

# 컴퓨터 바이오과학과 바이오 컴퓨터과학<sup>†</sup>

서울대학교 장병탁\*

## 1. Biocomputation

컴퓨터과학과 생명과학은 활발한 상호작용이 일어나고 있다. 1953년 왓슨과 크릭에 의해 DNA의 구조가 밝혀진 이래로 분자생물학과 유전공학이 급격히 발전하였으며 1990년에 인간유전체프로젝트가 시작되면서 대규모의 실험 데이터를 컴퓨터에 저장하고 검색하며 분석하는 바이오인포매틱스(bioinformatics) 연구의 필요성이 대두되었다. 최근 들어, 컴퓨터의 도입으로 인해 생명과학 연구의 패러다임이 변화하고 있으며, 개별적인 유전자의 기능을 연구하던 기존의 방식으로부터 많은 유전자들의 상호관계 또는 한 개체 내에 존재하는 모든 유전자들(genome)을 연구하는 유전체학(genomics) 연구로 옮겨가고 있다. 또한 단백질의 관점에서 세포 내에서 발현되는 전체 단백질들 간의 상호작용을 연구하는 단백질체학(proteomics)을 연구하고 있으며, 이러한 경향은 생물학을 시스템 관점에서 연구하는 시스템 생물학(systems biology)으로 발전하고 있다. 이제 생명과학자들에게 있어서 컴퓨터 기술 및 정보과학적인 사고는 선택의 문제라기 보다는 필수 요소가 되어가고 있다.

한편, 생물학은 이미 오래 전부터 컴퓨터과학에 새로운 아이디어와 정보처리모형을 제공해 왔다. Wiener는 1948년에 생명체에서의 자기조절 능력의 중요성을 발견하고 유기체를 닮은 기계와 인공 시스템을 연구하는 Cybernetics 연구를 제안하였으며[6], von Neumann은 1958년에 발간된 그의 책 *The Brain and the Computer*에서 디지털 컴퓨터와 신경계 시스템을 비교하여 논하였다[5]. 1970년을 전후하

여 Fogel, Rechenberg, Holland 등은 각각 독립적으로 다윈의 진화론과 분자 유전학에 기반한 계산 모델인 진화연산 모델을 제안하였다. 1987년에는 Langton 등에 의해 생명체를 컴퓨터 상에서 시뮬레이션함으로써 생명현상을 탐구하고 이를 문제해결에 응용하려는 인공생명(artificial life) 연구가 시작되었다. 1994년에는 Adleman이 DNA 기반의 분자컴퓨터가 실현 가능함을 실험적으로 제시하였다. 2000년대 들어, 세포 내에서 일어나는 유전자의 발현 조절 과정을 이용하여 생화학적으로 논리연산을 수행하는 toggle switch가 개발되었다. 20세기에 실리콘칩 컴퓨터와 소프트웨어기술이 IT 발전의 원동력이 되었던 것처럼 21세기에는 바이오칩 컴퓨터와 바이오인포매틱스가 BT 발전의 원동력이 될 것이다.

본 고에서는 이와 같이 급변하고 있는 생명과학과 컴퓨터과학 또는 BT와 IT의 융합기술 분야에서 컴퓨터과학자 및 공학자들이 맞이할 새로운 기회와 시너지 효과 및 수행해야 할 역할에 대하여 논하고자 한다. 이 토론을 보다 원활하게 수행하기 위하여, 그림 1에 관련 연구 분야를 포괄하는 분류체계를 하나 제시하였다. 생물학과 관련된 컴퓨터과학 분야 전

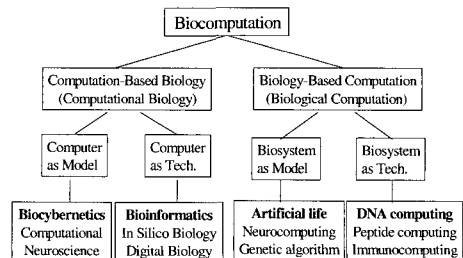


그림 1 Biocomputation 기술의 연구 분야 체계도

<sup>†</sup> 본 연구는 과학기술부 국가지정연구실 사업과 산업자립부 차세대신기술 사업에 의하여 지원되었음.

중신회원

체를 "biocomputation"이라고 하자면, 그 접근 방법에 따라 컴퓨터과학을 이용하여 생명과학을 연구하는 computation-based biology와 생명과학을 이용하여 컴퓨터과학을 연구하는 biology-based computation 분야로 구분할 수 있으며 이는 각각 "computational biology"와 "biological computation"으로 불릴 수 있을 것이다. 전자의 대표적인 예는 컴퓨터를 모델로 삼는 바이오사이버네틱스와 컴퓨터를 기술로 이용하는 바이오인포매틱스이며, 후자의 대표적인 연구 분야는 바이오시스템을 모델로 삼는 인공생명과 바이오시스템을 기술로 이용하는 DNA 컴퓨팅이다. 물론 이 밖에도 연구되었거나 현재 연구되고 있는 분야들이 존재하며 그 일부를 체계도에 포함하였다.

논문의 구성은 다음과 같다. 먼저 제 2절에서 분자생물학에 대한 기본적인 개념을 설명한 후, 제 3절에서는 컴퓨터 및 IT에 기반한 생명과학 연구로서 computational biology 분야의 연구 사례와 동향을 간략히 살펴본다. 제 4절에서는 반대로 생명과학에서의 발견과 생명공학적인 기술에 기반한 컴퓨터 및 IT 기술로서 biological computation 분야의 연구 사례와 동향을 살펴본다. 지면 관계상, DNA 분자를 정보저장 및 처리 매체로 사용하는 바이오 컴퓨터 기술을 연구하는 분자컴퓨팅(molecular computing) 분야에 관한 소개에 초점을 둔다. 마지막으로 제 5절에서는 21세기 생명공학과 나노기술의 시대에 컴퓨터의 발전 방향을 전망하고 컴퓨터과학자들의 역할에 대하여 토론한다.

