International journal of Advances in Electronics Engineering
Vol:1 Issue:1 ISSN 2278 - 215X

Identification of Characteristic Frequency in Proteins using Power Spectral Density

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

Dr. Sulochana Wadhwani
Reader: Department of Electrical Engineering
Madhav Institute of Technology and Science
Gwalior, India
E-mail:sulochana_wadhwani@rediffmail.com

***Abstract*— A technique for detection of characteristic frequency in proteins using power spectral density is described. The vast majority of proteins binds to other proteins at some time in their existence in order to perform various functions. Characteristic frequency is of utmost importance since it forms the basis for protein-target interaction, both protein and target must share common characteristic frequency for successful binding, thus an approach for identification of characteristic frequency using power spectral density (PSD) has been illustrated in this paper.**

***Keywords*—proteins,characteristic frequency, consensus spectrum, power spectrum density (PSD), electron-ion interaction potential (EIIP) , resonant recognition model (RRM).**

I. INTRODUCTION

The word protein comes from Greek language (prota) which means "of primary importance". This name was introduced by JonsJakob Berzelius in 1838. Proteins are the probably the most important class of biochemical molecules.Proteins are basis for the major structural components of animals and human tissue.Proteins are the building blocks of life and are essential for growth of cells and tissue repair. Protein is natural polymer molecule consisting of amino acid unit. All proteins are made up of different combination of 20 compound called amino acids. Depending upon which amino acid link together proteins molecules form enzymes, hormones, muscles, organs and many tissues in the body.

Proteins are polymers of amino acid joined together by peptide bond. There are 20 different amino acids that make up essentially all the proteins on earth. 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 distinguish one amino acid from another is its unique side chain that dictates an amino acid chemical property. [1]

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

Even though proteins can be imagined to be linear chain of amino acid, 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 enable 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.[2]

II. POWER SPECTRAL DENSITY (PSD)

Power spectral density shows the strength of variation (energy) as a function of frequency. In other words it shows at which frequency variations are strong and at which frequency variations are weak. Power spectral density describes how power of signal or time series is distributed with signal. PSD is very useful tool if you want to identify oscillatory signal in your time series data and want to know their amplitude. PSD is defined in terms of amplitude squared per hertz and thus is proportional to power delivered by signal in one-hertz band.

Computing PSD is done by computing autocorrelation function and then transforming it. Power spectral density or periodogram method is carried out by dividing the time signal into successive blocks and averaging the squared magnitude DFT's of signal block. Power spectral density for a sequence [x1,x2,…..xn] is given by formula

$$S_{(e^{jw})} = \frac{1}{2\pi N}\left|\sum_{n=1}^{N} xn.\, e^{-jwn}\right|^2 \quad (1)$$

w is in unit of radians per sample,
n is length of signal,
N is number of points in FFT.

III. RESONANT RECOGNITION MODEL

Proteins perform their biological function by interacting with other molecule known as target. These interactions are very selective in nature. The specificity of the interaction is by unique three dimensional (3-D) structure of protein molecules. These locations are known as active sites. For a successful protein-target interaction both protein and target must share the same characteristic frequency but have opposite phase. This 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 characteristic frequency is said to provide resonant recognition model (RRM) [9], [12], [15]. 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.

IV. ELECTRON ION INTERACTION POTENTIAL

Proteins are made of 20 amino acids and each amino acid is represented by a distinct alphabet, thus proteins can be represented by a character sequence. For the application of digital signal processing (DSP) to proteins, the protein character sequence needed to be mapped onto numerical sequences. The choice of numerals is base on some physical property that is relevant to biological function of amino acid. A successful attempt to assign numerical values to amino acid was made[3], [7] 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 physical property denoting the average energy of valence electrons [8] in the amino acids and is known to co-relate well with proteins biological properties.

The EIIP values for each amino acid was calculated using the following general model of pseudo potential [14]

$$(\overrightarrow{k+q}|w| = 0.25Z \sin(\pi\ 1.04Z)\ /(2\pi) \qquad (2)$$

Where q is change of momentum k of the delocalized electron in the interaction with w, while

$$Z = (\quad) /N \qquad (3)$$

where the number of valance electron of the *i*-th component of each amino acid and N is total number of atoms in amino acid.

