广西科学院学报
Journal of Guangxi Academy of Sciences

2010,26(2):130~139
Vol. 26,No. 2 May 2010

Creation and Application of Computational Mutation*
计算变异学创立背景及其用途

YAN Shao-min[1],WU Guang[2**]
严少敏[1],吴　光[2**]

(1. National Engineering Research Center for Non-food Biorefinery, Guangxi Academy of Sciences, Nanning, Guangxi, 530007, China; 2. Computational Mutation Project, DreamSciTech Consulting,Shenzhen, Guangdong,518054, China)
(1.广西科学院国家非粮生物质能源工程技术研究中心,广西南宁　530007;2.深圳市追梦科技咨询有限公司,广东深圳　518054)

Abstract:Computational mutation is a discipline developed according to the random principle that lies in the very heart of the nature. It overcomes the limitation of bioinformatics and computational biology, where the letters and measures are not subject to the sequence length, amino-acid composition and position, neighboring amino acids, etc. Three methods are developed, amino-acid pair predictability, amino-acid distribution probability and mutating probability, to quantify a whole protein or each amino acid in proteins, which provides living, dynamic measures to quantitatively analyze protein. Currently the computational mutation is applied to studying the protein evolution, diagnosing genetic disorder, estimating protein structure and function, designing drug target, predicting mutation and so on.

Key words: amino-acid pair predictability, amino-acid distribution probability, amino-acid mutating probability, protein, computational mutation

摘要:为了克服生物信息学和计算生物学中字母或数字不受序列长度、氨基酸组成和位置、相邻氨基酸影响的缺陷,根据自然界普遍存在的随机性原理,创立计算变异学。计算变异学用氨基酸对可预测性、氨基酸分布概率和变异概率3种方法量化整个蛋白质及每个氨基酸,用活的、动态的测量指标量化分析蛋白质。计算变异学方法可以应用于研究蛋白质进化、遗传病定量诊断,分析蛋白质结构与功能、药物设计和病毒变异预测等领域。

关键词:氨基酸对可预测性　氨基酸分布概率　氨基酸变异概率　蛋白质　计算变异学
中图法分类号:Q-03　　**文献标识码**:A　　**文章编号**:1002-7378(2010)02-0130-10

Since 1999, we have been developing a research approach that is now called the computational mutation. The computational mutation not only produces more than 70 research articles in international peer-reviewed journals including 50 articles indexed in SCI journals over last ten years[1~77], a chapter in a book[78] and a book[79], but also opens a new research front. Therefore, it is our duty to introduce the computational mutation to the scientific community inside China because all of our publications are in English without Chinese abstract.

In this mini-review, we would like to use the plain words to explain what the computational mutation is, where it comes from, what its advantage is over the current computational methods in biological sciences. Here, we use proteins to illustrate the computational mutation because our work exclusively concentrates on protein study although the computational mutation can be used for DNA and RNA studies.

1　Creation of computational mutation

1.1　Bioinformatics and computational biology

The challenge faced in post-human genome project is that humans have too much DNA/RNA/

收稿日期:2009-08-05

作者简介:严少敏(1958-),女,博士,研究员,主要从事定量诊断病理学和计算变异学研究。

*国际科技合作项目(2008DFA30710)、广西自然科学基金项目(广西重点实验室培育08-115-011和桂科自0991080)和广西科学院(桂科院研0701和09YJ17SW07)资助。

**通讯作者。

protein data to deal with, because the international open-access databanks are very rich resources, from which we can find far much more information.

In order to analyze these vast amounts of the data, not only computers are employed, but also more importantly several computational methods have been developed. Among these methods, bioinformatics and computational biology play the leading roles. Although there are many definitions on what bioinformatics and computational biology are, we would like to look at them from computational viewpoint.

As we know that DNA/RNA/protein sequences are represented using letters. For example, we generally use 20 letters to represent 20 amino acids, thus in fact a protein sequence is a sequence of different letters. Actually, we use a computer to deal with a sequence of letters when we would like to find any information from a DNA/RNA/protein.

In this context, the bioinformatics is somewhat similar to what we are doing when reading a book: we could meet unknown words or we could have interests on where a particular sentence comes from. In our reading, we can use a dictionary to find the meaning of unknown words and search various literatures to find where a particular sentence comes from. In biological fields, when we find a new protein, for instance, we can check it against protein databanks to see if there is any similarity between this new protein and proteins in databanks. Then we can define the possibly functional units by comparing each part of this new protein with known functional units of various proteins in databanks.

In this view, the bioinformatics mainly operates on the letters represented nucleotides or amino acids. Technically, the most programs in bioinformatics deal with comparing letters by commands of yes and no as well as how to align DNA/RNA/protein sequence in order to make it comparable with historical ones. There are two development stages in bioinformatics, the one was mainly involved with researchers specialized in program writing while the other is now mainly involved with researchers using these programs for biological studies.

But human efforts were not stopped at the stage of using computer to operate letters, because researchers also hope to bring more meaningful information rather than the letters into analysis on DNA/RNA/protein. This led to the computational biology, by which the physicochemical property of nucleotide/amino acid are used to replace the letters in DNA/RNA/protein sequence. For example, we can use the molecular weight of an amino acid to replace the letter of this type of amino acids in a protein, thus we get a molecular-weight sequence of a protein. Currently there are about 500 physicochemical properties used for this propose[80].