## 2. 세포, 단백질, DNA

모든 생명체는 세포(cell)라는 기본 단위로 구성되어 있다. 사람의 몸에는 1014개의 세포가 존재한다. 모든 세포는 각각 한 쌍의 유전체(genome) 즉 유전자들을 DNA에 가지고 있으며 이는 진핵세포의 경우 세포핵 내에 존재한다. DNA가 생명체를 만들기 위한 청사진이라면, 단백질(protein)은 생명체를 만들고 생명활동을 유지하는 작업자에 해당한다. 하나의 동물 세포는 약 79억개의 단백질 분자로 구성되어 있다. 동물 세포의 평균 크기는 15 마이크로미터( $10^6\text{m}$ )이며 무게는 3.5 나노그램( $10^9\text{g}$ ) 정도이다. 그림 2는 세포 및 바이오분자들의 크기를 기존의 반도체 소자와 비교하여 제시하고 있다.

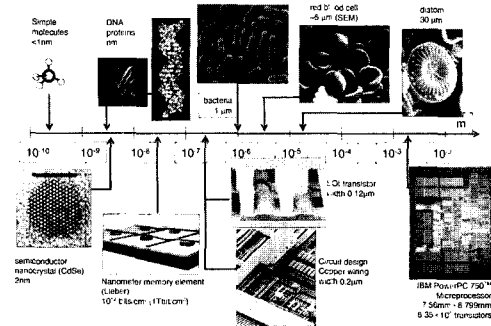


그림 2 세포, DNA, 단백질, 실리콘 소자의 비교

세포는 필요한 단백질을 합성해서 사용하며 합성될 단백질의 구조를 결정하는 정보 즉 설계도는 DNA가 가지고 있다(그림 3). 그러나 DNA는 한 카피만 존재하며 세포 활동은 많은 단백질을 필요로 하기 때문에 필요에 따라서 DNA를 복사해서 다양한 단백질을 만들며 이 때 DNA의 정보를 단백질 합성에 전달하는 분자가 mRNA이다. 이는 마치 책(DNA)이 도서관(세포핵)에는 한 질만 존재하고 사용자들은 필요한 내용(단백질)을 복사해서(mRNA) 사용하는 것과 유사하다. 단백질은 화학적으로는 아미노산(amino acid)들이 펩티드 결합을 통하여 형성되는 중합체(polymer)이며 보통 몇 백 내지 몇 천 개의 아미노산으로 구성되어 있다. 단백질 합성은 세포내의 리보솜에서 일어나며 이 때 tRNA는 mRNA에 있는 정보로부터 아미노산을

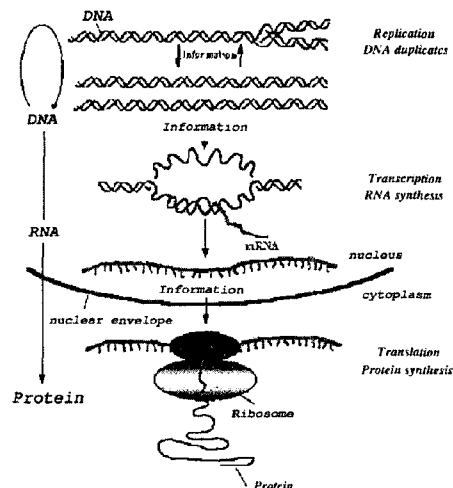


그림 3 DNA 유전 정보로부터의 RNA 전사 및 Protein 합성 과정

합성하는 매개체 역할을 한다. DNA, RNA, 단백질은 일반화학에서 다루는 분자들 보다는 아주 많은 수의 원자로 구성되어 있어 큰 분자량을 갖기 때문에 거대 분자(macromolecules)로 일컬어진다.

세포 내에서는 끊임없이 정보 활동이 일어나고 있다. 이는 크게 세 가지 즉 (1) 생명유지를 위한 대사 작용(metabolic networks) (2) 세포 내 유전정보의 전달(genetic networks), (3) 세포 외부의 자극에 대한 신호전달 과정(signaling networks)으로 요약할 수 있다. 대사 작용은 당분해 등 에너지 변환 및 전달 과정에 관여하며 기본적인 생명활동을 위해 필수적인 요소이다. 유전자망은 앞에서 언급한 DNA-RNA-Protein 간의 유전정보 전달에 관계되는 기작으로서, 실제로는 단순한 일방향이 아니라 합성된 단백질이 효소작용을 하여 다른 유전자의 발현을 조절하는 재귀적 자기조절 기작이 알려져 있으며, 이러한 유전자 조절 기작이 잘못될 경우 필요한 유전자가 발현되지 않거나 과다하게 발현되어 유전적인 질병을 유발할 수 있다. 세포는 생명체를 가진 가장 작은 단위로 독자적으로 외부의 자극에 대해 반응할 수 있는 능력을 가지고 있다. 즉 외부에서 어떤 화학 물질, 항원, 약물 등에 대해 세포막에서 수용체(receptor) 단백질이 반응을 보이며 이 신호들은 복잡한 기작을 통하여 내부로 전달된다. 특히 신경세포는 이러한 자극을 전달하여 정보처리에 관여하며 특히 반복적인 자극의 경우 분자수준에서의 기작에 의해 기억과 학습이 일어난다.

