Codon Usage in Hemoglobin Coding Genes

Li Gun[1], Tian Han[2], Ren Yumiao[3], Du Ning[4]

[1,2,3,4]Department of Biomedical Engineering, School of Electronics and Information Engineering,
Xi'an Technological University, Xi'an, Shaanxi Province, China

ARTICLE INFO	ABSTRACT
corresponding Author: **Li Gun**[1] Dept. of Biomedical Eng., School of Electronics and Information Eng., Xi'an Technological Univ., Xi'an, Shaanxi Province, China	Hemoglobin is a protein responsible for the transport of oxygen. The preference of hemoglobin coding gene is analyzed by CodonW. The relationship between the A3 / (A3 + T3) and G3 / (G3 + C3) and the relationship between GC_{12} and GC_3 are all investigated. Finally, the preference of hemoglobin coding gene is shown and discussed. The results show that codon use preference of CUG, CGU, UAA, ACC, GUG and GGC are greater than 2.5. This information facilitates an improved understanding of the structural, functional of hemoglobin

Introduction

Proteins are machines that perform a variety of important function in the body. Some proteins are composed of multiple peptide chains, each of which is called a subunit, and the specific spatial relationship between subunits is called the quaternary structure of the protein. Hemoglobin is a protein that makes the blood red and consists of four strands, two alpha chains and two beta chains. As the main carrier of genetic information, the number and order of the four bases of A, T, C and G in the DNA molecule contain very important information. Analyzing the DNA sequence characteristics is benefit to understanding the biological function of the sequence and to finding the characteristic information related to the specific function in the sequence. When the codon usage preference is concerned, many scientists studied vary gens recently, such as virus [1], Y-linked genes [2], CYB gene [3] and SRY gene [4], et al. In the virus gene area, most influential virus gene are all studied by scientist timely, such as Zika [1], classical swine fever virus [5], Japanese encephalitis virus [6], enterovirus D68 strains [7], and so on [8-10]. In fact, codon usage bias has would reveal many mechanism in biology [11-13] and affect by many factors [14-17]. In this paper, the codon usage in hemoglobin coding genes is studied via the CodonW, and the PR2-bias plot analysis: A3 / (A3 + T3) vs. G3 / (G3 + C3) and the neutral evolution analysis: GC_{12} vs. GC_3 are plotted and discussed.

Materials and Methods

This paper considers the most common type of hemoglobin, their coding sequences are download form the website http://www.ncbi.nlm.nih.gov, and

the accession numbers are: NP_000549.1 and NP_000509.1.

CodonW [18] is used to calculate the CAI, CBI, ENC and so on, the calculated results are further processed for getting the relative synonymous codon usage (RSCU), the PR2-bias plot analysis: A3 / (A3 + T3) vs. G3 / (G3 + C3), the neutral evolution analysis: GC_{12} vs. GC_3, et al.

Results and discussion

Codon system is used to predict the codon usage characteristics of hemoglobin, and the overall results are shown in Table 1. in the table 1 Codon Adaptation Index (CAI) ranges from 0 to 1, the greater the value, the stronger the sub-preference of the codon and the higher the gene expression level, the CAI value of hemoglobin are 0.4 in HBα1 and 0.36 in HBβ respectively. The effective codon number (ENC) reflects the degree of preference for the use of the codon, with the range of 20-61, the smaller the value, indicating that the stronger codon usage preference, the ENC value of hemoglobin are 29.84 in HBα1 and 34.01 in HBβ respectively. Frequency of Optimal Codons (FOP) refers to the percentage of the optimal codon used, the FOP value of hemoglobin are 0.64 in HBα1 and 0.49 in HBβ respectively. Codon Bias Index (CBI) is the component that reflects the expression of a superior codon in a particular gene, and this index has a good correlation with the ENC value generally.

Table 1. Nucleotide composition in the coding sequences of hemoglobin.

