Pertanika J. Sci. & Technol. 22 (2): 457 - 470 (2014)

PERTANIKA JOURNALS

SCIENCES & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Identification of Hot Spots in Proteins Using Modified Gabor Wavelet Transform

D. K. Shakya[1]*, Rajiv Saxena[2] and S.N. Sharma[3]

[1]*Department of Biomedical Engineering, Samrat Ashok Technological Institute, Chandrasekhar Marg, Vidisha-464001 (Madhya Pradesh), India*

[2]*Jaypee University, Anoopshahr, 203390 (Uttar Pradesh), India*

[3]*Department of Electronics and Communication Engineering, Samrat Ashok Technological Institute, Chandrasekhar Marg, Vidisha 464001 (Madhya Pradesh), India*

ABSTRACT

Identification of hot spots is an important issue in proteomics. Identifying hot spots using Digital Signal Processing (DSP) based methods is quite useful in newly discovered proteins as these methods do not require the structural information of proteins. In this paper, Modified Gabor Wavelet Transform (MGWT) was used to predict hot spots from primary amino acid sequence of protein. Incorporation of MGWT into Resonant Recognition Model (RRM) improves the prediction of the hot spots. The proposed method only requires tuning of MGWT to the characteristic frequency of the proteins' functional group, which is determined using RRM. This DSP-based technique is illustrated using several protein examples and the results are compared with the other recently reported digital signal analysis methods, viz. digital filtering and S-Transform based approaches. Relative procedural simplicity of this method over S-transform based approach and better prediction performance than digital filtering method are the novel features.

Keywords: Electron-Ion-Interaction-Potential, Hot spot, protein, resonant recognition model, wavelet transform

Article history:
Received: 2 February 2012
Accepted: 18 March 2013

Email addresses:
D. K. Shakya (devendrashakya@rediffmail.com),
Rajiv Saxena (rsaxena2001@yahoo.com),
S.N. Sharma (sanjeev_n_sharma@rediffmail.com)
*Corresponding Author

INTRODUCTION

Proteins are polymers built up from amino acids (Alberts *et al*., 1998). Although numerous different amino acids are theoretically possible, only 20 of them are commonly found in proteins, and all proteins are made up of combinations of these molecules. The 20 amino acids are represented in a protein sequence as a string

ISSN: 0128-7702

of alphabetical symbols with typical lengths ranging from 100 to 10000 (Anastassiou, 2001). Proteins are the main conductors and workforce in any living process. They play a vital role in body functioning as catalysts accelerating chemical reactions, as carrier and storage molecules in muscle contractions, as antibodies imparting immunity and as receptors in the nervous system generating and transmitting nerve impulses. These cellular processes are largely governed by different types of interaction between proteins, and the function of a protein can be better understood considering its interactions (Uetz *et al.*, 2000). By means of its three dimensional (3-D) structure, protein expresses its biological function. 3-D shape allows the protein to interact with other molecules known as targets. These interactions are very selective in nature. Studies on protein interfaces have revealed that energies are not uniformly distributed. Instead, there are certain critical residues called hot spots comprising only a small fraction of interfaces, yet accounting for the majority of the binding energy. The broad recognition of the importance of characterizing protein interactions in a cell has rendered the development of experimental and computational techniques to detect and predict hot spots with an objective to produce new and more efficient drugs and other biotechnological products.

