Inevitability of bio-molecules

Ken NAITOH

Waseda University, 3-4-1, Ookubo, Tokyo 169-8555 Japan
Email: k-naito@waseda.jp

Key Words: Origin, Inevitability, Asymmetry, Bases, Amino acids

Abstract

Inevitability of the bio-molecules which are related to nitrogenous-bases, amino-acids, DNA, RNA, and heavier molecules is clarified by using the molecular fluid-dynamics, i.e. the cyto-fluid dynamic theory. We try to solve the problem of why living beings use only five types of nitrogenous bases and twenty types of amino-acids. It should be stressed that the keyword is the mysterious number, the n-th root of n, which lies between 1.0 and the golden ratio. The present report may give the milestone for understanding the origin of life.

1. Introduction

There are some approaches in order to know the principle for designing life system.

Researchers on the origin of life have synthesized artificially amino-acids and nucleic-acids. [1] However, the chain reaction of the fundamental molecules is still in fog.

Chirality [2] is thought to be an important concept for understanding self-organizing process of life. However, this concept can not clarify all of the dynamical processes underlying life.

By using informatics and structure mechanics, Aristid Lindenmayer [3], D'Arcy Wentworth Thompson [4], and Donald E. Ingber [5] have showed the inevitability of large structures such as tree, shell, and cell. However, the inevitability of biological monomers and polymers has not been revealed so well, although the number of fundamental blocks is limited to only five bases and twenty types of amino acids.

Parameter or index is necessary to clarify the inevitability. DNA is composed by four bases of Adenine (A), Guanine(G), Cytosine(C), and Thymine(T). Frequency of usage of Guanine and Cytosine among these four bases (GC rate) is used world-widely as the important indexical property for controlling bio-chemical reaction processes. The hydrogen bond connection between Guanine and Cytosine is stronger than that between Adenine and Thymine. This index is related to the instability of molecules while increasing temperatures. Thus, the GC rate is important for optimizing the replication quality in the Polymerase chain reaction (PCR). [6]

On the other hand, we can classify four bases such as A, T, G, and C into two groups, purine and pyrimidine. Purines, i.e., A and G, have relatively large size, while pyrimidines, i.e., C and T, are small. The present index, the size ratio of purine and pyrimidine (YU rate) is related to the molecular size, i.e., weight. However, the YU rate is not employed for analyzing biological processes so well.

2. Analysis based on YU rate

2.1 Size of bases [7, 8, 9, 10, 11]

Asymmetric size ratios of purines and pyrimidines are around 1.50 in their hydrogen bonds within DNA and RNA, although the size ratio of 1.0, for hydrogen bond connections of identical sizes of bases, is often observed in RNA (Figs. 1, 2, and 3). Size ratios of 1.0 and approximately 1.5 can be described by the mysterious number of the n-th root of n (Fig. 3). This key number proposed here for solving inevitability of the symbiosis is unrelated to chirality.

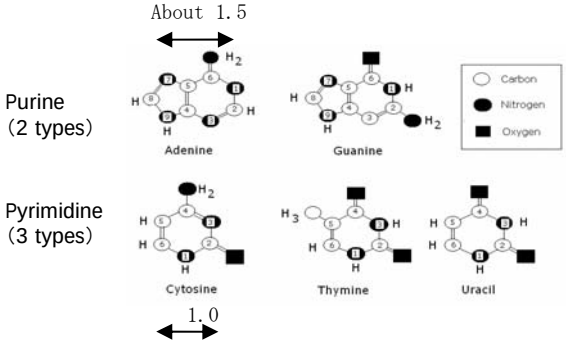

Fig. 1 Five types of nitrogenous bases for DNA and RNA.

Here, we will consider two nitrogenous bases, a purine and a pyrimidine, connected by hydrogen bonds in a large quantity of water. Owing to the influence of the nitrogenous bases, water molecules around the bases have different densities and arrays in comparison with those far away. (It is well-known that weight of hydrated water molecules per a base-pair inside DNA is of the order of that of the bases-pair.) Thus, we divide the water into two regions. A base and water molecules surrounding the base are defined as a continuum particle, parcel. Then, we

assume that the parcel size is proportional to the size of base. Further, we assume that the parcel acts as a flexible continuum particle.

