Deoxyribonucleic acid의 원리와 컴퓨팅 융합 기술 연구

정영진* , 진윤태**, 배영주**, 이준원**

요 약

Deoxyribonucleic acid (DNA)는 모든 살아있는 생물체와 일부 바이러스에서 발달과 기능에 사용되는 유전물질을 포함하는 핵산이다. 주요 역할은 유전정보의 장기보관이다. 단백질과RNA와 같은 다른 세포 구성성분 생성에 필요한 물질이다. 유전정보를 운반하는 DNA segment를 유전자라고 한다. DNA분자는 일반 컴퓨터의 실리콘을 바탕으로 한 게이트와 유사한 기본의 논리 게이트의 DNA의 짧은 가닥은 게이트의 입력과 출력 역할을 한다. 궁극적으로, 이러한 게이트들은 marker와 같이 분자들을 감지하거나 어떻게 반응되는지를 선택할 수 있는 역학을 할 수 있다. 이 DNA 게이트로부터 tic-tac-toe라는 게임을 조작함으로써 DNA에 의한 시스템 컴퓨터 조작의 능력을 설명할 수 있다. 따라서, 이중나선의 DNA를 통해 합성화학, 효소학, 구조나노기술, 컴퓨터 과학의 중개 역할과 융합연구의 기초가 되는 지식을 습득할 수 있다.

Computing Convergence Technology and Principles of Deoxyribonucleic Acid

Youn-Gjin Jung*, Yun-Tae Jin**, Young-Ju Bae**, Jun-Won Lee**

ABSTRACT

Deoxyribonucleic acid (DNA) is the genetic blueprint that encodes for the basic functions and development of all living organisms and many viruses. Its main purpose is the storage of genetic material required for the production of proteins, RNAs, and all cells within the body. These DNA segments that carry the genetic information are called genes. Similar to the ones used in modern silicone-based, electronic computers, DNA molecules can be used to create logic gates with short DNA strands acting as input and output signals. With further development, these gates may be used to detect certain molecules such as markers, and take appropriate, pre-programmed action against it. The power of DNA logic gates is demonstrated by a DNA computer that can play a game of tic-tac-toe against a human opponent. Thus, through the double-helix shaped DNA, valuable information and new insights can be gained in a new field of science that brings together synthetic chemistry, enzymology, nanotechnology, and computer science.

Key Words : Deoxyribonucleic acid, recombination, transcription, logic gate, computing

---

* 캐나다 토론토대학(✉fri13alex@gmail.com)
** 배재대학교 생명유전공학과
· 제1저자(First Author) : 정영진 · 교신저자(Correspondent Author) : 이준원
· 접수일(2010년 5월 24일), 수정일(1차 : 2010년 6월 23일), 게재확정일(2010년 6월 28일)

I. History of DNA

One of the shocking developments in Biology that emerged in the early 20th century was the discovery of the genetic material and the appearance of the gene at the molecular level. DNA was first discovered by Friedrich Miescher of Switzerland in 1896. Nitrogen and phosphorous containing material called "nuclein" has been extracted from the cell's nucleus before, but when DNA was first discovered, it was hypothesized that the genetic material would consist of proteins. In 1919, it was discovered by Phoebus Levene that the nucleotide unit is made up of base, sugar, and phosphate. The first X-ray diffraction was created by William Astbury in 1937. Evidence that DNA is the medium for genetic information was observed in a very simple microbiological experiment. In one of the early studies, which was carried out by Osward Avery in 1943, evidence was found that DNA could be the holder of genetic information. The results of that study is based on the experiment using pneumonia bacteria by England's bacteriologist Frederick Griffith.[1] Using the results of his study founded on Griffith's experiment, Osward Avery confirmed that DNA, which is involved in the formation of capsules, is spread from dead, contagious S-forms to the DNA of live, non-contagious R-forms which transforms its traits into the contagious S-form, and that these traits are passed off to the next generation.[2] In 1950, in an experiment using phage that infected colon bacillus, Alfred Hershey and Martha Chase were able to definitively prove that DNA is the genetic material. More information on the studies of DNA is shown in the timeline in table. 1.