Experimentally, hot spot residues are identified via Alanaine Scanning Mutagenesis (ASM), as described by Bogan and Thorn (1998). If a residue has a significant drop in binding affinity ($\Delta\Delta G$) when mutated to alanine, it is labelled as a hot residue. Thorn and Bogan (2001) deposited hot spots from the ASM experiments in the ASEdb. ASM is expensive, time consuming and requires a lot of efforts. Hence, simpler and less expensive computational techniques are required by biologists for estimating hot spot locations. Wet lab experiments can then be selectively performed by using the estimates obtained, resulting in a considerable saving of laboratory resources. Hot spot related databases/web servers have been complied by Tuncbag *et al*. (2009). Ofran and Rost (2007a), and Ofran and Rost (2007b) stated that all databases/servers require protein structure for prediction of hot spots except Interaction Sites Identified from Sequence (ISIS). ISIS predicts the hot spots from the primary sequence only and uses the physicochemical features, evolutionary and structural features of the protein through neural network model to predict the hot spots. However, for a newly discovered protein molecule, the only information initially available is its amino-acid sequence. Hence, the Digital Signal Processing (DSP) based methods play important roles in the analysis of these sequences as they do not need any structural information or training for estimating hot spots, apart from the primary amino-acid sequence (Cosic, 1994; Cosic, 2001; Cosic *et al.,* 2002; Vaidyanathan, 2004; Ramchandran & Antoniou, 2008). All the reported DSP-based methods first extract characteristic frequency using Resonant Recognition Model (RRM) (Cosic, 1994) and then apply DSP algorithms. The first DSP-based reported method by Cosic (1994) alters the amplitude at the characteristic frequency and the positions of the amino acids mostly affected by the change of amplitude are defined as hot spots. However, changing a single DFT coefficient affects all the elements of the protein's numerical sequence, making this particular method not reliable. Ramchandran *et al.* (2004) improved the performance of this method using short-time discrete Fourier transform (STDFT) and this improvement was

attained by employing digital filters (Ramchandran & Antoniou, 2008). In a recently reported work, Sahu and Panda (2011) used S-transform to predict hot spots with better accuracy than digital filtering. S-transform approach is relatively complex as it requires multiplication of the S-transform with the consensus spectrum in each time instant, followed by band limited filtering in time-frequency domain to select the characteristic frequency. The band limited filter is to be activated during the specific regions in the time-frequency plane. Application of the different wavelets' functions for their possible uses in the identification of active sites in the proteins has also been reported (see Cosic, 2001; Cosic *et al.,* 2002; Rao & Swamy, 2008). These wavelet based approaches successfully identified the areas of high energy regions that embrace the active sites but the exact identification of hot spot residues is missing in these works.

In this work, Modified Gabor Wavelet Transform (MGWT) reported by Mena-Chalco *et al.* (2008) was used to identify hot spots by tuning it to the characteristic frequency of the proteins' functional group. The prediction accuracy of this method has been compared with the digital filtering method introduced by Ramchandran and Antoniou (2008) and S-transform approach suggested by Sahu and Panda (2011).

The rest of the paper is organized as follows. In Section, II RRM is discussed. Section III describes MGWT and its application to identify hot spots in combination with RRM. The potentiality of the proposed method was assessed using a set of 10 proteins from different functional family selected from the standard databases. The protein sequences and the evaluation criteria used for the experimental study are discussed in Section IV. The experimental results of the proposed method are presented in Section V. A comparative study with S-transform and digital filtering approach is also elaborated in this section. Finally, the research is concluded in Section VI.

RESONANT RECOGNITION MODEL (RRM)

RRM (Cosic, 1994) is a physicomathematical approach to gain insights into selective protein interactions relevant to their biological function. This model explains selectivity of these interactions in terms of the resonant energy transfer between interacting molecules. RRM shows that certain periodicities within the distribution of energies of delocalized electrons along a protein molecule are critical for proteins biological function, i.e., the interaction with its target. RRM interprets information from protein sequences using signal analysis methods. It comprises of two stages: the first step involves the transformation of the amino acid sequence into a numerical sequence by assigning to each amino acid its Electron-Ion Interaction Potential (EIIP) value (Veljkovic *et al.,* 1985). EIIP of an amino acid is a physical property denoting the average energy of the valence electrons in the amino acid, and it is known to correlate well with a protein's biological properties (Lazovic, 1996). Thus, the resulting numerical series represents the distribution of the free electrons' energies along the proteins. The EIIP values for the 20 amino acids are listed in Table 1.