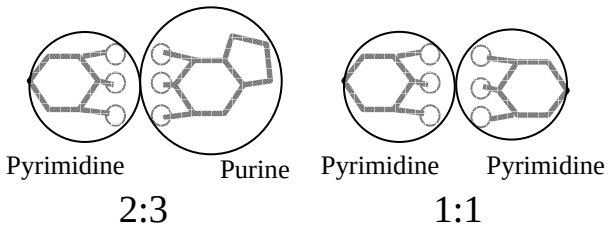

Fig. 2 Hydrogen-bonded bases.

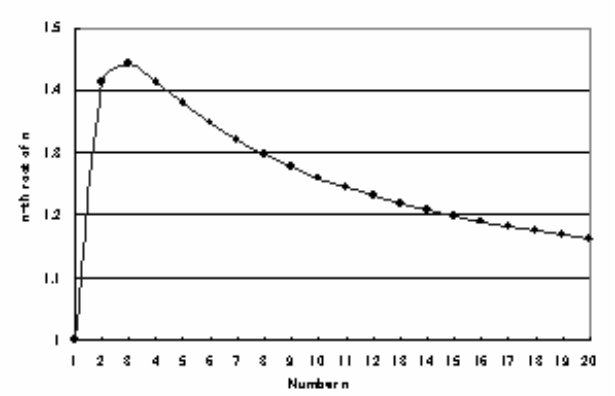

Fig. 3 Mysterious number, the n-th root of n, which appears in the size ratio of purines and pyrimidines.

Based on these assumptions, we derived the deterministic momentum equation describing particle deformation. The momentum equation shows that life prefers the size ratio of hydrogen-bonded nitrogenous bases taking the n-th root of n, which has the values of 1.0, 1.41, and 1.44 for n=1, 2 and 3, respectively (Fig. 3).

2.2 Two purines and three pyrimidines [9, 10, 11]

Asymmetric size ratio of purines and pyrimidines of around 1.50 results in different molecular weights. Molecular weight differences lead to "two" types of purines and "three" types of pyrimidines, because the larger purines produce fewer types. Mass balance leads to a frequency ratio of 1.5, i.e., two purines and three pyrimidines (Fig. 1).

2.3 Frequency ratios of purine and pyrimidine differ between tRNA and rRNA [11, 12, 13]

Let us examine hundreds of transfer RNA (tRNA). Averaged frequency ratio of purines and pyrimidines in tRNAs is about 1.10. The asymmetry can be confirmed, but the value is less than 1.50.

What is the frequency ratio of purines and pyrimidines in ribosomal RNA (rRNA)? The databases on rRNA show a value of around 1.30, which is relatively larger than that for tRNA.

This value less than 1.50 comes from three reasons. One is that the precise molecular weight is about 1.4 for the G-C pair and 1.2 for the A-U pair. Second reason is related to the fact that base pairs are joined to sugars and phosphoric acids by covalent bonds. Total molecular weight, including those of the nitrogenous base, sugar, and phosphoric acids, decides the frequency ratio of purines and pyrimidines in RNA (Fig. 4).

Third reason is that, because the frequency ratio of purines and pyrimidines in the stems of tRNA is 1.0, RNAs having longer stems bring the mean value closer to 1.0. Ribosomal RNA shows frequency ratio of around 1.30, which is larger than that of tRNA, because rRNA has a shorter stem than tRNA. (An extremely long stem-length results in DNA without leaves.)

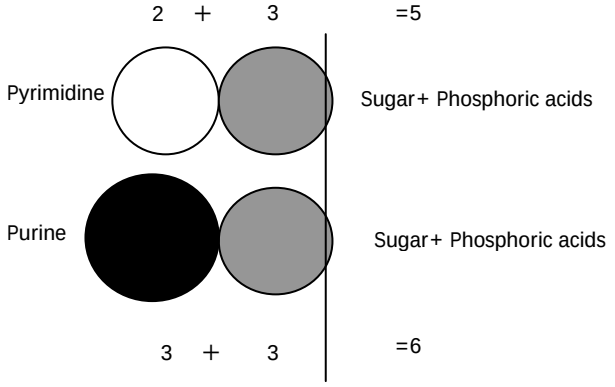

Fig. 4 Schematic diagram of nitrogenous bases connected by covalent bonds illustrating that sugars and phosphoric acids reduce the frequency ratio.