이와 같이 세포내의 정보처리 과정에 기반하여 볼 때 세포는 하나의 거대한 생화학 컴퓨터로 볼 수 있다[9]. 이 컴퓨터는 정보저장 및 전달 매체로 생체분자를 사용하며 정보처리를 위해서 열역학적 생화학 반응을 이용한다. 예를 들어, DNA와 RNA분자는 A, T (RNA에서는 U), G, C 개의 염기로 구성된 4진법의 정보를 저장하며, 단백질은 20가지의 아미노산으로 구성된 20진법의 정보를 저장한다. 세포 내에서 DNA 정보는 복제되기도 하고, 두 개의 DNA 가닥이 연결되어 새로운 정보를 생성하기도 하며, 기존의 DNA 가닥의 일부가 돌연변이를 통해 변경되기도 한다. 단백질의 경우 외부에서 오는 화합물과 결합하여 세포내로 신호 전달을 매개하기도 하고 칼슘이온 등을 공급 또는 배출함으로써 정보전달 및 통신을 수행한다. 효소(enzyme)들은 이러한 화학반응에 관여하여 자신은 변하지 않으면서 다른 화학반응의 속도를

조절하여 다른 정보활동을 돕는 단백질 분자로 된 정보처리 소자로 볼 수 있다.

### 3. Computational Biology

바이오인포매틱스(bioinformatics)는 생물학의 문제 해결을 위해서 바이오 데이터를 저장, 검색, 분석하기 위한 컴퓨터기술을 연구하는 분야로 정의될 수 있다. Computational biology는 (좁은 의미로) 컴퓨터를 도구로 사용하여 생물학적인 연구를 수행하는 분야를 일컫는 말로 주로 생물학자들의 입장에서 바이오인포매틱스를 가리킬 때 사용하는 용어이다. 광의의 computational biology는 바이오인포매틱스와 협의의 계산생물학을 포괄하며 in silico biology, digital biology 등으로 불리기도 한다.

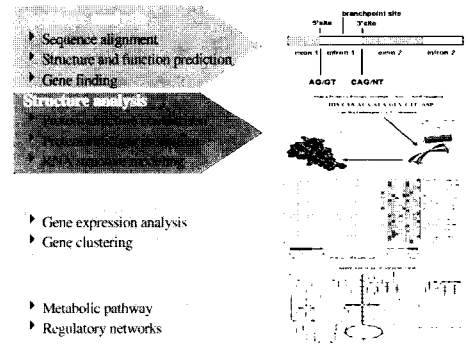


그림 4 바이오인포매틱스의 주요 연구 분야

바이오인포매틱스는 여러가지 방식으로 분류될 수 있으나, 한 가지는 대상으로 하는 바이오데이터의 종류에 따라서 분류하는 것이다. 이에 따르면 그림 4에 요약한 바와 같이, (1) DNA 서열에 기반하여 유전자를 찾고 그 기능을 밝히는 서열기반 생물정보학, (2) 단백질의 2차, 3차 구조를 밝히고 예측하는 구조기반 생물정보학, (3) DNA 칩이나 Protein 칩에서의 발현 양상을 분석하는 발현분석 생물정보학, (4) 유전자와 단백질의 상호작용을 분석하는 네트워크 생물정보학 등으로 나눌 수 있다.

바이오인포매틱스를 분류하는 또 다른 방식은 대상으로 하는 생명과학의 연구 분야로 구분하는 것이며, 유전체학(genomics), 약물유전체학(pharmacogenomics), 기능유전체학(functional genomics), 구조유전체학(structural genomics), 독성유전체학

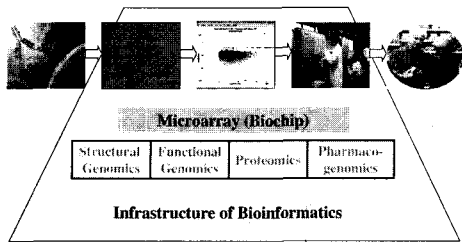


그림 5 바이오인포매틱스의 주요 응용 분야

(taxicogenomics) 등으로 나뉜다(그림 5). 그리고 최근에 와서는 유전체학뿐만 아니라 단백질체학(Proteomics), 대사체학(Metabolomics), 상호작용체학(Interactomics), 생리체학(Physiomics), 세포체학(Cellomics), 시스템생물학(Systems biology) 등의 용어로 기술되는 새로운 연구 분야들이 탄생하고 있다.

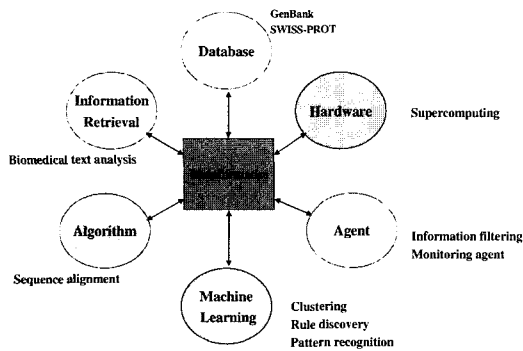


그림 6 바이오인포매틱스에 적용되는 컴퓨터 기술

바이오데이터를 분석하는 방법과 기술 관점에서 바이오인포매틱스는 컴퓨터과학의 모든 분야를 포괄한다(그림 6). 즉 대규모의 데이터를 저장 검색하기 위해 데이터베이스 기술이 기반이 되며, 서열데이터 간의 비교 정렬을 위해서 고속 패턴 매칭 알고리즘 기술이 필요하며, 이러한 알고리즘을 효율적으로 수행하기 위한 슈퍼컴퓨터 하드웨어 기술 및 전용 프로세서 개발 기술이 연구되고 있다. 인터넷 상에 분산되어 있는 많은 종류의 다양한 DB 및 문서데이터로부터 원하는 정보를 탐색하고 추출 분석하기 위한 정보검색 및 에이전트 기술과 데이터마이닝 및 기계학습 기술이 바이오인포매틱스를 위한 핵심기술로 사용된다. 그밖에 생물데이터 저장 및 분석 처리를 위한 인프라 구조로서 네트워크 및 통신, grid computing 기술 등이 활용된다.

