

## 분자정보처리기술 (Molecular Information Processing Technologies)

장 병 탁

서울대학교 컴퓨터공학부

### 요 약

약 1그램의 DNA는  $10^{21}$ 개의 DNA 염기를 가지며 이것은 10억 terabits의 정보저장 능력에 해당한다. 분자정보처리기술(molecular information processing technology)은 DNA나 단백질과 같은 거대분자의 초미세구조, 초고집적도, 초저에너지 소비특성에 기반하여, 용액상에서 아보가드로수( $6 \times 10^{23}$ ) 규모의 분자들이 3차원 화학 반응에 관여하는 초병렬분산 정보 저장 및 처리 능력을 활용하는 차세대 IT 기술이다. 이는 또한 생체분자의 특성을 나노수준에서 활용하는 정보기술이라는 점에 있어서 BT+NT+IT의 융합기술이며 그 기술적 산업적인 파급 효과가 아주 큰 특징이 있다. 본고에서는 현재 분자정보처리 기술의 원리 및 문제점을 설명하고 이의 해결을 위한 5가지 핵심 요소 기술을 도출한다. 최근 세계적인 연구 동향을 살펴보고, 국내의 기술 기반 및 국제 경쟁력을 분석하며, 궁극적인 경제성 확보를 위한 산업화 방안을 단기(3-5년), 중기(5-10년), 장기(10-20년)의 3단계로 나누어 제안한다.

### I. 서 론

생체분자를 사용하여 정보처리를 하려는 분자정보처리기술에 관한 연구는 이미 1970년대 후반부터 시도되어 왔다. 초기의 연구들이 대부분 하

나의 세포 안에서의 반응 기작을 활용하려는 시도나 생체대사작용과 관련된 생화학적 반응들을 모델로 하여 연산을 시도하였으나 큰 진전을 보지는 못하였다. 그러나, 1994년에 Adleman은 Science에 발표된 그의 논문에서 DNA를 사용하여 컴퓨터과학에서의 난제(NP-hard) 중의 하나인 해밀톤 경로문제를 분자생물학의 도구를 사용하여 초병렬적으로 해결할 수 있음을 실험적으로 보임으로써 분자컴퓨팅 및 분자정보처리기술에 대한 새로운 시대를 열게 되었다.

분자정보처리기술은 컴퓨터기술(IT)의 기반 위에 분자생물학(BT) 실험기술이 밀접히 접목되어야 하는 융합기술로, 최근까지 비교적 느린 발전을 보아 왔다. 왜냐하면 분자생물학 실험 장비를 갖춘 컴퓨터과학 연구실은 지금까지는 별로 존재하지 않았기 때문이다. 그러나 최근 들어 바이오기술(BT)에 대한 중요성이 부각되고 정부나 산업에서 이러한 융합기술에 대한 지원의지가 강해짐에 따라 BT와 IT의 융합기술을 탐구하는 새로운 연구 그룹들이 세계적으로 형성되고 있다. 특히 DNA 분자정보처리기술은 나노수준에서 분자를 다룰 수 있는 나노기술(NT)과 밀접히 관련되어 있다. 즉 지금까지의 정보기기, 산업 장비, 의료기기 등에 있어서 기존의 실리콘 기반의 디지털 컴퓨터가 맡아왔던 역할들을 향후에는 바이오기술, 나노기술과 관련된 분자수준의 초소형 장치들이 응용되는 모든 분야에서 생체기반의 분자컴퓨터가 맡아하게 될 가능성이 크다.

본고에서는 분자정보처리시스템 개발의 기술적 측면과 산업적 측면에서 최근의 연구 동향을 살펴본다. 먼저 2절에서는 분자정보처리의 기반이

되는 분자컴퓨팅의 기본 원리와 요소기술을 소개한다. 3절에서는 최근의 외국 연구 동향을 알아 보며, 특히 바이오산업과 나노산업과 관련된 연구 내용들을 살펴본다. 4절에서는 이 분야와 관련된 국내 기술 기반을 조사하고 국제적인 경쟁력을 검토한다. 5절에서는 분자정보기술의 단기적, 중장기적 산업화 방안을 제안하고 6절에서 결론을 맺는다.

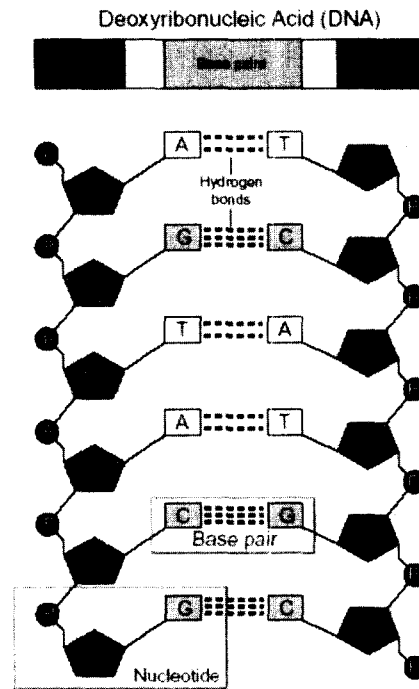
## II. 분자정보처리 원리

### 1. DNA 컴퓨팅

현재 가장 큰 실용화 가능성을 가지고 있는 분자컴퓨팅 기술은 DNA 분자를 이용하는 DNA 컴퓨팅 기술이다. DNA(deoxyribonucleic acids) 분자는 A(denine), T(hymine), G(uanine), C(ytosine)의 네 개의 뉴클레오티드가 반복적으로 이어진 사슬 구조를 하고 있다. 상온에서는 A와 T 그리고 G와 C가 Watson-Crick 상보성에 의해 결합된 이중나선(double helix) 구조로 존재한다<그림 1>.

DNA의 작은 조각을 올리고뉴클레오티드(oligonucleotide) 또는 올리고라고 하며 현재 바이오 텍 기술로 임의의 올리고 서열을 합성할 수 있다. 이들 합성된 올리고들을 하나의 시험관(test tube) 또는 반응기(reactor)에 넣으면, 단일가닥(single strand) DNA들이 Watson-Crick 상보성에 의해 결합함으로써 이중사슬구조(double strand) DNA를 형성한다. 이 반응은 용액상에서 초병렬적으로 일어난다. 또한 온도를 (예를 들어, 섭씨 90도 정도로) 높여주면 두 가닥으로 구성된 DNA가 한 가닥으로 분리된다.

