International Journal of Electronics and Electical Engineering

Volume 1 | Issue 1 Article 7

July 2012

Determination of Characteristic Frequency for Identification of Hot Spots in Proteins

Yashpal Yadav
Department of Electrical Engineering Madhav Institute of Technology and Science Gwalior, India, yash.yadav85@gmail.com

Sulochana Wadhwani
Department of Electrical Engineering, Madhav Institute of Technology and Science , Gwalior - 474005 (MP), India, sulochana_wadhwani1@rediffmail.com

Determination of Characteristic Frequency for Identification of Hot Spots in Proteins

Yashpal Yadav, Sulochana Wadhwani

Department of Electrical Engineering
Madhav Institute of Technology and Science , Gwalior - 474005 (MP), India
E-mail: yash.yadav85@gmail.com , sulochana_wadhwani1@rediffmail.com

***Abstract*— Identification of hot spots or protein-target binding sites in proteins using resonant recognition model requires the knowledge of characteristic frequency. For a successful protein target interaction, both the protein and the target signals must share the same characteristic frequency. The common characteristic frequency of a functional group of proteins is determined from the consensus spectrum obtained using DFT. In this work an alternative approach for identification of characteristic frequency using power spectral density is described. The performance of the proposed method is observed to be better than the DFT-based approach and is illustrated using simulation examples**.

Keywords: proteins; resonant recognition model (RRM); hot spots; Discrete Fourier Transform (DFT); Power Spectrum Density (PSD).

I. INTRODUCTION

Proteins are the most important class of biochemical molecules and are the basis for the major structural components of animals and human tissue. They are made up of different combinations of 20 compounds called amino acids. An amino acid consists of a carboxylic acid group, an amino group and a variable side chain all attached to central carbon atom (also called α carbon). The side chain is the only component that varies from one amino acid to another. Thus the characteristic that distinguishes one amino acid from another is its unique side chain that dictates an amino acid's chemical property [1]. Depending upon which amino acids link together, protein molecules form enzymes, hormones, muscles, organs and many tissues in the body.

Biologists distinguish four levels of organization in structure of proteins [2]. Primary structure is a linear arrangement of amino acids in a protein and the location of covalent linkages such as disulfide bond between amino acids. Secondary structure consists of area of folding or coiling within a protein, examples include alpha helices and pleated sheets, which are stabilized by hydrogen bonding. The final three dimensional (3-D) structure of proteins results from a large number of non-covalent interactions between amino acids. The quaternary structure arises from non-covalent interaction that binds multiple polypeptides into a single larger protein.

Even though proteins can be imagined to be linear chains of amino acids, they are not present as linear chains in reality. They fold into complex three dimensional (3-D) structures forming weak non-covalent bond between their own atoms. It is this folding ability that enables them to perform extreme specific functions. The information necessary to specify the three dimensional (3-D) shape of proteins is contained in its amino acid sequence. The 3-D structure of proteins is most stable form which a protein can attain and this 3-D structure is due to certain specialized regions in proteins known as hot spots [3]. Proteins perform their biological function by interacting with other molecules known as targets and the necessary binding energy for this protein-target interaction is provided by hot spots. Hot spots are small groups of amino acids which provide functional stability to proteins, so that protein can efficiently bind with a target and thus can perform its biological function.

The hot spots in proteins can be identified by the use of Resonant Recognition Model (RRM) [4], which correlates the biological functioning of the protein to the characteristic frequencies. These hot spots in proteins can be localized where the characteristic frequencies of the functional groups are dominant [5-7]. The signal processing techniques can be used to extract these characteristic frequencies in the protein sequences which are primarily based on the sequence information only [8]. In the earlier reported works [4-8], Discrete Fourier Transform (DFT) [9] has been used to determine the characteristic frequency. In this work determination of characteristic frequency using Power Spectrum Density (PSD) [9] is proposed. The rest of the paper is organized as follows. Section 2 describes the RRM. In Section 3 the proposed scheme of characteristic frequency determination using PSD is discussed. Simulation results are presented in Section 4. Finally the paper is concluded in Section 5.

