

DNA 이중나선구조에서의 오류 검출 및 복구방법연구

김석환*, 허창우**

*이화여자대학교 전자공학과, **목원대학교 전자공학과

A Study of Error Detection and Repair on DNA Duplicate Structure

Soke-Hwan Kim*, Chang-Wu Hur**

* Dept. of Electronics Engineering Ewha Womnas University,

**Dept. of Electronics Engineering Mokwon University

요 약

살아있는 생명체는 세포로 구성되며 성장 분열을 통해 스스로 복제할 수 있는 능력을 지녔다. DNA 상의 변이, 즉 돌연변이는 자손의 생존과 번식에 불리하게 작용할 수 있고 이점을 줄 수 있는 양면성을 지녔다. 본 연구에서는 DNA 이중나선은 복제 주형으로 사용되기 위해서는 먼저 이중나선이 열리고 단일 가닥으로 분리되어야 한다. 이중 나선구조결합에서의 결합의 오류부분의 위치를 찾아내고 복구하는 방법을 제시한다.

1. 서 론

살아있는 생명체는 세포로 구성되며 이 세포는 성장, 분열을 통하여 스스로 복제할 수 있는 능력을 지니게 된다. 세포는 생명의 기본적 단위로 일컫는다. 모든 생물에서 유전정보, 즉 유전자는 DNA분자에 저장되어 있으며 동일한 화학적 암호로 쓰여져 있고 똑같은 화학적 골격을 지니게 된다[1].

세포는 DNA 복제와 세포 분열을 통해 번식하며 DNA에 암호화 되어있는 유전정보와 동일한 복제물을 두 딸 세포에게 전달하게 된다. 이때 복제는 항상 완벽한 것이 아니어서 DNA상의 돌연변이는 자손의 생존과 번식에 불리하게 작용할 수 있게 된다. DNA를 복제한다는 것은 세포에서 질서를 유지하는 일로 유전 정보에 대한 지속적인 감시와 회복체계가 필요하게 된다[4].

생식세포에서 생긴 돌연변이는 이 세포에서 발생되는 다세포 생물의 모든 세포에 전달되기 때문에 중요하다. DNA mismatch 회복 시스템은 복제기구가 복구하지 못한 복제 실수를 제거한다[6]. 여기서 mismatch가 포함된 DNA 사슬 중

하나를 제거하고 결손된 DNA 사슬을 재합성하는데 오직 새로운 합성된 DNA 가닥만을 절제하여야 한다. 기존의 가닥을 절제하면 복제 실수의 교정이 아니라 그것을 보존하게 된다[1].

복잡한 생물학적 기능을 공학적인 모델로 적용하는 것은 매우 힘이 든다. 유전자에서 인공 게놈에서의 유전자 네트워크를 표현하고 분석하기 위해서는 복잡한 시스템 개발 기술을 요구한다[2].

오래전부터 컴퓨터의 기술이 발달함으로써 더 복잡한 모델을 분석하고자 하는 인간의 노력은 시도되고 있다. 현재 시도하고 있는 인공자기 복제구조의 연구는 시스템에 좌우되지 않으면 안된다. 인공세포에 대한 계산모델을 설계 하기 위한 실험 및 분석하기 위한 시스템으로 분석되어야 한다[3].

유전자에는 유전정보에 대한 순서가 있으며 유전자 게놈은 생물의 유전자형과 표현간의 진화과정에서 핵심적인 역할을 담당하고 있다. 생물이 진화하는 데 있어서 유전자 변이와 크로스오버는 매우 중요한 역할을 요소이다[5][7].

본 연구에서는 생물의 기본단위인 세포의 특성

을 모사하여 세포 분열 및 복제과정에서 발생하는 오류결합이 있을 경우 그 위치를 찾아내고 복구하는 방법을 공학적인 방법을 적용하고자 한다. 이에 관련 연구는 진화알고리즘 분야로 진행되어 왔다. 이에 본 연구에서는 단백질의 이중 나선 구조의 모델을 공학적인 방법으로 적용하여 결합구조의 형태가 다른 부분을 찾아내고 복구하는 방법을 제시한다.

II. 본론

1. 기존 계층분석방법

Artificial genome에서 결합구조의 위치를 찾아내고 상호간에 연결하기 위해서는 주어지는 genome에서 TATA box[0101]를 찾아내는 것이 우선이다. 그리고 다음에 같은 값을 지니는 값을 찾아야 한다. 그래서 결합되어 새로운 단백질을 생성하게 된다.