However, there is a limitation when using the physicochemical property to replace the letters in DNA/RNA/protein sequence, because the physicochemical property was created by physicists or chemists for their own proposes, which could be different from our proposes in biological fields. For example, the same type of amino acids may play different roles at different positions in a protein. However, there is no difference between any two amino acids of the same type if we use any physicochemical property to replace them because the physicochemical property is a constant value for a certain type of amino acids.

1.2 Development of new measures

A protein is alive and evolves through mutations. Therefore we aimed to find the measures that would be sensitive to mutation. This goal led to the creation and development of computational mutation.

On the other hand, we can find that the development of methods to measure natural phenomena appears in the earlier stages of most scientific fields if we look back the scientific history. For example, the geometry began from measuring land, and the relativity began from measuring speed of light. Thus, the famous French scientist, Henri Poincaré said that the important thing is not to know what it is, but how to measure it[81].

However, it is not difficult to note that there is no similar development of how to measure the living sign of DNA/RNA/protein in biological sciences. This fact again encourages us to search such measures.

1.2.1 Why we need measurements

Here, one may easily ask a question why we need to develop measures that convert letters into numbers. A simple answer to this question is that the replaced numbers can represent some particular property of DNA/RNA/protein. However, we would like to answer this question in such a way that the conversion of letters to numbers allows us to

observe DNA/RNA/protein in a domain different from the domain represented by the letters.

In fact, we frequently convert an issue of interest into another domain in order to find something that does not appear in the original domain. For example, when we use the molecular weight of amino acids to replace amino acids in a protein, we look at this protein from the domain of molecular weight.

Moreover, the conversion of DNA/RNA/protein into numerical domain from letter one not only helps us observe their patterns in numerical domain but also more importantly provides us with the opportunity of applying mathematical models to analyze the DNA/RNA/protein in numerical domain.

1.2.2 Difficulty in conversion from letters to numbers

Although a protein is generally composed of 20 types of amino acids, we cannot simply use numbers to replace these letters, say, to use 1 to 20 to replace 20 types of amino acids, not only there is no meaning in this conversion, but also we cannot deal with this type of data efficiently as our mathematical system is dealing with decimal, binary, octal and so on. This is the same for DNA and RNA.

Thus, we face two difficulties, (i) we need the converted numbers falling into our mathematical system with meanings, and (ii) we need the converted numbers to have a living sign of DNA/RNA/protein.

2 Methods of computational mutation

At first our attention paid to probability, because the probability would simply suggest the chance that a mutation is likely to occur if we associate this probability with a mutation. Besides, pure chance is now considered to lie at the very heart of nature[82]. Along this line of thought, we have developed three methods to measure the amino-acid difference in different compositions, in different lengths, at different positions and with different neighboring amino acids.

2.1 Method I: amino-acid pair predictability as a measure to analyze the composition of adjacent amino-acid pairs in a protein

2.1.1 Actual frequency and predicted frequency of amino-acid pair

As we compared the bioinformatics with our reading, we also initially considered a protein as an unknown text. For an unknown text, we can use the cryptology as a tool to analyze its meaning by counting the frequencies of a single letter, paired letters, and so on. Along this line of thought, we had counted the frequencies of amino-acid pairs, three-amino-acid sequences, four-amino-acid sequences, five-amino-acid sequences, and so on. These counted frequencies are their actual frequencies and finally we found that the amino-acid pairs are most suitable to be analyzed.

For analyzing an unknown text, we can use counted frequencies to compare with the natural frequencies of elements in a known human language. However, we have no protein language as a reference to compare the actual frequencies of amino-acid pairs in a protein. We therefore considered the frequency obtained by permutation as the predicted frequency, because there are 20 types of amino acids so theoretically there would be 400 (20×20) types of amino-acid pairs, which can serve as a reference for our comparison.

2.1.2 Amino-acid pair predictability

For each type of amino-acid pair, we can determine it belonging to predictable or unpredictable by comparing its actual frequency with the predicted one. For example, the human factor IX protein (accession number P00740) consists of 461 amino acids, which can be counted as 460 adjacent amino-acid pairs. The factor IX protein has 43 glutamic acids "E" and 32 asparagines "N": if the permutation can predict the appearance of amino-acid pair EN, which must appear three times ($43/461 \times 32/460 \times 460 = 2.98$), and it indeed appears 3 times, thus its appearance is predictable using permutation. By contrast, this factor IX protein has 16 tyrosines "Y" and 37 valines "V": if the permutation can predict the appearance of YV, which must appear once ($16/461 \times 37/460 \times 460 = 1.28$), however, it appears five times in realty, thus its appearance is unpredictable.

2.1.3 Predictable and unpredictable portions of protein

In this way, we can classify all amino-acid pairs in a protein as predictable and unpredictable, and calculate the predictable and unpredictable portions. For human factor IX protein, its predictable and unpredictable portions are 26.09% and 73.91% as the total is 100%. Both the predictable portion and the unpredictable can serve as a measure to represent

a protein.

There is a mutation of human factor IX substituting cysteine for arginine at position 407. Although this mutation is related to a single amino acid, its predictable and unpredictable portions become 25.65% and 74.35%. In this manner, each protein has a unique number, either predictable or unpredictable portion, to distinguish itself from others.

