

<http://dx.doi.org/10.7236/IIBC.2016.16.5.229>

IIBC 2016-5-34

## 블록순환 행렬에 의한 이중나선 DNA 구조 (Ⅱ)

### A Double Helix DNA Structure Based on Block Circulant Matrix (Ⅱ)

박주용\*, 김정수\*\*, 이문호\*\*\*

Ju-Yong Park\*, Jeong-Su Kim\*\*, Moon-Ho Lee\*\*\*

**요약** 본 논문에서는 4개의 유전자 핵염기 C, U(T), A, G를 행렬로 표시하고,  $4 \times 4$  RNA(ribose nucleic acid)에서  $8 \times 8$  DNA(deoxyribose nucleic acid)로의 행렬 구조에 대해 서술한다. BCHJM (block circulant Hadamard-Jacket matrix)에 의해 DNA 이중나선 구조(double helix)를 해석한다. 직교 BCHJM은 비대칭 쌍 상보성(complementary)을 보이고 있다. 블록순환(block circulant) RNA 쌍 손상(damage) 신뢰성(reliability)은 기존 이중나선 보다 우수함을 보이고 있다.  $k=4$ ,  $N=1$ 인 경우 블록 순환 상보 쌍 신뢰도는 93.75%이고,  $k=4$ ,  $N=4$ 인 경우 신뢰도는 98.44%로 기존 이중나선의 경우 보다 4.69% 개선된다.

**Abstract** In this paper, we present the four genetic nitrogenous bases of C, U(T), A, G to matrices and describe the structures from  $4 \times 4$  RNA(ribose nucleic acid) to  $8 \times 8$  DNA((deoxyribose nucleic acid) matrices. we analysis a deoxyribose nucleic acid (DNA) double helix based on the block circulant Hadamard-Jacket matrix (BCHJM). The orthogonal BCHJM is anti-symmetric pair complementary of the core DNA. The block circulant ribonucleic acid (RNA) repair damage reliability is better than the conventional double helix. In case of  $k=4$  and  $N=1$ , the reliability of block circulant complementarity is 93.75%, and in case of  $k=4$  and  $N=4$ , it is 98.44%. Therefore it improves 4.69% than conventional case of double helix.

**Key Words** : DNA, RNA, BCHJM, Double Helix, Orthogonal.

## 1. 서론

분자생물학(molecular biology)에서 상보성(complementarity)은 각각 lock-and key 원리를 따르는 두 구조간의 관계를 설명한다. 자연계에서 상보성은 DNA(deoxyribose nucleic acid) 복제 기본 원리이고, DNA나 RNA(ribose nucleic acid) 시퀀스(sequence) 간에 사용되는 성질을 표현하는 원리이며, DNA와 RNA가 서로 비대칭으로 할당되면 시퀀스의 각 위치에서 nucleotide

base는 거울에서 보는 것처럼 거꾸로 보이는 상보성이 된다. 이처럼 상보성 base를 쌍으로 만들면 세포(cell)들이 한세대에서 다음 세대로 정보를 복사하고, 심지어는 시퀀스에 저장된 정보를 인식하고 정보손상을 개선하기도 한다. 구체적인 유전 정보를 전달하는 분자유전 시스템의 기본 파라미터(parameter)에 BCHJM (block circulant Hadamard- Jacket matrix)을 적용하면 BCHJM<sup>[2,3]</sup>의 기존 장점들이 바이오정보학에 적용될 수 있음을 알게 된다.

\*정회원, 신경대학교 ICT융합학과

\*\*정회원, 숭실사이버대학교 컴퓨터정보통신학과

\*\*\*정회원, 전북대학교 전자정보공학부(교신저자)

접수일자 : 2016년 7월 7일, 수정완료 : 2016년 8월 25일  
게재확정일자 : 2016년 10월 7일

Received: 7 July, 2016 / Revised: 25 August, 2016 /

Accepted: 7 October, 2016

\*\*\*Corresponding Author: moonho@jbnu.ac.kr

Dept: Division of Electronic Engineering, Chonbuk National University, Korea

1953년에 Watson과 Crick은 화학적 생물학적 방법을 사용하여 DNA 이중나선을 증명하였고 1963년 노벨상을 수상했다<sup>[1]</sup>. DNA는 그림 1에 보인 것과 같이 같은 축 주변을 감고 있는 각 두 개의 나선 체인(chain)을 가지고 있다. 두 체인 모두 우측으로 선회하는 방식이나, 쌍으로 되어 있어서 두 체인내의 원자 시퀀스들은 서로 반대 방향으로 진행한다. 본 논문에서는 바이오정보학 신호처리 시스템에 의해 DNA 이중 나선 구조를 설명한다. 유전자 코드 문자중 네 개의 문자는 24 패(triplet)를 가지고 있고, 블록 순환 행렬의 7 subset 코드가 얻어질 수 있다.

