International Journal of Modern Electronics and Communication Engineering (IJMECE) ISSN: 2321-2152
Volume No.-7, Issue No.-1, January, 2019

Determination of Characteristic Period in Proteins using Ramanujan Fourier Transform

Abhishek Panchal
Electronics and Communication Engineering
Samrat Ashok Technological Institute
Vidisha, India
abhishekpanchal048@gmail.com

Abstract: Resonant Recognition Model (RRM) plays an important role in the field of genomic signal processing. Identification of protein-target binding sites in proteins using resonant recognition model requires the knowledge of characteristic frequency. For a successful protein-protein or protein-target interaction, both the protein and the target (protein) must share the same characteristic frequency. The characteristic frequency of a functional group of proteins is determined from the consensus spectrum obtained using DFT. In this work an approach for identification of characteristic period using Ramanujan Fourier Transform (RFT) is described. The characteristic period of a functional group of proteins is determined from the consensus spectrum obtained using RFT.

Keywords: Protein, Electron-Ion Interaction Potential, Resonant Recognition Model, Ramanujan Fourier transform, characteristic period.

I. INTRODUCTION

Genomic Signal Processing (GSP) is a discipline of signal processing that deals with the processing of genomic signals. The aim of GSP is to integrate the genomic theory and signal processing methods with the global understanding of genomics, placing special emphasis on genomic regulation. In the 21st century it is believed that many significant scientific and technological endeavors will be related to the processing and interpretation of the vast information that is currently revealed from genome sequence of many living organisms, including humans [1].

Proteins are the main building blocks of any living organism. A protein consists of many small, linked components called amino acids. In proteins there are 20 possible types of amino acids and these 20 amino acids are represented in a protein sequence as a string of alphabetical symbols with length ranging from 100 to 10000 [1]. They play an important role in body functioning as catalysts accelerating chemical reactions, as carrier and storage molecules in muscle contractions and as receptors in the nervous system generating and transmitting nerve signals or impulses. These cellular processes and their functions are largely governed by different types of protein-protein interactions or with its target [2]. The biological function of protein can be expressed by means of its three dimensional (3-D) structure. The 3-D shape of protein allows it to interact with other molecules known as targets whereas these interactions are very selective in nature. Many studies on protein interaction revealed that energies are not uniformly distributed. Instead, there are certain critical residues which are known as hot spots comprising only a small fraction of interfaces [3]. The recognition of the importance of characterizing protein interactions in a cell has provided the development of experimental and computational techniques to detect and predict these protein-target interactions or hotspots with an objective to produce more new efficient medicines and other biotechnological products.

For a newly discovered protein molecule, the only information initially available is its amino acids sequence [4]. Hence, the Digital Signal Processing (DSP) based methods play an important role in the analysis of these sequences [5] [6]. DSP methods do not need any structural information or training for detection of protein-protein interaction and they only use primary amino-acid sequence [7][8].

All the reported signal processing methods first extract the characteristic frequency using Resonant Recognition Model (RRM) and then apply the DSP algorithms. The DSP-based method reported by I.Cosic [4], provides the following criteria with respect to the characteristic frequency these are as follows:

1) For the same biological function only one peak exists for a group of protein sequences.

2) For biologically unrelated protein sequences no significant peak exists.

3) For different biological functions peak frequencies are different.

The protein-protein interaction in proteins can be identified by the use of Resonant Recognition Model (RRM), which correlates the biological functioning of the protein with the characteristic frequencies. These protein-protein interactions in proteins can be localized where the characteristic frequencies of the functional groups are dominant [7-9]. The digital signal processing techniques can be used to extract these characteristic frequencies in the protein sequences which are primarily based on the amino acid sequence information only

[10]. In the earlier reported works [8] [9] [10], Discrete Fourier Transform (DFT) [11] and Power Spectral Density (PSD) [12] have been used to determine the characteristic frequency of the protein families. In this work determination of characteristic period using Ramanujan Fourier Transform (RFT) is proposed. The rest of the paper is organized as follows. Section II describes the RRM. Section III describes RFT. In Section IV the proposed scheme of characteristic period determination using RFT is discussed. Simulation results are presented in Section V. Finally in section VI the paper is concluded.

II. RESONANT RECOGNITION MODEL

Proteins complete their biological function by interacting with other molecules known as targets and these interactions are very selective in nature. The specificity of the interaction is due to distinctive three dimensional (3-D) structures of protein molecules. For a successful protein-target interaction both protein and target must share the same characteristic frequency but with opposite in phase [4]. The concept of characteristic frequency corresponds to the fact that a peak which is present in energy distribution periodicity of the protein molecule must be matched with a corresponding trough in energy distribution periodicity of the target molecule and vice-versa. This matching of energy distribution periodicity resembles resonance and hence the model is termed as Resonant Recognition Model (RRM). On the basis of resonant recognition model we can predict that whether a particular protein will interact with arbitrary target molecule by examining whether or not the protein and target share a common characteristic frequency.