From the application viewpoint, we can use this approach to compare proteins from different families, subtypes, and species[30, 33, 38, 42, 47, 48, 74], and we can also use this approach to study the evolution of a protein family along the time course[41, 44, 48, 67, 71, 76, 77] when we consider each protein as a basic element for comparison.

2.1.4 Five attributes of amino-acid pair predictability

On the other hand, if our research interests concentrate on individual amino-acid pair rather than a whole protein, we can assign the predicted and actual frequency to each individual amino-acid pair, then we would have the difference between predictable and actual frequency. By comparing predicted frequency with actual one, we furthermore classify each individual amino-acid pair as (i) the randomly predictable present type of amino-acid pair with predictable frequency, (ii) the randomly predictable absent type of amino-acid pair, (iii) the randomly predictable present type of amino-acid pair with unpredictable frequency, (iv) the randomly unpredictable present type of amino-acid pair, and (v) the randomly unpredictable present type of amino-acid pair.

From the application viewpoint, we can use this approach to compare the difference in a particular amino-acid pair before and after mutation in order to determine the mutation patterns related to each individual amino-acid pair[21, 23~31, 34~36, 40, 65, 66, 70,75].

2.1.5 Meanings of amino-acid pair predictability

After developed this method conceptually, we have computed tens of thousands of proteins to determine if the amino-acid pair predictability is valid for different proteins. The results confirm that each protein has its unique predictable and unpredictable portions, and that the difference between predicted and actual frequency is very sensitive to mutation.

Meanwhile, the amino-acid pair predictability can have the following meanings: (i) the predictable amino-acid pairs suggest their construction with the maximal probability of occurrence, which needs the least time and energy; (ii) the unpredictable amino-acid pairs suggest that nature deliberately spend more time and energy for its construction no matter of what purpose is.

Still, the amino-acid pair predictability can reliably record the changes in proteins induced by gene mutation so it can measure the living proteins. The most important implication drawn from our studies is that protein evolution can be regarded as that Nature would like to minimize the difference between predicted and actual frequency of amino-acid pairs in a protein, which leads to mutations. However, any new mutation may create new difference between predicted and actual frequency of amino-acid pairs, which may lead Nature to minimize the new difference again through mutation. Such a process may continue without ending, which is one of reasons driving the evolution.

2.2 Method Ⅱ: Amino-acid distribution probability as a measure to analyze the complexity of amino-acid distribution

2.2.1 Amino-acid distribution probability

After developed the amino-acid pair predictability, we considered that we needed to develop a measure that could be sensitive to the positions of amino-acids in a protein because the amino-acid pair predictability is more relevant to protein length, composition and neighboring amino acids.

Initially, we assumed how we could guess an approximate position of an amino acid in a protein, whose answer is that an amino acid can be at any position of a protein. Then, we guess two approximate positions for two amino acids in a protein, naturally we could imagine dividing this protein into two partitions, then each part could have an amino acid, or a part could have two amino acids. Afterwards, we guess three approximate positions for three amino acids in a protein, and similarly we could imagine dividing this protein into three partitions, and so forth for more amino acids.

This led us to consider a similar situation in statistical physics, where there are Maxwell Boltzmann, Fermi Dirac, and Bose Einstein assumptions[83]. If we do not distinguish each partition, our situation is similar to the Maxwell-Boltzmann assumption. So we can view the

distribution pattern of a type of amino acids along protein sequence is analogous to the occupancy of subpopulations and partitions, thus we can calculate the amino-acid distribution probability using $\frac{r!}{r_1! \times r_2! \times \cdots \times r_n!} \times \frac{r!}{q_0! \times q_1! \times \cdots \times q_n!} \times n^{-r}$[84], where r is the number of amino acids, n is the number of partitions, r_n is the number of amino acids in the n-th partition, q_n is the number of partitions with the same number of amino acids, and ! is the factorial function.

For example, two amino acids have two possible distributions along a protein, that is, two amino acids distribute in each partition or in any one partition if we imagine dividing this protein into two partitions, for which we would have the distribution probabilities, $\frac{2!}{1! \times 1!} \times \frac{2!}{0! \times 2! \times 0!} \times 2^{-2} = \frac{2}{1 \times 1} \times \frac{2}{1 \times 2 \times 1} \times 0.25 = 0.5$, $\frac{2!}{2! \times 0!} \times \frac{2!}{1! \times 0! \times 1!} \times 2^{-2} = \frac{2}{2 \times 1} \times \frac{2}{1 \times 1 \times 1} \times 0.25 = 0.5$.

2.2.2 Characteristics of amino-acid distribution probability

In this way, we can computer the amino-acid distribution probability for each type of amino acids in a protein using the above equation. Here, we can notice several characteristics of amino-acid distribution probability.

First, there are two theoretical distributions for two amino acids in a protein as mentioned in the above section, and then there are three theoretical distributions for three amino acids in a protein. Hereafter, the thing is different, because there are five theoretical distributions for four amino acids, seven theoretical distributions for five amino acids, 11 theoretical distributions for six amino acids, 15 theoretical distributions for seven amino acids, and so on. Thus, the general rule is that the increase in the number of distributions is not proportional to the increase in the number of amino acids.

