

Full Length Research Paper

p53 oncogene mutations in head and neck cancer based on the weighted Euclidean distance

Kong Juanjuan^{1,2} and Zhu Ping^{1,2*}

¹School of Science, Jiangnan University, Wuxi 214122, People's Republic of China.

²Key Laboratory of Advanced Process Control for Light Industry, Jiangnan University, Ministry of Education, Wuxi 214122, China.

Accepted 13 December, 2011

In order to study the p53 mutations in head and neck cancer, we explored the relationship between the different positions of the bases and the amino acids' physical and chemical properties. In this paper, the Euclidean distance (d) was defined. Furthermore, by using improved variation coefficient method, the Euclidean named as weighted Euclidean distance (d^*) was weighted, and then the p53 oncogene mutations in head and neck cancer was analyzed by using d and d^* . The method based on the d^* can judge which position of the bases had changed more obviously. When d^* was more than 1.5, the second base, as well as the physical and chemical properties of amino acids, changed. In contrast, when d^* was less than 1.2, the first or the third base changed, but the physical and chemical properties of amino acids were not changed. The results therefore show that the second base has the greatest influence on amino acids, the first has smaller influence, and the third has less impact on the physical and chemical properties of amino acids.

Key words: Weighted Euclidean distance, head and neck cancer, p53, physical and chemical properties, amino acids.

INTRODUCTION

The p53 is a gene that codes for a protein that regulates the cell cycle and hence functions as a tumor suppressor. It is very important in maintaining genetic stability and regulating the cell cycle. However, when the wild p53 mutates, it can change the normal gene, whether human or animal, into the cancer gene. Thus it leads the human and animal to suffer from cancer. A lot of results (An et al., 2002; Joseph et al., 1999; Bodey et al., 1997) show that many human cancers are related to the p53 gene mutations, for example, lung cancer, throat cancer, breast cancer and so on.

The main aim of this paper was to explore the p53 mutations in head and neck cancer impacting on the physical and chemical properties of amino acids. A foundation research and application of the Codon Bias of Human p53 tumor protein had been analyzed (Kong and Ping, 2011). In addition, a lot of results about DNA sequences have been obtained by using algebra

(Sanchez et al., 2005; Sanchez et al., 2006; Sanchez and Grau, 2009; Yan and Ping, 2010; Zhu et al., 2011). First, in order to establish the algebraic model, Sanchez et al. (2005) proposed a new base 'O' which expressed insertion or deletion basic group. They not only analyzed the genetic code rule, but also conducted the gene mutation in a deep research. Secondly, Sanchez et al. (2006) proposed a novel Lie algebra of the genetic code over the Galois field of four DNA bases. They also suggested the genetic code Boolean lattice. According to Lie algebra of the genetic code over the Galois field of four DNA bases, they obtained the physicochemical property and biological information contained in DNA sequence.

Furthermore, in 2009, a new base 'D', which symbolizes one (or more) alternative hypothetical base(s) or dummy variables with non-specific pairings presented in primeval RNA and DNA molecules, was introduced (Sanchez and Grau, 2009) for establishing the evolution model. Then they described the mutation pathways in the molecular evolution process. In 2010, the foundations of the results researched by Sanchez et al. (2005, 2006)

*Corresponding author. E-mail: zhuping@jiangnan.edu.cn. Tel: 13861449741.

and Ping and Yan (2010) established a new algebraic structure. Bases 'O', 'P' and 'D', were introduced when the normal gene became the cancer one, where 'O': {G, C} → {A, C}, 'P': {A, T} → {G, C}, 'D': others. Then, the extended triplet set GF(7³) was established. Thus, the p53 mutation gene was studied by using the algebraic operation. Finally, Ping and Yan (2010) established the relationship φ: GF(7³) → C₃₄₃ and a bijection f: {D, A, C, O, G, T, P} → GF(7), where 'O': {G, C} → {A, C}, 'P': {A, T} → {G, C}, 'D': others. Meanwhile, the distance of codons was defined and the changes of the amino acids' physical and chemical properties when the normal gene became cancerous were pointed out.

