

*Mathematical Blockchain Modeling of
Peer-to-Peer Networks for
Gene Expression in Cell Population*

Jung In Choi

Seoul National University

jichoi@snu.ac.kr

(Title)

Mathematical Blockchain Modeling of Peer-to-Peer networks for
Gene Expression in Cell Population

(Abstract)

The study presents a mathematical modeling for supporting the premise that our body is composed of blockchain systems. A cell population is considered a space with peer-to-peer communication networks for applying blockchain protocol. Transaction is defined as gene expression that constantly occurs for protein synthesis in each cell. The transaction process proceeds according to the blockchain protocol with sharing and recording the data on the blockchain ledger. The premise comes from the clues of the previous research such that the cell population is a complex structured network system having cell-to-cell connections. Although individual cells exhibit stochastic nonlinear dynamic behavior, cell population shows consensus behavior that reaches to ensemble through interaction among cells. It is inferred that gene expression is not regulated by the corresponding cell only nor determined by external intelligence such as the brain. This is achieved through a consensus through stochastic interactions between cells over the whole cell population. These findings imply that mathematical blockchain modeling is a suitable for gene expression process.

The original contribution of the study is a methodology for applying mathematically the blockchain protocol to the real biological gene expression process. In other words, the DNA sequence is converted

into a numeral bit sequence that makes it possible to implement the blockchain protocol. A new DNA sequence scheme is proposed with adding methylated cytosine and adenine as the 5th and 6th bases for including epigenetic information which has profound effect on gene expression and regulation. The methodology was applied to the real biological synthesis process of protein samples. The protein is composed of amino acids that are encoded by triplet codons of 216 kinds with 6 base RNA sequence. The gene expression information is traced backward from a synthesized protein sample, amino acids of codons, RNA transcript up to DNA sequence. One of the results is a numeric value in the form of a bit sequence with which mathematical blockchain modeling is applicable. The cryptographic authentication and the consensus process are mathematically proven to work properly by the blockchain protocol. It implies that the same protein is synthesized, but with different epigenetic data, then protein's latent material properties will certainly be different.

Although the result has not been justified by the biologic experiment yet, it is sure that the biological hidden algorithm inside DNA sequence will be revealed by the binary bit-logic with physical on/off states which is mathematically proven. The research will greatly contribute to disease treatment and medicine development as well as epigenetics in the future.

(Keywords) Blockchain, P2P Networks, Nucleobase DNA/RNA sequence, Gene Expression, Cell Population, Epigenetics, Methylation, Consensus, Hash Algorithm, Triple Codons, Protein Synthesis, Cryptocurrency

1. Introduction

In the case of multicellular eukaryotic organisms such as humans, as one cell continues to divide into the same cell, it becomes a living organism consisting of 100 trillion cells. In the cell development stage, cells differentiate into functional cells according to their role and are made to match the properties of the organ. Nevertheless, all cells have the same genetic information in their DNA (deoxyribonucleic acid). Intercellular diversity is due to gene expression for the cell population that makes up the organ.

Why does a gene cause the expression? How does this gene know when it should be expressed? What kind of gene expression produces the corresponding protein synthesis? The answers to such questions still lie in the unknown area. The principle of gene expression has so far been unclear. Challenges are underway for research into gene expression. [1,2,3]

In this study, it is assumed that gene expressions in cell population are shared by data, regulated by consensus and stored in the form of a blockchain with cryptographic security over peer-to-peer (P2P) networks. [4,5]

This proposal is supported by some clues from previous studies as follows.

The first clue from previous research is that gene expression shows the nature of complex networks in the cell population. The basic research issue is to unravel the structure and dynamics of the networks. Some

research on gene expression speculated that gene regulatory networks govern the behavior of cells in multicellular organisms, such as proliferation or differentiation into specialized cell types. Other research has proposed that cellular networks have a critical role in biological functions such as tissue development or immune responses. This implies that a range of dynamic cell population behaviors, including cellular synchronization, delays, and bimodal responses, can emerge from secure decentralized cell-to-cell communication networks. [6,7,8] Blockchain supports its applicability to gene expression with its cryptographic security over decentralized P2P communication networks.

The second clue from the previous study shows that gene expression has the nature of stochastic behavior in cell populations. Gene expression, defined by a series of responses that control the abundance of gene products, affects most aspects of cell behavior. Gene products are susceptible to stochastic fluctuations due to complexity. Coordinated control of multiple genes must overcome the difficulties of substantial instability of gene products. To mitigate the complexity, some stochastic cryptographic remedy is presumably applied. The fundamental technique of blockchain is the hash algorithm, which is based on the knowledge of cryptography against stochastic complexity. [9,10,11]

The third clue from the previous research revealed that gene expression has the potential of 'Ensemble' in cell population. Many studies have already shown that the emergence of collective behavior, such as synchronization of processes, can arise due to the nonlinear regulations of complex network systems with environmental perturbations.

However, it remains unclear how the complex and dynamically evolving molecular networks found in biological systems can give rise to 'Ensemble', a globally coherent orchestrated response. Many mechanisms contribute to the regulation of gene expression to ensure coordinated cellular behaviors and fate decisions through ensemble in the cell population. For this point, the consensus algorithm of blockchain presents the nature of 'Ensemble' in cell population. [12,13,14]

In summary, this study attempts to explain the principle of gene expression in organisms by adopting blockchain concepts such as secure P2P networks, the cryptographic hash algorithm and the consensus algorithm in the cell population.

II. Mathematical Blockchain Modeling

The original contribution of the study is a methodology for applying mathematically the blockchain protocol to the real biological gene expression process. In other words, the DNA sequence is converted into a numeral bit sequence that makes it possible to implement the blockchain protocol.

II.1 Cellular Blockchain Concept

Blockchain began with the motivation of collective governance of a group of nodes with a decentralized P2P network system. This technology ensures transparency and reliability by data sharing and consensus with security. As an example, blockchain shows the possibility of replacing the legacy TTP (Trusted Third Party), a central bank, with cryptocurrency such as Bitcoin [15,16,17].

