DNA 이중나선에서의 오류위치 검출 방법 및 효율적인 복구 알고리즘 연구

김석환* · 허창우**

An research of the error detection method and efficient recovery algorithms in the DNA double helix

SokeHwan Kim* · Chang-Wu Hur**

요 약

세포에서 질서를 유지하기 위해서는 유전정보에 대한 지속적인 감시와 회복체계를 필요로 한다. DNA는 염기쌍의 결합으로 이루어지는데, 틀린 염기쌍이 정상적인 염기쌍보다 훨씬 낮은 빈도로 형성되지만, 이것이 수정되지 않고 DNA내에 축적될 경우 세포가 죽기도 한다. 본 연구에서는 DNA 복제 시 발생하는 실수, 손상된 부분을 회복하는 DNA 복구 기능을 모사하여 공학적인 개념을 도입한다. 기존에 발표 되었던 부분을 보완하여 여러 군데에서 발생한 오류 위치를 찾아내고 복구시키는 효율적인 알고리즘을 제시한다.

ABSTRACT

In order to maintain order in the genetic information at cells, it need ongoing monitoring and recovery system. DNA is accomplished by a combination of base pairs, Wrong base pairs is formed with a much more lower frequency than the normal DNA. if it does not modify and was accumulate, the Cells were died. In this study, mistakes of DNA replication and repair of the damaged part was introduced engineering concepts by mimicking DNA repair functions. It was presented recover the complementary part of the previously announced and presented an efficient algorithm at find and recover the complementary part.

키워드

자가복구, FPGA, 바이오 영감공학, 이중화구조

Key word

Self-repair, FPGA, Bio-inspired Engineering, DMR

* 정회원 : 이화여자대학교 전자공학과 (ksh63045@ewha.ac.kr)
** 종신회원 : 목원대학교 전자공학과

접수일자 : 2012. 10. 05
심사완료일자 : 2012. 10. 25

Open Access http://dx.doi.org/10.6109/jkiice.2012.16.11.2557

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(http://creativecommons.org/li-censes/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ⅰ. 서 론

DNA를 복제하는 것은 매우 복잡하고 이것을 분석하는 자체는 매우 어려운 일이다. 세포 분열을 통해 세포는 증식과 번식을 이루었고 후대 자손에게 생존 방법을 알려 주게 되었다. DNA를 복제한다는 것은 세포에서 질서를 유지하는 일로 유전 정보에 대한 지속적인 감시와 회복체계가 필요하게 된다[4]. 세포는 DNA 복제와 세포 분열을 통해 번식하며 DNA에 암호화 되어 있는 유전정보와 동일한 복제물을 두 딸 세포에게 전달하게 된다.

생물학에서의 세포는 매우 간단한 최소단위이면서 성장과 분열을 통해 증식할 수 있는 매우 복잡한 시스템이라 볼 수 있다. 이 복잡한 시스템에 분석을 위한 기술은 상호 작용하는 유전자의 네트워크의 관계를 먼저 이해해야 한다.

고도의 발달된 전자 시스템에서 생명공학의 원리를 모방하기 위한 영감공학 분야는 매우 무한하며 적용할 수 있는 모델의 수도 무한하다. 그러나 생명의 복잡성을 공학에 응용하기 위해서는 공학의 한계점에 부딪힐 수 있다.

본 연구에서는 DNA 복제 시 발생하는 실수, 손상된 부분을 회복하는 DNA 복구 기능을 모사하여 공학적인 개념을 도입한다. 기존에 발표 되었던 부분을 보완하여 여러 군데에서 발생한 오류 위치를 찾아내고 복구시키는 효율적인 알고리즘을 제시한다.

Ⅱ. 본 론

2.1. 생물학의 게놈

DNA 안에는 각 순서에 따른 유전 정보가 다르게 된다. 또한 유전자의 변이는 생물체가 진화하기 위한 과정에서 매우 중요한 역할을 수행한다.