It is clear that the codon has the close relation with the amino acids' physical and chemical properties. Many findings (Xue et al., 2001; Ma et al., 2003; Sun et al., 2003) had indicated the relationship of the codon evolution and the amino acid reciprocity, the inner link of the codon and the amino acids' physical and chemical properties, and the remarkable related or inverse correlation between the codon and amino acid. The amino acids having U at the second position of their codons are hydrophobic (Woese, 1965); the amino acids with A at the second position are hydrophilic (Crick, 1968); the amino acids with C at the second position have their polarities in middle; while the amino acids with G at the second position follow any regularity in their polarities (Grantham, 1974). Thus it can be seen that the codon mutation decides the amino acids' physical and chemical properties. The hydrophobicity of amino acids affects the structure and function of proteins. However, the structure and function of proteins (Dewu et al., 2009) is one of the hot topics in biology. Therefore, the reliable basis can be provided by investigating the situation of the gene mutation in cancers to design the new protein. Meanwhile, it provides the reasonable target to molecular structure for the new medicine design.

In this paper, the Euclidean distance of codons was defined. According to the weighted Euclidean distance (Smith, 2000), the Euclidean distance was weighted by using improved variation coefficient method (Wu, 2002). So the new weighted Euclidean distance was obtained, then the p53 gene mutations in head and neck cancer were analyzed by using these two kinds of distance. By comparing with the Euclidean distance, it was found that the new weighted Euclidean distance was more meaningful in studying the amino acids' physical and chemical properties.

THEORETICAL MODEL

Algebraic structure

The algebraic structures of genetic code have been established by using the extended DNA or RNA, such as {O, A, C, G, U} (Sanchez

et al., 2005), {D, A, C, G, U} (Sanchez and Grau, 2009) and {D, A, C, O, G, T, P} (Yan and Ping, 2010). Particularly, a new GF(7³) (Yan and Ping, 2010) over the set of extended DNA or RNA alphabet C={D, A, C, O, G, T, P} was defined, and the extended triplet set C₃₄₃ (Yan and Ping, 2010) of oncogene was also defined. Then the relationship φ: GF(7³) → C₃₄₃ and a bijection f: {D, A, C, O, G, T, P} → GF(7) were established (Zhu et al., 2011), where 'O': {G, C} → {A, C}, 'P': {A, T} → {G, C}, 'D': others. Explicitly for instance, for the extended ordered set {D, A, C, O, G, T, P} the bijection f was defined as:

$$f(D) = 0, f(A) = 1, f(C) = 2, f(O) = 3, f(G) = 4, f(T) = 5, f(P) = 6$$

Where, φ is defined by

$$\begin{aligned} \varphi(a_0 + a_1x + a_2x^2) &= (f^{-1}(a_1)f^{-1}(a_2) \\ f^{-1}(a_0)) &= (X_1X_2X_3). \end{aligned}$$

In order to explain some relations of the codon mutation and amino acid from the algebraic angle, we selected the p53 gene mutations in head and neck cancer (Peltonen et al., 2010). In the foundation of the establishment algebra theoretical model, the selected p53 gene was transformed to the codons of C₃₄₃ by the relationship φ: GF(7³) → C₃₄₃ (Yan and Ping, 2010). A changed regulation will be used, namely from DDD to UUU. Thus it will provide the correct and powerful foundation for the distance that was defined in this paper.

Distance between codons

In order to discuss and analyze the p53 mutations in head and neck cancer impacting on the physical and chemical properties of amino acids, we defined the Euclidean distance between codons in the established algebraic structure GF(7³) (Ping et al., 2011) and the Euclidean distance in algebra. The formula is as follows:

Definition 1: $\forall X_1Y_1Z_1 \in C_{343}, X_2Y_2Z_2 \in C_{343},$
 $d(X_1Y_1Z_1, X_2Y_2Z_2) = \sqrt{[f(X_1) - f(X_2)]^2 + [f(Y_1) - f(Y_2)]^2 + [f(Z_1) - f(Z_2)]^2}$

This is defined as the Euclidean distance between the two codons.