A cellular blockchain concept is motivated by the collective governance of gene expression in the cell population. In other words, a local organ is autonomously coordinating gene expression by a blockchain process rather than by a centralized intelligence, such as the brain.

Figure 1 shows the similarity between the digital blockchain formed in the internet space and the cellular blockchain formed in our body. It is assumed that the digital node corresponds to a cell and the CPU corresponds to DNA in the cellular blockchain.

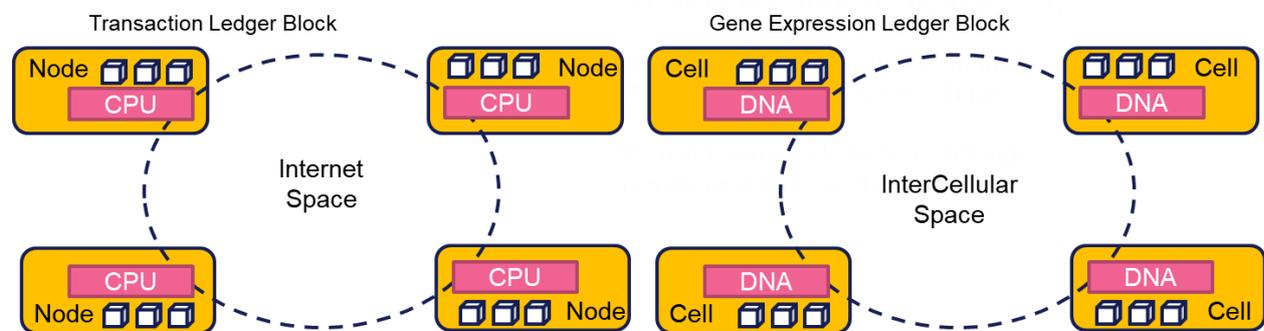


Figure 1 Similarity between Digital Blockchain and Cellular Blockchain

Mathematical modeling is based on the concept of DNA functional layers for cellular blockchain in Figure 2.

Hardware is divided into a virtual layer and a biophysical layer. In the hidden virtual layer, protocols and data are implemented by the bit logic from two states (on/off) of the basic element in a cell., while in the observable biophysical layer, they are implemented by the sequence of four kinds of nucleobases, such as A(adenine), G(guanine), C(cytosine), and T(thymine).[18]

	genetic layer (download by heredity)	epigenetic layer (update by experience)
virtual layer (two kinds of status : on/off)	system device binary numeral system to quaternary(a base of 4) (blockchain protocol)	memory device binary numeral system to senary(a base of 6) (blockchain data)
biophysical layer (4 kinds of molecule : AGCT)	4 nucleobase DNA sequence : A,G,C,T	6 nucleobase DNA sequence : A,G,C,T, ^m C, ^m A

Figure 2 Cellular DNA Functional Layers

Software is divided into a genetic layer that is immutable and an epigenetic layer that is updatable. The blockchain protocol is downloaded by inheritance in the genetic layer, while the blockchain data are updated by experience of gene expression resulting in methylation in the epigenetic layer.

In more detail, the blockchain protocol is implemented in the hidden virtual layer of hardware by a binary numeral system. It is converted to a quaternary numeral system for mapping into a 4 nucleobase DNA sequence in the biophysical genetic layer of DNA, which is observable and immutable. Additionally, the blockchain data are implemented in the hidden virtual layer of hardware by the binary numeral system. It is converted an octal numeral system for mapping into a 4 nucleobase DNA sequence and 4 more methylated nucleobase. However, in the real situation, only methylated cytosine and methylated adenine are observed, so the blockchain data can be formed by a 6 nucleobase DNA sequence in the biophysical epigenetic layer of DNA that is observable and updatable. The modelling of DNA functional layers is a core concept of cellular blockchain.

II.2 Cellular Blockchain Cryptograph

A purpose of blockchain is to provide the cell population with security over the cell-to-cell communication network by using cryptography, which is the process of encrypting and decrypting cellular data. The cellular public and private key pair is a crucial element of cryptography. This is how private and public keys work together. The private key is used to encrypt cellular block data and the public key is used to decrypt it to execute transactions such as gene expression. A single cell on the blockchain network in the cell population has its own private and public key pair and cellular address, which may be assigned in the stage of cell formation.

Figure 3 presents a real biophysical situation with a virtual scenario.

Step 1: The blockchain system generates and assigns a unique random number to the cell by the binary system of on/off status in the hidden virtual level. The binary number is converted into a 12-digit hexadecimal numeral value of private key (4ef17e7cab3e) through the hashing algorithm, which is described in further detail in the next section. The numeral private key is converted into a quaternary numeral system, and then mapped to the corresponding nucleobase AGCT to be biophysically observable in the DNA. Thus, each cell has its own cellular private key in a real 18-digit DNA sequence (CTGGTCATTGATAGAACC).

Step 2: The cellular private key in Step 1 creates its pairing public key by using complex functioning defined in the hidden virtual level. The resulting binary number is converted into a 12-digit hexadecimal

numeral hash value (633f01ec34cc) for the public key. The public key in the quaternary system is mapped to a real 18-digit DNA sequence (GCAGACTCGGATAACAACA) for the cellular public key in the observable biophysical layer. Thus, a cell can go through authentication for the blockchain network with its own private and public key pair.