TABLE 1
EIIP Values for 20 Amino-acids

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

In the next step, numerical series are analyzed by transforming them into frequency domain using discrete Fourier transform (DFT). The common frequency components for a group of protein sequences are determined by computing the cross-spectral function $S(e^{j\omega})$.

$$S(e^{j\omega}) = \left| X_1(e^{j\omega}) X_2(e^{j\omega}) \_\_\_\_\_ X_m \right| \quad (1)$$

where $X_1, X_2, \_\_\_\_\_ X_m$ in equation (1) are DFTs corresponding to M proteins. The consensus spectrum obtained from this product has a distinct peak at a certain frequency. This frequency is termed as characteristic frequency. A sufficient number of protein sequences are used to get a distinct peak in the consensus spectrum for a clear identification of the characteristic frequency. For a successful protein target interaction, both the protein and the target signals must share the same characteristic frequency but they must have opposite phase. The matching resembles resonance and so this model of the protein-target recognition has been termed as the resonant recognition model. After determining the characteristic frequency for a particular protein function, the hot-spot locations in a protein or target molecule can be marked by identifying the regions in the numerical sequence where this frequency is dominant. For locating these regions, DSP algorithms can now be employed as hot spot detection is now translated into a time-frequency analysis problem.

PREDICTION OF HOT SPOTS USING MGWT

Gabor wavelet was modified (Mena-Chalco *et al.*, 2008) for analyzing a signal in specific frequency and multiple scales. This has been achieved by varying the Gaussian standard deviation of the analyzing function, while its complex exponential frequency has been kept constant. The following relationship describes the analyzing function corresponding to MGWT:

$$\varphi_{MGWT}(t,n,a) = e^{\frac{-(t-n)^2}{2a^2}} e^{j\omega_0(t-n)} \quad (2)$$

In (2), *n* is the position index of the sequence to be analyzed, *a* is the scaling parameter, and ω_0 is the base frequency to which the MGWT is tuned. MGWT of a signal *u*(*t*) is given as:

$$U(n,a) = \int u(t) e^{\frac{-(t-n)^2}{2a^2}} e^{j\omega_0(t-n)} dt \tag{3}$$

The exponential term in (3) contains only a single frequency ω_0. This makes MGWT capable to capture frequency component ω_0 using different scales that are present at different locations along the position index. The spectrum of the sequence is obtained by computing the squared complex modulus of the MGWT coefficients as:

$$M(n,a) = \left| U(n,a) \right|^2 \tag{4}$$

This spectrum is then projected onto the position axis in order to detect the locations of the presence of a specific frequency w_0 which corresponds to the local maxima regions of the projection. For a sequence of length N, this projection spectra is obtained by summing up the MGWT coefficients for all the scales. Equation (5) describes this computation.

$$MGWT(n) = \sum_a M(n,a), \qquad n = 0, \_\_\_ N-1 \tag{5}$$

In this work, MGWT was used to analyze protein sequences for identifying the hot spot locations. Multi resolution analysis of the protein sequences to capture specific periodicity locations corresponding to the protein's characteristic frequency was carried out. This was achieved by fixing ω_0 at the characteristics frequency of the protein's functional group and then computing MGWT for different scales. Therefore, the MGWT approach combines the features of wavelet (Pirogova *et al.*, 2002) and digital filtering method (Ramchandran & Antoniou, 2008), resulting in an improved performance with less computational complexity. Following are the steps involved in identifying hot spots by using this method:

(a) Convert protein character sequences of the functional group of interest into numerical sequences by using the EIIP values listed in Table 1.

(b) Using (1), plot the consensus spectrum and determine the characteristic frequency.

(c) Tune MGWT by making ω_0 equal to the characteristic frequency.

(d) Using (3), compute the MGWT of the protein sequence of interest at different scales.

(e) Determine the projection spectrum of the sequence using (4) and (5).

(f) Plot the projection spectra and identify the hot spots by locating the energy peaks based on a suitable peak-to-average ratio (Ramchandran & Antoniou, 2008).