본 연구에서는 생물학적 단위의 세포를 사용하여, 우리는 세포 분열과 조합의 특성을 사용하여 결함 위치와 복구 알고리즘을 제안한다. 제시한는 알고리즘은 단백질 공학 접근 방식의 이중 나선 구조의 모델을 기준으로 한다.

2.2. 공학적 표현

본 연구에서는 생물학적 기본단위인 세포를 기준으로 살펴본다. 단백질은 이중나선 구조로 결합되었으며 A는 T와, G는 C와 결합을 한다. 기본적인 게놈에 대한 표현은 그림 1과 같다. 이중 나선의 구조를 십진수로 표현하여 각 자리를 더하면 3이 된다.

5′	A	G	C	T	A	A	T	G	G	C	C	3′
3′	T	C	G	A	T	T	A	C	C	G	G	5′
A=0, G= 1, C=2, T=3												
5′	0	1	2	3	0	0	3	1	1	2	2	3′
3′	3	2	1	0	3	3	0	2	2	1	1	5′
	3	3	3	3	3	3	3	3	3	3	3	

그림 1. 게놈의 디지털 표현
Fig. 1 Digital Expression of Genome

생물학적 모델을 공학 모델에 적용할 때는 십진수가 아닌 이진수 표현이 옳다. 이에 본 연구에서는 이중나선 결합 구조를 두 자리 비트표현으로 설명한다.

Ⅲ. 복구방법

3.1. 게놈의 비트 표현

본 연구에서 제시하는 방법은 결합된 게놈에 대해서 각 자리별로 비트로 표현하여 살펴보는 것이다. 그림 2에서와 같이 제시된 이중나선의 아미노산을 각각 비트 단위로 표현하였다. 상부와 하부의 비트를 각각 더하면 3(11)이 된다. 그러므로 모든 자리는 상보적인 관계가 있다. 그림 2의 예시는 11개의 결합 모델을 고려했다. 그러므로 상부의 앞자리 비트의 1의 개수와 하부의 뒷자리 비트의 1의 개수의 합은 11이다. 또한 상부의 뒷자리 비트의 1의 개수와 하부의 앞자리 비트의 1의 개수의 합은 11이다. 총 1의 개수는 22개이다. 이 결합에 문제가 발생하면 그 오류 위치를 찾아내고 정상적인 값으로 돌리는 알고리즘을 제시한다.

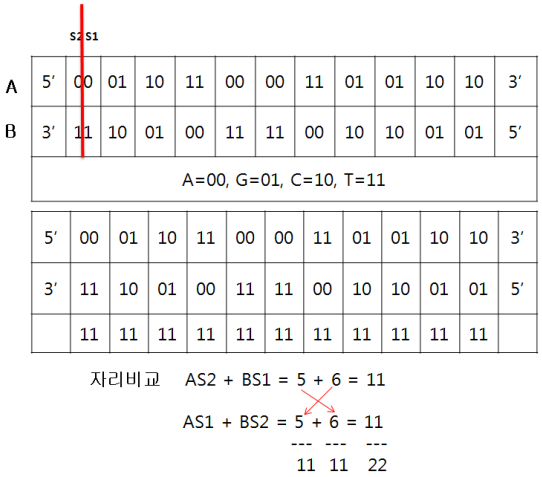

그림 2. 제안된 기본 게놈의 비트 표현
Fig. 2 Bit Expression of Proposed Genome

3.2. 알고리즘 분석

3.2.1. 1비트 오류의 검증

그림 3은 결합된 아미노산에서 한 개의 결합에 이상이 생긴 구조를 설명하고 있다. 정상적인 결합에서 각 비트 결합의 값은 3(11)이 되어야 하는데 2(10)가 되었다. 덧셈으로 값이 다르다는 것을 알지만 상부의 아미노산인지 하부의 아미노산의 결합 문제인지를 파악하는 것은 어렵다.