Table 1. Timeline of DNA

*1865: Gregor Mendel discovers through breeding experiments with peas that traits are inherited on specific laws (later to be termed "Mendel's laws")
*1869: Friedrich Miescher isolates DNA for the first time
*1871: The first publications describing DNA("nuclein") by Friedrich Miescher, Felix Hoppe-Seyler, and P. Plósz are printed.
*1889: Richard Altmann renames "nuclein" to "nucleic acid"
*1902: Theodor Boveri and Walter Sutton postulate that the heredity units (called "genes" as of 1909) are located on chromosomes.
*1902-1909: Archibald Garrod proposes that genetic defects result in the loss of enzymes and hereditary metabolic diseases.
*1928: Frederick Griffith postulates that a "transforming principle" permits properties from one type of bacteria (heat-inactivated virulent *Streptococcus pneumoniae*) to be transferred to another (live nonvirulent *Streptococcus pneumoniae*).
*1929: Phoebus Levene identifies the building blocks of DNA, including the four bases adenine(A), cytosine(C), guanine(G), and thymine(T).
*1949: Colette and Roger Vendrely and André Boivin discover that the nuclei of germ cells contain half the amount of DNA that is found in somatic cells. This parallels the reduction in the number of chromosomes during gametogenesis and provides further evidence for the fact that DNA is the genetic material.
*1952: Alfred Hershey and Martha Chase use viruses (bacteriophage T2) to confirm DNA as the genetic material by demonstrating that during infection viral DNA enters the bacteria while the viral proteins do not and that this DNA can be found in progeny virus particles.
*1953: James Watson and Francis Crick discover the molecular structure of DNA: a double helix in which A always pairs with T, and C always with G.

*1957: Francis Crick proposes the "central dogma"(information in the DNA is translated into proteins through RNA) and speculates that three bases in the DNA always specify one amino acid in a protein.

*1968-1970: Werner Arber, Hamilton Smith, and Daniel Nathans use restriction enzymes to cut DNA is specific places for the first time.

*1977: Frederick Sanger, Allan Maxam, and Walter Gilbert develop methods to sequence DNA.

*1982: The first drug (human insulin), based on recombinant DNA, appears on the market.

*1990: Sequencing of the human genome begins.

*1999: Sequence of the first human chromosome(22) is published.

*2000 : The complete sequences of the genomes of the fruit fly *Drosophila* and the first plan -*Arabidopsis*- are published.

*2001: The complete sequence of the human genome is published.

*2002: The complete genome sequence of the first mammalian model organism -the mouse- is published.

II. DNA structure

After the discovery of DNA as the genetic material, many scientists strived to discover the molecular structure of the DNA. In 1953, it was discovered by Watson and Crick that the DNA consists of two strands of polynucleotide wound together in a double-helical formation.[1-2] Before Watson and Crick, scientists found that there was an equal number of the DNA bases, A, T, and C, G.[2-3] Furthermore, the 3-D structure of the DNA was discovered through X-ray diffraction. With these two facts, Watson and Crick proposed a model for the structure of the DNA. (Fig. 1) The model contains the following features. [2]

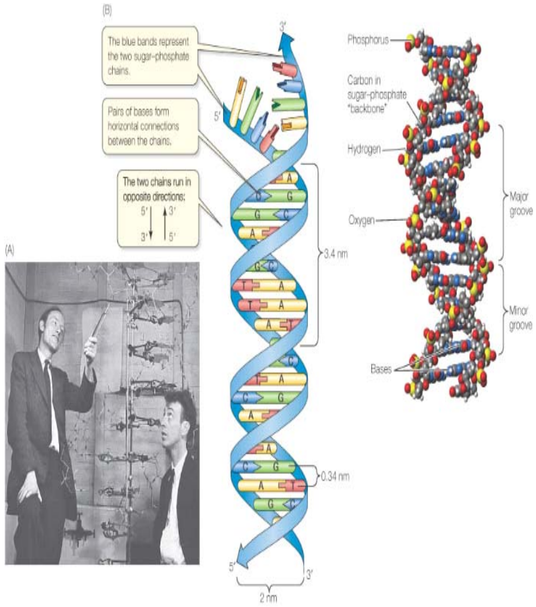

그림 1. 왓슨과 크릭의 DNA 이중나선구조 모델
Fig. 1. DNA double helix model of Watson and Crick [3]

1) Two long polynucleotide chains wind around the central axis to form a double-helix that winds to the right.

2) The two chains are parallel to each other, but run in the opposite direction.