Definition 2: $\forall X_1Y_1Z_1 \in C_{343}, X_2Y_2Z_2 \in C_{343},$
 $d^w(X_1Y_1Z_1, X_2Y_2Z_2) = \sqrt{2[\omega_1f(X_1) - \omega_2f(X_2)]^2 + 4[\omega_1f(Y_1) - \omega_2f(Y_2)]^2 + [\omega_1f(Z_1) - \omega_2f(Z_2)]^2}$

This is defined as the weighted Euclidean distance between the two codons; where, ω_{ij} (i=X, Y, Z, j=1, 2).

It is well-known that there is an association between the second-

position base and hydrophobicity (Qi et al., 2004). Chiusano et al. (2000) pointed out that the nucleotide frequencies in the second codon positions of genes are remarkably different for the coding regions that correspond to different secondary structures in the encoded proteins. Indeed, hydrophobic and hydrophilic amino acids were encoded by codons having U or A, respectively, in their second position. According to the codon degeneracy, the third base had less impact on the physical and chemical properties of amino acids. Therefore, according to the three positions of codon impacting on hydrophobic and hydrophilic of amino acids, the three positions' weights were selected as: 2:4:1. The following is to work out the weights of $f(D), f(A), f(C), f(O), f(G), f(T), f(P)$ by using the improved variation coefficient method. ($\omega_D, \omega_A, \omega_C, \omega_O, \omega_G, \omega_T, \omega_P$ stand for the weights).

Improved variation coefficient method

Variation coefficient method (Wu, 2002):

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ji} \text{ and } s_{ii} = \frac{1}{n-1} \sum_{j=1}^n (x_{ji} - \bar{x}_i)^2.$$

Where \bar{x}_i is the average of various targets, s_{ii} is variance, x_{ji} is the value of the j appraised object in the i . Thus the variation coefficient v_i of various targets is:

$$v_i = \frac{\sqrt{s_{ii}}}{\bar{x}_i}.$$

Since v_i is made in normalized processing, the weights ω_i of various targets can be obtained as:

$$\omega_i = \frac{v_i}{\sum_{j=1}^n v_j}.$$

In statistics, the information of target variation is weighed by the variance, but with the influence of various targets dimension and magnitude, various targets' variances have no comparison. Therefore, various targets coefficient of variation which could be comparable should be selected. To get the weights of various targets, we made various targets coefficients normalized processing. In this paper, v_i was improved to make the weight more objective, according to the internal variation information content and the extent of fluctuation which were provided by each target. Hence, $v_i = \frac{10|x_j - \sqrt{s_{ii}}|}{\bar{x}_i}$ was obtained and the weights

of $f(D), f(A), f(C), f(O), f(G), f(T), f(P)$ are $\omega_D = 0.34, \omega_A = 0.18, \omega_C = 0.03, \omega_O = 0.13, \omega_G = 0.29, \omega_T = 0.45$ and $\omega_P = 0.61$.

RESULTS AND DISCUSSION

The selected p53 gene mutations in head and neck cancer were analyzed by using d and d^* . The results can be seen from Table 1. From the data in the Table, it could be observed that the method based on the d^* can judge which position of the bases had changed more obviously as follows:

1. When the first position of the base changed, almost all the weighted Euclidean distance d^* of the mutated p53 in head and neck cancer were less than 1.2. Only the d^* in H&N20 and H&N3 reached 4.92, and both of these two mutations appeared on Exon 7. Moreover A, which is at the first position of the base, changed into C [$\ln GF(7^3)$] (Ping et al., 2011) changed into P]. In addition, the hydrophobicity of amino acid has never been changed though it can cause the corresponding macromolecule amino acid becoming a micromolecule (or the micromolecule into the macromolecule).
2. When the second position of the base changed, almost all the weighted Euclidean distance d^* of the mutated p53 in head and neck cancer were more than 1.5. Only the d^* in H&N63 was 0.66, which was less than 1. The mutation appears on Exon 6. Moreover C, which is at the second position of the base, was changed to T [$\ln GF(7^3)$] (Ping et al., 2011) changed into O]. When the second position of the base changed, the hydrophobicity of amino acid also changed, but when T changes into C (or C changes into T), the corresponding amino acid Val transforms into Ala (or Ala transforms into Val). Both Ala and Val are hydrophobic amino acids, thus the hydrophobicity was not changed. Moreover, when G changes into A, the corresponding amino acid is changed into hydrophilic amino acid.
3. When the third position of the base changed, all the weighted Euclidean distance d^* of the mutated p53 in head and neck cancer were less than 1.2, and the corresponding macromolecule amino acid becomes a micromolecule (or the micromolecule becomes a macromolecule). The hydrophobicity of amino acid was not changed as well. Through the calculation, the average d^* of the first base mutation which was more than the third base mutation was found. Though the first and the third position of the base of the mutated p53 gene in head and neck cancer have less impact on the hydrophobicity of amino acid, the first position has much influence than the third one.