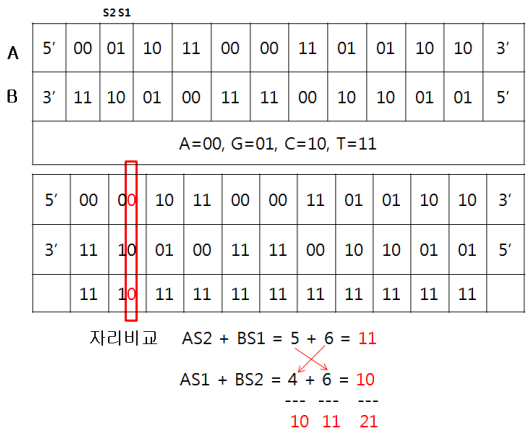

그림 3. 1비트 오류의 검증
Fig. 3 Detection of 1 Bit error

만일 찾았다 하더라도 어떤 아미노산의 값으로 돌리느냐가 핵심 문제가 된다.

그림 3과 같이 오류 위치는 찾았다. 이제 상부 하부 아미노산에 대한 비트를 검토하면 다음과 같다. AS2와 BS1의 1의 개수 합은 11개이며 AS2와 BS2의 1의 개수 합은 11개이므로 AS2는 정상이다. AS1의 경우 BS2와 결합 시 1의 개수는 10개 이므로 한 개가 다르다. 또한 BS1과 결합하면 1의 개수가 10개이다. 이것으로 AS1의 비트가 이상한 것임을 찾아내었다. 그러므로 A(00)를 G(01)로 변경시키면 정상적으로 복구가 된다.

3.2.2. 2비트의 오류의 검증

한 자리에서 오류가 발생한 후 복구하는 알고리즘을 이용하여 그림 4에서와 같이 두 군데 이상의 경우를 고려해 본다.

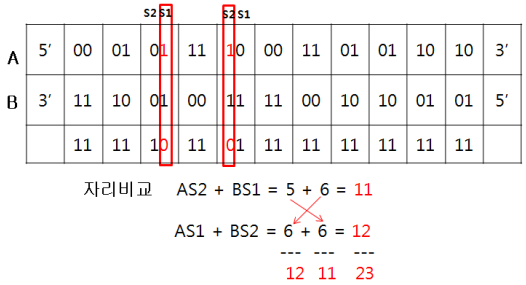

그림 4. 2비트 오류의 검증
Fig. 4 Detection of 2 Bit error

여기서 살펴보면 아래와 같다.

$$AS2 + BS2 = 11$$
$$AS1 + BS1 = 12$$

제시한 덧셈 방법에서는 1의 개수가 총 23개이다. 즉, 1개의 비트만 이상 있다는 것이다. 그러나 그림 4에서와 같이 첫 번째 자리의 오류를 보면 캐리로 인한 1의 값이 생겼음을 알 수 있다. 이 캐리를 제외한 덧셈 방법에 대한 추가 연구가 필요하다. 두 군데 오류 난 결합을 따로 떼어내어 정상적인 결합에 대한 경우를 살펴보면 4가지로 나뉘게 된다.

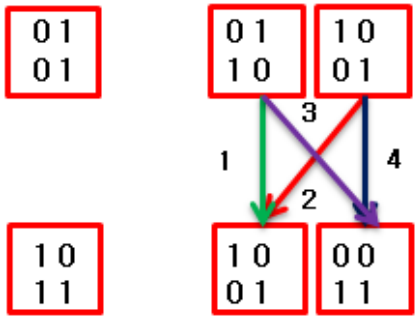

그림 5. 복구 비트의 경우의 수
Fig. 5 The number of Cases recovery bit

3.2.2.1. 1번의 경우

첫 번째 오류에서 상부는 정상으로 보고 하부가 틀렸을 경우, 두 번째 오류에서 상부는 정상으로 보고 하부는 틀렸을 경우의 복구한 결합이다.