3) The bases of the two chains have a flat structure and is located perpendicular to the axis. They face inwards.

4) The nitrogenous bases of the two opposing chains form a pairs using hydrogen bonds.

5) One complete rotation is 34Å long and 10 bases exist for each rotation of the chain.

6) Along the axis, the appearance of major grooves and minor grooves alternate.

7) The diameter of the double-helix is about 20Å long.

The basic unit of the DNA consists of deoxyribose, base, and nucleotide containing phosphate.[2-4]

There are two types of bases, purine and pyrimidine, and purine is further divided into adenine; A and guanine; G, while pyrimidine subdivides into cytosine; C, thymine; T, and uracil; U. Among these, uracil does not exist in the DNA, and can only be found within the RNA. When DNA forms the double-helix structure, each purine binds to a pyrimidine through hydrogen bonding.[2-4] (Fig. 2) Thus, A and T, C and G always exist as pairs. These hydrogen bonds stabilize the DNA's double-helix structure.

Three hydrogen bonds exist between G and C, and two bonds between A and T, therefore making the bond between G and C stronger. This combination of base pairs are said to be complementary. Complementarity plays a role in the information storage, replication, and transcription of DNA.

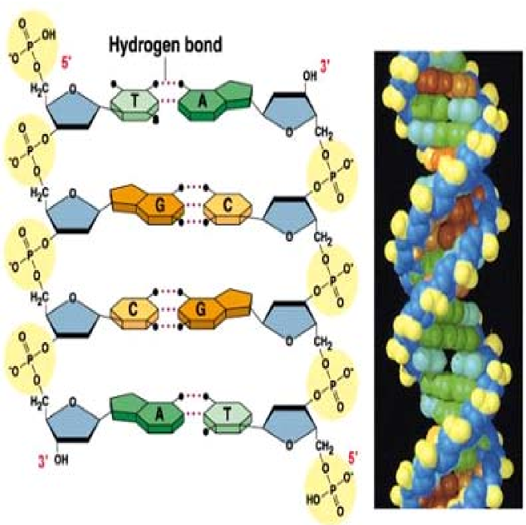

그림 2. 뉴클레오타이드의 구조
Fig. 2. Structure of Nucleotide [2]

Many types of DNA exist, including the most commonly known B-DNA, A-DNA which is more tightly wound, and Z-DNA which winds to the left. However, A-DNA and Z-DNA only form under special circumstances, and normally DNA exists in the B-DNA form.

III. Recombination

Bases are constantly mixed and rearranged through genetic exchange, and it mainly occurs during meiosis when the complementary pairs bond in the first nucleus division. These exchanges are traditionally called crossovers, and is one of the results of homologous recombination.

Homologous recombination is one of the essential cell phenomena, and is catalyzed by an enzyme created and adjusted specifically for this purpose. DNA recombination is crucial because it 1) induces genetic modification 2) restores the DNA base order when DNA is damaged by replacing the damaged part with an undamaged DNA from a homologous chromosome 3) provides a restarting mechanism if the replication process is stopped or damaged 4) adjusts gene expression by switching certain parts of the chromosomes, and transferring genes that have not been expressed by the cell to where they can be expressed 5) and its gene removal or insertion method while maintaining the entirety of DNA during genetic fabrication has proven to be a useful tool in in vivo development.[5] Recombination type is broadly divided into three classes - Homologous recombination, site-specific recombination, and transposition. [6,7]

IV. Transcription and Translation of Genetic Code

Genetic information in the DNA is stored in the form of base sequences, and is transcript by RNA to be translated into proteins. Thus, the arrangement of amino acids is derived directly from the base sequence of DNA. So what are the steps involved in turning DNA's base sequence into proteins? The sequence information is first transferred to the RNA through transcription. This messenger RNA, or mRNA translates the information into the production of amino acid chains using ribosomes. (Fig. 3) Proteins produced in this manner serve many special functions within the cell including structural support (e.g. collagen), transportation of oxygen in the bloodstream (e.g. hemoglobin), immune system (e.g. immune globulin), and as catalysts for various cell metabolism (e.g. enzymes). The RNA is always used as a medium to transfer the genetic material stored within the DNA.