Second, the largest distribution probability decreases as the increase in the number of amino acids, which are more than two. For instance, the largest distribution probability is 0.67 for 3 amino acids, 0.56 for 4 amino acids, 0.38 for 5 amino acids, 0.35 for 6 amino acids, and so on.

Third, the probability of uniform distribution is very small, that is, the chance is very small for each partition to have an amino acid.

Finally, sometimes different distributions can have the same distribution probability so we arrange all the distribution probabilities in descending order, which we call as the distribution rank for distinction. The bigger the distribution probability is, the smaller the distribution rank is. The distribution rank is the same for different distributions with the same distribution probability.

2.2.3 Actual and predicted amino-acid distribution probability

Theoretically there are many distributions for each type of amino acids in a protein, but a certain type of amino acids in a protein can adopt only one distribution in real-life, whose distribution probability is the actual amino-acid distribution probability.

According to the random principle, any event with the greatest probability will occur most likely. Therefore, the biggest amino-acid distribution probability can be considered as the predicted probability to serve as a reference, and we can compare the actual amino-acid distribution probability with the predicted one to analyze the complexity of amino-acid distribution in proteins.

2.2.4 Predictable/Unpredictable portion and difference between predicted and actual probability

As we do in Method I, we can also classify a protein as predictable and unpredictable portions in terms of amino-acid distribution probability. Still, we can assign the predicted and actual amino-acid distribution probabilities to each amino acid in a protein, which is somewhat different from what we do using the amino-acid pair predictability where we assign the predicted and actually frequencies to amino-acid pair. We can estimate the complexity of amino-acid distribution by the ratio of predicted versus actual probability[10, 15, 16, 19, 22, 37, 39, 45, 46].

2.2.5 Meanings of amino-acid distribution probability

The amino-acid distribution probability is a measure to estimate the spatial randomness in the protein primary structure. It can answer why a certain type of amino acids do not evenly distribute along a protein sequence but rather concentrate on different regions. From the random viewpoint, the probability is quite small for the uniform distribution, indicating that Nature requires the non-uniform distribution of amino acids along a protein sequence during the process of protein synthesis to form the

active site, which is the base for high-level structure and protein function.

The small distribution probability or the big distribution rank suggests that a protein composition is less random and more complicated. Such a protein is less stable and easy to mutate.

2.3 Method Ⅲ: amino-acid mutating probability as a measure to analyze the future amino-acid composition and would-be mutated amino acid

2.3.1 Translation probability between RNA Codons and Amino Acids

After developed the above two methods, which are not related to the time direction, we hope to develop a method, which is related to the time direction. Our attention at first paid to the relationship between RNA codons and translated amino acids.

There are unambiguous relationships between RNA codons and translated amino acids, and the same amino acid can be translated by different RNA codons because there are 64 RNA codons but 20 amino acids plus STOP signal, say, RNA codons are more than amino acids. For example, methionine is related with a single RNA codon (AUG), phenylalanine with two codons (UUU and UUC), isoleucine with three codons (AUU, AUC and AUA), proline with four codons (CCU, CCC, CCA and CCG), and leucine with six codons (UUA, UUG, CUU, CUC, CUA and CUG).

As a RNA codon is composed of any three out of four nucleotides (A, C, G and U), how does a point mutation at RNA codon affect the five amino acids mentioned above? We can infer that the order is methionine, phenylalanine, isoleucine, proline and leucine according to their affected extent. As methionine is only related with a single RNA codon, it will certainly be mutated into other amino acids; however, leucine is related with six codons so a point mutation at RNA codon will have a smaller impact on it. From the difference between the number of RNA codons and the number of amino acids, we can deduce the translation probability between RNA codons and translated amino acids, which is a time-orientated measure because the mutation is directed to the future, that is, this probability indicates which amino acid appears easily after a point mutation[43, 49, 78].

2.3.2 Amino-acid mutating probability

The translation probability between RNA codons and translated amino acids are related to two levels: RNA and protein. If we focus on studying variations in the protein level, we are more interested in the probability that an amino acid changes to other ones. Therefore, we sum up the mutating probabilities, by which we know the chance for an amino acid to mutate to another amino acid.

2.3.3 Amino-acid future composition

The amino-acid mutating probability reflects the time-orientated randomness, so we can calculate the future composition of amino acids in a mutant protein from the current wild protein. The difference between future and current composition of amino acids not only drives protein mutations, but also implies which type of amino acid has a greater chance appearing after mutation and which type of amino acid has a smaller chance.

More importantly, we know when a mutation can form the STOP signal so we can explain why some mutations will induce the truncated protein and its proportion. This is a remarkable feature that our approach is different from other methods.

2.3.4 Meanings of amino-acid mutating probability

The future amino-acid composition is based on large-scale and long-term statistics, which is governed by the translation probability between RNA codons and translated amino acids. To predict the would-be mutated amino acids, we cannot accurately predict what type of amino acids would form in the mutant, but we can know the probability of its formation.

In fact, the ratio of future versus current amino-acid composition indicates the mutation trend, that is, the bigger the ratio is, the larger the mutation trend is for given type of amino acids. From this ratio, we can know which type of amino acids is more likely to mutate, which is very important for knowing the mutation trend in each type of amino acids and predicting mutations, thus we can use this ratio as an indicator engineering mutation.