Cellular Address, Cellular Private Key, Cellular Public Key

Step1

1. random number generation : 101101011110001111010011000100001010 (binary)
2. hashing : 12-digit hexadecimal system → 4ef17e7cab3e (numeral private key in virtual level)
3. quaternary system conversion → 18-digit AGCT sequence

C	T	G	G	T	C	A	T	T	G	A	T	A	G	A	A	C	C
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Cellular Private Key
in biophysical layer

Step2

1. complex functioning : 11000010010111001010011001000001000 (binary)
2. hashing : 12-digit hexadecimal system → 633f01ec34cc (numeral public key in virtual level)
3. quaternary system conversion → 18-digit AGCT sequence

G	C	A	G	A	C	T	C	G	G	A	T	A	C	A	A	C	A
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Cellular Public Key
in biophysical layer

Step3

1. numeral public key 633f01ec34cc → double hashing : 1e4e18fde20b (numeral address in virtual level)
2. quaternary system conversion → 18-digit AGCT sequence

G	G	G	C	C	C	T	C	A	T	A	A	T	C	T	G	C	G
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Cellular Address
in biophysical layer

Figure 3 Blockchain Security over Cell-to-Cell Communication Network

Step 3: The cellular address is an identification number of the cell for the blockchain network. The address is derived by using the hashing process twice with a public key. The cellular address is designated to a 12-digit hexadecimal number (1e4e18fde20b) in a hidden virtual level and is observable with quaternary mapping to a real 18-digit DNA sequence (GGGCCCTCATAATCTGCG) in the biophysical layer. The receiver cell can obtain the sender's address from the public key.

DNA in a cell is programmed in the nucleobase sequence for gene

expression to be realized in protein synthesis. We still do not know how or at what time gene expression occurs in each cell. Additionally, it is not known who is regulating such a gene expression. In our hypothesis, gene expression that occurs in a cell is not regulated by the cell alone nor determined by the fate in an external intelligence. All members of cell population participate in the gene expression in which they work together, coordinate, and harmonize consensus. The application of blockchain is well suited for the modeling of gene expression in the cell population. That is why gene expression is considered a transaction on the blockchain network.

In this case, private and public keys alone are not enough to confirm gene expression. Another layer of security is required to verify and prove gene expression. This is the consensus process in the cell population, which is based on cryptography with the cellular hash algorithm and game theory [19,20,21,22,23,24,25,26]

II.3 Cellular Secure Hash Algorithm

The cellular hash algorithm is the process of taking the input nucleobase sequence of any length and turning it into a cryptographic nucleobase sequence output of the fixed length. It is a one-way irreversible encryption. A cryptographic hash algorithm needs to have several crucial qualities to be considered useful: Every hash is different from another. The same hash will always be generated for the same data. It is impossible to infer input data based on the hash output value. Even with a small change to the input, the whole hash will be changed.

The process of CSHA18 (Cellular Secure Hash Algorithm 18) is shown in Figure 4.

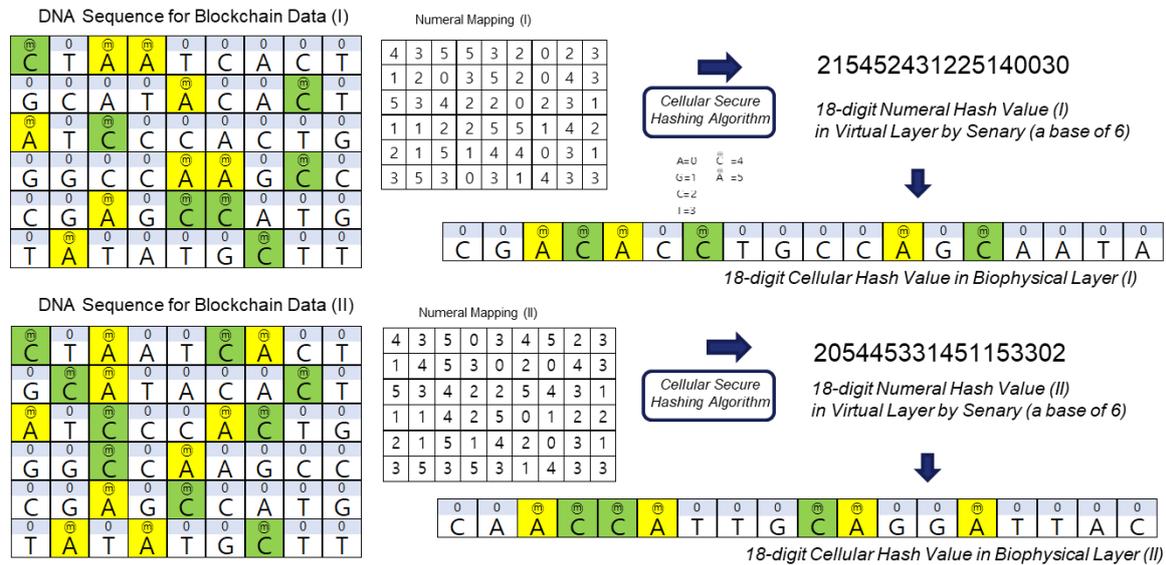
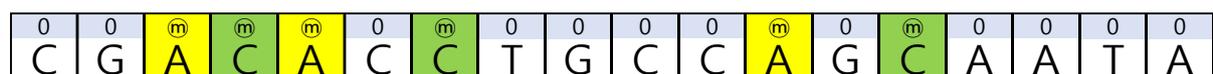


Figure 4 Cellular Secure Hash Algorithm 18 (CSHA18)

CSHA18 is defined as SHA of 18-digit length output. Any size of blockchain data input goes through CSHA18, resulting in a real 18-digit length nucleobase DNA sequence in the biophysical epigenetic layer. In reality, the blockchain data in the binary system go into CSHA18, and then the output comes out turning into a Numeral Hash Value of 18-digit number (2154524312251400302) by a 6-base senary system in the virtual epigenetic layer. In the biophysical layer, the numeral Hash Value is mapped into a **Cellular Hash Value (I)**, a real 18-digit nucleobase DNA sequence in the epigenetic layer. That is,



For the other table of DNA sequence (II), which is slightly changed by the methylation of cytosine and adenine, the outcome of CSHA18 is

completely different so that the original cannot be estimated at all. For example, a Numeral Hash Value (II), (205445331451153302), by a 6-base senary system is mapped to another 18-digit nucleobase DNA sequence of **Cellular Hash Value (II)**, in the biophysical epigenetic layer. That is,

0	0	Ⓜ	Ⓜ	Ⓜ	Ⓜ	0	0	0	Ⓜ	Ⓜ	0	0	Ⓜ	0	0	0	0
C	A	A	C	C	A	T	T	G	C	A	G	G	A	T	T	A	C

III. Methodology

In this study, a methodology for applying the blockchain protocol to the cell population has been illustrated by the process of transaction for gene expression.