바이오인포매틱스를 포함한 computational biology에 대한 보다 구체적인 연구 사례는 본 특집호의 다른 논문들에서 다루어지기 때문에 본고에서는 더 이상 다루지 않기로 한다. 아래에서는 computational biology와는 반대 방향으로 생명체에서의 정보처리 원리와 바이오소자의 특성을 활용하여 새로운 컴퓨터를 만들려는 시도인 biological computation 분야에 대해서 살펴보기로 한다.

#### 4. Biological Computation

실리콘 기반 디지털 컴퓨터는 계산을 반복적으로 정확히 수행하거나 대규모의 데이터를 저장하고 이를 정확히 검색하는데 뛰어난 성능을 보이고 있다. 반면에 모든 정보가 공간적인 위치를 가지고 있고 독립적으로 저장되며, 모든 정보처리 활동이 이 위치주소에 기반하여 이루어지기 때문에 정보 소자들 간의 대규모 상호작용에 기발한 정보처리에는 상당히 비효율적인 시스템이다. 현재 실리콘 컴퓨터로 잘 해결하지 못하는 문제들은 대부분 이러한 복잡계(complex systems)의 특성을 가진다. 한편, 앞의 2절에서 개괄한 바와 같이 세포 내에서 일어나는 정보처리 과정, 예를 들면 유전자 조절망은 많은 분자들이 액체 상태에서 3차원 화학반응에 의한 복잡한 상호작용을 통하여 이루어진다. 이러한 분자들의 망(molecular networks)에 의한 생화학적 계산 원리를 이해하고 이에 기반한 공학적인 기술을 개발하는 biological computation 연구를 통해서 현재의 실리콘 컴퓨터가 해결하지 못하는 많은 문제들을 해결할 수 있다. 표 1은 바이오분자 컴퓨터와 실리콘 컴퓨터의 특성을 비교하고 있다. 바이오컴퓨터의 가장 큰 특징 중의 하나는 생체정보를 직접 처리할 수 있다는 것이다. 이는 기존의 실리콘 기술이 바이오 정보를

표 1 분자 컴퓨터와 실리콘 컴퓨터의 특성 비교

	Molecular Computers	Silicon Computers
Material	Carbon	Silicon
Processing	Chemical (mobile)	Electronic (hardwired)
Medium	Liquid (wet)	Solid (dry)
Communication	3D diffusion	2D switching
Configuration	Amorphous (asynchronous)	Fixed (synchronous)
Parallelism	Massively parallel	Sequential
Speed	Fast (millisec)	Ultra-fast (nanosec)
Reliability	Low	High
Density	Ultrahigh	Very high
Reproducibility	Probabilistic	Deterministic

처리하기 위해서 전기화학적인 또는 광학적인 신호로 먼저 변환을 필요로 하는 것과 대조적이다.

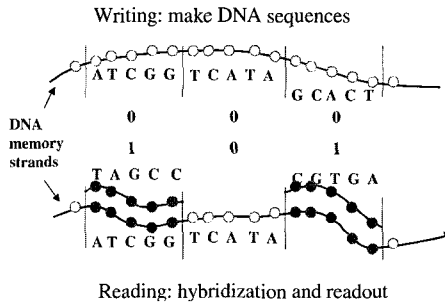


그림 7 메모리 소자로서의 DNA 분자

DNA를 사용한 분자 컴퓨팅이 가능함을 처음으로 보여준 사람은 (RSA 암호화 방식을 창안한 사람 중의 한 명인) Adleman이다. 그는 1994년 Science지에 발표한 논문 “Molecular Computation of Solutions to Combinatorial Problems”에서 DNA와 기초적인 분자생물학적인 실험방식을 사용하여 NP 문제인 Hamiltonian Path Problem을 polynomial time에 해결할 수 있음을 보였다[1]. 이 논문이 발표된 직후인 1995년부터 매년 ACM의 일부 지원하에 DNA Computing 국제학술대회가 개최되고 있으며 올해 2003년의 DNA9 학회는 Wisconsin 대학에서 열렸다. 현재 국제적으로는 미국, 독일, 영국, 일본, 캐나다, 이태리, 네덜란드 등에서 DNA 컴퓨팅 연구가 활발히 이루어지고 있으며[10], 국내에서는 1997년부터 서울대에서 기초 연구를 수행하였으며, 2000년부터는 산업자원부 차세대 신기술 사업의 수퍼지능칩 과제에서 서울대, 한양대, 서강대, (주)바이오니아, (주)바이오인포메틱스가 컨소시엄을 구성하여 DNA 컴퓨팅 연구(MEC: Molecular Evolutionary Computing)를 수행하고 있다. 오는 2006년에는 제12회 국제 DNA 컴퓨팅학술대회(DNA12)가 서울대학교에서 개최될 예정이다.