In summary, the p53 gene mutations in head and neck cancer are analyzed by using d and d^* . The method based on the d^* can judge which position of the bases has changed more apparently. Thus the relationship between the different positions of the bases and the physical and chemical properties of amino acids can be

Table 1. Missense mutations of the p53 in head and neck cancer.

Sample*	Exon(s)*	Mutated codon*	Mutation by sequencing*	In _{GF(7³)}	d	d*	AA change*	Change in properties*
H&N12	5	130	<i>CTC</i> → <i>ATC</i>	<i>CTC</i> → <i>OTC</i>	1	0.47	Leu → Ile	No change
	5	139	<i>AAG</i> → <i>TGG</i>	<i>AAG</i> → <i>DPG</i>	5.1	6.96	Lys → Trp	+Charged → aromatic
	7	225	<i>GTT</i> → <i>GCT</i>	<i>GTT</i> → <i>GPT</i>	1	2.82	Val → Ala	Hydrophobic → polar
H&N19	5	157	<i>GTC</i> → <i>TTC</i>	<i>GTC</i> → <i>OTC</i>	1	1.09	Val → Phe	Hydrophobic, small → aromatic, big
	7	258	<i>GAA</i> → <i>AAA</i>	<i>GAA</i> → <i>OAA</i>	1	1.09	Glu → Lys	-Charge, acidic → + charge
H&N20	7	243	<i>ATG</i> → <i>CTG</i>	<i>ATG</i> → <i>PTG</i>	5	4.92	Met → Leu	Leu aliphatic
	8	297	<i>CAC</i> → <i>TAC</i>	<i>CAC</i> → <i>OAC</i>	1	0.47	His → Tyr	No change
H&N31	5	171	<i>GAG</i> → <i>GAC</i>	<i>GAG</i> → <i>GAD</i>	4	1.16	Glu → Asp	Big → small
H&N29	7	259	<i>GAC</i> → <i>GAA</i>	<i>GAC</i> → <i>GAO</i>	1	0.33	Asp → Glu	Small → big
H&N32	5	171	<i>GAG</i> → <i>GAC</i>	<i>GAG</i> → <i>GAD</i>	4	1.16	Glu → Asp	Big→small
	8	271	<i>GAG</i> → <i>TAG</i>	<i>GAG</i> → <i>OAG</i>	1	1.09	Glu → stop	
H&N3	5	159	<i>GCC</i> → <i>ACC</i>	<i>GCC</i> → <i>OCC</i>	1	1.09	Ala → Thr	No change
	7	254	<i>ATC</i> → <i>GTC</i>	<i>ATC</i> → <i>PTC</i>	5	4.92	Ile → Val	Big → small
H&N16	5	130	<i>CTC</i> → <i>TTC</i>	<i>CTC</i> → <i>OTC</i>	1	0.47	Leu → Phe	Aliphatic → aromatic
	7	245	<i>GGC</i> → <i>GAC</i>	<i>GGC</i> → <i>GOC</i>	1	1.54	Gly → Asp	Polar → acidic
H&N4	8	283	<i>CGC</i> → <i>CCC</i>	<i>CGC</i> → <i>CDC</i>	4	2.32	Arg → Pro	+Charged → polar
H&N58	7	259	<i>GAC</i> → <i>GAA</i>	<i>GAC</i> → <i>GAO</i>	1	0.33	Asp → Glu	Small → big
H&N53	5	172	<i>GTT</i> → <i>GCT</i>	<i>GTT</i> → <i>GPT</i>	1	2.82	Val → Ala	Hydrophobic → polar
H&N64	7	238	<i>TGT</i> → <i>TCT</i>	<i>TGT</i> → <i>TDT</i>	4	2.32	Cys → Ser	Hydrophobic → polar
H&N63	6	189	<i>GCC</i> → <i>GTC</i>	<i>GCC</i> → <i>GOC</i>	1	0.66	Ala → Val	Polar → hydrophobic
H&N51	7	259	<i>GAC</i> → <i>GAA</i>	<i>GAC</i> → <i>GAO</i>	1	0.33	Asp → Glu	Small → big
H&N56	8	275	<i>TGT</i> → <i>TAT</i>	<i>TGT</i> → <i>TOT</i>	1	1.54	Cys → Tyr	Small → big, aromatic
H&N46	6	217	<i>GTG</i> → <i>GCG</i>	<i>GTG</i> → <i>GPG</i>	1	2.82	Val → Ala	Hydrophobic → polar
H&N43	5	155	<i>ACC</i> → <i>TCC</i>	<i>ACC</i> → <i>DCC</i>	1	0.25	Thr → Ser	No change
	5	175	<i>CGC</i> → <i>CAC</i>	<i>CGC</i> → <i>COC</i>	1	1.54	Arg → His	+Charged → big, aromatic

