DNA Molar Masses - Unsung Heroes of the Code of Life?

Sai Venkatesh Balasubramanian
Sree Sai Vidhya Mandhir, Mallasandra, Bengaluru-560109, Karnataka, India
saivenkateshbalasubramanian@gmail.com

Abstract:
While the revolutionary field of Genetics along with its immense applications needs no introduction, recent models of the Deoxyribonucleic Acid (DNA) reading process have proposed a complex system of mechanical oscillators giving rise to a solitonic transmission wave, dependent largely on the 'context', which consists of the structures and masses of the DNA Nucleotides. This observation logically raises the question of whether the masses of nucleotides in particular, contribute to the well-known high informational efficiency and capacity of the DNA. In this context, the present work purports to a simple experiment, where the genome sequences of three virus species are used to derive a sequence of the corresponding nucleotide molar masses. Along with the original nucleotide-molar mass associations, three fictitious associations are proposed, and for each of the four associations, the histograms of molar mass occurrences are plotted. It is seen that the original association alone leads to a distribution resembling the normal probability distribution, leading to the hypothesis that this preference affects choice of molar mass by nature, which then results in the efficiency of the solitonic transmissions and thus ultimately, the observed informational efficiency of the genetic codes.

Keywords: Genetic Code, DNA, Molar Masses, Normal Distribution, Genome Sequences

1. Introduction
The discovery of the double-helical structure of the Deoxyribonucleic Acid (DNA) Molecule by Watson and Crick has ushered in the era of Genetics, which is indeed a revolution in the truest sense of the term [1-2]. Apart from providing a whole new perspective of looking at inheritance, evolution and survival, genetics has spawned innumerous applications, of which forensics, genealogical migration and ancestry tracing and personalized medicine are just a few [2-6]. All these applications, given the fact that only around 5% of the human DNA is responsible for the coding of various proteins, and that the details and functionalities of the remaining 95%, called "Junk DNA", remains largely unknown, is a testimony of the informational capacity and power of the DNA based Genetic Code, which can aptly be called "The Code of Life" [7-9].

Perhaps, the best known way of visualizing the Information Processing Capacities of the DNA is the Genotype/Phenotype model, where the biochemical molecules involved in the DNA, RNA and other associated enzymes collectively form the "Hardware", and the Genetic Codes, along with the four codon alphabet (A, C, G and T) with appropriate start and stop codes form the "Software" of the Living Computing System [10-12]. Thus, part of the informational efficiency of the genetic code lies in the chemical molecules involved as well as the very structure of the DNA [12]. A noteworthy and relevant result from Gariaev et al. proposed the DNA as a basis for a Quantum Bio-Computer where, among other observations, the DNA reading process is modeled as a complex mechanical oscillator producing solitonic

wave transmissions, giving rise to a system of rotary pendulums, and depending on the "context" determined by the spatial structures and masses of the DNA Nucleotides [13-15].

It is then but logical to question whether the molar masses of the four nucleotides – Adenine, Cytosine, Guanine and Thymine, and the way they are arranged have anything to do with the enormous informational efficiency and power observed in the Genetic Code. This short article attempts to address this doubt by a very simple exercise. Specifically, the DNA of three organisms, namely the *Eschericia Coli* Virus (E.Coli), the Hepatitis-B Virus (HBV) and the *Streptomyces Acidiscabies* yeast (SA), are considered, and in each case, the four nucleotides A, C, G and T are replaced with their corresponding molar mass values. Following this, three fictitious alternative schemes are proposed for the molar masses of the nucleotides. For these three cases, along with the original molar mass association, the histograms are plotted. It is seen that for all the three cases, the histogram corresponding to the true molar-mass association resembles normal distributions, which is largely absent in the three fictitious associations.

Thus, it is hypothesized from this simple study that the choice of DNA Nucleotides by nature, according to their molar masses is an optimal one, driven by the preference towards normal distributions, and this might be one, if not the only factor accounting for the informational efficiency of the Genetic Code. It is also hoped that these preliminary results are substantiated by tailored experimental observations of the DNA Reading Mechanisms and computation of similar histograms of DNAs of other organisms. If such studies indeed prove the proposed hypothesis, then similar studies may be performed to gain deeper insight into the nature and natural preference, if any, for various Genetic Mutations.