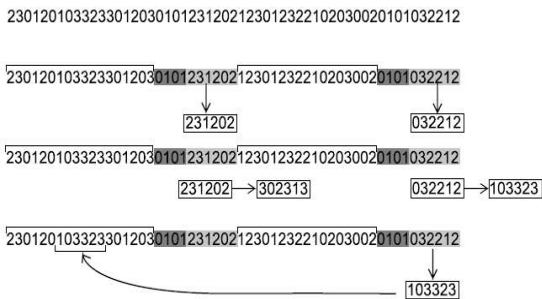


그림 1. gene 결합

그림1은 gene 결합구조의 방법의 예를 설명하고 있다. 여기서의 숫자는 단백질 아미노산의 4가지 요소인 아데닌(A), 구아닌(G), 시토신(C), 티민(T)을 각각 0, 1, 2, 3으로 표현한 것이다. 임의의 숫자들이 그림1의 첫 번째 숫자와 같다면 제일먼저 그 결합구조에서 결합을 시켜주는 TATA box를 찾아내야 한다. 두 번째 표시와 같이 "0101" 부분을 두 군데에서 찾았다. 이 바로뒤에 오는 6자리 숫자가 gene을 생성하기 위한 매개물로 값을 변경시키게 된다. 여기서의 첫

번째 "0101" 뒤에 있는 6자리 숫자 231202이고 뒤번째 "0101" 뒤에 있는 6자리 숫자는 032212이다. 세 번째 해야 하는 것은 이 숫자들을 변경하는 것이다. 즉 현재 값을 각각 하나씩 증가시킨다. 0 -> 1, 1 ->2, 2 -> 3, 3 -> 0으로 변경하며 "302313", "103323"으로 변경되었다. 이 결과숫자를 보고 "0101" 기준으로 앞 뒤에서 이와 같은 숫자를 찾아낸다. 첫 번째 "0101"앞에 "103323"이 동일함을 알 수있다. 이들은 서로 결합구조를 이루게 되고 새로운 단백질을 생성한다.

이 방법에 대한 연구는 여러 연구분야들이 TATA box를 이용하고 "0101" 값을 가지고 진행한다는 것에 거의 공통적인 부분이 있다. 그러나 세부적인 단백질 결합의 방법에서는 차이점을 보이지만 비슷한 맥락을 가지고 있다.

2. 디지털 시스템에서의 오류진단

공학적인 모델에서 적용되는 하드웨어 오류 위치를 찾아내기 위한 기본 방법은 하나의 회로 내용을 회로 기능의 최소 단위로 나누고 이중 모듈화(Double Modular Redundancy, DMR)로 설계를 한다. 또한 똑같은 기능을 하는 하드웨어 시스템을 세 개로 구성하여 삼중화 구조로 하는 경우도 있다. 그러나 이것은 매우 많은 비용 및 시스템의 크기가 증가하는 단점을 지니게 된다. 그림3은 이중화 구조의 오류 진단 방법 중에 하나의 예를 보여 준다.

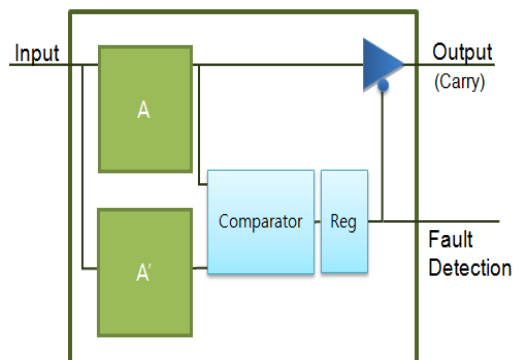


그림 2 . 이중화 구조

III. gene 결합 이중나선 구조

1. 오류 검출 및 복구

본 연구에서는 생물의 구성단위인 세포를 기준으로 하며 단백질의 합성과정 및 진화 과정에서 오류가 발생할 경우 복구하기 위한 방법을 제시한다. 기본적으로 단백질의 이중나선 구조로 하며 A는 T와 결합을 하고 G는 C와 결합한다.

각 4개의 아미노산은 숫자 0에서 3까지 부여한다. A는 0, T는 3, G는 1, C는 2의 값으로 부여했다. 각각의 결합에서 결합이 깨지는 현상을 부여된 비트 값이 다르게 되었을 경우 그 결합이 이상한 부분을 찾아내고 정상적으로 돌리는 방법을 제시한다.

5'	A	G	G	C	C	C	C	T	A	3'
3'	T	C	C	G	G	G	G	A	T	5'
A=0, G= 1, C=2, T=3										

5'	0	1	1	2	2	2	2	3	0	3'
3'	3	2	2	1	1	1	1	0	3	5'
	3	3	3	3	3	3	3	3	3	

그림 3. DNA 이중나선 구조의 표현

그림 3과 같이 쌍극성 DNA구조를 기준으로 설계를 한다. 이중 나선구조에서 각 자리에 값을 더하면 모두 공통적으로 3이 된다.