III.1 Cellular Blockchain Transaction Process

In Figure 5, as an example, the current block is a collection of transactions for gene expression in 6 cells. In the hidden virtual layer, the transaction table is a numeral 6x9 matrix by the senary system (0,1,2,3,4,5). It is the combination of the immutable genetic table by the quaternary (0,1,2,3) and the updatable epigenetic table by the binary (0,1). In the observable biophysical layer, the transaction for gene expression in a cell is a 9-digit DNA sequence of 6 kinds of nucleobase AGCT+^mC, ^mA (methylated cytosine and Adenine). In the conventional biology, the transaction is just a DNA sequence of 4 kinds of nucleobase AGCT, which cannot imply any information on epigenetic effects due to gene expression processes.

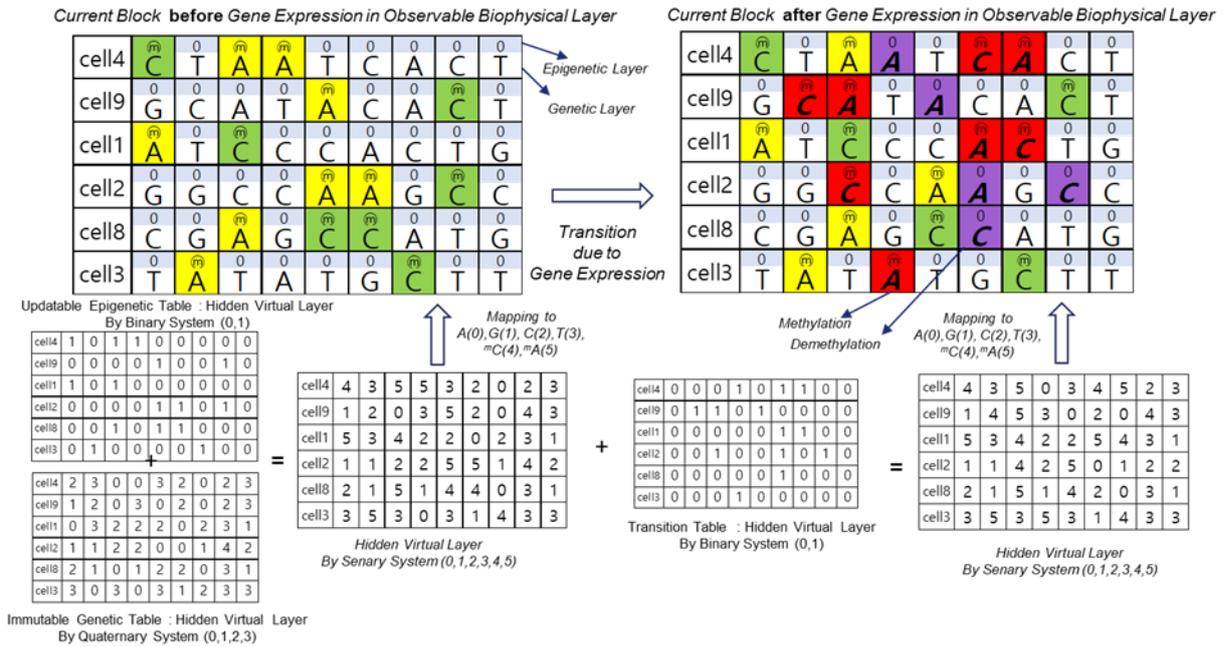


Figure 5 Current Block for Gene Expression in Cellular Blockchain

The transaction process includes a transition table for the update of the epigenetic table from before to after gene expression. The component of the table is '0', which means no change and '1', which means a change in epigenetic status. The transition table is reflected on the transaction in the ledger of the blockchain. In the biophysical layer, the epigenetic part of transaction is represented by the methylated nucleobase, such as adenine in yellow and cytosine in green. Then, the transition due to gene expression is implemented by the process of methylation in red and demethylation in purple. The nucleobase with such a process will be somewhat different, so it is recorded with extra metadata, in the ledger of transaction in the blockchain, such as the bold italic symbol in Figure 5. In reality, the difference may be identified by the degree of methylation plasticity that may imply the epigenetic history data of the cell.

In this way, a current block is formed and connected to the previous block. The number of blocks in connection is continuously increased over time to form a cellular blockchain. The blockchain is finally confirmed through the consensus process in the cell population for the final security stage. The details of this are described in the next section.

III.2 Cellular Blockchain Consensus Algorithm

This section describes the consensus algorithm for the cellular blockchain process. A new current block is assembled by gathering a scenario of qualified gene expressions that are required for protein synthesis in cells. The game participant cell forms its own scenario for the candidate of the current block. The next step is that such participant cells start the game called "cellular mining" to win the approval of the current block through a consensus process over the whole cell population. Following the blockchain protocol, the players continue to try the game until first reaching consensus with a correct answer. The winner cell shares the correct answer with the cell population and obtains approval when there are no objections. The point of this proposal is that gene expression scenarios are rehearsed in advance to function properly without any conflicts in cell population.

As an example shown in Figure 6, a sequence of gene expression is booked for cell 4, followed by the sequence of cells 9, cell 1, cell 2, cell 8 and cell 3. A transaction ledger of sequences is created in this way. The current block includes not only the transaction data but also the cellular hash value of the previous block and the arbitrary nonce matrix. The consensus algorithm is based on a game rule in which the player

cell changes the nonce matrix in a brute-force random manner to repeat the hashing process and wins the game if the resulting hash value of the sequence first reaches a target minimum, such as 6 consecutive adenines. That is,

0	0	0	0	0	0	m	0	0	0	m	0	0	m	0	0	m	0
A	A	A	A	A	A	C	G	T	G	C	A	T	A	C	A	C	T

The current block determined by the winner cell is connected to the previous block of the blockchain, thereby growing the blockchain. By tracking the blockchain backward, we may reach the data of the first Genesis Cell.