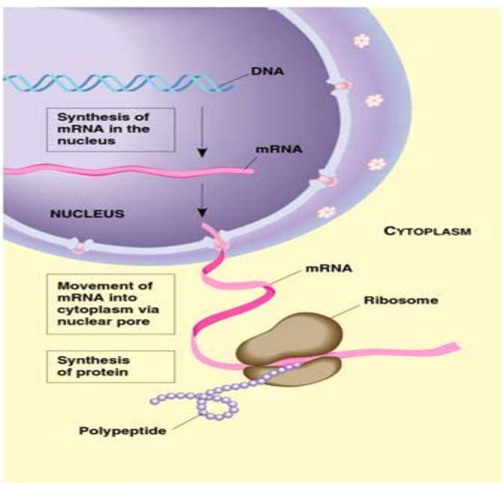

그림 3. DNA의 중요한 학설
Fig. 3. DNA central dogma [8]

As previously mentioned, the transfer of genetic information from transcription to translation proceeds in the order of DNA→RNA→proteins, and cannot occur in the reverse direction. This idea is encapsulated in a framework known as the central dogma of molecular biology, which states that once the genetic information is translated into proteins, it cannot be transferred back into other proteins or the nucleic acids, DNA or RNA. However, some exemplary cases that show that genetic information in the RNA can be used to create DNA, and that not all parts of the DNA are transcript: ① some types of virus called retrovirus contain RNA instead of DNA, and are known to use reverse transcriptase to create DNA from their RNA and insert it into the host's DNA. ② Some parts of the DNA are not transcript by RNA. Such parts are known to exist in fairly substantial amounts in eukaryotes, while they are almost non-existant in prokaryotes. [9]

V. Computing using DNAs

For years, scientists have been engaged in research that seeks to combine the power of the modern computer with DNA in order to create a computer that would be microscopic in size, yet compatible with the human biological system. In essence, a computer nothing more than a programmable machine capable of receiving and relaying input and output signals. The programs used in these computers are in turn, mere algorithms written in programming language. Biomolecular computers developed utilizing the DNA's ability to store information using its base sequence would work in a fashion similar to their silicone-based electronic counterparts. Molecular inputs will replace the peripheral input devices used in electronic computers such as CDs, and will be received by corresponding sensors within the molecular hardware. Molecular algorithms, or logic pathways, can be written to emulate the programs (ie. software) that are in use today. Fig. 4 illustrates the comparison between silicone-based and biomolecular computers.

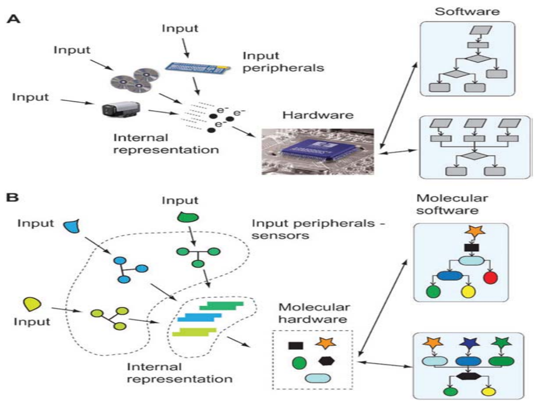

그림 4. DNA 컴퓨터와 실리콘 기반 컴퓨터의 비교
Fig. 4. Comparison between silicone-based and DNA computers [11]

In 1994, Leonard M. Adleman at the University of Southern California developed what is considered to be one of the earliest examples of a DNA computer. He used the tools of molecular biology to solve a puzzle known as the Hamiltonian path problem. All the possible solutions were encoded onto DNA molecules, and using a series of steps, the molecule containing the correct solution was isolated. [10] In 1998, Erik Winfree used a method called "tiling" to simulate a computer by using tiles composed of DNA to self-assemble into a two-dimensional structure. [12] More recently, using an enzyme made from a single-stranded DNA that can cleave other DNA strands (called deoxyribozymes), combined with molecular groups that act as sensors to either activate or inhibit the deoxyribozymes (called stem-loops), Stojanovic and colleagues were able to create a tic-tac-toe playing automaton. [13] A stem-loop is a strand of DNA that can fold over itself to create a "stem" of zipped DNA, and a "loop" of unzipped DNA. When the correct input strand is used, the input strand binds to the loop region of the stem-loop, prying open, or unzipping, the stem part in the process. By attaching the stem-loop to one of the arms of the deoxyribozyme, it can inhibit the enzyme's cleaving action when it is in its zipped formation by blocking the output strand from binding to the deoxyribozyme. (Fig. 5) So by using stem-loops to control the deoxyribozymes, powerful logic gates such as YES ("YES, input A is present"), NOT ("No, input A is NOT present"), AND ("Yes, both input A AND B are present"), and AND-AND-NOT gates ("Input A AND input B AND NOT input C are present") can be created.