3.Clover structure [11, 12, 13]

The next question is why RNA has a complex clover structure. This stems from the third reason mentioned above. Density asymmetry of purines and pyrimidines results in a leaf structure.

We examined the stochastic computer model simplified based on molecular dynamics. [11, 12, 13] First, we prepared seventy-six individual nitrogenous bases (purines and pyrimidines) in an initial pool. These bases were connected randomly by hydrogen bonds. If we employ a lot of pools, an enormous variety of sequence arrays can be generated. Computational experiment revealed that, when the frequency ratio of purines and pyrimidines in the initial pool was between 1.0 and 1.5, clover structures were produced with a relatively high probability.

The reason why a frequency ratio between 1.0 and 1.5 promotes clover structures is simple. An extremely large frequency ratio, say, far above 1.5, can not produce the stem in tRNA, because the presence of only one type of base cannot form the pairing of purine and pyrimidine.

More complex structures such as ribosomal RNA (rRNA) can also be explained by the above-mentioned dynamical mechanism, which brings fractal structure.

4. Inevitability of non-coding RNA

Recently, the research group in RIKEN [16] has reported

that there are lots of non-coding RNA in introns and junk, although functions of the non-coding RNA are in fog. Here, we will show a hypothesis related to inevitability of this non-coding RNA. Stochastic computer model in Section 3 generates small amount of clover structures. Sequences having clover structures close to tRNA are about 1.0 % in all of the possible sequences, which generated randomly. What are the substantial sequences of unstable structures of 99 %? These sequences of unstable structures may correspond to those generated in ancient pre-biotic processes. Then, these sequences of unstable structures may become the non-coding RNA.

Stochastic model [11, 12, 13] may tell us that we know only 1 % of molecules existing inside living cell.

5. Inevitability of five bases

Density ratio of purine and pyrimidine in tRNA having smaller leaves is around 1.10, while that in rRNA having larger leaves is about 1.30. Why will two different sizes of leaves exist? Mass conservation can not clarify the inevitability of two different sizes of leaves. Energy conservation is necessary for explaining this point.

In many cases, GC pairs are richer in tRNA than in rRNA. Amount of GC pairs is related to the difference of stem length. Actually, tRNA, which has relatively small leaves and long stems, has richer G-C pairs than rRNA. Hydrogen bond connection in G-C pair is stronger than that in A-U. Stronger hydrogen bond will stabilize longer stems.

Two types of base-pairs, G-C and A-U, are necessary to produce two types of RNAs, i.e., tRNA and rRNA.

Inevitability of four bases among five can be understood by the present discussion.

6. Twenty types of amino acids

Triplet inside the leaf of tRNA, anti-codon, permits sixty-four patterns at the maximum, because A, U, G, and C are possible for each locus among the three ones. Then, because the third locus inside anti-codon has relatively weak connection with codon, the minimum number of patterns of triplet is sixteen. This implies that the number of amino acids are between sixteen and sixty-four. As the result, the weak connection at third locus permits twenty types of amino acids, a little more than sixteen.

It should be emphasized that clarification of leaf size shown in the above sections leads us to the outline of inevitability of twenty amino acids.

7. Hydrophobic and hydrophilic amino acids

As is well-known, three nitrogenous bases correspond to one type of amino acid. Array of these three nitrogenous bases is codon. The database shows that the frequency ratio of purines and pyrimidines in codons is also about 1.30. [11, 12, 13] The frequency ratio of purines and pyrimidines for the center locus of anti-codon is also asymmetric. (See Fig. 5.)

Most of hydrophobic and hydrophilic amino acids correspond to whether the center locus in the anti-codon is a purine or a pyrimidine. [11, 12, 13]

Asymmetric frequency ratio of purine and pyrimidine in codons may lead to the frequency asymmetry of hydrophilic and hydrophobic amino-acids. Thus, the frequency ratio of purine and pyrimidine (YU rate) is also important for optimizing proteins and enzymes.