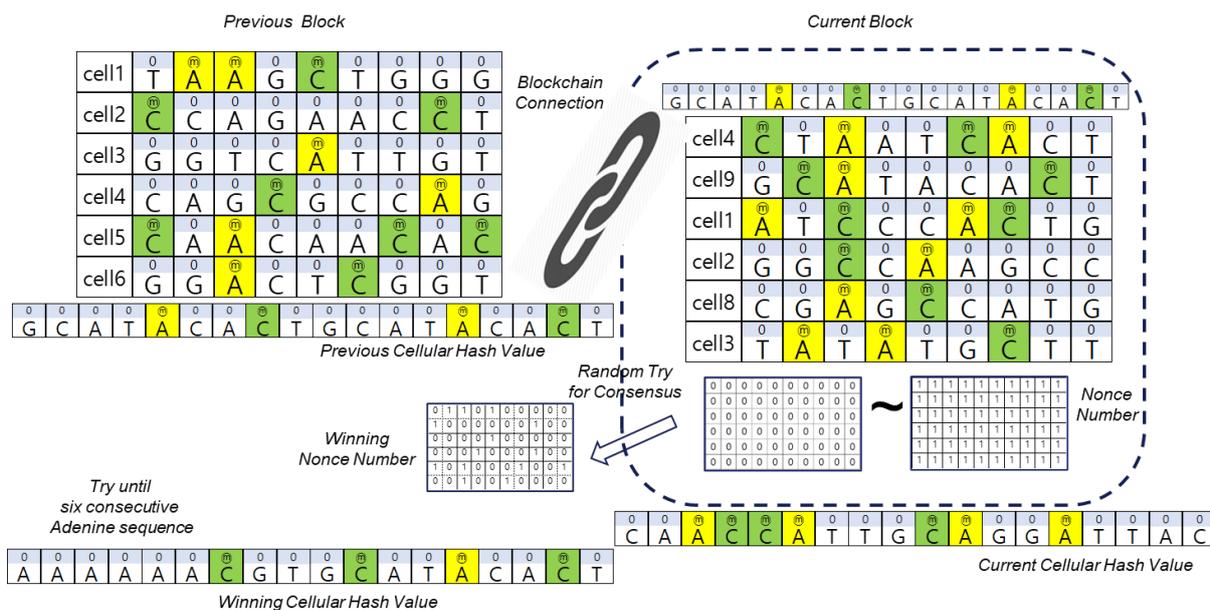


Figure 6 Consensus Algorithm for Current Block in Cellular Blockchain

IV. Application

The blockchain methodology has been applied to the real biological

DNA sequence for gene expression in a cell. How does the blockchain work biophysically for gene expression? As shown in Figure 7, the core algorithm for blockchain has been embedded from the beginning in the DNA genetic biophysical layer while blockchain data are presumably updated in the DNA epigenetic biophysical layer. The genetic layer provides firmware with the AGCT 4 nucleobase sequence, and the epigenetic layer provides data memory with AGCT+^mC, ^mA, and 6 nucleobase sequences in nuclear DNA of the cell. The real cell is embodied by a binary system in the virtual DNA layer, a quaternary numeral system for the genetic layer and a senary numeral system for the epigenetic layer.

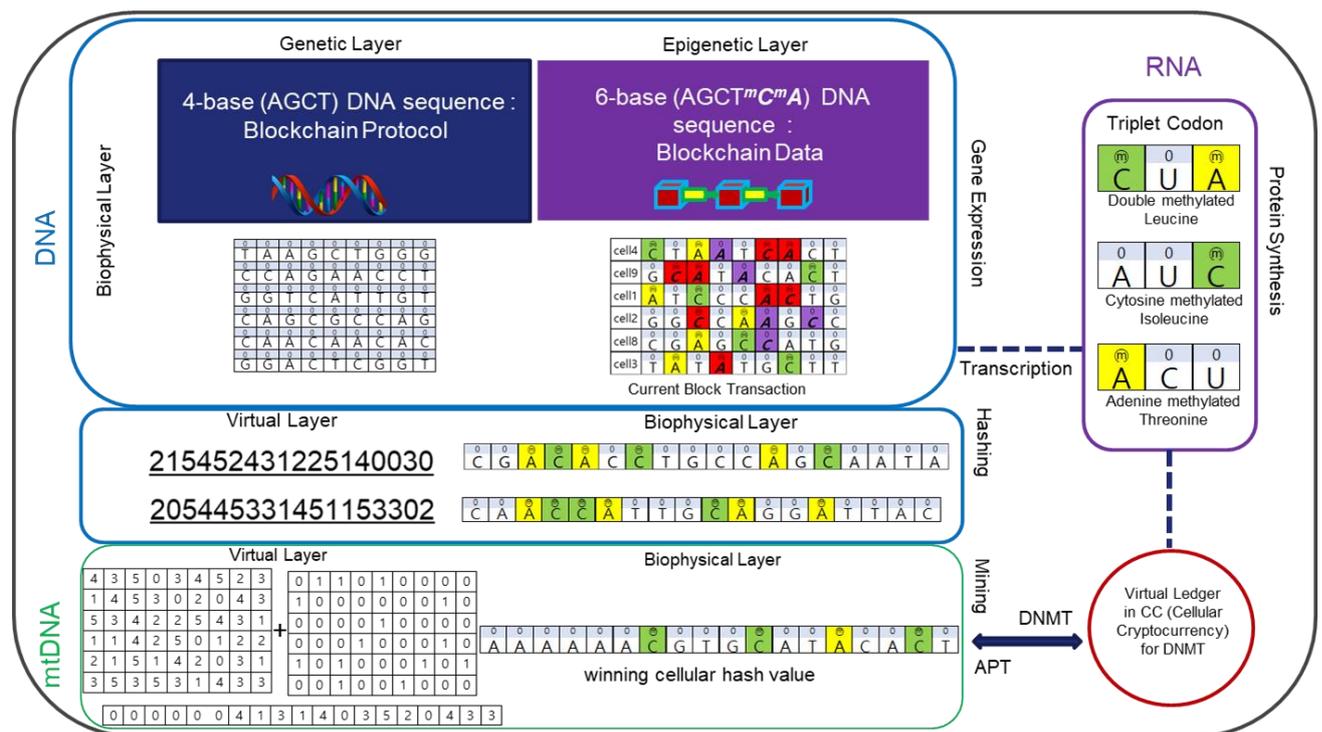


Figure 7 Biophysical Application of Cellular Blockchain in a Cell

The cellular hash algorithm is executed by a binary system in the hidden virtual layer and then cellular hash values are generated in the form of