분자컴퓨터는 자연에 존재하는 또는 합성된 생체 분자를 물리적인 정보저장 및 처리 매체로 사용하여 액체상에서 3차원의 열역학적인 반응을 통해 초병렬적으로 정보를 처리하는 화학적 컴퓨터이다. 분자컴퓨터의 기본 정보처리 메커니즘은 Lock-and-Key 메커니즘이다. DNA의 경우 염기 A는 T와 결합하고

염기 G는 C와 결합하므로 이에 기반한 서열인 ATCGG (lock)는 TAGCC (key)와 기본적으로 상보 결합하는 특징이 있다. 이러한 특성을 사용하여 Read/Write가 가능한 DNA 메모리도 구성할 수가 있다(그림 7). RNA의 경우도 유사하며 (T 대신 U) 단백질의 경우 Receptor-Ligand 반응이 역시 단백질의 구조에 따라서 패턴을 인식함으로써 정보처리를 한다. 항원-항체(antibody-antigen) 반응의 경우도 구조에 따른 분자 인식에 기반하여 있다. 또한 촉매 반응(catalytic reaction)인 경우 효소와 기질의 반응 또한 자물쇠-열쇠 메커니즘에 기반한 분자 인식(molecular recognition)에 기반하여 일어난다.

분자 컴퓨팅의 대표적인 예인 DNA 컴퓨팅은 합성 DNA 분자를 정보 소자로 사용하며 풀고자 하는 문제에 대한 해를 DNA 코드로 기술하고 주어진 DNA 조각들로부터 해를 화학적으로 합성하고 검출함으로써 답을 찾는다. 특히 최근 들어 DNA 칩의 개발에 이와 같은 DNA 컴퓨팅 기술이 여러모로 응용될 수 있음이 밝혀지고, 또한 반대로 DNA 칩을 사용하여 DNA 컴퓨팅을 할 수도 있음이 알려짐에 따라 DNA 컴퓨팅 기술을 산업화 하는 움직임이 일고 있다. 또한 DNA 컴퓨팅 기술은 최근 DNA 조각들의 초병렬적인 분자인식 능력에 기반한 자기조립(self-assembly)을 통해 나노구조물을 구성하려는 나노기술로 발전하고 있다. 미국에서는 NSF와 DARPA의 지원하에 Caltech의 Winfree 그룹, Lucent Technologies의 Yirke 그룹, Duke의 Reif 그룹, NYU의 Seeman 그룹에서 활발히 연구되고 있다[7, 10].

DNA기반의 컴퓨팅 과정을 실험적으로 구현하기 위한 절차를 요약하면 다음과 같다.

- Step 1. (Design) 문제 및 해를 표현하기 위한 DNA 코드워드(codeword) 서열을 설계한다.
- Step 2. (Synthesis) 각각의 코드워드에 대한 DNA를 다량(보통  $10^{16}$ 개 정도) 분자 합성한다.
- Step 3. (Hybridization) 합성된 DNA 분자들을 하나의 시험관에 넣고 화학반응을 시킨다.
- Step 4. (Detection) 문제에 대한 해답에 해당하는 DNA 서열이 존재하는지 확인한다.

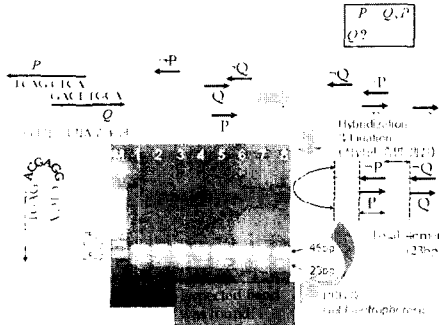


그림 8 DNA 컴퓨팅에 의한 Theorem Proving을 위한 코드워드의 설계 및 실험 결과의 Gel 사진[2]

- Nanoscale biomolecules as physical computational media
- Thermodynamics of computation  $P(x) = \frac{e^{-k(x)/kT}}{\sum_i e^{-k(x_i)/kT}}$ 
  - Energy
  - Temperature  $H(p) = -\sum p(x) \ln p(x)$
  - Entropy
- Information entropy  $D(r; p) = \sum r(x) \ln \frac{r(x)}{p(x)}$ 
  - Relative entropy  $H(r; p) = H(r) - D(r; p) = -\sum r(x) \ln p(x)$
  - Mutual information
- Statistical mechanics  $I(p, p_0) = H(p; p_0) - H(p; p) = \sum p(x) \ln \frac{p(x)}{p_0(x)}$

그림 9 분자컴퓨팅의 물리적, 정보이론적 기반

DNA 서열의 설계 단계는 현재 컴퓨터에서의 프로그래밍 단계와 유사한데, 아직 기술의 초기 단계이므로 여러가지 물리 화학적인 계산과정을 고려하는 저급언어(low-level language)에 의존한다(그림 8). 예를 들어서, 모든 DNA 분자반응이 비슷한 온도에서 일어날 수 있도록 하기 위해 모든 코드워드의 melting temperature  $T_m$ 가 같도록 코드워드를 설계한다(그림 9). 몇몇 연구 그룹에서 DNA 프로그래밍 시스템을 연구하고 있으며, 서울대에서는 1998년에 진화 연산 방식으로 코드를 최적화 하는 Molecular Programming 알고리즘을 제안하였으며[8], 최근 이에 기반한 설계 자동화 소프트웨어인 NACST (Nucleic Acid Computing Simulation Toolkit) 시스템을 개발하고 있다. 분자 합성은 분자합성기를 통하여 이루어진다. 반응시에는 polimerase, ligase, restriction enzyme과 같은 효소를 첨가하여 원하는 연산을 수행하거나 가속화 한다. 연산 결과는 PCR, gel electrophoresis, bead separation 등의 기술을 이용하여 광학적으로 확인하는데 이는 현재 컴퓨터의 출력 장치에 해당된다.