Characteristic frequency in proteins can be determined using Digital Signal Processing (DSP) techniques. For the application of DSP techniques, the protein character sequences in proteins are needed to be mapped into the numerical sequences. The choice of numerical mapping is based on some physical property that is relevant to biological function of the amino acids. A successful attempt to assign numerical values to amino acid has been made in [13], in which each amino acid is assigned by a numerical value called its electron-ion interaction potential (EIIP). The electron-ion interaction potential is a physical property of amino acid which denotes the average energy of valence electron in the amino acids and is known to co-relate well with the proteins biological properties [14]. The EIIP values for 20 different amino acids are listed in Table 1.

Table1. EIIP Valuesfor 20 Different amino acids

S.NO.	Amino-acid Name	EIIP Values
1	Leucine	0.0000
2	Isoleucine	0.0000
3	Asparagine	0.0036
4	Glycine	0.0050
5	Valine	0.0057
6	Glutamic acid	0.0058
7	Proline	0.0198
8	Histidine	0.0242
9	Lysine	0.0371
10	Alanine	0.0373
11	Tyrosine	0.0516
12	Tryptophan	0.0548
13	Glutamine	0.0761
14	Methionine	0.0823
15	Serine	0.0829
16	Cystrine	0.0829
17	Threonine	0.0941
18	Phenylalanine	0.0946
19	Arginine	0.0959
20	Asparatic acid	0.1263

III. RAMANUJAN FOURIER TRANSFORM

The classical Discrete Fourier Transform (DFT) is a traditional method of frequency spectrum analysis and it can also reveal the periodicity of the signals, it is represented as:

$$X(k) = \sum_{n=1}^{N} x(n) \exp\left(-\frac{i2\pi kn}{N}\right), k = 1,2,3, \ldots, N \quad (1)$$

Here, the basis function is represented by the exponential term and are obtained as multiples of a basis frequency (1/N).

The famous Indian mathematician SrinivasaRamanujan introduced a summation which is represented in trigonometric form, now known as the Ramanujan sum$C_q(n)$[15]. This summation is used as basis function in the Ramanujan Fourier transform.

This Ramanujan sum has the form

$$C_q(n) = \sum_{\substack{k=1 \\ (k,q)=1}}^{q} e^{j2\pi kn/q} \quad (2)$$

In this representation (k,q) represents the greatest common divisor (gcd) of k and q. Thus (k,q) = 1 means that k and q are co-prime numbers. For example if q = 10, then co-prime values of k are k = 1, 3, 7 and 9 so that

$$C_{10}(n) = e^{j2\pi n/10} + e^{j6\pi n/10} + e^{j14\pi n/10} + e^{j18\pi n/10} \quad (3)$$

Ramanujan sum $C_q(n)$is a real, symmetric, and periodic sequence in n. Here are the first few Ramanujan sequences, shown for one period $0 \le n \le (q-1)$.

c_1 (n) = 1.

c_2 (n) = 1, –1.

c_3 (n) = 2, –1, –1.

c_4 (n) = 2, 0, –2, 0.

c_5 (n) = 4, –1, –1,–1,–1.

c_6 (n) = 2, 1, –1, –2, –1, 1.

c_7 (n) = 6, –1, –1, –1, –1, –1, –1.

c_8 (n) = 4, 0, 0, 0, – 4, 0, 0, 0.

c_9 (n) = 6, 0, 0, 0, – 3, 0, 0, 0,-3, 0, 0.

c_{10} (n) = 4, 1, –1, 1, –1, – 4, –1, 1, –1, 1.

Notice that $C_q(n)$is always integer-valued function [15].

The Ramanujan sum has period q in terms of n, the quantity $C_q(n)$ is always integer valued, which is an attractive property of it. Therefore, this inspires signal processing researchers to use it for the detection of periodic components in a sequence. The most important point is that this tool is useful in extracting hidden periods in finite duration signal.

SrinivasaRamanujan's motivation in introducing this summation was to show that in the theory of numbers several arithmetic functions can be expressed as linear combination of Ramanujansum $C_q(n)$, that is

$$x(n) = \sum_{q=1}^{\infty} R(q)C_q(n), \quad n \geq 1. \qquad (4)$$

An arithmetic function is an infinite sequence defined for 1≤ n≤ ∞ which may be integer valued , for example the Mobius function μ(n), Euler totient function ϕ (n), the von Mangoldt function Λ(n), and the Rieman zeta function ζ(n) [16].