18-digit length 6 kinds of nucleobase DNA sequences in the observable biophysical layer. Similarly, the consensus algorithm is executed by a binary system in the hidden virtual layer. The process requires time and energy consumption because of brute force random trials of 'mining' for rewards. Thus, it is assumed that the mining process may be performed not in the nucleus DNA, but in the multiple mitochondrial DNA with spreading across the cytoplasm, allowing parallel processing. However, the approved data block with the transaction and hash values are recorded in the ledger of the cellular blockchain in the form of 18-digit length 6 kinds of nucleobase DNA sequences in the observable DNA biophysical layer.

V. Results

The study has presented the results of application to the real biological synthesis process of protein sample.

As shown in Figure 8, the protein is composed of amino acids that are encoded by triplet codons of RNA sequence. The gene expression information is traced backward from a synthesized protein sample, amino acids of codons, RNA transcript up to DNA sequence, which can be observed in the biophysical layer. For mathematical blockchain modeling, the gene expression data has been converted to the numeral values in the hidden virtual layer. One of the results is a numeric value in the form of a bit sequence with which mathematical blockchain modeling is applicable. The cryptographic authentication and the consensus process are mathematically proven to work properly by the

blockchain protocol.

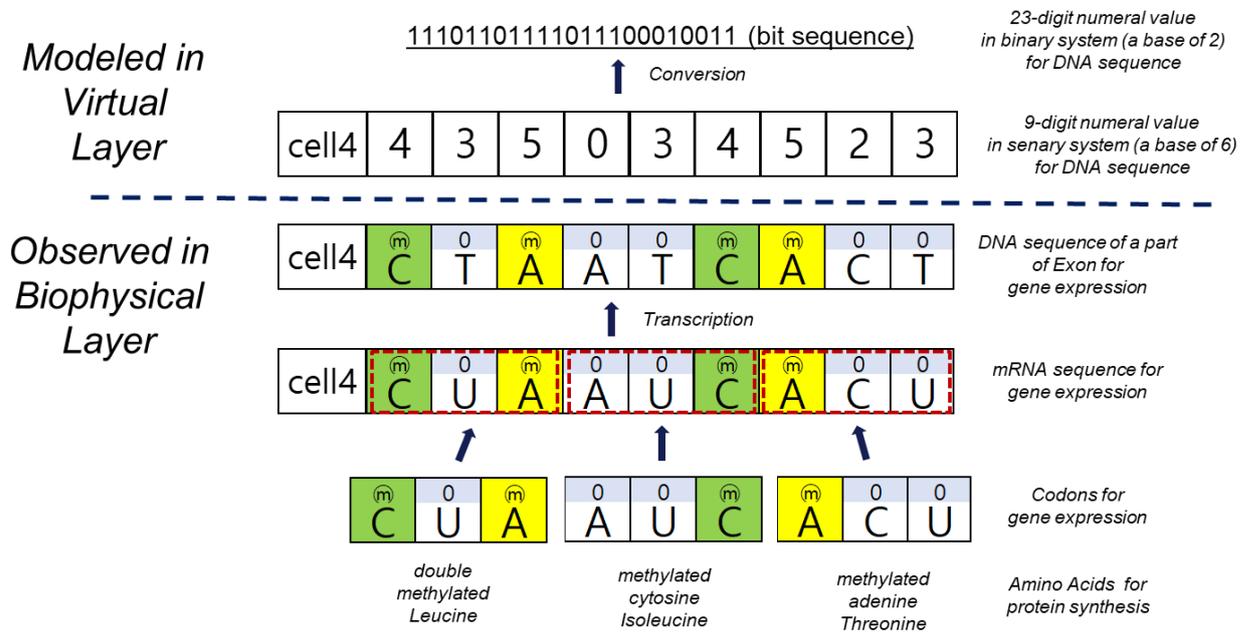


Figure 8 Mathematical Blockchain Modeling Results

VI. Further Discussion

A new triple codon scheme, which has been used in the results, is to be discussed furthermore. In addition, a cryptocurrency concept, which is the core application in the digital blockchain, is to be discussed for cellular blockchain ecosystem.

VI.1 Triplet Codons Scheme

The 9-digit DNA sequence is transcribed into 3 triplet RNA codons. In conventional biology [27,28], the possible case is 64 kinds of codons from the AGCU quaternary nucleobase RNA system. However, these codons do not distinguish the methylated nucleobase components. In

this study, by adding 2 more methylated RNA nucleobases, the 6 kinds of RNA nucleobase form the 216 kinds of triplet codon. In the previous result, Cell #4 is going to the gene expression process with a 9-digit nucleobase DNA sequence. It is transcribed to the RNA sequence, which is composed of 3 triplet codons.

^m C	⁰ U	^m A
----------------	----------------	----------------

 is Double methylated Leucine;

⁰ A	⁰ U	^m C
----------------	----------------	----------------

 is Cytosine methylated Isoleucine, and

^m A	⁰ C	⁰ U
----------------	----------------	----------------

 is Adenine methylated Threonine. These codons include information on epigenetic effects on the amino acids and proteins.