DNA 기반 분자컴퓨팅은 기본적으로 DNA 분자를 합성함으로써 정보를 코딩하여 저장하고, DNA가 가진 화학적 특성을 이용하여 정보를 처리하는 연산 방식이다. 약 1그램의 DNA는  $10^{21}$  개의 DNA 염기를 가지며 따라서 10억 terabits의 정보저장 능력을 지닌다. 1 Mole의 DNA 수



<그림 1> DNA 이중나선 구조

용액에는 아보가드로수 만큼의 즉  $6 \times 10^{23}$ 개의 분자를 가지고 있으며 이들은 용액상에서의 화학 반응에 의해 초병렬적 정보처리가 가능하다. 분자정보처리기술은 이러한 많은 수의 초미세구조의 연산 소자가 초고집적도로 모여서 정보를 저장하고 초병렬적으로 처리함으로써 기존의 실리콘 기술로서 불가능한 정보처리 능력을 발휘할

<표 1> The Power of Large Numbers

Big Number	설 명
$7 \times 10^4$	인간 유전체의 유전자 수
$2 \times 10^6$	인류의 출현(년)
$3 \times 10^9$	유간 유전체의 염기서열 수
$3.5 \times 10^9$	생명 세포의 출현(년)
$1.5 \times 10^{10}$	우주의 나이 (년)
$10^{11}$	인간 두뇌의 신경세포 수
$10^{14}$	인간 신체에 존재하는 세포 수
$6.02 \times 10^{23}$	아보가드로의 수 (분자수/mole)

수 있는 잠재력을 지니고 있다. <표 1>은 아보가드로 수가 얼마만큼의 큰 수인지를 다른 거대 숫자들과 비교하여 제시하고 있다.

DNA 컴퓨팅의 기본 원리를 요약하면 다음과 같다.

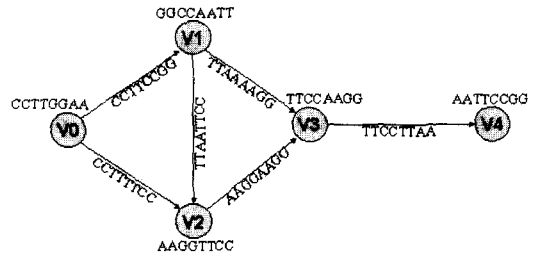
1. 풀고자 하는 문제에 대한 가능한 해답을 DNA 코드로 표현한다.
2. 이들 코드를 올리고 합성 기술을 사용하여 다량(보통  $10^{15}$  이상) 합성한다.
3. 각각의 성분 분자들을 합성기에 넣고 화학반응을 시킴으로써 가능한 모든 해를 생성한다.
4. 생성된 분자들 중에 찾는 해가 포함되었는지를 검사하여 답을 제시한다.

이러한 원리를 이용하여 해밀톤 경로 문제를 해결할 수 있음이 1994년에 Adleman에 의해 처음으로 실험적으로 증명되었고, 그 이후 여러 가지 계산 문제 및 다양한 응용에 대해 DNA 컴퓨팅이 연구되고 있다.

2. 문제해결의 예

해밀톤 경로 문제는 하나의 그래프 구조  $G=(V, E)$ 와 시작 정점(vertex)  $V_{in}$  및 도착 정점  $V_{out}$ 가 주어졌을 때, 모든 정점을 정확히 한번만 포함하며 전체 정점을 포함하는 경로가 있는지를 결정하는 문제이다. 즉, <그림 2>와 같이  $V_0, V_1, V_2, V_3, V_4$ 의 다섯 개의 정점으로 구성된 그래프에서  $V_{in}=V_0, V_{out}=V_4$ 라고 하면,  $(V_0, V_1, V_2, V_3, V_4)$ 는 하나의 가능한(이 경우는 유일한) 해밀톤 경로를 나타낸다.

이 문제를 풀기 위한 한 가지 방법은 모든 경로 가능성들을 생성한 후 이들이 주어진 그래프  $G$ 에서 가능한 하나의 경로인지를 검사한 후 하나라도 가능한 경로가 있으면 답은 yes이고 아니면 답은 no이다. 이러한 문제가 흥미있는 이유는, 문제의 크기 즉 정점의 수  $n$ 이 커지면 가능한 경로의 수가 지수적으로 증가하여 현재의 컴퓨터로는 실제적으로는 해결할 수 없는 NP의 복잡도를 갖는다는데 있다. 이에 대해 DNA 컴퓨팅은 한



<그림 2> 예제 그래프

가지 대안을 제시해 주고 있다. 즉, DNA 분자의 집적도를 활용하여 엄청난 수(적어도  $10^{15}$ 개 정도)의 경로를 용액상에서 병렬적으로(상수 시간에) 생성할 수 있는 가능성이 있다.

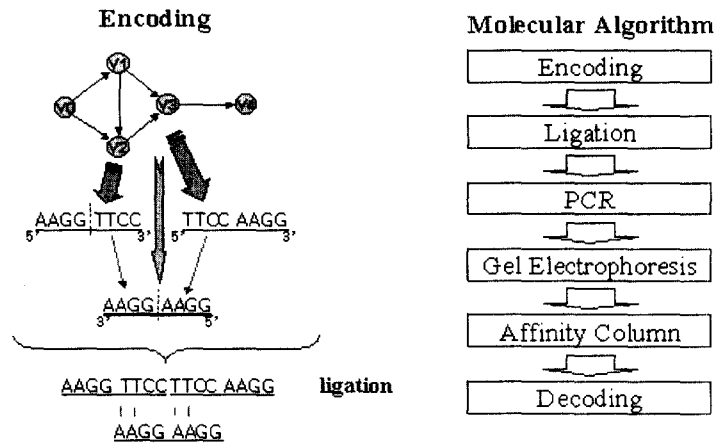
위의 그래프 문제를 해결하기 위한 구체적인 한 가지 방법은 다음과 같다(이 예는 설명의 편의를 위해 단순화 한 것이다. 실제 분자 실험을 고려한다면 여러 가지 추가 요인을 고려해야 할 것이다). 먼저 그래프의 각 정점  $V_i, i=0, \dots, 4$ 를 8개의 뉴클레오티드로 구성된 8mer의 올리고로 다음과 같이 코딩한다.