2.4 Comparison of three methods

All our three methods can quantify each amino acid in a protein sequence and the whole protein with a numerical datum. During 10 years, we have used them to study tens of thousands of various proteins, to observe their behaviors from the angle of time, space, and time and space. The results reveal that dynamic is the common feature for all of these measures. They are sensitive to the changes in

protein length, in amino-acid composition and position, and in neighboring amino acids. Thus, these methods are suitable for analyzing mutation.

Method Ⅰ deals with the various subsets of amino-acid pairs so it is sensitive to the changes in adjacent amino acids. Method Ⅱ deals with the linear space structure of amino acids so it is sensitive to the changes in position. Method Ⅲ deals with the time point of amino acids so it is sensitive to the would-be mutated amino acids.

Thus, these three methods reflect the protein randomness from different viewpoints, and provide new ways for further understanding of the biological evolution through mutations. The detail comparisons of the methods have been published in our previous works[20, 48, 79].

3 Applications of computational mutation

After developing the methods of computational mutation, we have been focusing on exploring their applications, in order to use them to explain various phenomena in biology and medicine, and to solve practical problems.

3.1 Analyzing mutation patterns in proteins

We have used the approaches of computational mutation to analyze tens of thousands of proteins, in order to reveal their mutation patterns. The results show that the mutation is highly likely to occur at the unpredictable amino-acid pairs. The majority of mutation-targeted pairs are characterized by one or both pairs whose actual frequency is larger than predicted one, meanwhile many mutations lead to one or both mutation-formed amino-acid pairs with their actual frequency smaller than predicted one. Thus, the mutation trend is to diminish the difference between predicted and actual frequency of amino-acid pairs[41, 44, 48, 67, 71, 76, 77].

3.2 Diagnosing genetic disorders quantitatively

We have analyzed various genetic diseases, firstly we use the amino-acid distribution probability to quantify the normal protein and its mutants, then we use the cross-impact analysis to determine the relationship between mutant proteins and their clinical outcome, and finally we use the Bayesian equation to calculate the probability that a certain disease occurs under a mutation.

In this way, we build a descriptively probabilistic method to determine the probability of occurrence of a single gene disorder when a new mutation is present. Our approach paves the way for analyzing quantitatively the relationship between genotype and phenotype, which will help the early diagnosis[61~64, 69, 72, 73].

3.3 Estimating protein structure and function

We have used the amino-acid distribution rank as a measure to quantify the normal and mutant hemoglobin chains to analyze their stability and oxygen affinity. The results show that if a mutation increases the amino-acid distribution rank, the mutant has a larger probability of increasing oxygen affinity, but also has a larger probability of instability.

These studies support our viewpoint: the larger the amino-acid distribution rank is, the more complex the protein structure is, the more powerful the protein function is, but the less stable the protein is. Our approach has benefit to quantitatively study the relationship between protein structure and its function[58, 79].

3.4 Designing potential targets for antiviral drug

From the probability viewpoint, we can choose some amino-acid pairs as potential targets of antiviral drug. These pairs will have a large chance of colliding with drug, and of linking closely to protein function, but will be less sensitive to mutations. This theoretical framework provides new ideas for designing antiviral drugs[32, 79].

3.5 Predicting mutations in influenza A viruses

In recent years, we have focused on the mutations of influenza A viruses because of the threat of flu pandemic. We have been explored how to systematically predict the mutations and developed the prediction strategies, for example, using the cross-impact analysis and Bayesian equation to calculate the probability of spontaneous mutations, using the logistic regression and neural network to predict the mutation positions, using the amino-acid mutating probability to predict the would-be mutated amino acids, using the Fast Fourier Transform to determine the periodicity, searching for natural factors that affect the virus mutations to time the outbreak of influenza, and so on[33, 39, 41, 42, 44, 46~48, 50~57, 59, 60, 67, 68, 71, 74, 76, 77].

4 Conclusions

Over the last 10 years, we have been fortunate to develop a discipline, which is called the computational mutation. This computational mutation has two important advantages over the other approaches: (i) the computational mutation

can measure the living sign of DNA/RNA/protein, which overcomes the limitation of bioinformatics and computational biology, and (ii) the computational mutation suggests that the evolution of DNA/RNA/protein is partially attributed to the fact that Nature has the intention to minimize the difference between predicted and actual values, which leads to mutations that create new difference between predicted and actual values, thus the evolution is a non-stop and continuous process.

References:

[1] Wu G. The first and second order Markov chain analysis on amino acids sequence of human haemoglobin α-chain and its three variants with low O_2 affinity[J]. Comp Haematol Int,1999,9: 148-151.

[2] Wu G. Frequency and Markov chain analysis of amino-acid sequence of human glutathione reductase [J]. Biochem Biophys Res Commun,2000,268:823-826.

[3] Wu G. Frequency and Markov chain analysis of amino-acid sequence of human tumor necrosis factor [J]. Cancer Lett,2000,153:145-150.

[4] Wu G. The first,second and third order Markov chain analysis on amino acids sequence of human tyrosine aminotransferase and its variant causing tyrosinemia type Ⅱ [J]. Pediatr Grenzgeb,2000,39:37-47.