The EIIP value for 20 different amino acid is listed in Table 1.Thus each and every amino acid in sequence can be represented by a unique number. Now digital signal processing algorithms or tools can be applied to the obtained numerical series.

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

V. DETERMINATION OF CHARACTERISTIC FREQUENCY

Previous successful attempts have been made for determination of characteristic frequency using fast Fourier transform (FFT) [7], [16]; here we propose a similar approach using power spectral density.

The first step for determination of characteristic frequency is selecting a protein functional group of interest. The number of proteins in a functional group may vary from case to case, suppose that we have M number of protein sequence in a functional group. The common characteristic frequency of a functional group of M proteins can be determined by first calculating the power spectral density (PSD) of all M proteins individually and then multiplying them in order to obtain consensus spectrum of group. The consensus spectrum of group has a peak at characteristic frequency. The number of protein sequence M required for a typical consensus spectrum varies from case to case. A sufficient number of protein sequences should be used to achieve a distinct peak at characteristic frequency in the consensus spectrum. Initially a set of two proteins sequence may be tried. If there is an ambiguity (there are several similar peaks in consensus spectrum), then one or more protein sequence from functional group of interest may be included in computation of consensus spectrum.

VI. ILLUSTRATIVE EXAMPLES

To demonstrate the power of proposed approach we show three distinct examples. We have chosen following protein sequences.

(1) Cytochrome C proteins.
(2) Lysozyme proteins.

Cytochrome C is a small heme proteinfound loosely associated with the inner membrane of mitochondrion. Cytochrome C is primarily known as electron carrying mitochondrial protein. The transition of cytochrome C

between the ferrous and ferric state within the cell makes it an efficient biological electron-transporter and it plays a vital role in cellular oxidation in both plants and animals. It is generally regarded as universal catalyst of respiration forming an essential electron-bridge between the respirable substance and oxygen.

Lysozyme, enzyme found in secretions (tears) of lacrimal gland of animals and in nasal mucus, gastric secretions, and egg white. Discovered in 1921 by Sir AlaxanderFleming,lysozyme catalyze the breakdown of certain carbohydrates found in cell wall of certain bacteria (eg. colic).

We have used four protein sequences for cytochrome C and five proteins sequence for lysozyme. The preliminary details pertaining to protein examples are shown in table 2.

VII. RESULTS AND DISCUSSIONS

For each examples, the characteristic frequency was determined from the consensus spectrum of sufficiently large set of protein sequences belonging to same functional group as the protein sequence of interest. The number of proteins sequences used to determine characteristic frequency is shown in table 2.

Figure 1(a) shows the graphical representation of corresponding numerical sequences obtained by replacing each amino acid with its EIIP values for cytochrome C tuna heart protein.
Figure 1(b) shows the power spectral density of cytochrome C tuna heart protein. There are many peaks hence it is impossible to clearly determine characteristic frequency.
Figure 1(c) shows the consensus spectrum for cytochrome C functional group using fast Fouriertransform (FFT) approach.
Figure 1(d) shows the consensus spectrum forcytochrome C functional group using power spectral density.
TABLE 2
PROTEINS DETAIL S FOR EXAMPLES

Example number	**Protein name**	**Sequence length**	**No. of seqs. used**	**Characteristic frequency**
1	Tuna cytochrome C	104	4	0.4727
2	Hen egg-white lysozyme	147	5	0.3255

A. Tuna Cytochrome C

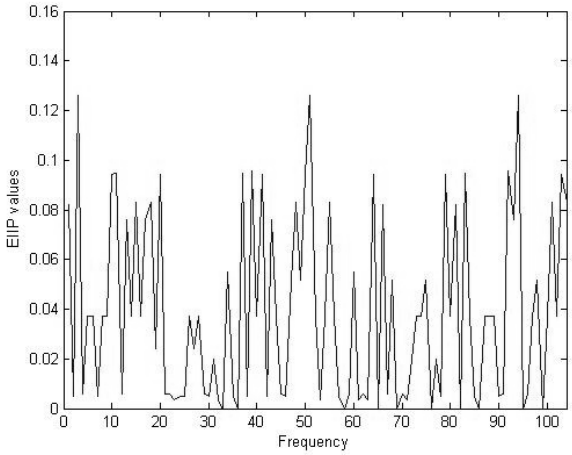

Figure 1(a)