- $V_0=(V_0pre, V_0post)=CCTT\ GGAA$
- $V_1=(V_1pre, V_1post)=GGCC\ AATT$
- $V_2=(V_2pre, V_2post)=AAGG\ TTCC$
- $V_3=(V_3pre, V_3post)=TTCC\ AAGG$
- $V_4=(V_4pre, V_4post)=AATT\ CCGG$

위에서  $V_{ipre}$ 와  $V_{ipost}$ 는 정점  $V_i$ 의 코드에 대한 전반부, 후반부의 각각 4mer를 가리킨다. 또한 존재하는 모든 연결선을 다음과 같이 정의한다.

- $E_{01}=(\neg V_0post, \neg V_1pre)=CCTT\ CCGG$
- $E_{02}=(\neg V_0post, \neg V_2pre)=CCTT\ TTCC$
- $E_{12}=(\neg V_1post, \neg V_2pre)=TTAA\ TTCC$
- $E_{13}=(\neg V_1post, \neg V_3pre)=TTAA\ AAGG$
- $E_{23}=(\neg V_2post, \neg V_3pre)=AAGG\ AAGG$
- $E_{34}=(\neg V_3post, \neg V_4pre)=TTCC\ TTAA$

이제 정점에 대한 5가지의 올리고 및 연결선에 대한 6가지의 올리고를 DNA 합성기를 사용하



〈그림 3〉 경로 생성의 예

여 합성한다. 이렇게 하면 총 11가지의 올리고 종류들이 각각 약  $10^{15}$ 개씩 합성된다. 이들을 하나의 시험관에 넣고 서로 반응을 시킴으로써 모든 가능한 경로들을 생성한 DNA 분자들이 생성된다. 이 예의 경우 〈그림 3〉과 같은 Hybridization 과정을 거쳐 다양한 종류의 DNA 분자들이 생성될 것이다.

- P1=(V0)
- P2=(V0, V1)
- P3=(V0, V1, V2)
- P4=(V0, V1, V2, V3)
- P5=(V0, V1, V2, V3, V4)
- P6=(V0, V1, V3, V4, V4)
- P7=(V0, V1, V2, V3, V4, V4)
- P8=(V0, V1, V3)
- P9=(V0, V1, V3, V4)
- P10=(V0, V2)
- P11=(V0, V2, V3)
- P12=(V0, V2, V3, V4)
- P13=(V1)
- P14=(V1, V2)
- P15=(V1, V2, V3)
- .....

위의 경로 생성 예 P1-P15에서 올바른 해밀톤

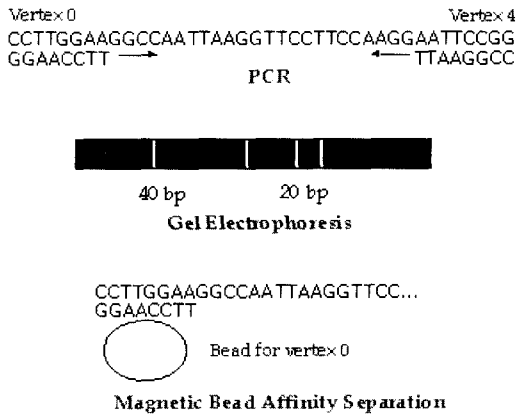
경로는 P5 하나 뿐이다. 또한 P6, P7과 같은 경로는 주어진 그래프에는 존재하지 않지만 화학적인 반응의 오류로 인해서 잘못 생성될 수도 있는 경로의 예를 나타낸다. P6의 경우, 원래 주어진 그래프에는 존재하지 않는 (V2, V4)와 (V4, V4)의 hybridization mismatch에 의해 틀리게 생성된 경로이며, P7의 경우 (V4, V4)의 연결선이 잘못 추가된 경로의 예를 나타낸다. 이러한 오류는 코드를 잘 설계함으로써 방지할 수 있으나 현재 기술 수준에서 완전히 제거하는 것은 쉬운 일이 아니다.

이와 같이 생성된 경로로부터 해밀톤 경로 문제에 대한 올바른 답을 다음과 같은 절차를 의해 검사한다.

- Step 1. 경로가  $V_{in}$ 로 시작하고  $V_{out}$ 로 끝나는지를 검사한다. 아닌 경로는 제거한다.
- Step 2. 경로의 길이(정점의 수)가 5인가를 검사한다. 아니면 제거한다.
- Step 3. 경로에 모든 정점이 포함되었는지를 검사한다.
- Step 4. 최종적으로 남은 DNA 분자가 있으면 답은 yes, 아니면 no라고 대답한다.

예로 들어, 경로 P2=(V0, V1)의 경우 Step 1에서 이미 제거된다( $V_{out}=V4$ 으로 끝나지 않

으므로). 경로  $P7=(V0, V1, V2, V3, V4, V4)$ 의 경우는 Step 1은 통과하지만 Step 2에 의해서 제거된다(경로에 포함된 정점의 수가 5가 아니라 6이므로). 경로  $P6=(V0, V1, V3, V4, V4)$ 은 Step 1과 Step 2를 통과하지만 Step 3에서 제거된다( $V2$ 가 경로에 포함되지 않았으므로).



〈그림 4〉 Bio-lab 실험 단계

현재 분자생물학 기법을 사용하여 위의 검사를 bio-lab 실험으로 수행하는 것이 가능하다. Step 1의 경우 원하는 서열 즉  $V_{in}$ 과  $V_{out}$ 에 대한 primer를 넣어 주는 중합효소연쇄반응(PCR)을 수행함으로써 시작점과 끝점을 포함한 DNA 분자만을 선택적으로 복제하는 실험 기술에 의해 구현될 수 있다. Step 2의 경우 전기영동법(Gel Electrophoresis)을 사용하여 분자량에 의해 정렬함으로써 길이가 정확히 40mer인 (8mer $\times$ 5개의 정점) 분자만을 선별함으로써 구현된다. Step 3은 Bead를 사용하여 검사할 수 있다. 즉, 각각의 정점에 해당하는 코드를 지닌 Magnetic Bead를 만들어 해당 코드가 있는 올리고를 자기적으로 선별해 냄으로써 모든 정점이 경로에 한 번 이상은 나타났는지를 검사할 수 있다. Step 4에 도달했을 때, 올리고가 test tube에 남아 있으면 이것은 해밀톤 경로에 해당한다. 따라서 답은 yes이고 아니면 답은 no이다.

### III. 핵심 요소기술의 도출

#### 1. 현재 기술의 문제점

앞 절에서 살펴본 바와 같이 분자정보처리는 자연 현상을 최대한 활용하는 기술이며 이론적으로 그리고 실험적으로 그 가능성이 증명된 방법이다. 그러나 실용화를 위해서는 아직 다음과 같은 몇 가지 기술적인 문제점들을 해결하여야 한다.