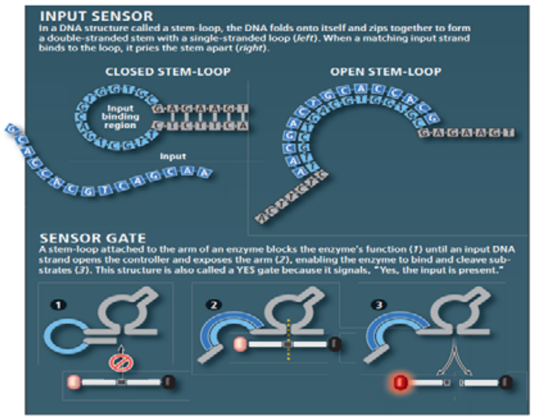

그림 5. DNA 고리구조를 이용한 커퓨팅 이론
Fig. 5. Stem-loop [13]

VI. MAYA-I and MAYA-II

The power and potential of DNA logic gates are demonstrated in MAYA, the tic-tac-toe-playing DNA computer. MAYA, short for Molecular Array of YES and AND-AND-NOT gates, makes use of 24 mutually exclusive parallel logic gates to play a perfect game of tic-tac-toe against a human opponent. (Fig. 6) Due to the nature of the game, if both parties only make the perfect moves, the result can only be a tie. Therefore the human opponent, who does not make the opening move, can only at best tie with MAYA. With these 24 logic gates, 19 different possible games can be played. One of these games results in a tie, while MAYA exploits the opponent's mistakes to win in the remaining 18.

Taking it a step further, a more user-friendly and unrestricted version called MAYA-II was created. The design of MAYA-II requires 128 logic gates, and it can play a total of 76 different possible games.

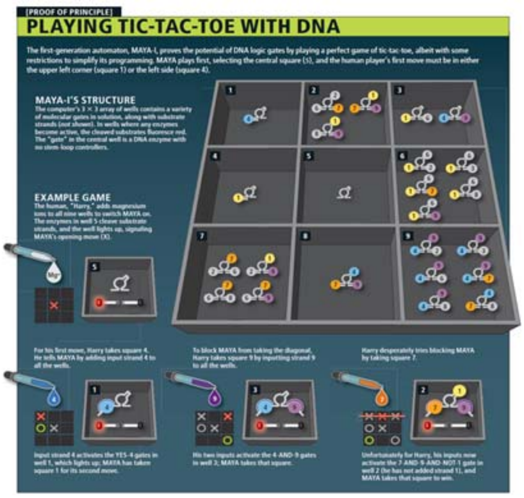

그림 6. DNA 컴퓨팅을 이용한 틱-탁-톡 게임
Fig. 6 Playing tic-tac-toe with DNA [13]

VII. Application and the future of DNA computers

Although the creators of MAYA have chosen a one-layered logic system where the logic gates are mutually exclusive to each other in order to reduce complications, it theoretically could have been a multi-layered system, where the output of one logic gate can serve as the input of another. This is possible because the basic material used both as the input and the output signals are the same - short segments of DNA, or oligonucleotides. Perhaps with further development, this multi-layer approach can be used to create "smart" machines that can perform multiple sequential actions with a single input, and be able to learn and self-develop.

One of the main advantages of DNA computers is that they are biocompatible. Ongoing development of in vivo computation suggests that due to the high level of complexity able to be achieved by in vivo biocomputers, they may be superior to other alternative means of computing. [11] Imagine a therapeutic agent in a person's bloodstream that can automatically sense the abnormalities in the body and treat it accordingly. For example, a molecular module in a diabetic's body could be used to constantly monitor the blood sugar level and release insulin at the appropriate time. DNA computers can also be programmed to target molecules containing specific markers. In present day therapies in treating leukemia, the antibodies indiscriminately attack any cells containing the characteristic marker associated with the disease. This leads to the destruction of many healthy cells along with the diseased ones, because the antibodies cannot analyze the different types of markers and decide if a cell is truly harmful or not. DNA analysis is another field that the molecular modules could become successful in, searching for mutations in the genes or identifying pathogens.