PROTEIN SEQUENCES AND PERFORMANCE EVALUATION

In a recent publication, where DSP based identification of hot spots was reported by Sahu and Panda (2011), 10 proteins belonging to different functional families were selected from the standard databases for the experimental study. With an objective to compare the results of the proposed method with S-transform and digital filtering approaches, the same set of protein

sequences (Sahu & Panda, 2011; Ramchandran & Antoniou, 2008) was used in this work. The sequence length, characteristic frequency and PDB ID of these protein sequences are listed in Table 2. The amino acid sequences for these proteins were obtained from the freely available protein data bank (PDB) [http://www.rcsb.org/] and Swiss-Prot [http://us.expasy.org/sprot/]. A benchmark to compare the hot spots identified by the proposed approach had also been generated by combining the results of ASEdb by Thorn and Bogan (2001), and Robetta interface alanine scanning (Robetta-Ala) reported by Kortemme *et al.* (2004) and Kortemme and Baker (2002). In ASEdb, an interface residue is considered as a hot spot if its corresponding $\Delta\Delta G$ is equal to or higher than 2.0 kcal/mol. As for Robetta-Ala, the interface residues with $\Delta\Delta G$ more than 1.0 kcal/mol are taken as hot spots.

TABLE 2
Protein Sequences Investigated and their Characteristics Frequency

S.No.	Organism	Protein Name	PDB ID	Swiss-port ID	Sequence Length	Characteristics frequency
1	Human	basic fibroblast growth factor (bFGF)	4fgf	P09038	146	0.904
2	Human	Growth hormone (hGH)	3hhr	P01241	190	0.270
3	Human	Growth hormone binding Protein (hGHbp)	3hhr	P10912	205	0.270
4	Human	Interleukin (IL4)	1rcb	P05112	129	0.587
5	Human	Human alpha hemoglobin	1vwt	P69905	141	0.023
6	Bacteria	Barnase	1brs	P00648	110	0.321
7	Bacteria	Barstar	1brs	P11540	89	0.321
8	Bacteria	Tryptophin RNA-binding attenuator protein(TRAP)	1wap	P19466	75	0.247
9	*E.Coli*	Colicin-E9 immunity protein (IM9)	1bxi	P13479	86	0.190
10	*C.fumi*	Endoglucanse C	1ulo	P14090	152	0.093

Sahu and Panda (2011) identified the hot spots in the protein sequences by comparing the energy in the regions that contributed to characteristic frequency with a reference energy level. Evaluation criteria of peak to average ratio, proposed by Ramchandran *et al.* (2008), had been followed by Sahu and Panda (2011). In the current work, the author also used the same evaluation criteria with the purpose of establishing a comparative study of the results with the earlier reported DSP approaches. The average value of the projection spectra was computed and used as a reference level for indicating the hot spots in protein sequence. In

order to control the resolution of this method, the ratio of the peaks of the projection spectra to the average value was set as threshold (t_p). The efficiency of the method in identifying the hot spots can be varied by increasing or decreasing the threshold value.

EXPERIMENTAL RESULTS AND COMPARISON STUDY

The MGWT has been applied to protein sequences with transform calculated for different scales *a* and frequency ω_0 (characteristic frequency of the protein). The results in this work were obtained by using 40 analyzing functions corresponding to 40 scale values exponentially separated between 0.2 and 0.7. The lengths of these functions have been truncated to 15 sequence points. As a sample plot, the spectrogram for human basic fibroblast growth factor (ω_0 = 0.904) is shown in Fig.1(a). In Fig.1(b), projection of the spectrum values onto the position axis is plotted. The peaks at certain interface residues shown in Fig.1(b) correspond to the hot spot locations identified by MGWT. These locations were detected by taking 90% of the average energy as threshold to locate the hot spots. Sahu and Panda (2011) also used the same threshold value in their work for the comparative study.

The hot spot prediction performance of MGWT for the proteins listed in Table 2 was compared with the results of the S-transform technique, digital filtering technique and the alanine scan computed from both ASEdb and Robitta-Ala (Sahu & Panda, 2011). The comparative detection results for these methods are given in Table 3. To assess the prediction performance using these results, the following measures (Baldi *et al.*, 2000) were used in this paper:

(i) **Accuracy (A)** – Accuracy is the ratio of the number of correctly predicted residues to the number of all the predicted residues, formulated as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{6}$$

where TP, FP, TN, and FN stand for the number of true positives (correctly predicted hot spot residues), number of false positives (non-hot spot residues incorrectly predicted as hot spots), number of true negatives (correctly predicted non-hot spot residues) and number of false negatives (hot spot residues incorrectly predicted as non-hot spots), respectively.

(ii) **Recall (R)** – Recall or sensitivity is the proportion of the number of correctly classified hot spot residues to the number of all hot spot residues.