2. Results and Discussion

The DNA Molecule is essentially a macro-molecule consisting of two biopolymer strands, known as "polynucleotides" coiled around each other to form a double helix structure [1-2]. The polynucleotides are made up of simpler units, called nucleotides, which are composed of a nitrogenous base, five-carbon sugars called deoxyribose and atleast one phosphate group [16]. In DNA, two of the four nucleotides, namely Adenine (A) and Guanine (G) are Purines, which are heterocyclic aromatic organic compounds consisting of pyrimidine rings fused to imidazole rings [16]. The other two nucleotides, Cytosine (C) and Thymine (T) are Pyrimidines, which are one of three diazines (aromatic six-membered heterocyclic compounds with two nitrogen atoms in the ring) [16]. In Ribonucleic Acid (RNA), Thymine is replaced with Uracil. It is known that in the double helical DNA structure, Adenine always pairs with Thymine, whereas Guanine always pairs with Cytosine [16]. The structures of the four DNA nucleotides are shown in Fig. 1.

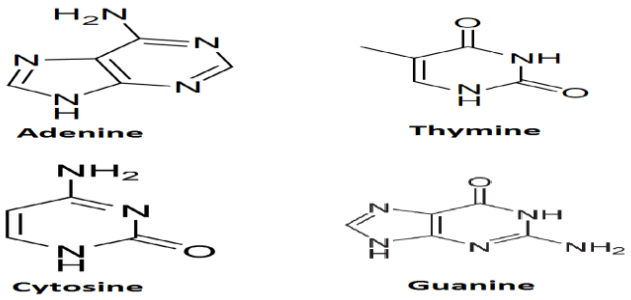

Figure 1 Block Diagram illustrating the proposed Genome Data Compression Technique

The Molar Masses of A, C, G and T nucleotides are given as 135.13 g/mol, 111.10 g/mol, 151.13 g/mol and 126.12 g/mol respectively [16]. The primary motive behind this article is to investigate if this choice of molar masses by nature has a causal connection to the informational efficiency of the genetic code, and in accordance with this, three other fictitious nucleotide-molar mass association schemes are proposed, as in Table 1. As seen, Scheme A is the original association mentioned above, whereas Scheme B is a fictitious association of molar mass by alphabetic order, Scheme C is a fictitious association with a different order, and Scheme D is a fictitious association of molar masses by nucleotide type (purine/pyrimidine).

Table 1 Original and Fictitious Nucleotide-Molar Mass Association Schemes

Scheme	Scheme A	Scheme B	Scheme C	Scheme D
Nature	Original	Fictitious	Fictitious	Fictitious
Adenine	135.13	100	100	100
Cytosine	111.10	110	101	120
Guanine	151.13	120	103	120
Thymine	126.12	130	102	100

For each of the four Schemes, the histograms showing the occurrences of molar masses are plotted, for three genome sequences, namely Eschericia Coli, Bepatitis-B Virus and Streptomyces Acidiscabies, in Fig. 2-4. The Genome Sequences are obtained from the genome database of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov).

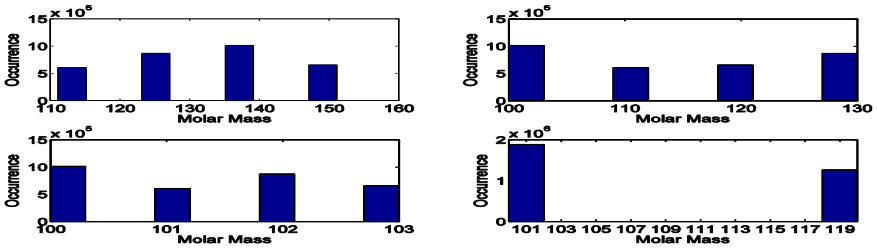

Figure 2 Histograms of E.Coli Molar Masses according to Schemes A (top left), B (top right), C (bottom left) and D (bottom right)

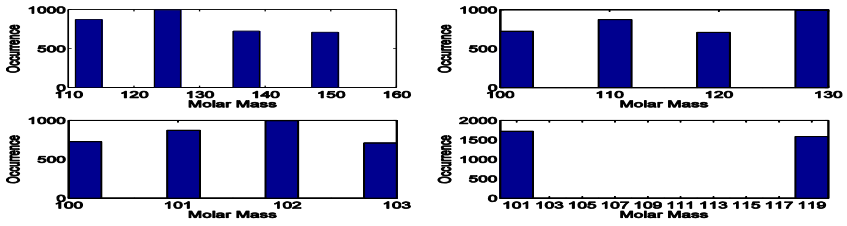

Figure 3 Histograms of Hepatitis B Molar Masses according to Schemes A (top left), B (top right), C (bottom left) and D (bottom right)

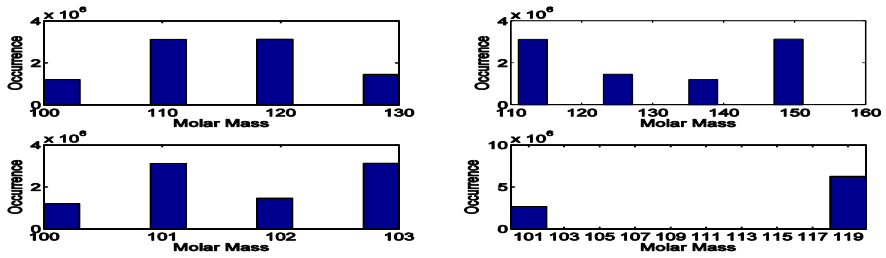

Figure 4 Histograms of Streptomyces Acidiscabies Molar Masses according to Schemes A (top left), B (top right), C (bottom left) and D (bottom right)

From the plots, it is seen that for all three genome sequences, the Scheme A (original association) follows a Distribution resembling Normal Distribution, whereas the other three distributions neither resemble normal distributions, or are consistent across species.