In the above equation $R(q)$represents the RFT coefficients given by

$$R(q) = \frac{1}{\phi(q)} \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} x(n)\ C_q(n) \quad q=1, 2, ., N \qquad (5)$$

This equation is known as Ramanujan Fourier Transform (RFT) [17]. Here, $\phi(q)$denotes the Euler totient function and $C_q(n)$ denotes Ramanujan sums.

IV. Determination of Characteristic Period Using RFT

Previously lots of attempts were made for determination of characteristic frequency using DFT. Here propose an approach using RFT. Step by step procedure for determination of characteristic Period for proteins using RFT is given below.

(1) Select any two proteins from the functional group of interest.

(2) Convert amino acid character sequences into numerical sequences using EIIP values.

(3) Determine RFT of numerical sequences obtained in the step (2) and compute consensus spectrum by multiplying them.

R (q) = $|R_1(q)| * |R_2(q)|$

(4) If a distinct peak is observed in the consensus spectrum, R (q), record the corresponding period as the characteristic period.

(5) If the peak present in the consensus spectrum is not distinct, increase the number of protein from family function group in step 1 and repeat steps 1 to 4 till a distinct characteristic period is obtained.

V. Results

Functional groups of proteins were selected from Swiss-Prot (UniProt) [18] to demonstrate the performance of the proposed approach and are described in table 2. The characteristic period of fibroblast growth factor (FGF) family in which nine protein sequences are used to draw its consensus spectrum and is shown in figure 1.

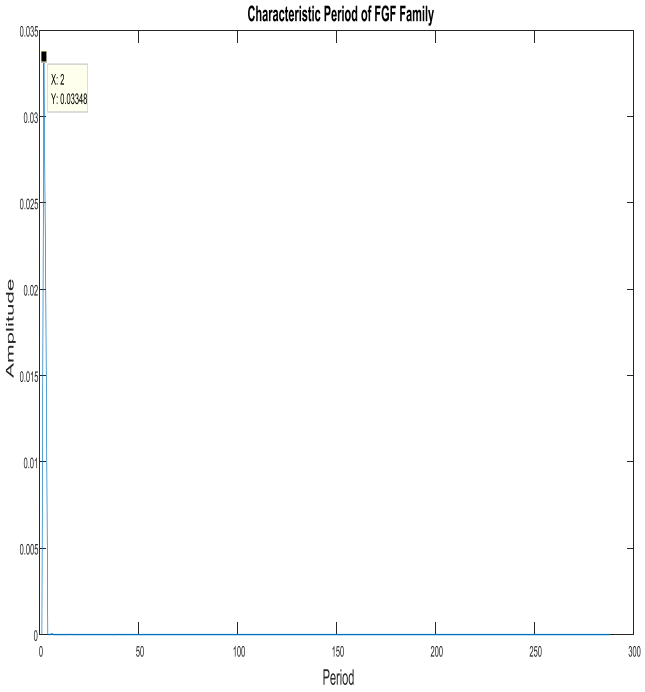

Figure1. Characteristic Period of FGF family

As shown in figure 1 the peak present at position 2 indicate the characteristic period of FGF family.Table 2 provides the details of different proteins along with their protein ID andcharacteristic periodsof different protein families that are used in this work.

Table2. Different Protein families with their ID's and characteristic period

Protein Name	No. of Seq.	Swiss-port ID	Char. period
Fibroblast Growth Factor (FGF)	9	P09038,P05230,P15656, P55075,O15520,O54769, Q9EPC2,Q9HCT0, Q9QY10	2
Human Growth Hormone (HGH)	8	P16882,P10912,P16310, P19941,Q9JI97,Q9TU69, Q02092,O46600	3
Human Growth Hormone Binding (HGH Binding)	6	P79194,P79108,Q9XSZ1, Q95JF2, Q95ML5,Q28575	24
TRAP	6	Q2RHB9,P19466,P48064, Q8EQB3, C5D3E7,Q9X6J6	6
Colicin-E9 Immunity (IM9)	3	P13479,P15176,B9VMA0	6
Human Alpha Hemo-globin	9	P68048,P68050,P68871, P69905,P01942,P01946, P01958,P02062,P60524	2

VI. CONCLUSION

I applied RFT a methodology which determine characteristic period of protein family by using amino acid sequence. I applied this methodology to different protein sequences having different lengths. Previously only characteristic frequency can be determined but in this work characteristic period of the protein families can be obtained. In summary, RFT is observed to be a promising approach to determine characteristic period of different proteins in human genome. In future the relation between characteristic frequency and characteristic period may also provide some other characteristic of proteins.