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

(iii) **Specificity (S)** – Specificity is the proportion of the number of correctly predicted non-hot spot residues to the number of all the non-hot spot residues.

$$Specificity = \frac{TN}{TN + FP} \tag{8}$$

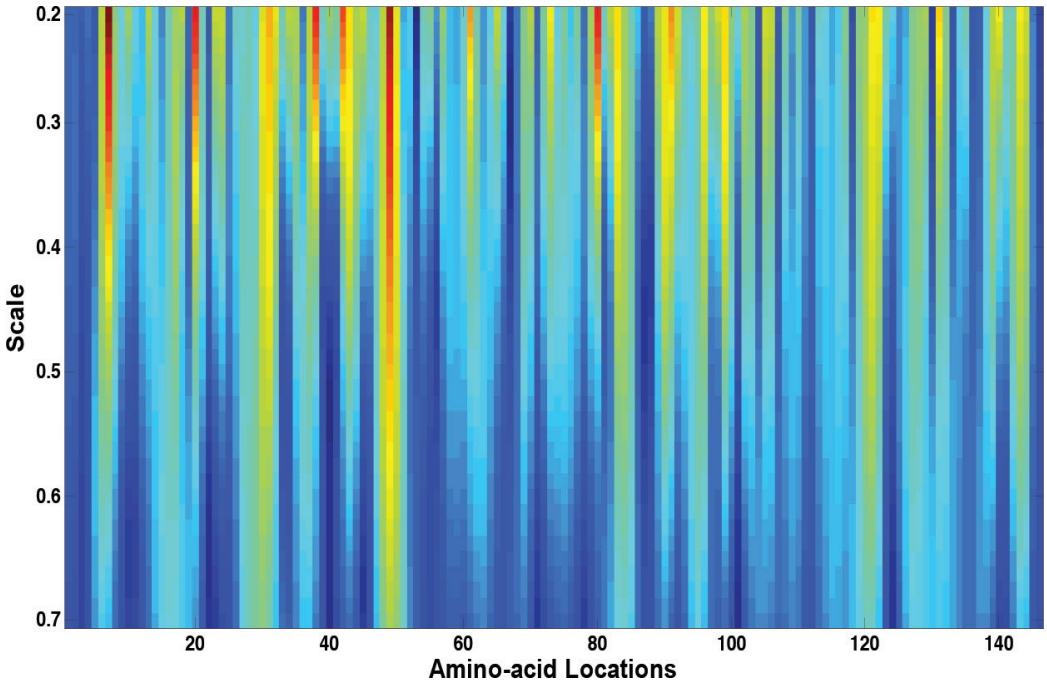

Fig.1(a): Spectrogram for human basic fibroblast growth factor

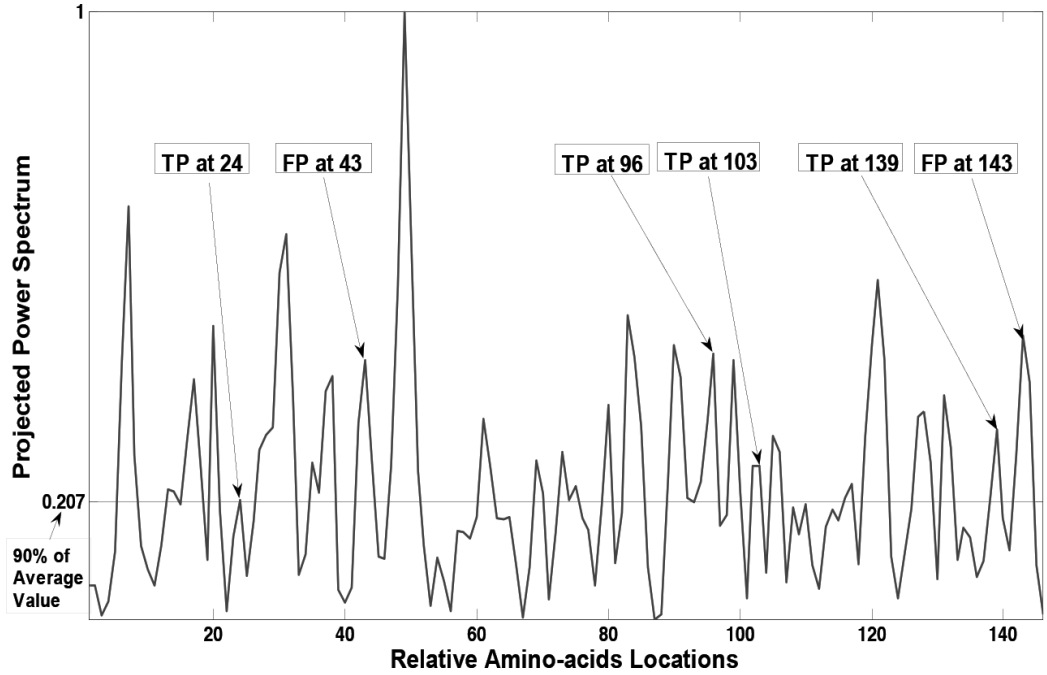

Fig.1(b): Projection spectra for human basic fibroblast growth factor

(iv) **Precision (P)** – Precision is the ratio of number of correctly classified hot spot residues to the number of all residues classified as hot spots.

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

(v) **F-measure (F)** – By using F-measure, we checked the balance between precision and recall, which is formulated as follows:

$$F - Measure = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{10}$$

(vi) **Matthews Correlation Coefficient (MCC)** – While there is no perfect way of describing true and false positives and negatives by a single number, the Matthews correlation coefficient is generally regarded as being one of the best such measures (Baldi *et al.*, 2000). The MCC can be calculated directly using this formula:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{11}$$

By using the detected hot spots and interface residues listed in Table 3, the average values of the six performance evaluation measures for the ten proteins have been obtained and are shown in Table 4. It is evident from the results in Table 4 that all the performance measures for MGWT are superior to the digital filtering technique. The values of TP and FN for MGWT are less than the corresponding values for S-transform, resulting in comparatively small value of Recall. However, the number of false positives identified by S-transform is greater than those identified by MGWT. Time-frequency filtering operation has been cited (Sahu & Panda, 2011) as the reason behind the large number of false positives. Therefore, the Specificity and Precision values for MGWT are better than the S-transform method. Meanwhile, the performance of S-transform is slightly better than MGWT in terms of F-measure and MCC. However, the Accuracy for S-transform and MGWT methods is the same.

