Characterization of Enzyme Structure by Informational Complexity

Takayuki Kamei
kamei@calla.dse.ibaraki.ac.jp

Yasuo Yonezawa [1]
yonezawa@life01.dse.ibaraki.ac.jp

Division of Systems Engineering, Graduate School of Science and Engineering, Ibaraki University
4-12-1 Nakanarusawa-cho, Hitachi, Ibaraki, 316, Japan

Abstract

It is well-known that enzymes is very important as reaction factor in life systems activity. But the properties based on information theory are not yet enough in biological studies. Then, we examined correlation the complexity at amino acid sequences with its function of Enzymes by informational measure, in order to elucidate the informational properties of sequence structure. Also, power spectrum of enzyme complexity are obtained specific profile by Fourier Transform(FT) method. At results, correlation at sequence complexity, the sequence of enzyme Proteins are given complexity more than non-enzyme Proteins. Moreover, FT profile are given typical pattern at complexity of enzyme Protein sequences. This result are suggested that the new view-point for Protein analysis by information Science.

1 Introduction

It is important to estimate the properties of the sequence at the view-point of informational methods. For the furthermore study of properties at sequential description of Proteins, we are evaluated with complexity and correlation of complexity in sequences. In this paper, we are examined with adoption the informatical measure known as Shannon's Informatical theory to analyze the sequential complexity. That is a effective method for dealing with the distribution of characteristic informations represented as sequences. By this method, we evaluated the differences of complexity between functional and non-functional Proteins. At the previous research in Genome Informatics, the correlation of DNA sequence are examined by FT(Fourier Transform). Also, informatical properties concerned with Protein function is detected by FT for the elucidate of inter-relation with Complexity and Protein action. If the correlation was examined, the wave slope of power spectrum shows one of the specific profile. Therefore, we are tried with the correlation analysis with amino acid sequences for Complexity properties by FT.

2 Methods

We calculated the complexity of sequential information according to following Shannon's Entropy equation,

$$S = -\sum_{i=1}^{k} p_i log_2 p_i$$

where, $p_i = n_i/N$ represent appearance probabilities of i-th type amino acid at given position. Also, term n_i is the total number of i-th type amino acid and the term N is the length of sequence at given position. In case of Proteins, term k is up to 20. According to appear amino acids close to all types equally, the complexity measure S go in the direction of high variety. On the other hand, the smallest varied state of sequences is occurred when the given position is occupied by single type amino acid, then p_i equal to 1/1. Then, plotted the complexity term S correspond to each position and estimated complexity differences between Enzymes and non-Enzymes.

[1] All correspondence to Yasuo Yonezawa

3 Results & Discussion

3.1 Complexity of amino acid sequences & Power spectrum profile

The result was shown in Fig.1(a),(b). From this result, complexity of Enzymes((1)Catalase bovine) appear at higher region than non-Enzymes((2)Collagen Alpha 1(I) bovine). It suggest that the sequential structure of Enzymes are constructed by including randomness at informatical properties. Also, the understanding point here is that these variable state in sequences could be necessary for expression of Protein function. In case of non-Enzymes, the complexity was relatively low region because of its order properties in sequence level.

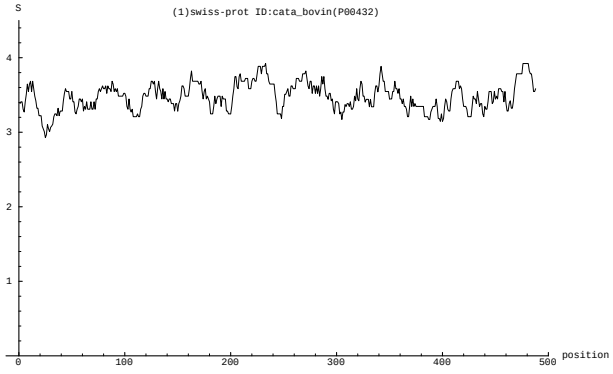

Fig.1(a)Examples of fluctuation of complexity

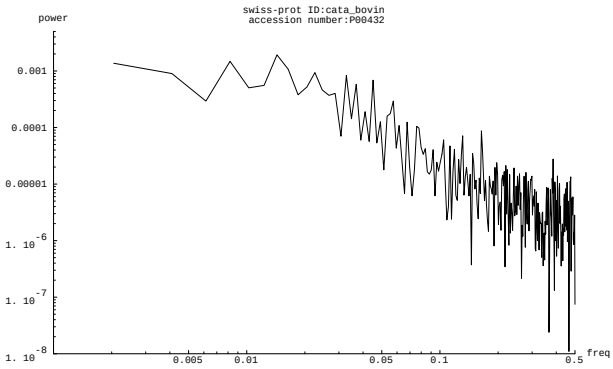

Fig.1(b)Power spectrum of fluctuation of complexity

Power spectrum which is FT of autocorrelation function was used in order to detect the profile of correlations in complexity fluctuation. If the correlations is in existence between its fluctuations, the slope of power spectrum indicate inverse power-law ,that is, $1/f^{\alpha}$ spectra. On the other hand, the slope approximates to white noise(α=0) if the correlation is negative. Namely, the exponent α represent strength of correlations. The FT equation is as follows,

$$S(f) = \left|\frac{1}{N}\sum_{i=1}^{N} fi \cdot e^{-j(2\pi/N)ki}\right|^2$$

where, the frequency:f=k/N, N:the total number of sample data, k=1,2,3, ...,N/2, f_i:i-th data
From the results(Fig.1(b)), typical pattern of peaks at high frequency region are obtained in non-Enzymes. But in Enzymes, spectra behaves as gentle slope without specific peak. Namely, informational properties of Enzymes and non-Enzymes are evaluated by detecting the typical pattern of complexity correlations. And it is understood the complexity of Protein sequences are depended on the function created with structural form like Enzymes. Therefore, we are mentioned description as below.In the first understanding point, the complexity of sequences at Enzymes are more large than at non-Enzymes. In second point, the sequential arrangement of Protein complexity are obtained randomness as non-cyclic order patterns. This suggest that the power spectrum of complexity under non-Enzyme's sequences indicate white-noise like spectra at low frequency region and also the specific peaks was not obtained at Enzymes. These understanding points suggest the necessary on the Informational complexity for creation of function at Enzymes. Concerned with above understandings, the complexity included the two order information(Specific Protein pattern) are given by these examinations.

References

[1] Wentian Li,Thomas G Marr and Kunihiko Kaneko,"Understanding Long-range correlation in DNA sequences", *Physica D* ,Vol 75,pp392-416,1994.

[2] Peter S.Shenken,Batu Erman and Lucy D Mastrandrea", Information-Theoretical Entropy as a measure of Sequence Variability," *PROTEINS:Structure,Function and Genetics*,Vol11,pp297-313,1991.