This leads one to hypothesize that preference of Nucleotide Molar Masses towards normal distributions might lead to efficient solitonic transmissions in the DNA reading process, finally resulting in the informational efficiency of the DNA.

3. Conclusion

The observation of complex mechanical oscillators producing solitonic wave transmissions in the DNA reading process naturally leads one to question whether the natural choice of DNA Nucleotide structures and masses, forming the context of such transmissions, are related to the informational efficiency and capacity of the resulting Genetic Codons. It is in this light that the present article pertains to a simple experiment, where the genome sequences of three different species are considered, and in each case, four schemes of nucleotide-molar mass associations are proposed – one original and three fictitious. For each of the four schemes, the histograms of molar mass distributions across the entire genome sequence are plotted. It is seen that in all the three genome sequences, the original scheme alone maintained a normal-like distribution, leading one to hypothesize that this preference affects choice of molar mass by nature, which then results in the efficiency of the solitonic transmissions and thus ultimately, the informational efficiency of the genetic codes.

In conclusion, this hypothesis must at best be taken as a preliminary observation, an impetus to perform more rigorous studies of DNA molar mass distributions in genome sequences corresponding to higher species, thus proving or disproving the proposed hypothesis. However, if the hypothesis is indeed proved to be true, the next step would be to investigate if preferences of molar masses also somehow influence the occurrence, nature and preferences towards genetic mutations.

References

[1] Watson, James D., and Francis HC Crick. "Molecular structure of nucleic acids." Nature 171, no. 4356 (1953): 737-738.

[2] Klug, William S., and Michael R. Cummings. Essentials of genetics. No. Ed. 3. Prentice-Hall Inc., 1999.

[3] Smith, John Maynard. Evolutionary genetics. Oxford University Press, 1989.

[4] Butler, John M. Forensic DNA typing: biology, technology, and genetics of STR markers. Academic Press, 2005.

[5] Otterstrom, Samuel M., and Brian E. Bunker. "Genealogy, migration, and the intertwined geographies of personal pasts." Annals of the Association of American Geographers 103, no. 3 (2013): 544-569.

[6] Ginsburg, Geoffrey S., and Jeanette J. McCarthy. "Personalized medicine: revolutionizing drug discovery and patient care." TRENDS in Biotechnology 19, no. 12 (2001): 491-496.

[7] Crick, Francis, Leslie Barnett, Sydney Brenner, and Richard J. Watts-Tobin. General nature of the genetic code for proteins. Macmillan Journals Limited, 1961.

[8] Ohno, Susumu. "So much “junk” DNA in our genome." In Brookhaven Symp Biol, vol. 23, pp. 366-370. 1972.

[9] Biémont, Christian, and Cristina Vieira. "Genetics: junk DNA as an evolutionary force." Nature 443, no. 7111 (2006): 521-524.

[10] Chanock, Stephen J., Teri Manolio, Michael Boehnke, Eric Boerwinkle, David J. Hunter, Gilles Thomas, Joel N. Hirschhorn et al. "Replicating genotype–phenotype associations." Nature 447, no. 7145 (2007): 655-660.

[11] Lodding, Kenneth N. "The hitchhiker's guide to biomorphic software." Queue 2, no. 4 (2004): 66.

[12] Abbott, Derek, Paul CW Davies, and Arun K. Pati. Quantum aspects of life. Imperial College Press, 2008.

[13] Gariaev, Peter P., Peter J. Marcer, Katherine A. Leonova-Gariaeva, Uwe Kaempf, and Valeriy D. Artjukh. "DNA as basis for quantum biocomputer." DNA Decipher Journal 2011c 1 (2011): 025-046.

[14] Gariaev, Peter P., and E. A. Leonova. "The strange world of wave genetics." DNA Decipher Journal 4, no. 1 (2014): 39-56.

[15] Gariaev, Petr Petrovich. "Another Understanding of the Model of Genetic Code Theoretical Analysis." Open Journal of Genetics 5, no. 02 (2015): 92.

[16] Berk, Arnold, and S. Lawrence Zipursky. Molecular cell biology. Vol. 4. New York: WH Freeman, 2000.