TABLE 3
Comparative Results of the Detected Hot-Spots Using Experimental and Different Computational Methods

Protein Name	Interface Residues	Actual Hot Spots (ASEdb)			Detected Hot-Spots		
		ASEdb (2kcl/mol)	*Robetta Alanine (1kcl/mol)*	*ASEdb+ Robetta Alanine*	Digital Filter	S-Transform	MGWT
basic fibroblast growth factor (bFGF)	22, 24, 26, 44, 46, 96, 97, 101, 103, 107, 109, 110, 111, 113, 114, 140, 142	24, 96, 103, 140	----	24, 96, 103, 140	24, 26	24, 96, 103, 140	24, 96, 103, 139
Growth hormone (hGH)	2, 3, 4, 8, 9, 12, 15, 16, 18, 19, 21, 22, 25, 26, 29, 42, 45, 46, 48, 51, 56, 62, 63, 64, 65, 68, 164, 167, 168, 171, 172, 174, 175, 176, 178, 179, 182, 183, 186	172, 175, 178, 176	18, 25, 42, 45, 46, 64, 168, 171, 175, 179	18, 25, 42, 45, 46, 64, 168, 171, 172, 175, 178, 179	26, 41, 45, 64, 168, 171, 175, 178, 179	18, 25, 42, 47, 65, 168, 172, 175, 178, 180	17, 41, 45, 47, 63, 65, 167, 170, 172, 176, 179
growth hormone binding protein (hGHbp)	43, 44, 72, 76, 77, 80, 98, 102, 103, 104, 105, 108, 120, 121, 122, 124, 126, 127, 164, 165, 166, 167, 169	43, 104, 105, 165, 169	43, 76, 104, 127, 169	43, 76, 104, 105, 127, 165, 169	43, 105, 127, 164, 165, 169	43, 103, 105, 127, 165, 170	43, 126, 164
interleukin (IL-4)	5, 6, 8, 9, 11, 13, 15, 16, 19, 77, 78, 81, 82, 84, 85, 88, 89, 91	9, 88	----	9, 88	9, 88	9, 88	87
human alpha haemoglobin	5, 18, 22, 36, 43, 59, 76, 109, 111, 131	18, 22, 36, 43, 59	----	18, 22, 36, 43, 59	37, 60	22, 36, 60	37, 42
Barstar	29, 35, 39, 42, 76, 80	29, 35, 39	29, 35, 39, 42, 76	29, 35, 39, 42, 76	35, 38, 42	29, 35, 38, 42	36, 40, 43, 76
Barnase	27, 58, 59, 60, 73, 87, 102	27, 58, 59, 73, 87, 102	27, 59, 60, 83, 87, 102	27, 58, 59, 60, 73, 87, 102	27, 59, 73, 87, 102	27, 58, 60, 3, 86,102	27, 57, 60, 73, 87, 102