PCR은 DNA를 복제하는 기술로 해에 해당하는 DNA 분자의 개수가 적어서 검출이 어려울 경우 이를 검출하기 쉽도록 시그널을 강화하는 목적으로 사용된다. Gel electrophoresis는 DNA 분자들을 그 분자량에 따라서 정렬하는 생화학적 기술이다. DNA 분자가 음전하를 띠고 있는 성질을 이용하여 겔 상태의 판 위에 DNA 조각을 올려놓고 양의 전기를 걸어서 DNA가 겔 상태의 물질을 통과하게 만든다. 이 때 분자량이 작은 DNA 조각은 분자량이 큰 조각에 비해 빨리 움직여서 멀리 가는 성질을 이용하여 분자량에 따라서 DNA 조각을 분류한다. 분류시 이미 분자량을 아는 DNA 조각(marker)을 함께 이동시킴으로써 어느 위치에 어떤 크기의 DNA 조각이 있는지를 알 수 있다. 특기할 것은 이러한 정렬은 물리적인 이동을 필요로 하기 때문에 시간이 걸리기는 하지만 분자의 개수가 아무리 많아도 상수 시간에 연산이 수행될 수 있다는 것이다. 보통 다루는 분자의 수가  $10^{16}$  정도 되므로 이를 기존의 컴퓨터로 정렬하는 것은 실제적으로는 불가능하나 분자컴퓨팅 실험으로는 몇 분만에 가능하다. 최근에는 초소량의 시료를 사용하여 이러한 분석을 가능하게 하는 microfluidics 기반의 MEMS 및 Lab-on-a-Chip 기술들이 등장하고 있다[10]. 이러한 기술들이 바이오인포매틱스와 통합될 경우 DNA 컴퓨팅 기반의 지능형 차세대 바이오칩으로 발전할 수 있다(그림 10).

위에서 살펴본 바와 같이 분자컴퓨팅은 개념적으로 기존의 실리콘 컴퓨터 기술과는 대조적이다. 실리콘 기술은 반도체 기판 위에 복잡한 회로를 구성하고 전자를 원하는 시간에 원하는 곳으로 이동시키며 정

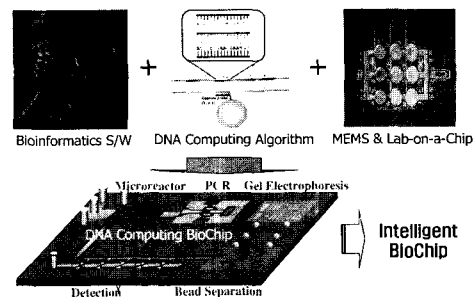


그림 10 Computational Biology, Biological Computation, Lab-on-a-Chip 기술을 융합한 DNA 컴퓨팅 기반의 차세대 지능형 바이오칩의 개념도 [10]

보를 처리하는 고체 컴퓨터인데 반해서 분자컴퓨터는 3차원 용액 상에서 반응이 일어나는 액체 컴퓨터이다. 고체 컴퓨터는 매우 정확한 계산을 잘 수행하고 재현성이 좋다는 장점이 있으나 반면 많은 메모리 소자 또는 정보처리 소자간에 동시다발적인 상호작용을 실현하기 어렵다는 단점이 있다. 이에 비해서 분자컴퓨터의 액체상에서의 확률적인 용액 반응은 정확하지는 않지만 초병렬적인 상호작용에 기반한 연산 수행에 적합한 장점이 있다. 이러한 특징은 많은 변수와 제약조건이 주어지는 공학적인 최적화 문제의 해결이나 복잡계 특성을 갖는 물리현상, 생명현상, 인지정보처리 과정 등을 시뮬레이션 하는 도구로서의 유용성을 시사한다.

분자 컴퓨터는 개념적으로 뿐만 아니라 정보저장 매체나 실현 기술에 있어서 기존의 전자기반 디지털 컴퓨터와는 다르다. 특히 아주 많은 수의 정보 입자들간의 초병렬적인 확률적인 상호작용을 통한 열역학적인 계산 방식은 이러한 특성을 이용할 경우 기존의 컴퓨터로는 수행하기 어려운 새로운 정보처리 패러다임을 생성할 수 있다. 예를 들면, 전통적인 기계 학습 방식을 DNA 분자로 구현한 학습능력을 갖는 바이오분자기계를 만들 수 있다[3, 11]. 또한 분자 컴퓨터를 신경과학이나 인지과학의 연구에 활용할 수 있다. 하나의 시냅스에서 일어나는 신경전달 물질에 의한 학습 현상을 분자 컴퓨터로 시뮬레이션 하는 방법이 연구되었으며[9], 이러한 방법이 구현될 경우 연상 기억(associative memory) 능력을 갖는 초고집적도의 나노메모리 시스템을 실현할 수 있다. DNA 컴퓨팅 기술을 이용하여 인간의 anagram 문제 해결 과정을 시뮬레이션 하는 시스템이 설계되었으며[9], 시뮬레이션 결과 이러한 모델은 기존의 실리콘 컴퓨터상에서 수행하기 어려운, 많은 규칙들의 비선형적인 복잡한 상호작용에 의한 언어처리 과정을 모델링 하는 도구로서 유용한 것으로 판명되었다.