[5] Wu G. Frequency and Markov chain analysis of the amino-acid sequence of sheep p53 protein[J]. J Biochem Mol Biol Biophys,2000,4:179-185.

[6] Wu G. Frequency and Markov chain analysis of the amino acid sequence of human alcohol dehydrogenase α-chain[J]. Alcohol Alcohol,2000,35:302-306.

[7] Wu G. The first,second,third and fourth order Markov chain analysis on amino acids sequence of human dopamine β-hydroxylase [J]. Mol Psychiatry,2000,5: 448-451.

[8] Wu G. Frequency and Markov chain analysis of amino-acid sequences of mouse p53[J]. Hum Exp Toxicol, 2000,19:535-539.

[9] Wu G,Yan SM. Prediction of two- and three-amino-acid sequences of Citrobacter Freundii β-lactamase from its amino acid composition [J]. J Mol Microbiol Biotechnol,2000,2:277-281.

[10] Wu G,Yan S. Prediction of distributions of amino acids and amino acid pairs in human haemoglobin α-chain and its seven variants causing α-thalassemia from their occurrences according to the random mechanism [J]. Comp Haematol Int,2000,10:80-84.

[11] Wu G,Yan SM. Prediction of two- and three-amino acid sequence of human acute myeloid leukemia 1 protein from its amino acid composition [J]. Comp Haematol Int,2000,10: 85-89.

[12] Wu G,Yan SM. Frequency and Markov chain analysis of amino-acids sequence of human platelet-activating factor acetylhydrolase α-subunit and its variant causing the lissencephaly syndrome [J]. Pediatr Grenzgeb, 2000,39:513-526.

[13] Wu G,Yan SM. Prediction of presence and absence of two- and three-amino-acid sequence of human monoamine oxidase B from its amino acid composition according to the random mechanism[J]. Biomol Eng, 2001,18:23-27.

[14] Wu G,Yan SM. Frequency and Markov chain analysis of amino-acid sequences of human connective tissue growth factor[J]. J Mol Model,2001,5:120-124.

[15] Wu G,Yan SM. Analysis of distributions of amino acids,amino acid pairs and triplets in human insulin precursor and four variants from their occurrences according to the random mechanism [J]. J Biochem Mol Biol Biophys,2001,5:293-300.

[16] Wu G,Yan S. Analysis of distributions of amino acids and amino acid pairs in human tumor necrosis factor precursor and its eight variants according to random mechanism[J]. J Mol Model,2001,7:318-323.

[17] Wu G,Yan SM. Prediction of presence and absence of two- and three-amino-acid sequence of human tyrosinase from their amino acid composition and related changes in human tyrosinase variant causing oculocutaneous albinism [J]. Pediatr Grenzgeb, 2001, 40:153-166.

[18] Wu G,Yan SM. Random analysis of presence and absence of two-and three-amino-acid sequences and distributions of amino acids,two- and three-amino-acid sequences in bovine p53 protein[J]. Mol Biol Today, 2002,3:31-37.

[19] Wu G,Yan SM. Analysis of distributions of amino acids in the primary structure of tumor suppressor p53 family according to the random mechanism[J]. J Mol Model,2002,8:191-198.

[20] Wu G,Yan S. Randomness in the primary structure of protein: methods and implications [J]. Mol Biol Today,2002,3:55-69.

[21] Wu G,Yan S. Determination of amino acid pairs sensitive to variants in human low-density lipoprotein receptor precursor by means of a random approach[J]. J Biochem Mol Biol Biophys,2002,6:401-406.

[22] Wu G,Yan SM. Analysis of distributions of amino acids in the primary structure of apoptosis regulator Bcl-2 family according to the random mechanism[J]. J Biochem Mol Biol Biophys,2002,6:407-414.

[23] Wu G,Yan SM. Estimation of amino acid pairs sensitive to variants in human phenylalanine hydroxylase protein by means of a random approach [J]. Peptides,2002,23:2085-2090.

[24] Wu G,Yan S. Analysis of amino acid pairs sensitive to variants in human collagen α5(Ⅳ) chain precursor by means of a random approach [J]. Peptides,2003,24: 347-352.

[25] Wu G,Yan S. Determination of amino acid pairs sensitive to variants in human β-glucocerebrosidase by means of a random approach[J]. Protein Eng,2003, 16:195-199.

[26] Wu G,Yan SM. Determination of amino acid pairs in

human haemoglobulin α-chain sensitive to variants by means of a random approach [J]. Comp Clin Pathol, 2003,12: 21-25.

[27] Wu G, Yan S. Determination of amino acid pairs sensitive to variants in human Bruton's tyrosine kinase by means of a random approach [J]. Mol Simul, 2003, 29: 249-254.

[28] Wu G, Yan S. Determination of amino acid pairs sensitive to variants in human coagulation factor IX precursor by means of a random approach [J]. J Biomed Sci, 2003, 10: 451-454.

[29] Wu G, Yan S. Determination of amino acid pairs in human p53 protein sensitive to mutations/variants by means of a random approach [J]. J Mol Model, 2003, 9: 337-341.

[30] Wu G, Yan S. Prediction of amino acid pairs sensitive to mutations in the spike protein from SARS related coronavirus [J]. Peptides, 2003, 24: 1837-1845.