〈표 2〉 현재 분자정보처리기술의 주요 문제점

- |   |
|---|
| <ul style="list-style-type: none"> <li>- 분자반응의 정확도 및 재현성</li> <li>- 목표 분자들의 정확한 검출</li> <li>- 대규모 문제 해결 능력</li> <li>- 수작업 바이오 실험의 한계</li> <li>- 데이터 입출력 방법</li> </ul> |
|---|

첫 번째는 정보처리의 정확도 문제이다. 이는 생체분자들의 화학반응과 관계가 있으며 DNA의 혼성화 반응의 경우 mismatch가 일어나지 않도록 하는 기술을 개발하여야 한다. 한 가지 방법은 분자들이 올바르게 반응하지 못하도록 DNA 코드를 설계하는 것이다. 연산의 정확도를 개선하는 또 다른 방법은 PCR, GE 등 실험 기술을 개선하거나 새로운 실험 방법을 개발하는 것이다.

두 번째의 해 검출 문제는 분자 및 원자 수준에서의 물질 조작 기술과 관계가 있다. 현재 DNA 분자를 관찰하거나 조작하기 위해서는 원자현미경(AFM), 전자현미경(SEM, TEM) 등이 주로 사용되는데 이러한 관련 기술의 발전을 필요로 한다. 또한 용액이나 표면상에서의 광학적인 검출을 위해서 fluorescence, nanophore 등이 활용되는 방법이 최근에 등장하고 있다.

세 번째, 대규모 문제 해결 능력 향상은 새로운 실험 기술의 개발을 필요로 한다. 이론적으로 DNA 컴퓨팅을 활용하여 Turing Machine을 구현할 수 있음이 증명되었다. 그러나, 실험적으로 대규모의 문제를 다루기 위해서는 임의 크기의 DNA를 합성하고 이를 PCR 등을 통해서 조작할 수

있는 기술이 필요한데 이는 기존의 분자생물학 실험 기법만으로는 부족하다.

네 번째는 정보처리의 속도 문제이다. 이는 진정한 의미에서 분자연산의 병렬성을 활용하기 위해 중요하다. 즉 분자정보처리는 각 실험 단계에서는 분자의 초병렬적 연산 특성을 활용할 수 있으나, 단계간의 작업이 아직은 주로 수작업으로 이루어짐에 따라서 전체 연산 시간은 오래 걸린다. 이를 해결하는 한 가지 방법은 실험 과정을 자동화하는 것이다. 또한 일부 실험 기술은 아직 많은 시간을 필요로 하는데(예를 들어, PCR), 이들 실험 장치들의 속도를 향상하는 기술이 개발되어야 한다.

다섯 번째는 입출력 문제이다. 현재 데이터를 넣어주기 위해서는 올리고를 offline으로 합성해서 반응기에 넣어주어야 한다. 또한 결과의 출력을 위해서는 주로 Gel이나 형광물질을 사용하는데 이들 장비들은 정보처리 모듈과 online으로 연결되어 있어 편의성과 효율을 저해하는 요인이 되고 있다.

2. 핵심 요소기술

앞 절에서 논의한 여러 가지 문제점들을 해결하기 위한 요소 기술들은 다섯 가지 정도로 요약할 수 있다. 즉 모델링 기술, 실험, 제작, 분석, 응용 기술이며 이를 <표 3>에 정리하였다.

첫 번째의 모델링 기술은 생체분자들의 정보처

리 과정을 시뮬레이션하고 이의 모델을 구축하기 위한 소프트웨어 기술이다. 이러한 기술은 전통적으로 생화학이나 분자생물학 분야에서 순수 자연과학적인 목적으로 지금까지 연구되어 왔으나, 최근 들어서 공학적인 측면에서 활용 가능성을 많이 시사하고 있다. 특히 DNA 올리고 서열들의 hybridization 과정에 관련된 열역학적, 생화학적 모델링을 통해 정확한 분자반응을 하는 코드를 디자인할 수 있다.

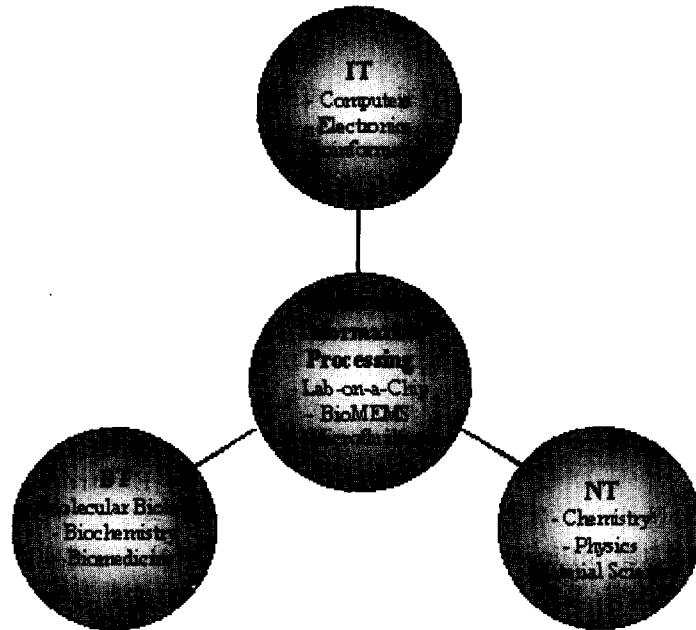
두 번째 요소기술로는, 분자정보처리를 위해서 원하는 분자들을 분석하고 조작하며 추출할 수 있는 실험 기술을 들 수 있다. 현재 사용되고 있는 PCR, Gel Electrophoresis, Beads, 분자형광기술 등의 용량과 정확도를 개선하고 궁극적으로는 단일분자 수준에서 개개의 분자들을 분리해 낼 수 있는 실험 장치 및 기법을 개발하여야 한다. 특히 이 기술은 원자 현미경 등을 다루는 나노기술과도 밀접한 관련이 있다.

세 번째로, 분자정보처리를 하는 마이크로시스템 제작 기술이다. 분자정보처리 기술이 기존의 실리콘 기반 정보처리 기술과 구별될 수 있는 한 가지 특징은 초소형화 가능성이며 이 특징을 최대한 활용하여 산업화 하기 위해서는 생물학적 실험 기술들은 궁극적으로 바이오칩 형태로 구현하는 기술이 연구되어야 한다. 구체적으로, 마이크로 PCR, 마이크로 Capillary Electrophoresis, 마이크로 비드 제작 기술 및 Microfluidics 칩 등을 포함한다. 칩화하는 것은 또한 실험 시간을 단축하고 자동화하기 위해서도 중요하다.