II. RESONANT RECOGNITION MODEL

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

The characteristic frequency in proteins can be determined using Digital Signal Processing (DSP) techniques. For the application of DSP techniques to proteins, the protein character sequences are needed to be mapped into numerical sequences. The choice of numerals 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 [10], where each amino acid is assigned a numerical value called its electron-ion interaction potential (EIIP). The electron-ion interaction potential of amino acid is a physical property denoting the average energy of valence electrons in the amino acids and is known to co-relate well with proteins biological properties. The EIIP values for 20 different amino acids are listed in Table 1.

TABLE 1

EIIP VALUES FOR 20 AMINO ACIDS

Amino acid	**EIIP value**	**Amino acid**	**EIIP value**
Leucine	0.0000	Tyrosine	0.0516
Isoleucine	0.0000	Tryptophan	0.0548
Asparagine	0.0036	Glutamine	0.0761
Glycine	0.0050	Methionine	0.0823
Valine	0.0057	Serine	0.0829
Glutamic acid	0.0058	Cysteine	0.0829
Proline	0.0198	Threoinine	0.0941
Histidine	0.0242	Phenylalanine	0.0946
Lysine	0.0371	Arginine	0.0959
Alanine	0.0373	Aspartic acid	0.1263

Thus each and every amino acid in a sequence can be represented by a unique number.

DFT of EIIP sequences of proteins belonging to a particular functional group reveal an unique spectral component shared by whole functional group having same biological functions. The common characteristic frequency of a functional group of K proteins can be computed by the cross-spectral function defined in (1) [4].

$$S(\omega) = |X_1(\omega)|\,|X_2(\omega)|\,|X_3(\omega)|\,\ldots..\,|X_K(\omega)| \quad (1)$$

Peak frequency in the consensus spectrum denotes the characteristic frequency for all the proteins analyzed. If a protein performs more than one biological function then there may be different characteristic frequency each corresponding to different biological functions. A new method for determination of characteristic frequency using PSD is described in Section 3. This method provides more distinct peak at the characteristic frequency with improved signal to noise (S/N) ratio.

III. DETERMINATION OF CHARACTERISTIC FREQUENCY USING POWER SPECTRAL DENSITY (PSD)

Previous attempts were made for determination of characteristic frequency using DFT [4]. Here we propose a similar approach using PSD. Step by step procedure for determination of characteristic frequency for proteins using PSD is given below.

(1) Select two proteins from the functional group of interest.
(2) Convert protein character sequences into numerical sequences using EIIP values.
(3) Determine PSD of numerical sequences obtained in step (2) and compute consensus spectrum by multiplying them.

$$P(\omega) = |X_1(\omega)|^2\ |X_2(\omega)|^2 \quad (2)$$

(4) If a distinct peak is observed in the consensus spectrum, P(ω), record the corresponding frequency as the characteristic frequency.
(5) If the peak in the consensus spectrum is not distinct, increase a protein in step 1 and repeat steps 1 to 4 till a distinct characteristic frequency is obtained.

IV. ILLUSTRATIVE EXAMPLES

Functional groups of proteins were selected from Swiss-Prot Protein Knowledgebase [11] & Protein Data Bank [12] to demonstrate the performance of the proposed approach and are described in Table-2. These proteins have been used in [6] and [7].