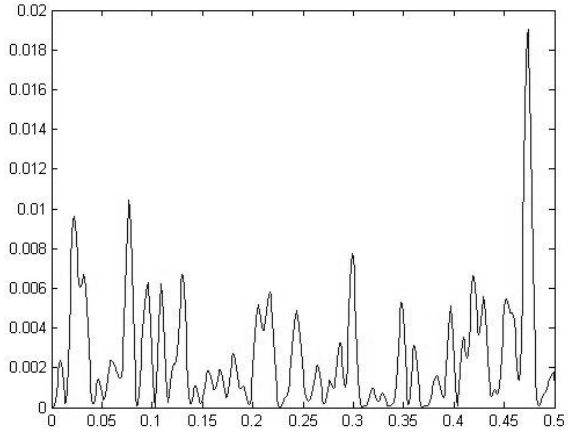

Figure 1(b)

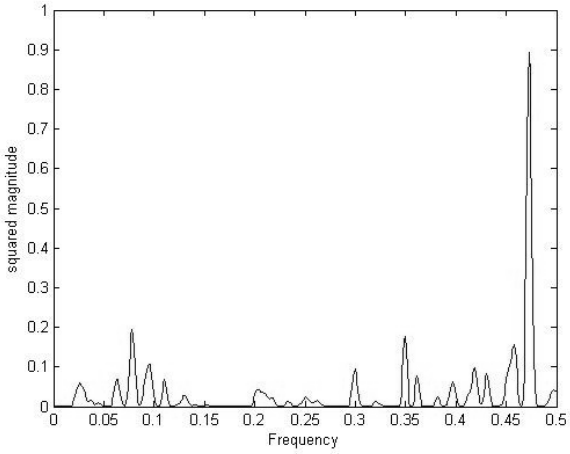

Figure 1 (c)

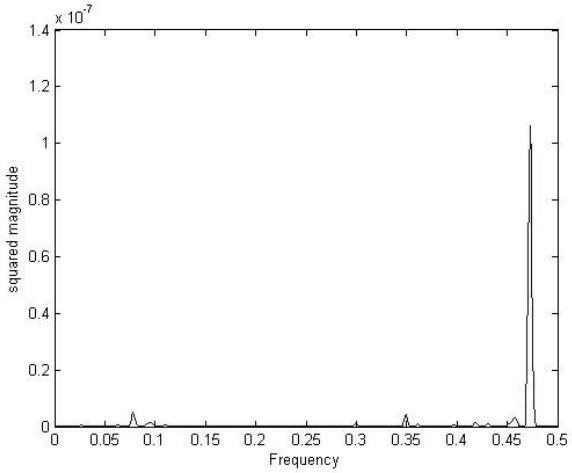

Figure 1(d)
Consensus spectrum of cytochrome C using PSD. The peak corresponds to characteristic frequency.

B. Hen egg-white lysozyme

Figure 2(a) shows the graphical representation of corresponding numerical sequences obtained by replacing each amino acid with its EIIP values for hen egg-white lysozyme.

Figure 2(b) shows the power spectral density of hen egg-white lysozyme. There are many peaks hence it is impossible to clearly determine characteristic frequency.

Figure 2(c) shows the consensus spectrum for lysozyme functional group using fast Fouriertransform (FFT) approach.

Figure 2(d) shows the consensus spectrum forlysozyme functional group using power spectral density. The distinct peak corresponds to characteristic frequency.

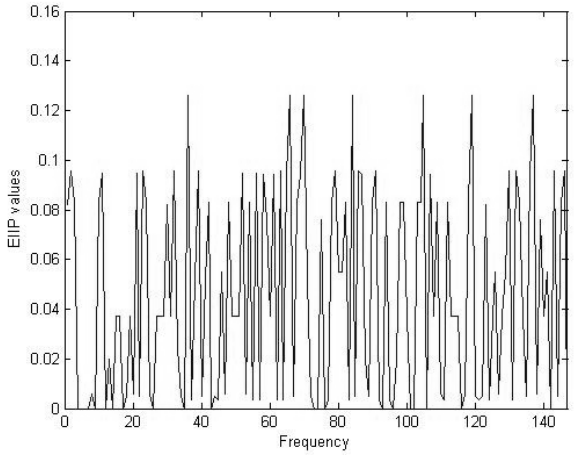

Figure 2(a)