ACKNOWLEDGMENT

I am thankful to the Department of Electronics and Communication Engineering of SATI, Vidisha, India for providing the necessary facilities for the successful completion of this work.

REFERENCES

[1] Dimitris Anastassiou, "Genomic signal processing," IEEE Signal Processing Magazine, pp. 8–20, July 2001.

[2] P.Uetz, L.Giot, G.Cagney, T.A.Mansfield, and R.S.Judson, "A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae,"Nature, pp.623–627.

[3] Andrew A. Bogan and Kurt S. Thorn, "Anatomy of hot spots in protein interfaces," Journal of Molecular Biology,280, pp.1–9, 1998.

[4] I.Cosic, "Macromolecular bioactivity: is it resonant interaction between macromolecules? –Theory and applications, " IEEE Transaction on Biomedical Engineering, vol. 41, pp. 1101–1114, 1994.

[5] I.Cosic, "Analysis of HIV proteins using DSP techniques,"Proceedings of the 23rd Annual EMBS International Conference,Istanbul, Turkey, pp. 2886–2889,2001.

[6] E.Pirogova, Q. Fank, M. Akay and I. Cosic. "Investigation of structural and functional relationship of oncogene proteins," Proceeding of the IEEE, vol. 90, pp. 1859–1867, 2002.

[7] P.Ramachandran, A.Antoniou, and P.P.Vaidyanathan, "Identification and location of hotspots in proteins using the short time discrete Fourier transform," Proceeding 38th Asilomer Conference Signals, Syatems, Computers,pcific Grove,CA, pp. 1656–1660, 2004.

[8] P.Ramachandran and A. Antoniou, "Identification of hotspots locations in proteins using digital filters," IEEE Journel of Selected Topics in Signal Processing, vol. 2, No. 3, pp. 378–389, June 2008.

[9] S.S.Sahu and G.Panda, "Efficient localization of hotspots in proteins using a novel S-transform based filtering approach," IEEE Transaction on Computational Biology and Bioinformatics, vol. 8, no. 5, pp. 1235–1246, 2011.

[10] P. P. Vaidyanathan and B–J. Yoon, "The role of signalprocessing concepts in genomics and proteomics," Journal of the Franklin Institute,vol. 341, pp. 111-135, 2004.

[11] Stoica, P. and R.L. Moses,"Introduction to Spectral Analysis", Prentice-Hall, 1997, pp. 24-26.

[12] Yashpal Yadav, Sulochana Wadhwani, "Determination of Characteristic Frequency for Identification of Hot Spots in Proteins," International Journal of Electrical and Electronics Engineering (IJEEE) ,Volume-1 ,Issue-1 ,2011.

[13] K. D. RAO and M. N.S Swamy, "Analysis of Genomic and Proteomics Using DSP Technique", IEEE Transactions on Circuits and Systems-1: papers, Vol. 55, No. Regular 1, Feb 2008.

[14] J.Lazovic, "Selection of amino acid parameters for Fourier transform-based analysis of proteins," Computer Application Bioscience, vol. 12, no. 6, pp. 553–562, 1996.

[15] P.P.Vaidyanathan, "Ramanujan Sums in the context of Signal Processing – Part I: Fundamentals", IEEE Transactions on Signal Processing, Vol.62, No.16, Aug 2014.

[16] P.P.Vaidyanathan, "Ramanujan Sums in the context of Signal Processing – Part II: FIR representaion and applications", IEEE Transactions on Signal Processing, Vol.62, No.16, Aug 2014.

[17] Jian Zhao, Jiasong Wang, Wei Hua and Pingkai Ouyang, "Algorithm,applications and evaluation for protein comparision by Ramanujan Fourier Transform", ELSEVIER: Molecular and Cellular Probes, Aug. 2015.

[18] Swiss-Prot Protein Knowledgebase. Swiss Inst. Bioinformatics (SIB).[Online]. Available: http://us.expasy.org/sprot/.

AUTHOR'S BIOGRAPHY

Abhishek Panchal received his Bachelor of Engineering degree in Eletronics and Communication Engineering from Rajiv Gandhi Proudyogiki Vishwavidyalaya Bhopal, M.P., India, in 2014. He is pursuing M-Tech in Electronics and Communication Engineering from Samrat Ashok Technological Institute (SATI) Vidisha, M.P.,India. His research interests are Genomic signal processing, image processing and Bio signal processing.