Table 1. Contd.

H&N61	5	171	<i>GAG</i> → <i>GAC</i>	<i>GAG</i> → <i>GAD</i>	4	1.16	Glu → Asp	Big → small
H&N60	5	172	<i>GTT</i> → <i>CTT</i>	<i>GTT</i> → <i>DTT</i>	4	1.64	Val → Leu	Small → big
	7	259	<i>GAC</i> → <i>GAA</i>	<i>GAC</i> → <i>GAO</i>	1	0.33	Glu → Asp	Big→small
H&N69	5	148	<i>GAT</i> → <i>GAG</i>	<i>GAT</i> → <i>GAP</i>	1	1.41	Asp→Glu	Small→big
	6	221	<i>GAG</i> → <i>GAC</i>	<i>GAG</i> → <i>GAD</i>	4	1.16	Glu→Asp	Big→small

* Youwei (2002).

analyzed thoroughly. When d^* is more than 1.5, the second position of base changes, and at the same time, the hydrophobicity of amino acid is changed. When d^* is less than 1.2, the first or the third position of base changes without changing the hydrophobicity of amino acid. However, the average d^* of the first base mutation was more than the third base mutation by calculation. Hence, we concluded that the second base has the greatest influence on amino acids; that is to say, it can change the physical and chemical properties of amino acids. The first base has a little influence on amino acids. It, however, has much influence than the third one, though it cannot change the physical and chemical properties of amino acids.

Conclusions

In this paper the Euclidean distance named d was defined on the basis of the established algebraic structure $GF(7^3)$ (Ping et al., 2011). Furthermore, the Euclidean distance was weighted by using improved variation coefficient method, named weighted Euclidean distance d^* . We pointed out that the method based on the d^* can judge which position of the bases has changed more

obviously. The hydrophobicity of amino acid of the mutated p53 in head and neck cancer was analyzed by calculating the weighted Euclidean distance d^* . When the second position of base changed, it could change the hydrophobicity of amino acid. However, when the first or the third position of base changed, it could not change the hydrophobicity of amino acid. The results therefore show that the second base had the greatest influence on amino acids, the first had smaller influence, while the third had less impact on the physical and chemical properties of amino acids.

Furthermore, by analyzing the hydrophobicity of amino acid, the change of the protein's structure in head and neck cancer could be obtained when the normal gene becomes cancerous. This result provides a reasonable target molecular structure to the new medicine member design and has a positive value in the treatment of cancer. The findings of this paper therefore have certain significance, especially in biology and medicine.