However, for all its complexity, DNA computers have a few critical shortcomings. First, since computations are carried out using biological mechanisms, the process is hard to reverse and the logic pathway cannot be reset. Second, unlike the fast computations performed by electronic computers, the chemical reactions used in DNA computers take time to complete, making large systems containing numerous logic gates prohibitively slow. Third, while the digital mechanism of the electronic computer only allows for gates to be in the ON or OFF position (1 or 0) according to their threshold voltages, DNA gates can take on any intermediate form between ON and OFF. This allows the system to be "partially ON", depending on the number of inputs. Lastly, DNA computers lack the tools or the methods to "debug" its system, in case something does not work as intended. [11, 13] These problems do not yet pose a real threat in the development stage, but for DNA computers to find widespread use and display their full potential, these challenges should be addressed and overcome.

Molecular computation is fast becoming a popular topic of study, one that could potentially have an unlimited economic value. With consistent research and further development, we may witness the emergence of a new revolutionary interdisciplinary area that is to become the future of bioengineering, nanotechnology, and medical and pharmaceutical science.

Reference

[1] D. R. Friedrich, “Friedrich Miescher and the discovery of DNA“, *Developmental Biology*, Vol. 278, pp. 274－288, Review, 2005.

[2] KIug, William R. "Essential genetic. 5th ed", *Prentice Hall*, pp. 211-230.

[3] R. S. Richard, "DNA structure and function. 2nd ed", *Academic press*, pp. 3-30, 1994.

[4] S. Arnott, "Historical article: DNA polymorphism and the early history of the double helix", *Trends Biochem Sci*, Vol. 31, No. 6, pp. 349-354, 2006.

[5] L. Serrano, L. Liang, Y. Chang, L. Deng, C. Maulion, S. C. Nguyen, and J. A, Tischfield, "Homologous Recombination conserves DNA sequence integrity throughout the cell cycle in embryonic stem cells", *Stem Cells Dev*, Vol. 21, 2010.

[6] P. A. Muniandy, J. Liu, A. Majumdar, S. T. Liu, and M. M. Seidman, "DNA interstrand crosslink repair in mammalian cells: step by step", *Crit Rev Biochem Mol Biol*, Vol. 45, No. 1, pp. 23-49, 2010.

[7] X. Li, and W. D. Heyer, "Homologous recombination in DNA repair and DNA damage tolerance", *Cell Res*, Vol. 18, No. 1, pp. 99-113, 2008.

[8] http://www.detectingdesign.com/images/Abiogenesis

[9] S. Cooper, "The central dogma of cell biology", *Cell Biol Int Rep*, Vol. 5, No. 6, pp. 539-549, 1981.

[10] L. M. Adleman, "Molecular computation of solutions to combinatorial problems", *Science*, Vol. 266, pp. 1021-1024, 1994.

[11] Y. Benenson, "Biocomputers: from test tubes to live cells", *Mol BioSyst* Vol. 5, pp. 675-685, 2009.

[12] E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman, "Design and self-assembly of two-dimensional DNA crystals", *Nature* Vol. 394, pp. 539-544, 1998.

[13] J. Macdonald, D. Stefanovic, and M. N. Stojanovic, "DNA computers for work and play", *Sci Am* Vol. 299, pp. 84-91, 2008.


정영진 (Youngjin Jung)
2009~ Attending University of Toronto Undergraduate program in Life Sciences (B. Sc.)

※ Majoring in Anthropology and Human Biology


진윤태 (Yuntae Jin)

서울 한영외국어 고등학교

현재 배재대학교 생명유전공학과 인턴 연구원
※ 관심분야 : 약제 개발


배영주 (Youngju Bae)
2002 배재대학교 유전공학과(이학사)
2005 배재대학교 유전공학과(이학석사)

2009년 ~ 현재 배재대학교 생명유전공학과 박사학위과정
※ 관심분야 : 생리활성대사산물, 골질병 약제 개발


이준원 (Junwon Lee)

2001 연세대학교 생명공학과(공학박사)
2006 전남대학교 의과대학 연구교수

2008년~ 현재 배재대학교 교수
※ 관심분야 : 생리활성대사산물, 골질병 약제 개발