$$AS2 + BS2 = 11$$
$$AS1 + BS1 = 11$$

자리비교 AS2 + BS1 = 5 + 5 = 10

AS1 + BS2 = 6 + 6 = 12

--- --- ---

11 12 22

각 자리에 대한 합계는 모두 3(11)이며 전체 1의 개수는 22개이다. 그러나 제안한 알고리즘을 적용하면 상부영역(A)의 상위비트(AS2)와 하부영역(B)의 하위비트(BS1)의 합은 1의 개수가 한 개 부족하였다. 또한 상부영역(A)의 하위비트(AS1)와 하부영역(B)의 상위비트(BS2)의 1의 개수 합은 한 개가 기준보다 많았다. 첫 번째 경우는 각 자리에 대한 복구가 틀렸음을 알 수 있다.

3.2.2.2. 2번의 경우

첫 번째 오류에서 상부는 정상으로 보고 하부가 틀렸을 경우, 두 번째 오류에서 하부는 정상으로 보고 상부는 틀렸을 경우의 복구한 결합이다.

$$AS2 + BS2 = 11$$
$$AS1 + BS1 = 11$$

자리비교 AS2 + BS1 = 4 + 5 = 9

AS1 + BS2 = 6 + 7 = 13

--- --- ---

11 11 22

각 자리에 대한 합계는 모두 3(11)이며 전체 1의 개수는 22개이다. 그러나 제안한 알고리즘을 적용하면 상부영역의 상위비트와 하부영역의 하위비트의 합은 1의 개수가 두 개 부족하였다. 또한 상부영역의 하위비트와 하부영역의 상위비트의 1의 개수 합은 두 개가 기준보다 많았다. 두 번째 경우는 각 자리에 대한 복구가 틀렸음을 알 수 있다.

3.2.2.3. 3번의 경우

첫 번째 오류에서 하부는 정상으로 보고 상부가 틀렸을 경우, 두 번째 오류에서 상부는 정상으로 보고 하부는 틀렸을 경우의 복구한 결합이다.

$$AS2 + BS2 = 11$$
$$AS1 + BS1 = 11$$

자리비교 AS2 + BS1 = 6 + 6 = 12

AS1 + BS2 = 5 + 5 = 10

--- --- ---

11 11 22

각 자리에 대한 합계는 모두 3(11)이며 전체 1의 개수는 22개이다. 그러나 제안한 알고리즘을 적용하면 상부영역의 상위비트와 하부영역의 하위비트의 합은 1의 개수가 한 개가 많았다. 또한 상부영역의 하위비트와 하부영역의 상위비트의 1의 개수 합은 한 개가 기준보다 적었다. 두 번째 경우는 각 자리에 대한 복구가 틀렸음을 알 수 있다.

3.2.2.4. 4번의 경우

첫 번째 오류에서 하부는 정상으로 보고 상부가 틀렸을 경우, 두 번째 오류에서 하부는 정상으로 보고 상부는 틀렸을 경우의 복구한 결합이다.

$$AS2 + BS2 = 11$$
$$AS1 + BS1 = 11$$

자리비교 AS2 + BS1 = 5 + 6 = 11

AS1 + BS2 = 5 + 6 = 11

--- --- ---

11 11 22

각 자리에 대한 합계는 모두 3(11)이며 전체 1의 개수는 22개이다. 그러나 제안한 알고리즘을 적용하면 상부영역의 상위비트와 하부영역의 하위비트의 합은 1의 개수가 11개로 정상이었다. 또한 상부영역의 하위비트와 하부영역의 상위비트의 1의 개수 합은 11개로 정상이었다. 이 네 번째 경우는 각 자리에 대한 복구가 정상임을 알 수 있다. 네 가지의 경우로 두 군데 결합에 오류가 발생한 경우 정확히 그 위치를 찾아내었고 정상적인 결합으로 변경 시켰다.