네 번째는 실험 및 제작 장치와 관련된 분석 소프트웨어를 개발하는 것이다. 이는 분자컴퓨터의 입출력 장치의 개발과 관련이 있다. 현재까지 사용되는 분자생물학 기술은 생명과학 연구를 위한 것이었으며 이것은 분자컴퓨팅과 같은 정보처리 관점에서의 기술로는 부족한 점들이 많이 있다. 일례로, 생물학자들은 많은 실험의 결과를 Gel Electrophoretogram을 눈으로 확인함으로써 확인하는데, 이러한 출력 장치를 보다 편리하게 하는 기술을 개발할 필요가 있다. 마찬가지로 분자정보처리칩을 구현하기 위한 기술과 관련하여 새

<표 3> 분자정보처리 요소 기술

기술 분야	구체적인 요소 기술의 예
모델링 기술	DNA 분자 시뮬레이션, 분자 Probe/Library 설계
실험 기술	단일분자 분석/검출/조작, DNA 자가조립, 분자 프로그래밍
제작 기술	Micro PCR/CE/Bead, Microfluidics, 집적 및 패키징
분석 기술	실험 과정(PCR) 및 출력 결과(Gel, 형광) 분석 소프트웨어
응용 기술	분자 진단/의약학, 식품환경에너지, Gene/Drug Screening



〈그림 5〉 BT+IT+NT 융합기술로서의 분자정보처리기술

로운 분석 소프트웨어의 개발을 필요로 할 것이다.

다섯 번째는 산업화 및 상품화와 관련된 응용 콘텐츠 개발 기술이다. 분자정보처리기술이 궁극적으로 경제적인 가치를 발휘하기 위해서는 하나의 독립적인 상품으로 구체화될 필요가 있으며 이를 위해서는 의약학, 식품, 환경, 에너지 등 생명공학의 전 분야에 걸쳐 분자정보처리기술을 응용하기 위한 콘텐츠 개발 기술이 병행하여 연구되어야 한다. 특히 인간유전체프로젝트 연구 결과로 생성된 많은 분자 및 유전자 데이터베이스와 문헌 정보로부터 유용한 지식을 추출하기 위한 바이오인포매틱스 기술이 함께 개발되어야 할 것이다.

분자정보처리기술의 특징 중의 하나는 다양한 연구분야의 기술을 필요로 하는 복합 및 융합 기술이라는 점이다.〈그림 5〉 문제 자체의 성격으로 볼 때, 정보기술(IT)이 근간이 되며, 생체분자를 기반으로 하고 분자생물학의 실험 기술을 다루는 점에서 바이오기술(BT)이며, 궁극적으로 나노단위에서 단일 DNA나 단백질 등의 분자들을 설계, 조작, 검출하여야 하는 점에서 나노기술(NT)

이다. 이들 기술이 융합되어 구현된 최종적인 분자정보처리시스템은 하나의 Lab-on-a-Chip과 같은 형태가 될 것이다.

#### IV. 최근 연구 동향

분자정보처리와 관련된 최초의 연구 과제는 미국 과학재단(NSF)이 1996-2000년까지 지원한 Biomolecular Computation (BMC) 프로젝트이다. 이 과제에서는 Caltech, USC, Princeton, Duke, NYU 등의 대학이 공동으로 학제적인 컨소시엄을 형성하여 분자컴퓨팅의 실험기술과 응용을 개발하는 연구를 하였다. 이 과제는 2001년부터 다시 4년간 2단계 연구를 지원 받았으며 여기서는 DNA Nanostructure의 Self-assembly를 이용한 분자연산 및 정보처리기술 개발 및 응용에 주력하고 있다. 최근에 이 기술을 Nanorobotics와 Cryptography에 응용하는 연구를 시도하고 있다. MIT 공대에서는 2001년 미생물

〈표 4〉 분자정보처리기술의 주요 연구 과제 및 발표 결과

년도	주요 연구 과제 및 결과
1994	DNA 분자컴퓨팅 가능성 실험적으로 제시 (Science지 발표)
1995	- Science지 다수의 DNA 컴퓨팅 논문 게재 및 토론 - 제1차 국제 DNA-Based Computer Workshop 개최
1996	미국 NSF 지원 Biomolecular Computing (BMC) 프로젝트 시작 (1996-2000)
1997	일본 상무성 지원 Molecular Computer Project (MCP) 시작 (1997-2000)
1998	DNA Self-assembly 기반 범용 분자연산 가능성 검증 (Nature지 발표)
1999	독일 과기부 GMD 국립연구소 내에 BioMIP 신설 및 DNA 컴퓨팅 프로젝트 (DNACom) 지원
2000	한국 산자부 차세대신기술 Molecular Evolutionary Computing (MEC) 과제 시작
2001	- 일본 동경대 연구팀 Science지에 DNA 컴퓨팅 논문 발표 - 미국 NSF 지원 Biomolecular Computing 프로젝트 2단계 착수 - 일본 상무성 지원 Molecular Computer Project 2단계 착수 - 이스라엘 Hebrew대 연구팀 Nature지에 DNA Computer 논문 발표
2002	- Olympus와 동경대 공동연구팀 최초의 진단용 DNA 컴퓨터 발표 - 제8차 국제 DNA-Based Computer Workshop 개최

을 통해 분자컴퓨팅 방식으로 식물의 유전자를 프로그래밍함으로써 지시된 시기에 프로그램된 행동을 하는 식물을 개발하는 연구를 발표하였다. Princeton 대에서는 DES 암호를 해독하는 DNA 컴퓨터를 개발하고 있다.

유럽에서는 1997년부터 EU의 분자정보처리 컨소시엄으로 ECMC라는 조직을 결성하였다. 현재 네덜란드의 Leiden 대학에 본부를 두고 있

으며 영국, 독일, 네덜란드, 핀란드 등이 이 핵심 구성원으로 이 컨소시엄에 가입해 있다. 독일의 경우 독일국립정보기술연구소(GMD) 내에 1999년에 BioMIP(Biomolecular Information Processing) Institute를 신설하고 정부에서 향후 5년 동안 DNA Computing 연구를 지원하고 있다. BioMIP (2001년부터 Fraunhofer Institute로 조직 개편)에서는 Microfluidics 및 MEMS 기술을 사용하여 DNA 기반의 분자진단용 Lab-on-a-Chip 컴퓨터를 제작하는 연구를 수행하고 있다. Leiden 대학의 LCNC에서는 섬모층의 Gene-Assembly 과정에서 일어나는 분자정보처리 기작을 이용하여 DNA 컴퓨팅을 하려는 연구를 하고 있다.