Title	Accession	T3s	C3s	A3s	G3s	CAI	CBI	Fop	Nc	GC3s	GC
HBα1	NP_000549.1	0.11	0.61	0.02	0.47	0.4	0.38	0.64	29.84	0.89	0.65
HBβ	NP_000509.1	0.35	0.36	0.07	0.45	0.36	0.13	0.49	34.01	0.65	0.56

The parity rule states that if there is no mutation in the two complementary strands of DNA or the bias on the selection effect, the base content should have A = T and G = C. This method was used to analyze the PR2 bias occurring at the third place of the codon. The center of the graph follows the PR2 principle, A = T and G = C, that is the vertical and horizontal coordinates are 0.5. The vector from this center represents the degree and direction of the PR2 bias. The parity preference is mainly used to analyze the relationship between pyrimidine and purine at 3rd base of a codon. The content of pyrimidine (C + T) at the third position of the hemoglobin codon gene is greater than the purine content (A + G) (Figure 1).
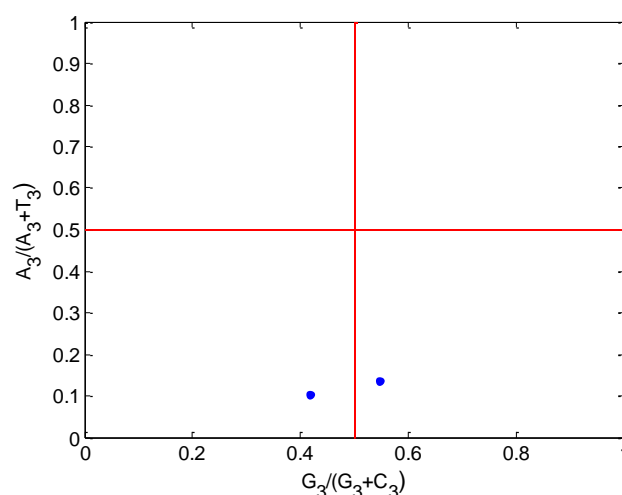

Figure 1. PR2-bias plot analysis

In neutrality plotting, GC_{12} is as the ordinate, and GC_3 is as the abscissa, each point in the figure represents the independent gene. Hemoglobin codon gene neutrality plotting result is shown in the

Figure 2, the result shows that the distribution of GC_{12} is 56.73% and 55.83% in HBα1 and HBβ respectively, GC_3 is 81.41% and 59.51% in HBα1 and HBβ respectively, GC 12 distribution was significantly smaller. These values are distributed diagonally in the vicinity, and for these regions, base mutations may be a major factor influencing codon preference (Figure 2).
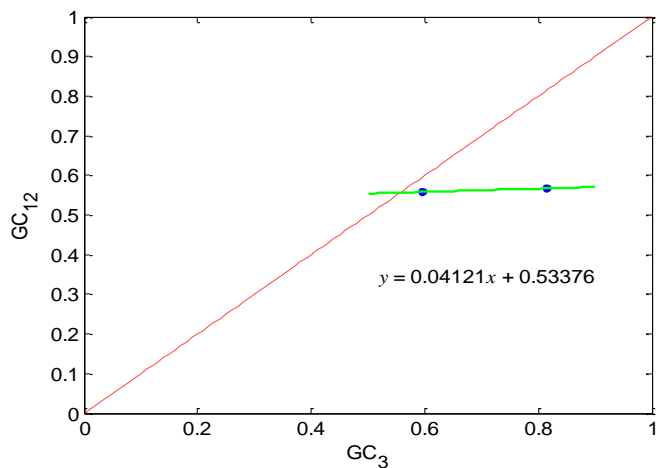