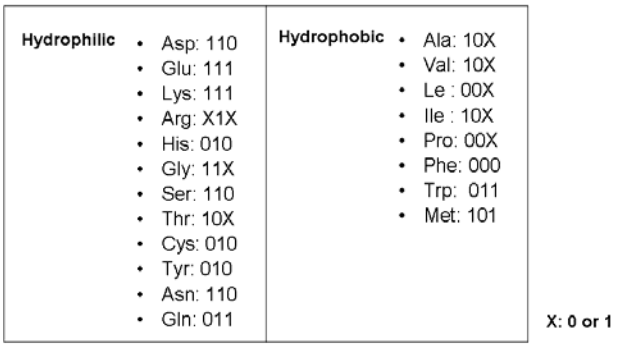

Fig. 5 Twenty types of amino-acids and anti-codon.

8. Structure of proteins

As the center locus inside codon mainly decides whether the amino acid corresponding is hydrophilic or hydrophobic, both of hydrophilic and hydrophobic amino acids exist in proteins. Symbiosis of hydrophilic and hydrophobic molecules brings the complex shapes of proteins, because hydrophobic parts are often folded inside proteins.

9. Causual chain of bio-molecules

Several types of molecules such as nitrogenous bases, amino-acids, RNA, DNA, and proteins are related to each other, through the principle of the n-th root of n and the conservation laws.

Although the inevitability of size ratio of purine and pyrimidine is explained by momentum conservation in the former section, mass conservation law explains that many sizes and frequencies of bio-molecules are inevitable in the causual chain of molecules. Will the explanations shown with mass conservation law be correct? This question is related to whether cell or earth is considered. If the system is earth, mass conservation law will be dominant, because there are less materials coming from the space, although energy comes from outside of the earth, the sun. Here, we think the earth as control volume, not cell. This relates to the food-chain.

Unbalance of purine and pyrimidine in living beings will lead to instability of life without doubt, because the pool of only purine can not generate clover structures such as tRNA and rRNA.

10. Origin

In foregoing sections, we examined the inevitability of several lengths of biological molecules, from base to protein. However, we do not explain how the molecular

length increases during 4600 million years.

Let us see the DNA length of representative species plotted against time. (See Fig. 6.)

The Earth might be born about 4600 million years ago and first life might emerge 3600 million years ago. T. Oshima thought that the first primitive cell had DNA of the order of 1.0 x 10^5 base pairs. [15] (The first cell can be generated with the DNA length of 1.0 x 10^5 base pairs, about 300 types of enzymes. Each enzyme can be coded with 300 base pairs of DNA.) Then, about 3600 million years after the first cell bring human beings of 3.0 x 10^9 bp.

Let us examine further on this. First, we assume that the DNA length increase with the rate of ten times per 1000 million years. As the DNA length was 1.0 x 10^5 base pairs 3600 million year ago, about 10^9 base pairs, the order of human being, can be possible now, because about 4000 million years pass from the first cell. This increasing rate of DNA length implies that DNA should have the order of 10^4 base pairs 4600 million years ago. (See Fig. 6.)

The problem is whether the synthesis of DNA with 10^4 base pairs could be possible or not, 4600 million years ago. The possibility is shown by Yanagawa et al. [1] that polymer macromolecules such as protein or cell membrane can be produced by synthesizing amino acids under high pressure and high density of metal during a few weeks. The other data which supports 10^4 bp of DNA 4600 million years ago is that too many cycles of PCR during hours produces longer DNA sequences. [23]

Discussion shown above does not give the clear answer to the question of whether life was born on the earth or not. DNA length plotted against time should be analyzed further.

11. Conclusion

Inevitability of the bio-molecules such as nitrogenous-bases, amino-acids, DNA, RNA, and heavier molecules is clarified, although there is also eventuality around the inevitable molecules.

References

[1] Yanagawa, H. (1991), Seimeiwa Ikani Tsukuraretaka, TBS Britaniaca.