TABLE 2

PROTEINS OF FUNCTIONAL FAMILY USED FOR COMPUTATION OF CONSENSUS SPECTRUM

Protein Name	**Proteins of functional family used for computation of consensus spectrum**	
	Swiss-prot ID	**PDB ID**
Cytochrome C	P00025,P62894,P99999, P00008	
Human Hemoglobin	P60524,P01958,P02062, P68871,P69905,P68050, P01942,P01946,P68048	
Lysozyme	P04421,P67977,P00698, P61626,P16973	
Barnase	B7M0V1,C4ZK78,C6UF64, C6XRM1,C9NF27,D0KFB0, P00648,P10912	
FGF		1FGA, 1AFC

DFT based results are compared with the proposed PSD approach in Figs. 1 and 2 for Lysozyme and Barnase proteins respectively. In [4], Signal to noise (S/N) ratio for each peak is defined as a measure of similarity between sequences analyzed.

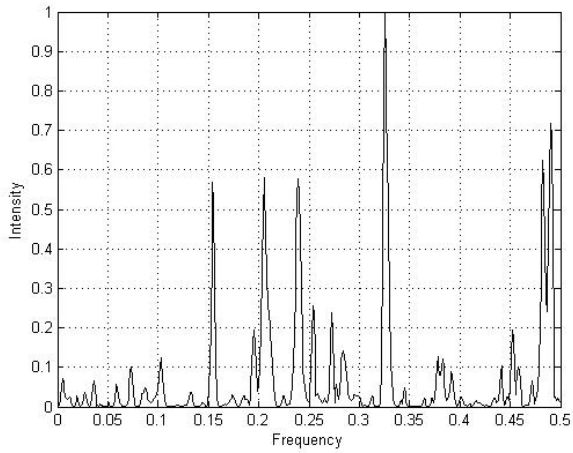

Figure 1 (a)

Consensus spectrum of Lysozyme using DFT

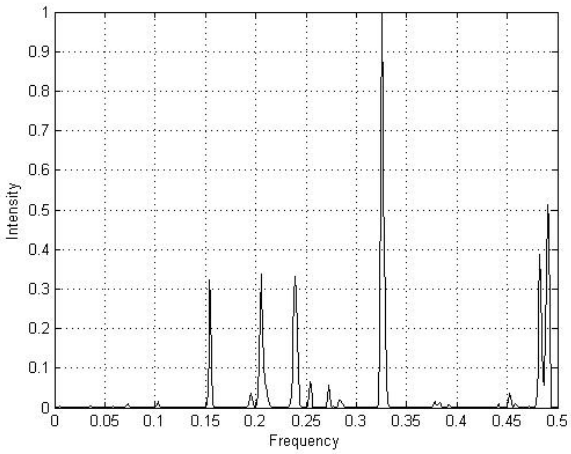

Figure 1 (b)

Consensus spectrum of Lysozyme using PSD

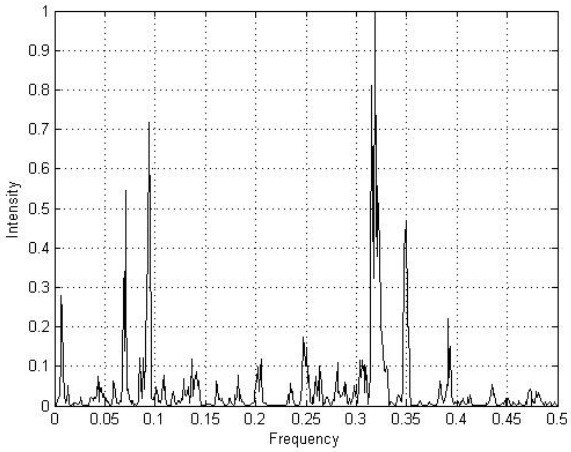

Figure 2 (a)

Consensus spectrum of Barnase using DFT

S/N has been calculated as ratio between signal intensity at the particular peak frequency and the mean value over the whole spectrum. In [4] S/N ratio of at least 20 is considered significant. S/N ratio for DFT and PSD approach for the proteins are compared. PSD approach clearly indicates a significant improvement in S/N ratio over DFT approach. The characteristic frequency obtained and number of sequences required for computation of consensus spectrum is shown in table 3. S/N ratio for both DFT and PSD approaches are also investigated in table, hence the result obtained clearly indicate that PSD approach is better than DFT based approach.