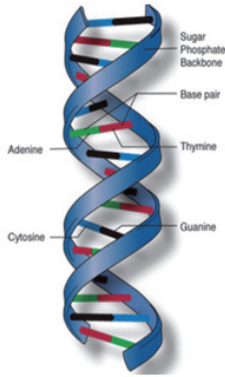


그림 1. DNA 이중 나선 구조<sup>[1]</sup>  
Fig. 1. DNA double helix structure<sup>[1]</sup>.

본 논문의 구성은 다음과 같다. II장에서는 BCHJM 행렬에 대해 서술하고, III장에서는 RNA 유전자 4×4 순환 행렬에 대해 서술하며, IV장에서 RNA 신뢰도에 대해 다루고 마지막으로 V장에서 결론을 맺는다.

## II. BCHJM(Block Circulant Hadamard -Jacket Matrix)

BCHJM은 [3]에 의해 다음과 같이 주어진다.

$$C_4 = \begin{pmatrix} C_0 & C_1 \\ C_1 & C_0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 \end{pmatrix} \quad (1)$$

그런데 BCHJM 시퀀스  $C$  는 다음과 같이 다시 쓸 수 있다.

$$C = I_0 \otimes C_0 + I_1 \otimes C_1 \quad (2)$$

여기서  $I_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ ,  $I_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ ,  $C_0 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ ,  $C_1 = \begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix}$  이다.

순환 부행렬(submatrix)의 크기는 2×2 이다. 즉 모든 행은 첫째행의 주기적천이(cyclic shift)가 이루어지며, 모든 열은 첫 번째 열의 주기적 천이가 이루어지는 성질을 각 부행렬이 갖도록 한다. 이러한 부행렬을 순환 블록(circulant block)이라 한다.

## III. RNA 유전자 4×4 순환 행렬

식 (1)과 유사한 방법으로 유전자 행렬이  $[C \ U; A \ G]$ <sup>[3]</sup>인 RNA 시퀀스의 예가 다음과 같이 [4]에 의해 주어진다.

$$P = \begin{pmatrix} C & U \\ A & G \end{pmatrix}, P^2 = \begin{pmatrix} C & U \\ A & G \end{pmatrix} \otimes \begin{pmatrix} C & U \\ A & G \end{pmatrix}, P^3 = \begin{pmatrix} C & U \\ A & G \end{pmatrix}^2 \otimes \begin{pmatrix} C & U \\ A & G \end{pmatrix} \quad (3)$$

여기서  $\otimes$  은 Kronecker product를 의미하고,  $C$ 는 시토신(cytosine),  $U$  는 우라실(uracil),  $A$  는 아데닌(adenine),  $G$  는 구아닌(guanine)을 나타낸다. 그림 2처럼 생명체가 계절에 따라서 나비가 되었다가 유충이 되고 애벌레, 그러다가 번데기, 나비가 다시 되어도 RNA는 변하지 않는다. 왜냐하면 eigenvector가 있고 그의 역은 같기 때문이다. 무잡음 코딩(noise-immunity coding)으로 유추하여 mosaic gene 행렬  $[C \ U; A \ G]$ 은 음과 양을 각각 -1 과 +1로 나타내는 주역의 64패로 대체 표시할 수 있다. 따라서  $[R]_8$  은 Rademacher singular 행렬로서 다음과 같이 나타낼 수 있다.

$$[R]_8 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 \\ -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 \end{pmatrix} \quad (4)$$

또한 다음과 같이 정의된다.

$$R_8 \square I_0 \otimes C'_0 \otimes P_2 + I_1 \otimes C_1 \otimes P_2 \quad (5)$$

여기서  $I_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ ,  $I_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ ,  $C'_0 = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ ,  $C_1 = \begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix}$  이고,  $P_2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$  와 같이 double stochastic 치환행렬이다.

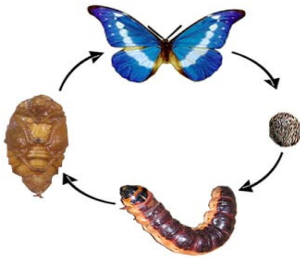


그림 2. 나비→알→애벌레→번데기 : 생명체의 변화. 외형은 계절(겨울→봄→여름→가을)에 따라 변하지만 원형 cell은 변하지 않고 Circulant 행렬  $\begin{pmatrix} C_0 & C_1 \\ C_1 & C_0 \end{pmatrix}$  과 같다.

Fig. 2. Butterfly→egg→larva→pupal: change of life. Appearance is changing according to season(winter→spring→summer→autumn), but original cell is not changing and is same to circulant matrix  $\begin{pmatrix} C_0 & C_1 \\ C_1 & C_0 \end{pmatrix}$ .