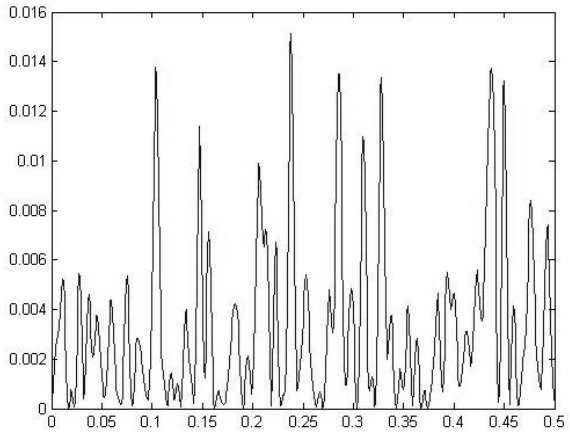

Figure 2(b)

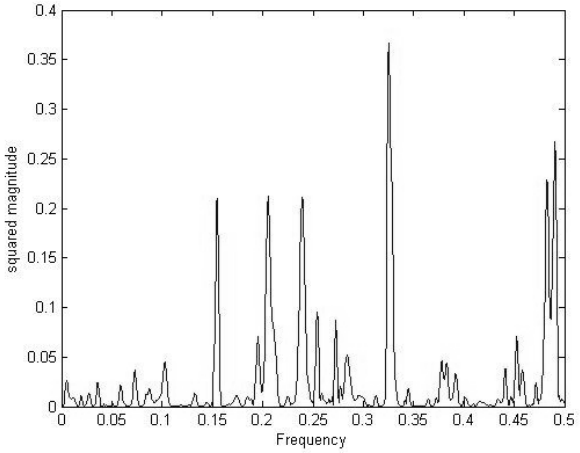

Figure 2 (c)

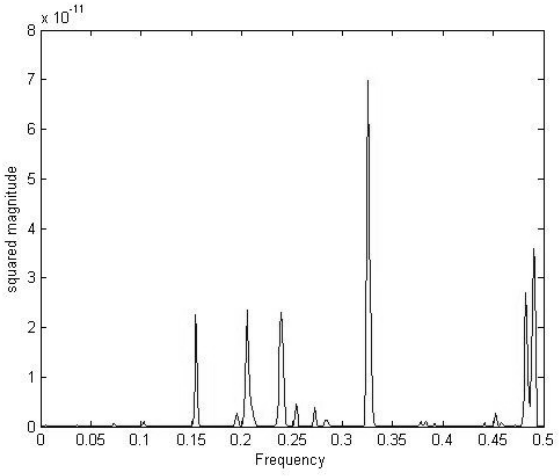

Figure 2 (d)

Consensus spectrum of lysozyme using PSD. The peak corresponds to characteristic frequency.

VIII. CONCLUSION

An approach for determination of characteristic frequency has been proposed using power spectral density (PSD). A significant peak exists at characteristic frequency for biologically related protein sequence which can be obtained from consensus spectrum using a number of proteins sequences from same functional group. The peak corresponds to characteristic frequency and only one peak exists for a group of protein sequences sharing same biological functions.

ACKNOWLEDGMENT

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.

REFERENCES

[1] B.Alberts,D.Bray,A.Johnson,J.Lewis,M.Raff,K.Roberts,andP. Walter,*EssentialCellBiology*,GarlandPublishing,NewYork,19 98.

[2] A.A.BoganandK.S.Thorn,"Anatomyofhotspotsinproteininterf aces,"*JournalofMolecularBiology*,vol. 280,pp.1-9, 1998.

[3] V.Veljković, I.Cosić,B.Dimitrijević,andD.Lalović,"Isitpossible toanalyzeDNAandproteinsequencesbythemethodsofdigital signalprocessing?,"*IEEETrans.Biomed.Eng.*,vol.BME-32,no.5,pp.337–341,May1985.

[4] T.ClacksonandJ.A.Wells,"Ahotspotofbindingenergyinahormo nereceptorinterface,"*Science*,vol. 267,pp. 383-386, Jan.1995.

[5] T.KortemmeandD.Baker,"Asimplephysicalmodelforbindinge nergy hotspotsinprotein-proteincomplexes,"*PNAS*,vol.99,no.22,pp.14116-14121, Oct. 2002.