Thus far, it is understood that if the nucleobase sequences of the genetic layer are the same, the results of protein synthesis are the same. However, if the data of the epigenetic layer are different, the results must be different. The role of epigenetic data has not yet been identified in the process of protein synthesis. In this study, it is postulated that the epigenetic data may be used for the regulation of gene expression and for the sustainable cellular blockchain ecosystem with cryptocurrency.

VI.2 Cryptocurrency Concept for Cellular Blockchain Ecosystem

To maintain the blockchain ecosystem continuously, digital cryptocurrency is introduced. [29,30] Then, what is the concept of cryptocurrency in the cellular blockchain?

Cryptocurrency should be able to exchange and store some measurable value for transactions. In this study, a transaction in the cellular blockchain is gene expression. In other words, the transaction means the nucleobase sequence includes the changes in the epigenetic layer

through gene expression. If the changes are quantifiable in cellular cryptocurrency (CC), the transaction can be recorded with the change of CC in the blockchain ledger.

Then, what is the value of cellular cryptocurrency? This is related to the regulation of gene expression, and it must be closely related to epigenetic influence. The results produced through gene expression can be quantified by the number of methylation or demethylation. In this study, the value of 1 CC is defined as the amount required for methylation of 1 nucleobase. For example, as shown in Figure 5, if 2 methylation (in red) and 1 demethylation (purple) occur through gene expression in the 4th cell, the cell has exhausted 2 CC and earned 1 CC. As a result, 1 CC is reduced in the balance. Alternatively, in the 2nd cell, the methylation is 1 CC and the demethylation is 2 CC, so the balance is increased by 1 CC.

What is the biomaterial that embodies biophysically cellular cryptocurrency? In this study, it is assumed that one of the biomaterials is a DNMT (DNA methyltransferase) [31] that activates methylation. Once the amount of DNMT corresponding to 1 CC is determined, the amount of DNMT accumulated in the cell can be converted into CC, which will be the current balance in CC. As such, the balance of cryptocurrency is booked in the blockchain ledger as a result of the transaction of gene expression. The balance represents epigenetic plasticity that can be inherited by future generations. [32]

In more detail, in order to activate competition to reach consensus, more cells should participate in the game, depleting ATP (adenosine triphosphate) in the mitochondria. [33] If so, what is the motivation of

the cell to induce it? It may be a selfish characteristic of cells. [34,35] In other words, it may be such a cloning instinct that accumulates the effects of gene expression occurring in the cell population through epigenetic plasticity to strengthen its inherent empirical traits and transmit it to the next generation.

Such transactions with cellular cryptocurrency are used to sustain the blockchain ecosystem. In addition to transactions, a cell can earn the cryptocurrency by the consensus game. The cell that has won a consensus game is granted a certain amount of cryptocurrency as a reward. Instead, the cell must consume much energy in the mitochondria with ATP during the game for mining process. To compensate for ATP depletion, the cells buy ATP from the others in CC. Therefore, in the blockchain ecosystem, it is necessary to regulate and maintain the exchange rate between the appropriate ATP and CC within the cell group for sustainability. In the actual biophysical layer, DNMT and ATP are exchanged.

VII. Conclusion

The conclusion of this study is such a premise that our body is a blockchain system. This means that all activities in our body are being processed through the blockchain protocol on the peer-to-peer communication network in the cell population. All cells have the same genetic data in their DNA, but intercellular diversities are due to gene expression, from which epigenetic data are postulated to be shared, regulated by consensus and stored in the form of blockchain with

cryptographic security over the cellular network.

According to the clues from previous research, the cell population in which numerous gene are continuously expressed is a complex structured network system with cell-to-cell connections. Although individual cells exhibit stochastic nonlinear dynamic behavior, the cell population shows consensus behavior that reaches to ensemble through interactions among cells. It is inferred that gene expression is not regulated by the corresponding cell only nor determined by external intelligence such as the brain. This is achieved through a consensus agreement through stochastic interactions between cells in the cell population. The consensus algorithm in blockchain is provided for gene expression transactions with mathematical proof.

The original contribution of the study is a methodology for applying mathematically the blockchain protocol to the real biological gene expression process. In other words, the DNA sequence is converted into a numeral bit sequence that makes it possible to implement the blockchain protocol. Mathematical modeling is based on the concept of DNA functional layers for cellular blockchain. Hardware is divided into a virtual layer and a biophysical layer. In the hidden virtual layer, protocols and data are implemented by bitwise logic of a binary system, while in the observable biophysical layer, they are embodied by nucleobase DNA sequences. Software is divided into a genetic layer and an epigenetic layer. The existing DNA sequence for gene expression is composed of only 4 bases, AGCT in the genetic layer, while in this study, a new DNA sequence scheme is presented with adding two more

nucleobases considering epigenetic layer data. In other words, by adding methylated cytosine and adenine as the 5th and 6th bases, the DNA sequence is composed of 6 bases. In the hidden virtual layer, the new DNA sequence scheme is operated by the senary numeral system rather than the quaternary system. In the transcription process for protein synthesis, 216 types of RNA triple codons are presented rather than 64 types in conventional biology.

Thus far, it is understood that if the nucleobase sequences of the genetic layer are the same, the results of protein synthesis are the same. However, if the data of the epigenetic layer are different, the results must be different. The role of epigenetic data has not yet been identified in the process of protein synthesis. It is postulated that epigenetic data may be used for the regulation of gene expression and for the sustainable cellular blockchain ecosystem with cryptocurrency.

Cryptographic private and public key pair provides the cell with a complete authentication process over blockchain network. Additionally, the cellular secure hash algorithm is defined for the process of taking the input nucleobase sequence of any length and turning it into a cryptographic nucleobase sequence output of the fixed 18-digit length.

For the consensus algorithm, it is postulated that stochastic interactions among the cells create the random game of competition and that the selfish nature of the cell which is its sustainable influence on the population, provides the motivation to participate in the game more. The winning reward is suggested to be a degree of epigenetic plasticity such as the amount of DNMT in cellular cryptocurrency.