[2] Kuroda, R. (1992) Seimei Sekai no Hitaishousei, Chuukou Shinsho.

[3] Prusinkiewics P. And Lindenmayer A. (1990) The algorithmic beauty of plants, Springer-Verlag.

[4] Thomson, D. (1961) On Growth and Form, Cambridge University Press.

[5] Ingber, D.E. (1998) The Architecture of Life. 52. Scientific American.

[6] Mullis K.B. et al (1994) PCR, BIRKHAUSER.

[7] Naitoh, K. (1999). Cyto-fluid Dynamic Theory for Atomization Processes, *Oil&Gas Science and Technology*, Vol. 54, No.2 205-210.

[8] Naitoh, K. (2000). Mesoscopic Kinetic Theory for Fluid Phenomena and Biochemical Reaction Processes. *J. of Advanced Science*, 12-3, 223 (2000).

[9] Naitoh, K. (2001). Cyto-fluid Dynamic Theory, *Japan Journal of Industrial and Applied Mathematics*, Vol.18, No.1.

[10] Naitoh, K. (2000). Cyto-fluid Dynamic Theory of the Origin of Base, Information, and Function. *Proc. of 6th Int. symp. on Artificial life and Robotics 6th*.

[11] Naitoh, K. (2002). Cyto-fluid Dynamic Theory of the Origin of Base, Information, and Function. *Artificial Life and Robotics*, **6-1 & 2,** 82.

[12] Naitoh, K. (2005) Self-organising mechanism of biosystems, *Artificial life and robotics*, 9, 96-98.

[13] Naitoh, K. (2005) Seimei no Kihon Bunshi wo Tsuranuku Patan (Patterns underlying bio-molecules), *NIKKEI SCIENCE*, 58-65.

[14] Naitoh, K. (2006) Gene Engine and machine Engine, Springer Japan.

[15] Oshima, T. (1995) Seimeiwa Nessuikara Hajimatta, Tokyo Kagaku Doujin.

[16] RIKEN Genome Exploration Research Group and Genome Science Group. (2005) Antisense Transcription in the Mammalian Transcriptome, *SCIENCE*, 309.

[17] Lowe, T.M., Eddy, S.R., (1997). *Nucleic Acids Res.*, **25**, 955. (available at http://rna.wustl.edu/tRNAdb/.)

[18] Benson, D.A., Boguski,M.S., Lipman, D.J., and Ostell, J., Ouellette, B. F. F., (1998). *Nucl. Acids Res.*, **26**, 1.

[19] Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Wheeler, D. L., (2003). *Nucleic Acids Res.*, **31**, 23.

[20] Nakamura, Y., Gojobori, T., Ikemura, T., (2000). *Nucl. Acids Res.*, **28,** 292. (available at http://www.kazusa.or.jp/codon/.)

[21] DNA Data Bank of Japan, http://www.ddbj.nig.ac.jp/.

[22] JCM On-line catalogue, Japan Collection of Microorganisms, RIKEN (The Institute of Physical and Chemical Research), http://www.jcm.riken.go.jp/

[23] Naitoh, K. (2005) DNA computing based on actual biological sequences and accurate reaction control, *Proc. of Sixth Int. Nobeyama Workshop on the new century of computational fluid dynamics*, 2003, 60-66.

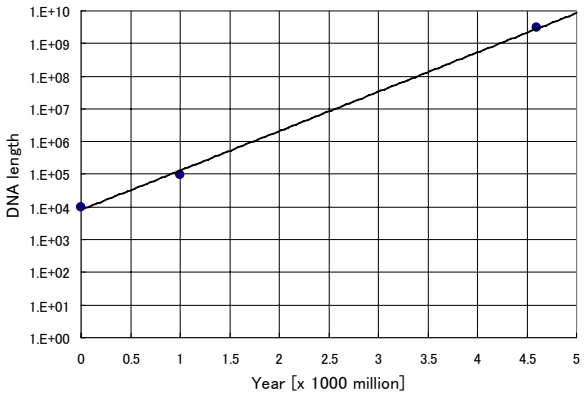

Fig. 6 DNA length plotted against year.