REFERENCES

An Q, Liu Y, Gao Y, Huang J, Fong X, Liu L, Zhang D, Zhang J, Cheng S (2002). Deletion of tumor suppressor genes in Chinese non-small cell lung cancer. *Cancer Lett.* 184: 189-

195.
 Bodey B, Bodey Jr. B, Groger AM, Luck JV, Siegel SE, Taylor CR, Kaiser HE (1997). Immunocytochemical detection of proteins in the P170 multidrug resistance (MDR) and the p53 tumor suppressor gene proteins in human breast cancer cells: clinical and therapeutical significance. *Anticancer Res.* Mar-Apr, 17(2): 1311-1318.
 Crick FHC (1968). The Origin of the Genetic Code. *J. Mol. Biol.* 38: 367-369.
 Dewu D, Yanrui D, Yujie C, Wenbo X (2009). Structural and Functional Analysis of the GSC of *Bacillus thuringiensis* Metabolic Network. *Braz. J. Microb.* 40(2): 411-416.
 Grantham R (1974). Amino Acid Difference Formula to Help Explain Protein Evolution. *Science*, 185: 862-864.
 Joseph S, Chomehai MD, Wei Du PhD, Fazlul H, Sarkar PhD, (1999). Prognostic significance of p53 gene mutations in laryngeal cancer. *Laryngoscope.* 109: 455-459.
 Kong J, Zhu P (2011). Research and application of the Codon Bias of Human p53 tumor protein. *Appl. Res. Computers*, 28: 2987-2990.
 Ma F, Wu Y, Xu X (2003). Correlation Analysis of Some Physical Chemistry Properties among Genetic Codons and Amino Acids. *J. Anhui Agric. Uni.* 30(4): 439-445.
 Maria Luisa C, Alvarez-Valin F, Massimo Di G, Giuseppe D, Gaetano A, Giovanni C, Giorgio B (2000). Second codon positions of genes and the secondary structures of proteins. Relationships and implications for the origin of the genetic code. *Gene.* 261: 63-69.
 Peltonen K, Henni MH, Paavo P, Taina T-H, Kirsi H V (2010). P53 in head and neck cancer: Functional consequences and environmental implications of TP53 mutations. *Head Neck Oncol.* 2: p. 36.
 Qi L, Wang Y, Zhan Y (2004). Relation between base distribution in protein-coding regions and stop codons. *J. Shandong University Technol.* (Science and

- Technology).18(2): 95-98.
- Sanchez R, Perfetti LA, Grau R, Morgado E (2005). A New DNA sequences vector space on a genetic code Galois field. Match Commun. Math. Comput. Chem. 54: 3-28.
- Sanchez R, Grau R (2009). An algebraic hypothesis about the primeval genetic code architecture. Math. Biosci. 221: 60-76.
- Sanchez R, Grau R, Morgado E (2006). A novel Lie algebra of the genetic code over the Galois field of four DNA bases. Math. Biosci. 202: 156-174.
- Smith SA (2000). Aggrawal Management of Multi-item Retail Inventory Systems with Demand Substitution. Operations Res. 48(1): 50-64.
- Sun L, Yu F, Xu X (2003). Some Important Properties of Genetic Codons and Amino Acids and Their Correlation Relationship. J. Nanjing Normal University (Natural Science Edition). 26(2): 74-77.
- Woese C (1965). On the Evolution of the Genetic Code. Proc. Natl. Acad. Sci. 54: 1546-1552.
- Wu Y (2002). Experiment Design and Data Processing. Suzhou University Press. M. pp. 1-418.
- Xue X, Wang S, Jin Y (2001). Information Connotation of Genetic Codes and Their Corresponding Relationship with Amino Acids. J. Nanjing Normal University (Natural Science Edition). 24(2): 63-66.
- Yan Y, Zhu P (2010). Extended triplet set C_{343} of DNA sequences and its application to p53 gene. The Chinese Physics B. 20(1): 701-709.
- Zhu P, Yan Y, Kong J (2011). Application of oncogene using distance and Y-inner product between codon. J. Jiangsu University (Natural Science Edition). p. 32.