[31] Wu G, Yan S. Determination of amino acid pairs in Von Hippel-Lindau disease tumour suppressor (G7 protein) sensitive to variants by means of a random approach [J]. J Appl Res, 2003, 3: 512-520.

[32] Wu G, Yan S. Potential targets for anti-SARS drugs in the structural proteins from SARS related coronavirus [J]. Peptides, 2004, 25: 901-908.

[33] Wu G, Yan S. Fate of 130 hemagglutinins from different influenza A viruses [J]. Biochem Biophys Res Commun, 2004, 317: 917-924.

[34] Wu G, Yan S. Amino acid pairs sensitive to variants in human collagen α-1(Ⅰ) chain precursor [J]. EXCLI J, 2004, 3: 10-19.

[35] Wu G, Yan S. Determination of amino acid pairs sensitive to variants in human copper-transporting ATPase 2 [J]. Biochem Biophys Res Commun, 2004, 319: 27-31.

[36] Wu G, Yan S. Susceptible amino acid pairs in variants of human collagen α-1(Ⅲ) chain precursor [J]. EXCLI J, 2004, 3: 20-28.

[37] Wu G, Yan S. Determination of sensitive positions to mutations in human p53 protein [J]. Biochem Biophys Res Commun, 2004, 321: 313-319.

[38] Wu G, Yan S. Reasoning of spike glycoproteins being more vulnerable to mutations among 158 coronavirus proteins from different species [J]. J Mol Model, 2005, 11: 8-16.

[39] Wu G, Yan S. Prediction of mutation trend in hemagglutinins and neuraminidases from influenza A viruses by means of cross-impact analysis [J]. Biochem Biophys Res Commun, 2005, 326: 475-482.

[40] Wu G, Yan S. Amino acid pairs susceptible to variants in human protein C precursor [J]. Protein Pept Lett, 2005, 12: 491-494.

[41] Wu G, Yan S. Timing of mutation in hemagglutinins from influenza A virus by means of unpredictable portion of amino-acid pair and fast Fourier transform [J]. Biochem Biophys Res Commun, 2005, 333: 70-78.

[42] Wu G, Yan S. Mutation features of 215 polymerase proteins from different influenza A viruses [J]. Med Sci Monit, 2005, 11: BR367-BR372.

[43] Wu G, Yan S. Determination of mutation trend in proteins by means of translation probability between RNA codes and mutated amino acids [J]. Biochem Biophys Res Commun, 2005, 337: 692-700.

[44] Wu G, Yan S. Searching of main cause leading to severe influenza A virus mutations and consequently to influenza pandemics/epidemics [J]. Am J Infect Dis, 2005, 1: 116-123.

[45] Gao N, Yan S, Wu G. Pattern of positions sensitive to mutations in human haemoglobin α-chain [J]. Protein Pept Lett, 2006, 13: 101-107.

[46] Wu G, Yan S. Timing of mutation in hemagglutinins from influenza A virus by means of amino-acid distribution rank and fast Fourier transform [J]. Protein Pept Lett, 2006, 13: 143-148.

[47] Wu G, Yan S. Fate of influenza A virus proteins [J]. Protein Pept Lett, 2006, 13: 377-384.

[48] Wu G, Yan S. Mutation trend of hemagglutinin of influenza A virus: a review from computational mutation viewpoint [J]. Acta Pharmacol Sin, 2006, 27: 513-526.

[49] Wu G, Yan S. Determination of mutation trend in hemagglutinins by means of translation probability between RNA codons and mutated amino acids [J]. Protein Pept Lett, 2006, 13: 601-609.

[50] Wu G, Yan S. Prediction of possible mutations in H5N1 hemagglutinins of influenza A virus by means of logistic regression [J]. Comp Clin Pathol, 2006, 15: 255-261.

[51] Wu G, Yan S. Prediction of mutations in H5N1 hemagglutinins from influenza A virus [J]. Protein Pept Lett, 2006, 13: 971-976.

[52] Wu G, Yan S. Improvement of model for prediction of hemagglutinin mutations in H5N1 influenza viruses with distinguishing of arginine, leucine and serine [J]. Protein Pept Lett, 2007, 14: 191-196.

[53] Wu G, Yan S. Improvement of prediction of mutation positions in H5N1 hemagglutinins of influenza A virus using neural network with distinguishing of arginine, leucine and serine [J]. Protein Pept Lett, 2007, 14: 465-470.

[54] Wu G, Yan S. Prediction of mutations in H1 neuraminidases from North America influenza A virus engineered by internal randomness [J]. Mol Divers. 2007, 11: 131-140.

[55] Wu G, Yan S. Prediction of mutations engineered by randomness in H5N1 neuraminidases from influenza A virus [J]. Amino Acids, 2007, 34: 81-90.

[56] Wu G, Yan S. Prediction of mutations initiated by internal power in H3N2 hemagglutinins of influenza A virus from North America [J]. Int J Pept Res Ther, 2008, 14: 41-51.