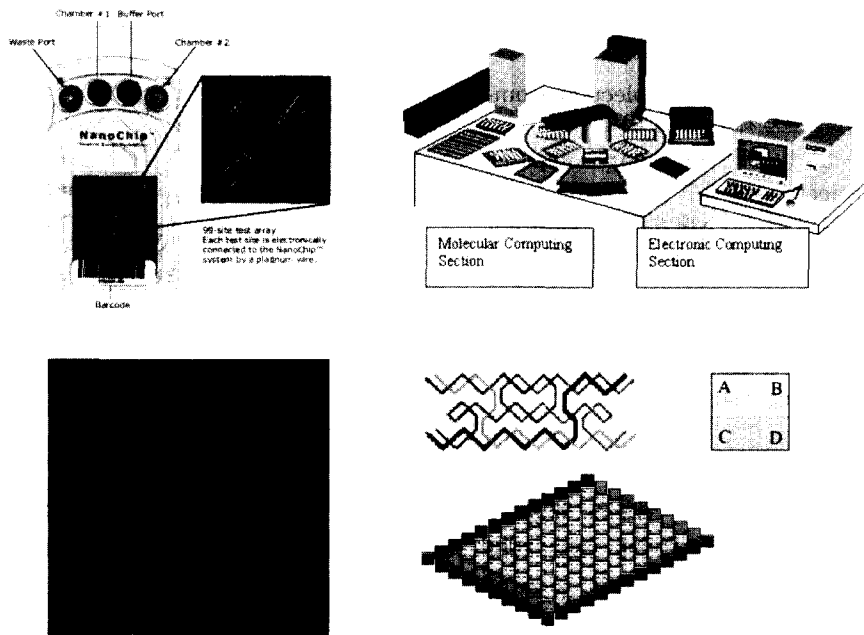
일본에서도 1997년부터 2000년까지 상무성 지원하에 동경대를 중심으로 한 산학컨소시엄에서 MCP(Molecular Computer Project) 과제를 시작하였으며 연구 결과가 2000년 Science지에 발표되었다. 이 과제는 2001년 2단계 연구에 착수하였으며 나노기술과의 접목을 시도하고 있다.

최근 들어 분자정보처리기술을 기반으로 파생된 새로운 아이디어를 사업화 하는 벤처들이 나타나고 있다. 광학기기 회사로 잘 알려진 Olympus는 동경대의 DNA 컴퓨팅 기술을 이전 받아 온라인으로 유전자 발현을 분석하여 분자진단을 가능하게 하는 진단용 DNA Computer를 개발하였다고 발표하였다. 지능형 DNA 칩이라고 부르는 이러한 DNA 컴퓨터 칩은 기존의 Microarray 기반의 DNA 칩 기술을 대체할 수 있는 혁신적인 기술로 평가되고 있다. 또 다른 예로, 독일의 Informium은 Dortmund 대학에서 DNA 컴퓨팅 및 Artificial Chemistry 기반 분자정보처리 연구 중에 고안된 Programmable Polymer를 이용하여 자연 DNA를 labelling함으로써 암호화하는 기술을 개발하여 DNACrypt라는 제품명으로 보안 시장에서 시판하고 있다. 미국에서는 Wisconsin 대학에서 Surface Chemistry 기반의 분자 Computing Chip을 연구하여 상용화를 앞두고 있다. NanoGen은 분자정보처



〈표 5〉 분자정보처리기술 연구 그룹

연구 기관 또는 그룹	대표연구자	연구 주제
ATR HIP 그룹	Shimohara	Evolvable Molecular Hardware
Caltech	Winfrey	DNA 나노구조 자가조립 연산
Duke U.	Reif	초대용량 분자소자 데이터베이스 개발
ECMC	EU	유럽연합 분자컴퓨팅 컨소시엄
Georgia Tech	Lipton	DNA기반 암호기술
Fraunhofer I. BioMIP	McCaskill	분자컴퓨팅 Reconfigurable Lab-Chip
LCNC	Rozenberg	Gene-assembly
Lucent Technologies	Yurke	DNA 나노머신
MIT	Knight	Amorphous Biological Computing
SNU Biointelligence Lab	Zhang	DNA 분자진화컴퓨팅 (MEC)
Tokyo U. & Olympus	Suyama	지능형 DNA Chip
U. Dortmund	Banzhaf	Artificial Chemistry
U. Southern California	Adleman	분자컴퓨팅
Wisconsin U.	Smith	Surface Chemistry DNA 컴퓨팅 칩



〈그림 6〉 최근 연구 기술

리기술의 개념에 입각하여, DNA 분자의 Hybridization 반응을 전자적으로 제어 및 Progra-

mming 할 수 있는 범용의 DNA “프로그래밍” 칩을 출시하고 있다.〈그림 6〉

### V. 국내 기술 기반 및 국제 경쟁력

분자정보처리기술은 BT, IT, NT 및 MEMS 등을 기반으로 하며 따라서 생명공학, 응용화학, 컴퓨터공학 및 기계전자공학 분야의 기술이 총체적으로 융합되어야 발전할 수 있는 학제적인 연구 분야이다. 여기서는 이 분야에 대한 국내의 기술 기반을 조사하고 이 분야에 대한 국제 경쟁력을 점검하고자 한다. 이를 위해 요소 기술 개발을 다시 한 번 <표 6>에 요약하였다. 그리고 이와 관련된 BT, IT, NT 및 MEMS 기반의 국내 연구 그룹들의 예를 <표 7>에 열거하였다.

분자 모델링 소프트웨어 기술의 경우, 주로 외국의 제품이나 프리 소프트웨어에 의존해 오고 있어 이 분야의 국내 기술 기반은 취약하다고 볼 수 있다. 이러한 취약성은 이 기술이 고도의 지식 집약 기술인데 반해서 수요가 대학이나 연구소의 실험실 정도로 제한되어 있어서 경제성이 부족한데 원인이 있는 것으로 보인다. 그러나 최근 들어, DNA칩이나 단백질칩을 개발하는 바이오센서와 생명공학산업에서 이러한 소프트웨어 툴의 필요성을 호소하고 있으며, 앞으로는 IT 산업의 관점에서 많은 수요가 창출될 것으로 보인다.