## 5. 바이오 컴퓨터과학의 미래

생명현상에 관한 과학기술은 순수한 학문적인 호기심에서 뿐만 아니라 21세기 생명공학의 시대를 맞이하여 인류복지 향상에 기여하고 국가경제 성장의 견인차 역할을 할 중요한 경쟁요인이다. 본 고에서는 생명현상의 이해와 관련되는 BT 분야의 발전이 필요로 하는 IT 기술과 이러한 BT의 발전을 기반으로 어떻게 다시 IT가 혁신할 수 있는지에 관해서 살펴

보았다. BT와 IT의 융합기술인 바이오정보기술(BIT)을 논함에 있어서 한 가지 주의할 점은, 생명과학 분야에서 연구하는 사람들에게는 컴퓨터(또는 바이오인포매틱스)가 마치 과거에 현미경이 생물학 연구에서 행하였던 역할과 같은 도구 성격으로 이해된다는 것이다. 즉 현미경이 없으면 새로운 발견이 이루어지지 못하는 것은 사실이지만 중요한 발견을 하려면 현미경 기술 못지 않게 의미 있는 생물학적인 문제를 제기할 수 있는 능력을 필요로 한다. 이 사실은 컴퓨터과학자들(또는 bioinformatician들)이 어렵겠지만 받아들여야 할 현실이다.

그렇다면 급격히 발전하고 있는 바이오정보기술 분야에서 컴퓨터과학자들은 무엇을 할 것인가? 혹자는, 여전히 IT 분야에서 할 일이 많이 있으므로 새로운 기술의 등장과는 상관없이 지금까지의 전통기술을 더욱 발전시키고 응용하자고 주장할 것이다. 그러나 지금부터 20년쯤 후에 컴퓨터공학은 지금의 기계공학과 같은 시대를 맞이하게 될지도 모른다. 현재 기계공학에서 어떤 변화가 일어나고 있는지를 관찰해 본다면 컴퓨터과학의 장기적인 발전 방향에 대한 문제가 생각보다 심각하다는 것을 깨닫게 될 것이다. 학문이 발전할수록 대학에서 수행하던 연구의 많은 부분이 점점 기업으로 넘어가게 되고, 따라서 대학에서의 연구는 완성도를 높이는 기술 개발 보다는 새로운 길을 개척하는 연구를 수행하는 것이 전체 산업 및 국가 발전에 도움이 된다는 것은 이미 구미 선진국에서 몇 백년의 과학기술 발전의 역사를 통하여 입증된 사실이다.

컴퓨터과학자들에게 보다 큰 용기와 창의적 아이디어와 성실성이 요구되는 때이다. 즉 현존 컴퓨터 기술을 주로 응용하는 소극적인 bioinformatician 보다는 생명과학의 문제해결을 위해서 새로운 IT 기술을 개발하려고 시도하고, 더 나아가서 컴퓨터과학적인 사고와 도구로 생물학의 핵심 문제를 풀고자 하는 computational biologists가 더 많이 나와야 할 것이다. 일부 컴퓨터과학자들은 바이오인포매틱스 기술 자체가 의미가 있고 중요하므로 생물학과 생명현상에 적극적인 관심을 갖기보다는, 잘 정의된 문제를 생물학자들이 제기해 주기만을 기다린다. 그러나 생물학의 많은 문제가 그 특성상 정보과학적인 용어로 형식화해서 정의하기가 용이하지 않고 또한 그렇게 될 수 있다고 하더라도 이를 해 낼 수 있는 생물학자는 아직 얼마 되지 않는다. 물리학이나 화학과는 달

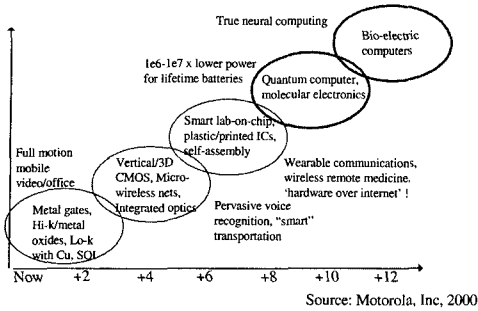


그림 11 미래 IT 기술의 발전 방향과 바이오분자 컴퓨터

리 생명과학은 명확한 이론이 제기되고 이를 철저히 검증하는 방식으로 발전하는 hard science의 성격이 아직은 미약하다. 그러나 현재의 발전 속도와 데이터 생성량의 증가 추세로 본다면 조만간 생명과학의 연구를 정량화 하고 형식화 하는 연구가 가능해 질 것이다. 따라서 정보과학적인 새로운 틀을 적극적으로 개발하고, 생명현상 자체의 규명과 그 응용 산업을 발전시키려는 용기 있는 컴퓨터 과학자와 사업가가 많이 필요하다. 그 반대 방향의 시도는 이미 이루어지고 있다. 즉 많은 생명과학자들이 최근 들어 컴퓨터 기술과 이론을 배우고 있다.

또 한편 생명현상의 이해와 이에 기반한 생명공학 기술을 다시 컴퓨터과학에 적용함으로써 궁극적으로 IT의 발전을 추구하는 biological computer scientists의 활약이 기대되는 바이다. 바이오분자로 구성된 유기체는 지능적인 정보처리 시스템의 존재하는 증명이며 이의 응용은 Post-Silicon 시대의 첨단 신기술로서 “바이오컴퓨터과학”이라고 부를 수 있는 새로운 IT 영역을 창출할 것이다. DNA와 단백질은 뉴클레오타이드와 아미노산을 각각 구성 요소로 하여 나노 수준에서 정보를 저장하고 조작하며 이는 nanotechnology(NT)와도 밀접한 관련을 갖는다. 뿐만 아니라 앞에서 살펴본 바와 같이 분자 컴퓨터는 wet 상태의 데이터를 실리콘 기술의 도움 없이 바로 처리할 수 있는 장점이 있어 의료진단, 신약개발 및 생명과학의 기초 연구 등 BT 분야에 새로운 응용 및 산업을 개척할 수 있을 것이다. 디지털 컴퓨터가 지금과 같이 폭넓게 활용되리라고는 1940~50년대에는 아무도 상상하지 못했던 것과 마찬가지로, 분자 컴퓨터에 대해서도 향후 20~50년 후에 어떤 응용 영역이 개척될지를 예측하는 것은 쉬운 일이 아니다. 그러나 한가지