5'	0	1	1	2	2	2	2	3	0	3'
3'	3	2	2	1	1	1	1	0	3	5'
	3	3	3	3	3	3	3	3	3	

5'	0	1	1	2	2	3	2	3	0	3'
3'	3	2	2	1	1	1	1	0	3	5'
	3	3	3	3	3	4	3	3	3	

그림 4. 1비트 오류의 경우

임의의 한 자리에 두 개의 이중 나선 중 하나의 값이 다른 경우를 고려한다. 그림 4의 상단과 같이 더한 값이 3이 되어야 하는데 하단의 한 부분에 4가 되었다. 이 경우 쉽게 위치를 찾을 수가 있으며 Parity check를 통해 복구가 가능하다.

이 구조는 오류 위치를 찾을 수 있지만 복구시키기 위한 상 하단의 위치를 찾지 못하게 된다. 첫 번째 경우는 한자리에 두 개의 오류가 동시에 발생할 경우이다. 이에 해당되는 내용은 그림 5와 같다.

5'	0	1	1	2	2	2	2	3	0	3'
3'	3	2	2	1	1	1	1	0	3	5'
	3	3	3	3	3	3	3	3	3	

5'	0	1	1	2	2	3	2	3	0	3'
3'	3	2	2	1	1	2	1	0	3	5'
	3	3	3	3	3	5	3	3	3	

그림 5. 한자리에 상호간 비트 오류 경우

두 번째 경우는 그림 6과 같이 두 개의 자리에서 각 한 개씩 동시에 발생할 경우이다. 더한 합의 값은 달라서 위치를 찾을 수 있지만 정확히 위아래 중 어느 것인지 그림 5와 같이 알 수가 없다.

5'	0	1	1	2	2	2	2	3	0	3'
3'	3	2	2	1	1	1	1	0	3	5'
	3	3	3	3	3	3	3	3	3	

5'	0	1	1	2	2	3	2	3	0	3'
3'	3	2	3	1	1	1	1	0	3	5'
	3	3	4	3	3	4	3	3	3	

그림 6. 두자리에 동시에 비트 오류 경우

IV. 결론

생물의 기본 최소단위인 세포를 기준으로 DNA 이중 나선구조 결합에서 오류가 발생할 경우 그 위치를 찾아내서 복구하는 방법을 제시하고자 한다. 단백질의 아미노산의 4가지 요소를 각각 숫자 0에서 3까지 신호를 부여하였다. 이를 기준으로 이중 나선구조로 배열 할 경우 각각의 합한 값은 3이 되도록 하였다.

이중 나선 결합구조에서 어느 임의의 하나의 위치에서 하나의 값이 다르게 되었다면 본 연구에서 제시한 바와 같이 정확히 그 위치와 변경되어야 할 값을 알 수가 있었다. 그러나 임의의 하나의 위치가 오류가 발생할 경우에 두 개의 이중나선의 값이 동시에 오류가 발생을 하게 되면 합계의 값이 다르므로 어느 위치인지 알 수가 있지만 Parity Check를 하게 되면 정상적으로 인식되기 때문에 정상상태로 복구하기는 힘들게 된다.

각 이중나선 구조에서 하나의 부분에서 오류가 났다 하더라도 임의의 두 군데에서 오류가 발생할 경우는 각 자리의 값이 3이 아니기 때문에 찾을 수 있지만 Parity Check 값 체크 후 정상상태이므로 복구는 할 수가 없게 된다.

본 연구에서 제시한 방법에서 임의의 하나의 위치에서 이중나선 중 하나의 경우는 정확히 찾아내고 복구할 수 있었다. 그러나 두군데 이상에서 오류가 발생하거나 하나의 위치에서 이중나선 두 개가 동시에 값이 다르게 되면 위치는 찾아나 복구할 수는 없게 됨을 알 수 있었다. 이에 본 연구에서는 추후 다른 방법에서는 이런 부분을 보완하는 새로운 알고리즘 방법을 제시하고자 한다.

참고 문헌

- [1] Alberts, Bray저, 박상대 역 "필수 세포생물학," 교보문고, 2010년
- [2] KaiWilladsen. ; JanetWiles ; "Dynamics of Gene Expression in an Artificial Genome",
- [3] Moshe Sipper, " Fifty Years of Research on Self-Replication : An Overview " Artificial Life 4, pp. 137-257, 1998.
- [4] Asso Yamauchi Toshihiro, Nakashima, " Evolvability of random polypeptides through functional selection within a small library, " Protein engineering vol. 15 no.17 pp 619-626, 2002.
- [5] M Srinivians ; Lalit M, "Genetic algorithms a survey", IEEE, pp.17 - 26 , 1994
- [6] Bedau; John S. McCaskill, " Open Problems in Artificial Life", Artificial Life 6 : pp 363-376, 2000
- [7] Jong-Hwan Kim, Kang-Hee Lee, and Yong-Duk Kim , "The Origin of Artificial Species: Genetic Robot", International Journal of Control, Automation, and Systems, vol. 3, no. 4, pp. 564-570, December 2005