Figure 2. Neutrality plot of GC_{12} vs. GC_3

The relative synonymous codon usage is an estimate of the preference for the usage of synonymous codons, defined as the ratio of the observed value of the number of synonymous codons used to the expected number of occurrences of the codon. If the codon usage is not preferred, the RSCU value is 1; if a codon is used frequently, the RSCU value is greater than 1; if a codon usage frequency is low, the RSCU value is less than 1. The RSCU value intuitively reflects the preference for codon usage in hemoglobin (Figure 3). The results show that codon usage preference of CUG, CGU, UAA, ACC, GUG and GGC are greater than 2.5. Because it is simple to calculate and intuitively reflects the preference of codon usage, it is usually used as a measurement of the codon usage preference analysis.
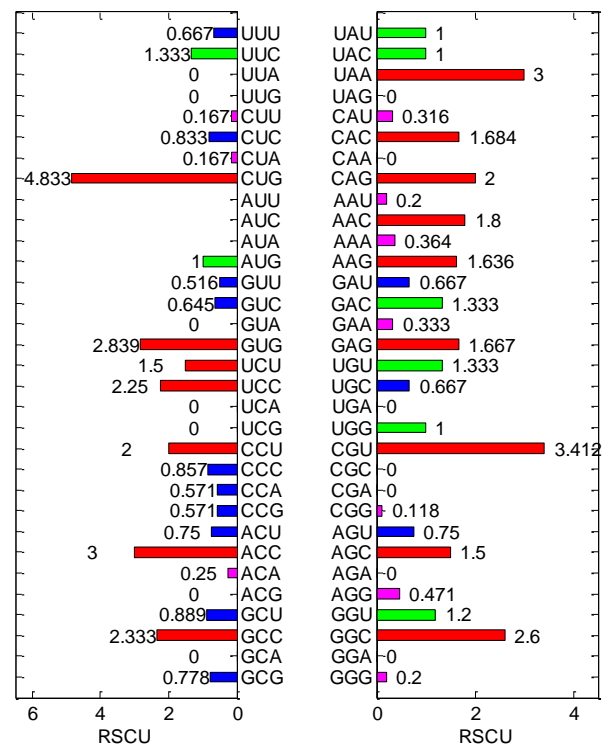

Figure 3. RSCU of the hemoglobin coding gene

Codons as a link between nucleic acids and proteins, it can be used to characterize the evolutionary relationships within the genome or between genomes [19, 20]. For example, the same genes of different kinds of creatures or the different genes of the same creature have different characteristics of codon usage [21]. The closer the genetic relationship is, the closer the codon preference is, the closer the genetic relationship is, and the less difference in codon preference [22]. One kind of a gene of different species or the same species may have different characteristics in codon usage [23]. However, the frequencies of the different codons that encode the amino acids for the same species may different [24]. Different codon usage could results differences frequency in amino acids encode even in the same amino acid, resulting in quite different differences between different species genes codon usage.

Conclusion

Codons are the basic units of life information transmission. In a certain species or genotype the phenomenon of unequal use of synonymous codons is common, where the synonymous codons that are preferentially used are called optimal codons. There are significant differences in codon usage preferences in different organisms, meanwhile, codon usage preference not only plays an important regulatory role map at the level of gene expression, but also helps to improve the accuracy and efficiency of translation. In this paper, the preference of hemoglobin coding gene is analyzed, such as the relationship between the $A3 / (A3 + T3)$ and $G3 / (G3 + C3)$, the relationship between GC_{12} and GC_3 are all studied and the preference of hemoglobin coding gene is also discussed. The results show that codon usage preference of CUG is 4.833, which is the largest, codon usage preference CGU is 3.412, codon usage preference UAA is 3, and so on. All these information are important for studying the structural, functional of hemoglobin.

References

1. Snawar Hussain, Sahibzada Tasleem Rasool. Analysis of synonymous codon usage in Zika virus. Acta Tropica, 2017, 173: 136-146

2. Monisha Nath Choudhury, Arif Uddin, Supriyo Chakraborty. Codon usage bias and its influencing factors for Y-linked genes in human. Computational Biology and Chemistry, 2017, 69: 77-86

3. Arif Uddin, Supriyo Chakraborty. Codon usage trend in mitochondrial CYB gene. Gene, 2016, 586(1): 105-114

4. Supriyo Chakraborty, Arif Uddin, Monisha Nath Choudhury. Factors affecting the codon usage bias of SRY gene across mammals Gene, 2017, 630: 13-20

5. Xin Xu, Dongliang Fei, Huansheng Han, Honggui Liu, Hua Zhang. Comparative characterization analysis of synonymous codon usage bias in classical swine fever virus. Microbial Pathogenesis, 2017, 107: 368-371

6. Niraj K. Singh, Anuj Tyagi, Rajinder Kaur, Ramneek Verma, Praveen K. Gupta. Characterization of codon usage pattern and influencing factors in Japanese encephalitis virus. Virus Research, 2016, 221: 58-65