[57] Wu G, Yan S. Prediction of mutation in H3N2 hemagglutinins of influenza A virus from North America based on different datasets [J]. Protein Pept

Lett,2008,15: 144-152.
[58] Wu G,Yan S. Building quantitative relationship between changed sequence and changed oxygen affinity in human hemoglobin β-chain[J]. Protein Pept Lett,2008,15: 341-345.
[59] Wu G,Yan S. Three sampling strategies to predict mutations in H5N1 hemagglutinins from influenza A virus[J]. Protein Pept Lett,2008,15: 731-738.
[60] Wu G,Yan S. Prediction of mutations engineered by randomness in H5N1 hemagglutinins of influenza A virus[J]. Amino Acids,2008,35: 365-373.
[61] Yan S,Wu G. Quantitative relationship between mutated amino-acid sequence of human copper-transporting ATPases and their related diseases[J]. Mol Divers, 2008,12: 119-129.
[62] Yan S,Wu G. Quantitative relationship between mutated structure of human glucosylceramidase and Gaucher disease status[J]. Int J Pept Res Ther,2008, 14: 263-271.
[63] Yan S,Wu G. Connecting mutant phenylalanine hydroxylase with phenylketonuria [J]. J Clin Monit Comput,2008,22: 333-342.
[64] Yan S,Wu G. Connecting KCNQ1 mutants with their clinical outcomes[J]. Clin Invest Med,2009,32: E28-E32.
[65] Yan S,Wu G. Determination of mutation pattern in human androgen receptor by means of amino-acid pair predictability [J]. Protein Pept Lett, 2009, 16: 289-296.
[66] Yan S,Wu G. Determination of mutation patterns in human ornithine transcarbamylase precursor[J]. J Clin Monit Comput,2009,23: 51-57.
[67] Yan S,Wu G. What these trends suggest[J]. Am J Appl Sci,2009,6: 1116-1121.
[68] Yan S,Wu G. Prediction of mutation position,mutated amino acid and timing in hemagglutinins from North America H1 influenza A virus[J]. J Biomed Sci Eng, 2009,2: 117-122.
[69] Yan S,Wu G. Descriptively probabilistic relationship between mutated primary structure of von Hippel-Lindau protein and its clinical outcome[J]. J Biomed Sci Eng,2009,2: 190-109.
[70] Yan S, Wu G. Mutation patterns in human menin [M]. IEEE Xplore,2009,ISBN: 978-1-4244-2902-8.
[71] Yan S,Wu G. Describing evolution of hemagglutinins from influenza A viruses using a differential equation [J]. Protein Pept Lett,2009,16: 794-804.
[72] Yan S,Wu G. Descriptively quantitative relationship between mutated N-acetylgalactosamine-6-sulfatase and mucopolysaccharidosis IVA[J]. Biopolymers: Pept Sci,2009,92:399-404.
[73] Yan S,Wu G. Coupling of mutations with their clinical outcomes in antithrombin Ⅲ [J]. J Guangxi Acad Sci, 2009,25:183-186.
[74] Yan S,Wu G. Determination of inter- and intra-subtype/species variations in polymerase acidic protein from influenza A virus using amino-acid pair predictability[J]. J Biomed Sci Eng,2009,2:273-279.
[75] Yan S,Wu G. Mutation patterns in human α-galactosidase A[J]. Mol Divers,2010,14:147-154.
[76] Yan S,Wu G. Trends in global warming and evolution of matrix protein 2 family from influenza A virus[J]. Interdiscip Sci: Comput Life Sci,2009,1:272-279.
[77] Yan S,Wu G. Trends in global warming and evolution of polymerase basic protein 2 family from influenza A virus[J]. J Biomed Sci Eng,2009,2:458-464.
[78] Wu G,Yan S. Translation probability between RNA codons and translated amino acids,and its applications to protein mutations. In: Leading-Edge Messenger RNA Research Communications[M]. New York: ed. Ostrovskiy M H Nova Science Publishers, 2007, Chapter 3:47-65.
[79] Wu G,Yan S. Lecture notes on computational Mutation[M]. New York: Nova Science Publishers, 2008.
[80] Fasman GD. Handbook of biochemistry: section D physical chemical Data[M]. 3rd ed. London and New York: CRC Press,1976.
[81] Poincaré H. Science and hypothesis[M]. London and Newcastle-on-Cyne: Walter Scott,1902.
[82] Everitt BS. Chance Rules: an informal guide to probability, risk, and statistics [M]. New York: Springer,1999.
[83] Van der Lubbe JCA. Information theory[M]. Cambridge:Cambridge University Press,1997.
[84] Feller W. An introduction to probability theory and its applications[M]. 3rd ed. New York: Wiley,1968: 34-40.

(责任编辑:尹 闯)

(上接第129页)

3 结束语

本研究结果显示菌草无粪栽培蘑菇和常规栽培蘑菇的营养成分和营养价值并无本质区别,在营养成分的某些指标上菌草无粪栽培的蘑菇还略高于常规栽培的蘑菇。在资源利用上菌草无粪栽培蘑菇占有很大优势,而且可以避免农药残留等的困扰,食用更加安全卫生。菌草无粪栽培蘑菇是可行的,开辟了蘑菇栽培新的原料途径。

参考文献:

[1] 林占熹,林辉. 菌草学[M]. 北京:中国农业科学出版社,2003:124-132.
[2] 林树钱. 中国药用菌生产与产品开发[M]. 北京:中国农业出版社,2000:211-217.

(责任编辑:韦廷宗)