This study helps to understand the principle of gene expression in cell populations. The conclusion has not been verified by biological experiments, but it will be clearly confirmed mathematically. It is sure that the biological hidden algorithm inside the DNA sequence will be revealed from the binary bit-logic with physical on and off states which is mathematically proven. The results will contribute to epigenetics, including future disease treatment and medicine development.

Reference

- 1] Hoopes, L., Introduction to the gene expression and regulation topic room. *Nature Education* 1(1):160, 2008
- 2] The GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature* 550, 204–213, 2017
- 3] Darren J. Burgess, Principles of gene regulation across tissues, *Nature Reviews Genetics* 18, 701, 2017
- 4] Sergi Delgado-Segura ,et al, Cryptocurrency Networks: A New P2P Paradigm, Review Article Open Access ID 2159082, Page 16, 2018
- 5] S. Seebacher and R. Schüritz, Blockchain Technology as an Enabler of Service Systems: A Structured Literature Review, in *Exploring Services Science*, vol. 279 of *Lecture Notes in Business Information Processing*, pp. 12–23, Springer International Publishing, Cham, 2017.
- 6] Strogatz SH1, Exploring complex networks, *Nature* 410(6825): 268-76, 2001
- 7] Huang S, Eichler G, Bar-Yam Y, and Ingber DE, Cell fates as high-dimensional attractor states of a complex gene regulatory network, *Phys Rev Lett.* 2005
- 8] Kevin Thurley, Lani F. Wu, and Steven J. Altschuler, Modeling Cell-to-Cell Communication Networks Using Response-Time Distributions, *Cell Systems* 6, 355–367, 2018
- 9] Jonathan M. Raser, and Erin K. O'Shea, Noise in Gene Expression: Origins, Consequences, and Control, *Science*, 309(5743): 2005

- 10] Masa Tsuchiya, Alessandro Giuliani, Midori Hashimoto, Jekaterina Erenpreisa, and Kenichi Yoshikawa, Self-Organizing Global Gene Expression Regulated through Criticality: Mechanism of the Cell-Fate Change, PLoS One. 11(12), 2016;
- 11] Tsuchiya M, et al., Local and global responses in complex gene regulation networks. Physica A 388:1738–1746, 2009
- 12] Tsuchiya M., et al., Collective Dynamics of Specific Gene Ensembles Crucial for Neutrophil Differentiation: The Existence of Genome Vehicles Revealed, PLoS One. 5(8), 2010
- 13] Scott D. Pope, and Ruslan Medzhitov, Emerging Principles of Gene Expression Programs and Their Regulation, Molecular Cell, Volume 71, Issue 3, 389-397 2018
- 14] Gibcus JH, Dekker J Connecting the genome: Dynamics and stochasticity in a new hierarchy for chromosome conformation. Mol Cell 49: 773–782. 2013
- 15] Nakamoto, Satoshi "Bitcoin: A Peer-to-Peer Electronic Cash System", 24 May 2009
- 16] "White Pape" ethereum/wiki Wiki · GitHub", Archived from the original on 28 March 2015
- 17] Sergey Smetanin et al., Blockchain Evaluation Approaches: State-of-the-Art and Future Perspective, Sensors (Basel); 20(12), 2020
- 18] Haitao Yang, Weixian Xi, Nucleobase-Containing Polymers: Structure, Synthesis, and Applications, Polymers (Basel); 9(12),) 2017
- 19] M. Raikwar et al., SoK of Used Cryptography in Blockchain,

IEEE/ACCESS 2946983, 2019

20] T. T. A. Dinh, et al., Untangling blockchain: A data processing view of blockchain systems, *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 7, pp. 1366–1385, 2018

21] Lamport, L., et al., The Byzantine Generals Problem, *ACM Transactions on Programming Languages and Systems*, 4 (3): 382–401, 1982

22] D. Kraft, Difficulty control for blockchain-based consensus systems, *Peer-to-Peer Networking and Applications*, vol. 9, pp. 397–413, 2016

23] Myerson, Roger B. *Game Theory: Analysis of Conflict*, Harvard University Press, p. 1. Chapter-preview links, pp. vii–xi, 1991

24] Eleanor R. Brush, et al., A Family of Algorithms for Computing Consensus about Node State from Network Data, *PLoS Comput Biol.* 9(7): 2013

25] Design and Implementation a new security hash algorithm based on MD5 MD5 AND SHA-256

26] <https://blog.wetrust.io/why-do-i-need-a-public-and-private-key-on-the-blockchain-c2ea74a69e76>

27] Paweł Błazej, et al., Optimization of the standard genetic code according to three codon positions using an evolutionary algorithm, *PLoS One.* 13(8), 2018

28] Anton A. Komar, The Yin and Yang of codon usage, *Hum Mol Genet.*

25(R2): R77–R85, 2016

29] Matthias Stuermer, et al., Digital sustainability: basic conditions for sustainable digital artifacts and their ecosystems, *Sustain Sci.* 12(2): 247–262. 2017

30] Luisanna Cocco, Michele Marchesi, Modeling and Simulation of the Economics of Mining in the Bitcoin Market, *PLoS One.* 11(10), 2016

31] Monique GP van der Wijst, et al., Local chromatin microenvironment determines DNMT activity: from DNA methyltransferase to DNA demethylase or DNA dehydroxymethylase, *Epigenetics.* 10(8): 671–676. 2015

32] Alessandro Fontana, Cell Tracking: Genesis and Epigenesis in an Artificial Organism, *European Conference on Artificial Life, ECAL, Advances in Artificial Life* pp 163-171, 2007

33] Justin C. St. John, Structure and function of the mitochondrial genome, *Principles of Cloning (Second Edition)*, 2014

34] Orgel LE, Crick FH, Selfish DNA: the ultimate parasite, *Nature*, 284 (5757): 604–7, 1980

35] Doolittle WF, Sapienza C, Selfish genes, the phenotype paradigm and genome evolution, *Nature*, 284, 5757, April 1980;