7. Uladzimir U. Karniychuk. Analysis of the synonymous codon usage bias in recently emerged enterovirus D68 strains. Virus Research, 2016, 223: 73-79

8. Yu Xu, Renyong Jia, Zhilong Zhang, Yanyan Lu, Anchun Cheng. Analysis of synonymous codon usage pattern in duck circovirus. Gene, 2015, 557(2): 138-145

9. Sudhir Morla, Aditi Makhija, Sachin Kumar. Synonymous codon usage pattern in glycoprotein gene of rabies virus. Gene, 2016, 584(1): 1-6

10. Gonzalo Moratorio, Andrés Iriarte, Pilar Moreno, Héctor Musto, Juan Cristina. A detailed comparative analysis on the overall codon usage patterns in West Nile virus. Infection, Genetics and Evolution, 2013, 14: 396-400

11. Arup Kumar Malakar, Binata Halder, Prosenjit Paul, Supriyo Chakraborty. Cytochrome P450 genes in coronary artery diseases: Codon usage analysis reveals genomic GC adaptation. Gene, 2016, 590(1): 35-43

12. Changyuan Hu, Jing Chen, Lulu Ye, Renpin

Chen, Xiangyang Xue. Codon usage bias in human cytomegalovirus and its biological implication. Gene, 2014, 545(1): 5-14

13. Abhishek Subramanian, Ram Rup Sarkar. Comparison of codon usage bias across Leishmania and Trypanosomatids to understand mRNA secondary structure, relative protein abundance and pathway functions. Genomics, 2015, 106(4): 232-241

14. Max L. Nibert. Mitovirus UGA(Trp) codon usage parallels that of host mitochondria. Virology, 2017, 507: 96-100

15. Devi Lal, Mansi Verma, Susanta K. Behura, Rup Lal. Codon usage bias in phylum Actinobacteria: relevance to environmental adaptation and host pathogenicity. Research in Microbiology, 2016, 167(8): 669-677

16. Mei He, Chun-Bo Teng. Divergence and codon usage bias of Betanodavirus, a neurotropic pathogen in fish. Molecular Phylogenetics and Evolution, 2015, 83: 137-142

17. Sheng-Lin Shi, Yi-Ren Jiang, Rui-Sheng Yang, Yong Wang, Li Qin. Codon usage in Alphabaculovirus and Betabaculovirus hosted by the same insect species is weak, selection dominated and exhibits no more similar patterns than expected. Infection, Genetics and Evolution, 2016, 44: 412-417

18. Peden, J.F., 1999. CodonW. (PhD Thesis). University of Nottingham, Nottinghamshire.

19. Yang Liang, Mei He, Chun-Bo Teng. Evolution of the vesicular stomatitis viruses: Divergence and codon usage bias. Virus Research, 2014, 192: 46-51

20. Arghya Kamal Bishal, Rashmi Mukherjee, Chandan Chakraborty. Synonymous codon usage pattern analysis of Hepatitis D virus. Virus Research, 2013, 173(2): 350-353

21. Laura Kubatko, Premal Shah, Radu Herbei, Michael A. Gilchrist. A codon model of nucleotide substitution with selection on synonymous codon usage. Molecular Phylogenetics and Evolution, 2016, 94: 290-297

22. Jasmine Saini, Uri Hershberg. B cell Variable genes have evolved their codon usage to focus the targeted patterns of somatic mutation on the complementarity determining regions. Molecular Immunology, 2015, 65(1): 157-167

23. Tonghai Yu, Jinsong Li, Yang Yang, Liu Qi. Codon usage patterns and adaptive evolution of marine unicellular cyanobacteria Synechococcus and Prochlorococcus. Molecular Phylogenetics and Evolution, 2012, 62(1): 206-213

24. Ratna Prabha, Dhananjaya P. Singh, Swati Sinha, Khurshid Ahmad, Anil Rai. Genome-wide comparative analysis of codon usage bias and codon context patterns among cyanobacterial genomes. Marine Genomics, 2017, 32: 31-39