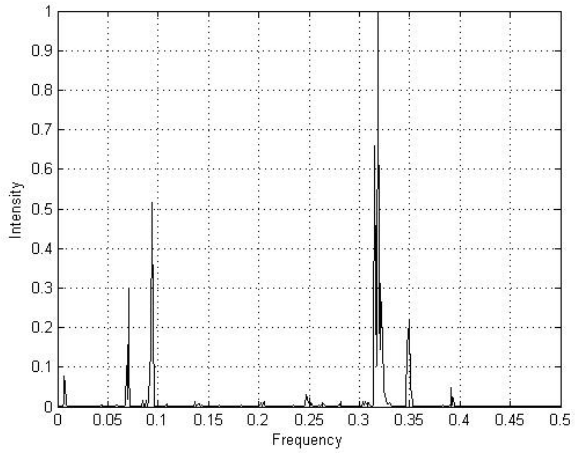

Figure 2 (b)

Consensus spectrum of Barnase using PSD

TABLE 3

THE PROTEIN SEQUENCES INVESTIGATED

S. No	Protein Name	No. of Seq. used	Charac-teristic Freque-ncy	S/N	
				DFT	PSD
1	Cytochrome C	4	0.472	27.61	92.64
2	FGF	2	0.452	7.17	28.78
3	Hemoglobin	9	0.023	96.18	179.2
4	Lysozyme	5	0.325	14.73	41.95
5	Barnase	8	0.318	22.08	81.84

V. CONCLUSION

A new approach for determination of characteristic frequency for hot spot identification is proposed using PSD. In this approach the unwanted peaks of different magnitudes present in the DFT method are suppressed. Also there is a considerable improvement in S/N ratio compared to DFT. Hence this approach can be very useful for correctly identifying the characteristic frequency which can be useful for hot spots detection.

ACKNOWLEDGEMENT

The authors are pleased to acknowledge the support by department of Electrical Engineering, Madhav Institute of technology and Science (MITS), Gwalior, MP, India for providing literature review and experimental facility.

REFRENCES

[1] B. Alberts, D. Bray, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P.Walter, *Essential Cell Biology,* Garland Publishing, New York, 1998.

[2] M. O Dayhoff, *Atlas of Proteins Sequences and Structures*, 5 th ed. Washington: Nat. Biomed. Res. Fund. 1978, suppl. 3. pp 345-352.

[3] A. A. Bogan and K. S. Thorn, "Anatomy of hot spots in protein interfaces," *Journal of Molecular Biology,* vol. 280, pp. 1-9, 1998.

[4] I. Cosic, "Macromolecular bioactivity: is it resonant interaction between macro-molecules? – theory and applications," *IEEE Trans. on Biomedical Engr.,* vol. 41, no. 12, pp. 1101-1114, Dec. 1994.

[5] P. Ramachandran, A. Antoniou, and P. P. Vaidyanathan, "Identification and location of hot spots in proteins using the short-time discrete Fourier transform," in Proc. 38th Asilomar Conf. Signals, Systems, *Computers*, Pacific Grove, CA, Nov. 2004, pp. 1656–1660.

[6] P. Ramachandran and A. Antoniou, "Identification and location of hot spots in proteins using digital filters," IEEE Journal of selected topics in signal processing. Vol .2, No. 3,june 2008.

[7] S. S. Sahu and G. Panda, *Efficient Localization of Hot spots in Proteins Using A novel S-Transform Based Filtering Approach*, accepted for publication in IEEE Transations on Computational Biology and Boinfromatics, doi 10.1109/TCBB.2010.109.

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

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

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

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

[12] Protein Data Bank (PDB), Research Collaboratory for Structural Bioinformatics (RCSB). [Online]. Available: http://www.rcsb.org/pdb/.

.