분자 실험 기술의 경우 분자생물학 기술이나 분석화학 기술이 기반이 되어 연구되어 왔다. 이 분야의 기술은 많은 바이오센서들이 가지고 있으므로 이를 분자정보처리기술로 전환한다면 조기에 IT용 BT 기술 확보가 가능할 것으로 보인다. 다만 현재까지는, BT 연구자들의 IT 기술에 대한 관심 부족으로 이러한 융합 기술이 발전의

<표 6> 분자정보처리 요소기술 개발 과제

- 분자정보처리 모델링 및 시뮬레이션 소프트웨어 기술
- 분자정보처리를 위한 단일분자 분석/조작/추출 실험 기술
- 분자정보처리용 MEMS 및 Fluidics 바이오 칩 제작 기술
- 분자정보처리 실험 분석 및 칩제작 지원 소프트웨어 기술
- 분자정보처리 응용컨텐츠 개발 바이오정보 수집/분석/추출 기술

지연되어 온 것으로 보인다. 마찬가지로, 최근 물리나 화학 또는 재료공학 기반의 나노기술 연구자들이 새로운 정보기술 개발에 참여하기 시작하였다. 예를 들어서, 전통적인 형광법에 의한 분자 검출 기술 대신 여러 가지의 nanoparticle, nanopore 및 nanobiosensor 기술이 연구되고 있는데 이는 분자정보처리를 위한 핵심기술로 활용될 수 있다. 최근, 나노기술관련 국가 과제들에서 이러한 연구를 수행하기 시작하였다.

바이오칩 제작 기술은 MEMS 연구자들이 주로 연구하고 있으며, 이 분야는 국내 기반기술이 비교적 잘 확보되어 있는 영역이다. 특히 국내 기술의 강점인 반도체 개발에 사용되던 실리콘 기반의 마이크로칩 기술을 기반으로 분자정보처리 시스템 개발에 필요한 BioMEMS, Fluidics 및 Lab-on-a-Chip 기술을 확보할 수 있다. 이미 관련 연구로, 과학기술부에서 추진하고 있는 지능형 마이크로시스템 프론티어사업이 있고, 산업자원부의 차세대신기술로 지원되는 DNA 칩 프

<표 7> 분자정보처리 요소기술 관련 국내 연구 그룹

기술 분야	국내 연구 그룹
Biotechnology	생명공학연구원, SK 생명공학, LG 화학, 녹십자, 바이오니아, 인바이오넷, 판제노믹스, 바이오메드랩
Nanotechnology	나노바이오연구소(부산대), 나노과학기술연구소(과기원), 초미세생체전자연구센터(서울대)
BioMEMS	ETRI, 삼성종합기술원, LG중앙연구원, BeadTech, DBT, 마이크로바이오칩센터(한양대), 마이크로시스템기술센터(서울대)
Bioinformatics	마크로젠, 디지털지노믹스, 아이디알, 바이오인포메틱스, 바이오정보기술연구센터(서울대)

로젝트와 Protein 칩 프로젝트가 있다. 이러한 과제들을 통해 현재 산학연에서 이 분야의 핵심 원천 기술이 연구되고 있기 때문에 이들 기술을 활용할 경우 분자정보처리기술 분야에서도 국제 경쟁력 확보가 가능하다.

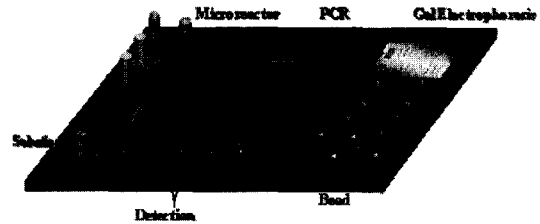
바이오 실험 분석 및 칩제작 지원을 위한 소프트웨어 기술의 경우 국내 연구가 가장 취약한 부분중의 하나이다. 근본적인 이유는 이러한 연구가 아직 국내에서 많이 수행된 바가 없다는데 있다고 볼 수 있으며, 위에서 언급한 분자 시뮬레이션 및 분자 모델링 소프트웨어와 마찬가지로 이러한 소프트웨어는 BT, NT의 기반 및 응용 연구를 위해서 앞으로 많은 수요가 창출될 것이므로 새로운 IT 사업 부문으로 떠오를 것이다.

분자정보처리시스템 응용 콘텐츠 개발의 경우, 최근 들어 정부와 IT 산업 분야에서 바이오인포매틱스 분야에 많은 연구비가 투자되고 있는 상황이므로 앞으로 국제 경쟁력을 확보할 수 있을 것으로 보인다. 특히 진단 키트 등의 콘텐츠가 궁극적인 산업화, 제품화, 상품화를 통해 경제성을 얻을 수 있는 분야이기 때문에 분자정보처리기술의 상품화를 선도하는 역할을 할 수 있다. 또한 국내 IT 산업의 국제적인 경쟁력이 강한 점을 고려할 때 큰 잠재력을 가지고 있다고 볼 수 있다.

## VI. 산업화 방안

위에서 살펴본 바와 같이, 부분적으로 국내 기술 기반이 취약한 면도 있으나, 일부 요소 기술에 있어서는 상당히 경쟁력을 갖춘 연구소, 벤처기업, 대학 연구실들이 존재하고 또 정부와 산업 부문에서 현재 이러한 융합 기술에 대한 투자를 늘리고 있기 때문에 산업화 전망을 밝다고 볼 수 있다. 또한 국제적인 수준에서 볼 때 분자정보기술은 아직 초보 단계에 있다.

따라서, 제한된 자원을 최대한 활용하여 제품화 산업화를 염두에 둔 분자정보처리 핵심원천기술을 개발한다면 앞으로 파급 효과와 경제성이



〈그림 7〉 DNA 컴퓨터 Lao-on-a-Chip

아주 큰 산업 핵심기술을 확보할 수 있게 될 것이다. 그리고 연구 목표의 설정에 있어서 선진 외국의 연구보다는 산업화를 염두에 두고 연구 역량을 집중하되, 연구개발 과정에서 장기적인 국제 경쟁력을 확보할 수 있는 핵심기반 원천 기술을 개발하는 방향으로 연구를 수행하여야 할 것이다.

