물전달과 같은 효과를 가진다. 이 방식은 DNA 컴퓨터로 분자 진단 및 약물 전달을 할 수 있다는 개념을 처음으로 실험적으로 보였다는 점에 있어서 의의가 크다. 또한 컴퓨터과학의 기반 계산 구조인 유한상태기계 오토마타를 구현하였으며 이를 중간의 간섭 없이 자율적으로 수행할 수 있는 자율적 연산의 개념을 제시한 점에서 의의가 있다.

### DNA Computing for Information Technology: DNA 분자 기반 연상기억 정보처리



[그림 5] DNA 나노 조립에 의한 순차적 논리 연산

필자의 연구 그룹에서는 최근에 DNA 컴퓨팅을 이용하여 인간의 연상기억 정보처리를 수행하는 모델을 개발하였으며 이를 실험적으로 구현하는 연구를 수행하고 있다[6]. 하

이퍼망 모델이라고 명명된 이 연상기억 장치는 DNA 분자 조각들로 메모리 조각을 표현하고 아주 많은 수의 메모리 조각들의 혼성화 반응을 통해서 연상작용을 이용하여 기억을 인출해 낸다. 이 모델을 필기체 숫자 패턴 인식 문제에 적용하여 시뮬레이션 한 결과 기존의 기계학습 모델 보다 우수한 인식 성능을 얻을 수 있었다. 이 모델은 인간의 기억현상에 대해서 지금까지 밝혀진 정보 저장 및 인출 특성과 유사하다는 점에서 아주 흥미로우며 대규모의 하이퍼망을 DNA 컴퓨터로 실현할 경우 인공지능 및 기계학습 연구에 새로운 장을 열 수 있을 것으로 기대되고 있다. 그러나 현재 대규모의 하이퍼망을 DNA 분자 조작 기술로 실현하기에는 여러 가지 기술적인 제약이 따르며 이를 극복하기 위해서는 추가의 연구가 필요하다. 예를 들면, 다양한 종류의 DNA 분자들이 혼합되어 있는 경우 특정 성분을 가진 분자들을 특이적으로 복제한다든지 특정 분자들을 변별적으로 분리하는 기술 등에 정밀성을 향상할 필요가 있다. 한편, 하이퍼망이 기억을 저장하고 이를 인출하는 과정에 필요한 정보처리 능력은 주로 메모리 접근 연산에 기반하며 복잡한 수치적인 계산을 수행하지 않는다. 따라서 이러한 메모리 프로세서는 현재의 메모리 반도체 기술로서 구현이 가능하며 이는 연상작용을 필요로 하는 메모리 프로세서의 새로운 응용 분야를 개척할 수 있을 것으로 기대된다. 현재 대규모의 패턴인식 문제 해결과 바이오데이터 분석에 적용하기 위한 하이퍼망 실리콘 칩을 구현하고 있다[7].

### DNA 컴퓨터의 미래

순차적 정보처리에 기반한 전통적인 폰노이만 방식으로부터 벗어난 새로운 컴퓨터를 개발하려는 시도는 오래전부터 있어 왔다. 폰노이만 자신도 이미 1956년에 병역에서 집필한 저서 "The Brain and the Computer"에서 새로운 패러다임의 컴퓨터에 대해서 고민하고 있었다. 비고전적 컴퓨팅(Nonclassical Computation), 비전형적 컴퓨팅(Unconventional Computation), 내추럴 컴퓨팅(Natural Computation) 등으로 불리는 이러한 새로운 패러다임의 컴퓨팅 모델로는 신경망 컴퓨팅, 진화 컴퓨팅, 분자/DNA 컴퓨팅, 양자 컴퓨팅 등이 있다. 신경망 컴퓨팅과 진화 컴퓨팅은 자연현상을 실리콘 컴퓨터상에 시뮬레이션 하는데

그친데 반해서 분자 컴퓨팅과 양자 컴퓨팅은 매체의 물리적인 특성을 활용하여 정보처리를 수행한다는 점이 기존의 다른 내추럴 컴퓨팅 모델과 구별된다. DNA기반 분자 컴퓨터는 초병렬적인 반응을 기반으로 한다는 점에서 양자 컴퓨팅과 유사하나, 양자 컴퓨팅은 원자 수준에서의 반응 제어와 결과 검출을 위한 고난도의 양자 소자를 개발해야 하는데 비해서, 분자 컴퓨터는 분자 수준에서의 제어와 검출에 기반하기 때문에 기술적인 실현 가능성이 훨씬 높다. DNA 컴퓨터는 최근 들어 나노기술(NT), 바이오기술(BT), 정보기술(IT)과 접목된 융합기술로 발전하고 있다. 나노기술과 접목되어 나노 패터닝이나 나노구동장치의 프로그램 역할을 수행하는데 DNA 컴퓨터가 사용되고 있다[7]. DNA

는 또한 다른 DNA를 비롯한 RNA나 단백질 등의 바이오 데이터와 직접 반응하여 정보를 조작할 수 있는 특징이 있기 때문에 생명공학과 의약학 분야에서 기존의 실리콘 기술의 개입 없이 정보를 바로 처리 할 수 있는 장점이 있다. 또한 DNA는 화학적으로 대량 합성이 가능하여 공학적인 이용이 이미 가능한 단계에 도달하여 있고 제조 가격도 급속히 하락하고 있다. DNA 분자들의 분자인식 특성에 기반한 나노수준에서의 대규모 초병렬 자기조립 현상을 이용하여 연상기억 패턴 정보처리와 같은 비고전적인 컴퓨팅 패러다임을 실현하는 새로운 기술로서도 DNA 컴퓨터는 그 가능성을 제시해 주고 있다.

#### [ 참고 문헌 ]

- [1] Adleman, L., (1994) Molecular Computation of Solutions to Combinatorial Problems, *Science*, 266(5187):1021-1024.
- [2] Mao, C., LaBean, T., Reif, J., and Seeman, N., (2000) Logical Computation Using Algorithmic Self-Assembly of DNA Triple-Crossover Molecules, *Nature*, 407:493-495.
- [3] Rothmund P.W., (2006) Folding DNA to Create Nanoscale Shapes and Patterns, *Nature*, 440(7082):297-302.
- [4] Benenson, Y., Gil, B., Ben-Dor, U., Adar, R., and Shapiro, E., (2004) An Autonomous Molecular Computer for Logical Control of Gene Expression, *Nature*, 429:423-429.
- [5] Shapiro, E. and Benenson, Y., (2006) Bringing DNA Computers to Life, *Scientific American*, 294:45-51.
- [6] Zhang, B.-T. and Kim, J.-K., (2006) DNA Hypernetworks for Information Storage and Retrieval, *Proc. 12th Meeting on DNA Computing, LNCS 4287: 298-307.*
- [7] Kim, J.-K., Kim, B.S., Kwon, O.H., Hwang, S.K., Ha, J.W., Park, C.H., Chung, D.J., Lee, C.H., Park, J., and Zhang, B.-T., (2007) A DNA Computing-Inspired Silicon Chip for Pattern Recognition, *Proc. 13th Meeting on DNA Computing*, p. 373.