Ⅳ. 결 론

세포에서는 질서를 유지하기 위해서 유전정보에 대한 지속적인 감시와 회복체계를 필요로 한다. DNA는 염기쌍의 결합으로 이루어지는데, 틀린 염기쌍이 정상적인 염기쌍보다 훨씬 낮은 빈도로 형성되지만, 이것이 수정되지 않고 DNA내에 축적될 경우 세포가 죽기도 한다.

본 연구에서는 DNA 복제 시 발생하는 실수, 손상된 부분을 회복하는 DNA 복구 기능을 모사하여 공학적인 개념을 도입 하였다. 기존에 발표 되었던 부분을 보완하여 여러 군데에서 발생한 오류 위치를 찾아내고 복구시키는 효율적인 알고리즘을 제시하였다. 단백질은 이중나선 구조로 결합되었으며 A는 T와, G는 C와 결합을 한다. 설계된 모델에서는 상부와 하부의 비트를 각각 더하면 3(11)이 되는데 모든 자리는 상보적인 관계가 있다.

본 연구에서 제시한 11개의 결합 모델에서 상부의 앞자리 비트의 1의 개수와 하부의 뒷자리 비트의 1의 개수의 합은 11이다. 또한 상부의 뒷자리 비트의 1의 개수와 하부의 앞자리 비트의 1의 개수의 합은 11이다. 총 1의 개수는 22개이다. 이 결합에 문제가 발생하면 그 오류 위치를 찾아내고 정상적인 값으로 돌리는 알고리즘을 제시하였다.

하나의 결합에서 발생한 오류에 대해서는 정확히 그 위치를 찾아내고 정상적인 결합에 대한 아미노산 값을 찾아내었다. 임의의 두 군데에서 발생한 오류에 대해서는 쉽게 위치를 찾아내었다. 그러나 정상적인 결합의 아미노산 값으로 돌리기 위해서는 4가지의 경우가 있었다. 각 경우에 대해서 1의 전체적인 개수는 22개가 되지만 제시된 알고리즘에 적용한 각 자리의 합계는 한 경우만 남기고 달랐다. 그러므로 정확하게 오류 위치를 찾아서 복구 시킬 수 있었다. 이후 본 연구는 세 군데 이상의 오류에 대한 복구를 살펴볼 것이다.

참고 문헌

[1] Ricard V. Sole, “Evolution and Self-assembly of protocells,” *The International Journal of biochemistry & Cell Biology 41,* pp. 274-284, 2009

[2] Bedau; John S. McCaskill, ″Open Problems in Artificial Life″, Artificial Life 6 : pp 363-376, 2000.

[3] V.P Shcherbakov, S.T. Sizova, T.S. Shcherbakova, I.E. Granovsky, and K. Yu. Popad, “In vivo Study of Fidelity of DNA Double Strand Break Repair in bacterio phage T4”, Russian Journal of Genetics Vol.44, No.9, pp1025-1030, 2008.

[4] 김석환, 허창우, ″ DNA 이중나선 구조에서의 오류 검출 및 복구, ″ 한국정보통신학회논문지, vol. 15, no. 11, pp. 2500-2504, 2011.11.

[5] 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.

저자소개

김석환(SokeHwan Kim)

2003.8 : 목원대학교 전자공학과 공학박사
2003.8～2005.2 : (주)제니텔 정보통신 선임연구원
2005.3～2006.1 : (주) 인터포스 책임연구원
2006.2～2008.4 : (주) 한백전자 책임연구원
2008.5～2010.8 : KAIST 박사 후 연구원
2001～현재 : 한국 해양정보통신학회 정회원
2009.1～현재 : 한국 해양정보통신학회 이사
2010.9～현재 : 이화여자대학교 전자공학과 연구교수
※관심분야 : Bio-inspired engineering, Self-repair, Gene Expression, Mutation,, FPGA 설계

허창우(Chang-wu Hur)

1991.2 : 연세대학원 전자공학과 공학박사
1986.9 ~ 1994.2 : 금성사 중앙연구소
1994.3 ~ 현재 : 목원대학교 전자공학과 정교수
※주관심분야 : 반도체공학 및 VLSI 설계