가능한 예측은 지금의 실리콘 기반의 디지털 컴퓨터 기술은 분자 컴퓨터 기술과 함께 계속 발전할 것이며 분자 컴퓨터는 이를 개선하거나 대처하는 incremental technology라기 보다는 기존의 컴퓨터가 잘 하지 못하는 새로운 분야에서 두각을 나타내는 disruptive technology로서 더욱 발전할 것이다(그림 11). BT 산업의 “마이크로프로세서”로서의 분자 컴퓨터 칩과 인간의 기억과 학습 능력을 분자수준에서 모사하는 도구로서 바이오지능 컴퓨터는 이러한 예에 속할 것이다(그림 12).

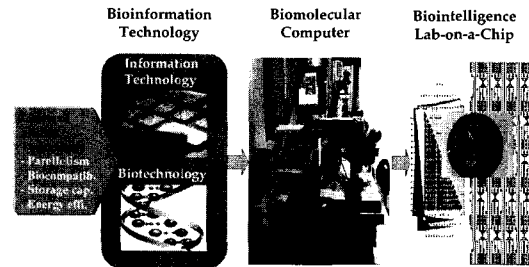


그림 12 바이오지능 컴퓨터 및 지능형 바이오칩 개발 방향

현재의 컴퓨터공학 교육 및 연구 체제에서 biological computer scientists를 배출하기는 computational biologists가 나오기를 기대하는 것만큼이나 어려운 일인지도 모른다. 그러나 지난 20세기에 컴퓨터와 IT에 대한 지식이 보편화 되었던 것처럼 오는 21세기에 생명과학과 BT에 대한 일반인의 지식수준이 올라갈 경우, 앞으로의 세대들은 컴퓨터와 생명과학을 동시에 공부하는데 큰 부담을 가지지 않을 것이다. 또한 최근 연합전공이나 협동과정과 같은 학제적인 교육 프로그램이 확산되고 있는 추세를 볼 때 이러한 인력들이 조만간 배출될 것이다. 따라서 현존하지 않는 새로운 분야에서 개척자적인 연구를 훌륭히 수행하고 그 결과를 선진국형의 고부가가치 산업으로 발전시킬 수 있는 “진정한” 과학기술자와 벤처사업가들이 국내에서도 많이 탄생될 수 있기를 기대해 볼 만하다.

### 참고문헌

[ 1 ] Adleman, L., Molecular computation of solutions to combinatorial problems, Science, 266, 1021-1024, 1994.  
 [ 2 ] Lee, I.-H., Park, J.-Y., Jang, H.-M., Chai,



Y.-G., and Zhang, B.-T., DNA implementation of theorem proving with resolution refutation in propositional logic, *Lecture Notes in Computer Science*, 2568, 156-167, 2003.

[3] Lim, H.-W., Jang, H.-M., Chai, Y.-K., Yoo, S.-I., and Zhang, B.-T., A lab-on-a-chip module for bead separation in DNA-based concept learning, *Proc. of 9th Annual Int. Conference on DNA Computing*, pp. 9-18, Madison, Wisconsin, 2003.

[4] Maley, C. C., DNA computation: theory, practice, and prospects, *Evolutionary Computation*, 6(3), 201-229, 1998.

[5] von Neumann, J., *The Brain and the Computer*, Princeton University Press, 1958.

[6] Wiener, N., *Cybernetics: or Control and Communication in the Animal and the Machine*, The MIT Press, 1948.

[7] Yurke, B., Turberfield, A.J., Mills, A.P. Jr, Simmel, F.c., and Neumann, J.L., A DNA fuelled molecular machine made of DNA, *Nature* 406, 605-608, 2000.

[8] Zhang, B.-T. and Shin, S.-Y., Molecular algorithms for efficient and reliable DNA computing, *Proc. Third Annual Genetic*

*Programming Conf. (GP-98)*, Koza, J.R. et al. (Eds.), Morgan Kaufmann, pp. 735-742, 1998.

[9] Zhang, B.-T., *Biomolecular Computation*, <http://bi.snu.ac.kr/Courses/>, 2003.

[10] 장병탁, 분자 정보처리 기술, *대한전자공학회지*, 29(4), 270-277, April, 2002.

[11] 장병탁, 학습하는 바이오분자 기계, *대한전기학회지*, 52(2), 44-52, February, 2003.

[12] 장병탁, 바이오분자 컴퓨터 기술, *한국물리학회지*, pp. 13-19, May, 2003.

---

장 병 탁



1986. 2 서울대학교 컴퓨터공학과(학사)  
 1988. 2 서울대학교 컴퓨터공학과(석사)  
 1992. 7 독일 Bonn 대학교 컴퓨터과학  
 과(박사)  
 1992. 8~1995. 8 독일국립정보기술연구  
 소(GMD) 연구원  
 1995. 9~1997. 2 건국대학교 컴퓨터공  
 학과 조교수  
 1997. 3~현재 서울대학교 컴퓨터공학부  
 조교수, 부교수

2001. 1~현재 서울대학교 바이오정보기술연구센터(CBIT) 센터장  
 2002. 6~현재 과기부 국가지정 바이오지능연구실 실장  
 주관심 분야 : Biointelligence, Molecular Computation, Learnable  
 and Evolvable Machines  
 E-mail : btzhang@cse.snu.ac.kr

---