식 (4)는 중복행이 반복되어 있으므로 중복을 제거하면, Rademacher 행렬  $[R]_8$  로부터 다음 식을 얻는다.

$$R'_8 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 \end{pmatrix} \quad (6)$$

그리고 열 중복을 제거하면 식 (7)과 같은 singular 행렬이 얻어 진다.

$$R'_4 = \begin{pmatrix} 1 & 1 & 1 & -1 \\ -1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 1 \end{pmatrix} = \begin{pmatrix} C_0 & C_1 \\ C_1 & C_0 \end{pmatrix} \quad (7)$$

여기서  $C'_0 = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ ,  $C_1 = \begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix}$  이다.  $C'_0$  는 고유치  $\lambda_1^{(1)}, \lambda_2^{(1)} = 1 \pm i$  와 고유벡터  $\zeta_1 = (1-i)^T, \zeta_2 = (1+i)^T$  를 얻고,  $C_1$  또한 고유치  $\lambda_1^{(2)}, \lambda_2^{(2)} = \pm\sqrt{2}$  와 고유벡터  $\zeta_1 = (-1 + \sqrt{2}, 1)^T$ ,  $\zeta_2 = (-1 - \sqrt{2}, 1)^T$  을 얻는다. 따라서 일반식은 다음과 같이 얻을 수 있다.

$$R_{4 \times 2^k} = R_4^n \otimes \underbrace{P_2 \otimes P_2 \dots}_k, k = 0, 1, \dots \quad (8)$$

이러한 자기복제(self-replication) 행렬은 직교 Hadamard 행렬로 가능하나 역시 핵 DNA의 비대칭 쌍 상보성이 있다. 이러한 RNA의 복제 과정이 그림 3에 보인 바와 같이  $[R]_8$  까지 확장이 가능하다.

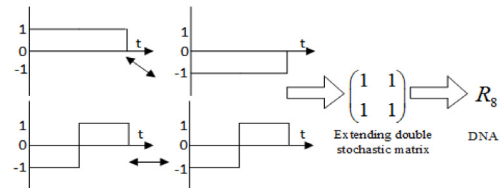


그림 3. 블록 다이어그램:  $4 \times 4$  RNA에서  $8 \times 8$  DNA행렬로의 구조  
 Fig. 3. Block Diagram: The construction from  $4 \times 4$  RNA to  $8 \times 8$  DNA matrices.

표 1. RNA  $4 \times 4$  쌍 유전자 행렬

Table 1. RNA  $4 \times 4$  pair genetic matrix.

RNA Class	RNA $4 \times 4$ Matrix	Pair Complementary: Sequence of Continuous Function
$\begin{pmatrix} C & U \\ A & G \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 1 & -1 \\ -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 1 \end{pmatrix}$	
$\begin{pmatrix} U & C \\ G & A \end{pmatrix}$	$\begin{pmatrix} -1 & 1 & 1 & 1 \\ -1 & -1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & -1 & -1 \end{pmatrix}$	
$\begin{pmatrix} A & G \\ C & U \end{pmatrix}$	$\begin{pmatrix} -1 & -1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 \end{pmatrix}$	
$\begin{pmatrix} G & A \\ U & C \end{pmatrix}$	$\begin{pmatrix} 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & 1 \\ -1 & -1 & 1 & -1 \\ -1 & 1 & 1 & 1 \end{pmatrix}$	

표 2. 상보성 RNA의 1/2 쌍  
Table 2. Half pair of complementary RNA.

RNA Class	RNA 4×4 matrix ( $C_0, C_1, C_0, C_1$ )	Half pair of complementary: sequence of continuous function
$\begin{pmatrix} C & U \\ A & G \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 \end{pmatrix}$	
$\begin{pmatrix} A & C \\ G & U \end{pmatrix}$	$\begin{pmatrix} -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 \end{pmatrix}$	
$\begin{pmatrix} G & U \\ A & C \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & -1 & -1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \end{pmatrix}$	
$\begin{pmatrix} C & A \\ U & G \end{pmatrix}$	$\begin{pmatrix} 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ -1 & -1 & 1 & 1 \end{pmatrix}$	

쌍 상보성 핵 RNA 커널(kernel)은 뉴클레오타이드 (nucleobase) 간의 (아데닌, 우라실, 구아닌, 시토신) 간의 분명한 상호작용에 의해서 달성된다. 그림 3에 보인 바와 같이 00에서 11로 01에서 10으로 쌍 상보성 RNA 커널이 있다. 표 1로부터 블록 순환하는 쌍 상보성의 4 종류 RNA 쌍 유전 행렬을 볼 수 있다. 반면에 블록 순환 RNA 1/2 쌍 상보성이 표 2에 제시되어 있다.