[6] P.P.VaidyanathanandB–J.Yoon,"Theroleofsignal-processingconceptsingenomicsandproteomics,"*JournaloftheF ranklinInstitute*,vol. 341,pp. 111-135, 2004.

[7] I.Cosic,"Macromolecularbioactivity:isitresonantinteractionbet ween macro-molecules?–theoryandapplications,"*IEEETrans.onBiomedicalEngr.*,vol. 41,no. 12,pp. 1101-1114,Dec. 1994.

[8] J.Lazovic,"SelectionofaminoacidparametersforFouriertransfor m basedanalysisofproteins,"ComputerApplicationsintheBiosciences,vol. 12,no. 6,pp. 553-562, 1996.

[9] I.Cosic,"Proteindesignbymethodsofdigitalsignalanalysis,"*Pro c. AnnualConf.IEEEBiomed.Eng.*,vol.2,NewYork:IEEEPubl.Ser-vice,pp. 901-905,1988.

[10] A.V.OppenheimandR.W.Schafer,*Discrete-timeSignalProcessing*,PrenticeHallInc.NJ,1999.

[11] I. CosicandM.T. W.Hearn,"Proteinactivesitesaredefinedasresonantspotsinprote in3Dstructure,"EngineeringinMedicineandBiologySociety199 2vol.14ProceedingsoftheAnnualInternationalConference*oftheI EEE*,vol. 1,pp. 206-207,Oct. 1992.

[12] E.Pirogova,Q.Fang,M.Akay,andI.Cosic,"Investigationofstruct ural andfunctionalrelationshipsofoncogeneproteins,"*Proc.oftheIEE E*,vol. 90,no. 12,pp.1859-1867, Dec.2002.

[13] A.Antoniou,DigitalFilters:Analysis,Design,andApplications, McGraw-Hill,NewYork,1993.

[14] V.VeljkovićandI.Slavić,"Simplegeneral-modelpseudopotential,"*Phys.Rev.Lett.*, vol.29,no.2,pp.105–107,Jul.1972.

[15] I.Cosic,E.Pirogova,andM.Akay,"Applicationoftheresonantr ecognitionmodeltoanalysisofinteractionbetweenviralandtu morsup-pressorproteins,"inProc. 25thAnnu.Int. Conf.IEEEEMBS,Cancun,Mexico,Sep.17–21,2003,pp.2398–2401.

[16] P.Ramachandran,A.Antoniou,andP.P.Vaidyanathan,"Identi ficationandlocationofhotspotsinproteinsusingtheshort-timediscreteFouriertransform,"inProc.38thAsilomarConf.Sig nals,Systems,*Computers*,PacificGrove,CA,Nov.2004,pp.16 56–1660.

[17] ProteinDataBank(PDB),ResearchCollaboratoryforStructur alBioinformatics(RCSB).[Online].Available: http://www.rcsb.org/pdb/.

[18] Swiss-ProtProteinKnowledgebase.SwissInst.Bioinformatics(SIB) .[Online].Available: http://us.expasy.org/sprot/.

[19] AlanineScanningEnergeticsDatabase(ASEdb),[Online].Av ailable:http://nic.ucsf.edu/asedb/index.php.

[20] ------"The relationship of resonant recognition model to effect of low-intensity light on cell growth," *Int. J Radiat Biol* vol. 56, pp 179-191, 1989.

[21] D.DresslerandH.Potter,*DiscoveringEnzymes*.New York: ScientificAmericanLibrary,1991.

[22] C.Tanford and J.Reynolds, Nature's Robots: A History of Proteins. Oxford,U.K.:OxfordUniv.Press,2001.

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

[24] Welch, P.D, "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms," *IEEE Trans. Audio Electroacoustics, Vol. AU-15 (June 1967)*, pp.70-73.

[25] Oppenheim, A.V., and R.W. Schafer, *Discrete-Time Signal Processing*, Prentice-Hall, 1989, pp.730-742.

[26] http://wiki.answers.com,www.**wikipedia**.org,

[27] P.Ramachandran,A.Antoniou,"Identificationandlocationof hotspotsinproteinsusingdigital filters,"IEEE Journal of selected topics in signal processing. Vol .2, No. 3,june 2008.