〈표 8〉에 분자정보처리기술의 산업화를 3단계로 나누어, 단기적(3-5년 후), 중기적(5-10년 후), 장기적(10-20년 후)인 관점에서 역량을 집중할 수 있는 핵심 기술의 예가 제안되어 있다.

단기적으로 볼 때, 분자정보처리기술은 바이오 산업에서 제품화될 가능성이 제일 높다. 특히 분자 진단 시장이나 신약 스크리닝 분야에서 분자 컴퓨팅에 기반한 정보처리기술을 이용한 Lab-on-a-Chip 개념의 진단 키트는 지금부터 3년 내지 5년 이내에 상품화될 수 있을 것이다.

〈표 8〉 분자정보기술의 단기, 중장기적 실용화 전망

단계	분자정보처리 기술
단기 (3-5년)	- 파이프라인형 분자 진단 - 온라인 유전자 발현 분석 - 계층적 신물질 스크리닝 - 프로그램 가능한 Lab-on-a-Chip - 단일분자 진단
중기 (5-10년)	- 분자 전달을 위한 나노 우체국 - 파이프라인형 나노엔지니어링 - 나노머신 DNA 프로그램 - 나노 자가조립 제어
장기 (10-20년)	- 초미세 생체 컴퓨터: 연산, 제어, 처리, 동작 - 초저에너지형 초병렬 바이오수퍼컴퓨터 - 자기증식형 나노바이오컴퓨터

5-10년 앞을 내다 본 중기적인 시각에서 볼 때, 분자정보처리 기술은 나노기술과 접목될 것이다. 분자정보기술은 한편으로는 나노기술에 기반하여 이를 이용하기도 하지만, 반대로 DNA 분자를 프로그램으로 사용하여 나노구조체를 조작 제어할 수 있어 기술적으로 시너지 효과를 가져올 것이다. 특히 체내에서 신약을 전달하기 위한 나노머신을 분자프로그램으로 제어하거나 나노구조 자가조립을 제어하기 위한 DNA 프로그램 등의 개념을 응용한 새로운 제품을 개발할 수 있을 것이다. 이는 현재의 기계 및 전자 장치들에서 컴퓨터가 하는 역할을 초미세 분자장치들에서의 컴퓨터 역할로 볼 수 있다. 특히 나노기술이 앞으로 제조업과 의료 분야에 혁신을 가져올 수 있는 잠재력을 생각해 볼 때 지금의 개인용 컴퓨터나 PDA와는 비교할 수 없는 엄청난 수의 분자컴퓨터칩의 수요가 발생하게 될 것이다.

장기적으로는 10-20년 후에 앞에서 개발된 BT 응용 기술과 NT 응용 기술 개발과정에서 확보한 기반 기술을 이용하여 BT, NT 기술을 융합한 새로운 분자정보처리를 위한 컴퓨터를 개발할 수 있을 것이다. 그 것의 궁극적인 모습이 무엇이 될지 지금으로서는 확실하지 않지만 현재의 연구 경향으로 미루어 보아, 아마도 생체 특성을 살린 초미세 생체 컴퓨터나, 분자반응의 초병렬성을 최대한 활용한 바이오수퍼컴퓨터, 또는 유기분자들의 자기증식 특성을 살린 나노바이오컴퓨터 형태가 될 것으로 전망된다.

## Ⅶ. 결 론

본 고에서는 분자정보처리기술의 현재 기술 수준과 연구 동향 및 산업화 방안에 대하여 논하였다. 현재와 같은 기술의 급변기를 맞이하여 점차 과학기술의 발전주기가 짧아짐에 따라 점점 핵심

원천기술의 개발 보다는 단기적인 성과만을 바라보는 기술개발을 강조하는 경향이 있다. 그러나 이러한 근시안적인 연구개발전략은 궁극적으로 국내 과학기술과 학문의 장기적인 발전에 큰 장애 요인으로 작용할 것이며, 따라서 적어도 10년 정도 후의 제품화를 염두에 둔 Technology Roadmap을 가지고 제한된 자원과 연구 역량을 집중할 필요가 있다. 분자정보처리기술은 이러한 측면에서 볼 때, 앞 6절의 산업화 방안에서 제시한 바와 같이, 단기적인 수익성과 중장기적인 핵심기술의 확보 및 산업화를 잘 조화시킬 수 있는 연구개발 분야이다.

분자정보처리기술의 또 다른 특징은, BT 및 NT와의 접목을 통해 상호 시너지 효과를 낼 수 있어 그 기술적인 파급 효과가 큰 차세대 IT 기술이라는 점이다. 조만간 실리콘 반도체 기술은 물리적 한계를 맞이하게 될 것이며 생체분자에 기반한 나노스케일에서의 새로운 정보처리기술은 이 한계를 극복할 수 있는 새로운 가능성을 열어 줄 것이다. 실제로, 이미 DNA Self-assembly에 의한 분자컴퓨팅 연구를 통해서 이러한 방향의 연구가 시작되었다.

또 한편, IT 사업은 현재 수익성 면에 있어서 한계에 부딪히고 있으며, 유수의 IT 업체들도 보다 수익성이 높은 새로운 사업 영역을 탐색하고 있는 단계이다. IBM은 “인터넷이 아니라” 생체 분자모델링을 위한 Blue Gene 컴퓨터개발 계획을 추진중이며, Motorola와 같은 실리콘 반도체 업체도 Micro PCR과 Micro Capillary Electrophoresis와 같은 분자생물정보처리 칩 개발에 투자하고 있다. 이러한 기술은 머지 않아 기존의 DNA 마이크로어레이 칩과는 다른 개념의 새로운, 예컨대, “온라인 파이프라인 분자진단용 바이오컴퓨팅 Lab-on-a-Chip”과 같은 새로운 개념의 바이오칩 제품을 시장에서 볼 수 있게 할 것이다.

## 저 자 소 개



張炳卓

1986년 2월 서울대 컴퓨터공학과 학사, 1988년 2월 서울대 컴퓨터공학과 석사, 1992년 7월 독일 Bonn 대학교 컴퓨터과학과 박사, 1992년 8월~1995년 8월:

독일국립정보기술연구소(GMD) 연구원, 1995년 9월~1997년 2월: 건국대학교 컴퓨터공학과 조교수, 1997년 3월~현재 서울대학교 컴퓨터공학부 조교수, 부교수, 2001년 1월~현재: 서울대학교 바이오정보기술연구센터(CBIT) 센터장, <주관심 분야: 인공지능, 기계학습, Bioinformatics, Bio-computing>