#### IV. RNA 신뢰도

상보성 시퀀스 종류에는 DNA나 RNA를 감지하거나 감지하지 못하는 두 가지가 있다. 감지하지 못하는 경우는 감지 시퀀스와 상보 관계에 있다. 따라서 손상을 보상하는 신뢰도(R)은 다음 식 (9)와 같이 주어진다.

$$R = 1 - \frac{1}{N} \prod_{i=1}^k (1 - p) \quad (9)$$

여기서  $p$  는 신뢰의 확률이고,  $k$  는 반복 행의 수이며,  $N$  은 블록 수이다. 식 (4)의 예로서 이중나선으로 인하여 가능한 신뢰도의 확률은 50%이고, N은 1과 4의 두 경우만 존재하므로,  $k = 4, N = 1$  인 경우는 다음과 같다.

$$R_{double} = 1 - \prod_{i=1}^4 (1 - 0.5) = 1 - (0.5)^4 = 93.75\% \quad (10)$$

제안한  $k = 4, N = 4$  인 블록순환 상보 쌍 신뢰도는 식 (11)과 같다.

$$R_{block} = 1 - \frac{1}{4} \prod_{i=1}^4 (1 - 0.5) = 1 - \frac{1}{4} (0.5)^4 = 98.44\% \quad (11)$$

따라서 손상 개선 신뢰도는 4.69% 개선된다.

#### V. 결론

본 논문에서는 BCHJM 특성을 이용한 DNA 이중나선에 대해 서술하였다. 이는 RNA와 상보 쌍을 이룬다. 神은 인간을 창조할 때 심장, 간, 등 인체의 중요부분을 오직 하나만 갖게 했다. 그러나 단백질 유전자가 이중나선 구조를 갖게 해 손상 개선 신뢰도(repair damage reliability)를 향상시켰다. 본 논문에서 제안한 BCHJM에 의하면 블록 순환 RNA 손상 개선 신뢰도는 기존 이중나선의 경우 보다 4.69% 개선됨을 알 수 있다.

#### References

- [1] Watson, J. D., Crick, F. H. C., "Molecular structure of nucleic acids," Nature, vol. 171, no. 4356, pp. 737-738, April 1953.
- [2] Lee, M. H., Hou, J., "Fast block inverse Jacket transform," IEEE Signal Processing Letters, vol.13, no.8, pp. 461-464, 2006.
- [3] Lee, M. H., Hai, H., Zhang, X. D., MIMO Communication Method and System using the Block Circulant Jacket Matrix, USA Patent 14/688.524, 03/03 2016.
- [4] Matthew H., Petoukhov, S., Mathematics of Bioinformatics, Wiley, 2011.
- [5] Sung Kook Lee, Ju Yong Park and Moon Ho Lee, "A Double Helix DNA Structure Based on the Block Circulant Matrix ( I )," The Journal of the Institute of Internet, Broadcasting and Communication(JIIBC), Vol. 16, No. 6, October 2016.
- [6] Jae Seung Yang, Ju Yong Park and Moon Ho Lee, "Jacket Matrix in Hyperbola," The Journal of the

Institute of Internet, Broadcasting and Communication  
(JIIBC), Vol. 15, No. 3, June 2015.

## 저자 소개

### 박 주 용(정회원)



- 1994년 : 전북대학교 전자공학과 박사
  - 1991년 ~ 2007년 : 서남대학교 전기전자공학과 교수
  - 2007년 3월 ~ 현재 : 신경대학교 ICT 융합학과 교수
- <주관심분야 : 무선이동통신, 통신이론, Molecular communication>

### 김 정 수(정회원)



- 1998년 : 전북대학교 정보통신공학과 석사
  - 2003년 : 전북대학교 컴퓨터공학과 박사
  - 2002년 6월 ~ 현재 : 숭실사이버대학교 컴퓨터정보통신학과 부교수
- <주관심분야 : 이동통신>

### 이 문 호(정회원) 교신저자



- 1984년 : 전남대학교 전기공학과 박사, 통신기술사
- 1985년 ~ 1986년 : 미국 미네소타 대학 전기과 포스트닥터
- 1990년 : 일본동경대학 정보통신공학과 박사
- 1970년 ~ 1980년 : 남양MBC 송신소장

- 1980년 10월 ~ 2010년 2월 : 전북대학교 전자공학부 교수
  - 2009년 4월 ~ 2013년 8월 : WCU-2 연구책임교수
  - 2015 : 국가연구개발 우수성과 100선
  - 현재 : 전북대학교 전자공학부 초빙교수
- <주관심분야 : 무선이동통신, 통신이론, Molecular communication>

※This work was supported by MEST, 2015R1A2A1A05000977, NRF, BK, Korea.