TABLE 3 (continue)

TRAP	20, 22, 28, 30, 32, 37, 42, 50, 51, 56, 58, 60	37, 40, 56, 58	56, 58	37, 40, 56, 58	37, 40, 56	40, 56, 59	36, 40, 59
colicin-E9 immunity protein (IM9)	23, 24, 27, 28, 29, 30, 33, 34, 37, 38, 41, 48, 49, 50, 51, 55, 56	33, 34, 41, 50, 51, 55	30, 33, 38, 50, 55	30, 33, 34, 38, 41, 50, 51, 55	34, 41, 50, 51, 55	33, 41, 50, 51, 55	30, 39, 41, 49, 52, 55
Endoglucanase C	19, 50, 75, 80, 81, 84, 87, 90, 124, 128	19, 50, 84	----	19, 50, 84	50	50, 84	No Detection

TABLE 4
Comparative Performance of Different Computational Methods

Performance Evaluation Measures in Percentage	Digital Filtering	S-Transform	MGWT
Accuracy	60	67	67
Recall	67	79	70
Specificity	56	60	65
Precision	46	52	53
F-Measure	54.5	62.7	60.3
MCC	21.65	37.29	33.49

Another dimension in which a comparison can be established is the computational load of the three methods. The computational complexity of S-transform is relatively more as it requires the following additional processing:

S-transform spectrum is to be multiplied with the consensus spectrum in each time instant to suppress the noisy frequencies and to boost up energy at the characteristic frequency.

(i) After multiplication by the consensus spectrum distinct energy concentrated areas in the time-frequency plane where the characteristic frequency is dominated are obtained. In order to separate the frequency of interest, a band limited time-frequency filter is to be designed and activated during the specific regions in the time-frequency plane.

In this context, digital filtering (Ramchandran & Antoniou, 2008) and the proposed method are at par, as multiplication by consensus spectrum and time-frequency filtering is not needed in these methods. Tuning of MGWT to the protein characteristic frequency is analogous to anti-notch filtering of the proteomic signal, with anti-notch frequency set to the characteristic frequency. Because of this filtering, multiplication by consensus spectrum in each time instant to boost the characteristic frequency component is not required. Multi-resolution analysis then identifies the hot spot locations and thus eliminates the need of filtering in time-frequency domain to locate the regions where characteristic frequency is dominant.

CONCLUSION

This paper proposes a simple and efficient method for the identification of hot spot residues in proteins using MGWT. The proposed method predicts hot spots from the amino acid sequence only and it does not requires structural information of the protein or any kind of prior training using other protein features. Hence, this method can be quite useful in estimating hot spot residues prior to performing wet lab experiments in a newly identified protein, for which the only information initially available is its amino acid sequence. The performance of this method has been evaluated on protein sequences selected from the standard protein data bases using sufficient number of prediction performance measures. A comparative study with the recently reported signal analysis methods, based on digital filtering and S-transform, was also carried out. Prediction accuracy of all the measures for the proposed method is observed to

be superior to digital filtering method. When compared with S-transform, MGWT gives a better performance in terms of specificity and precision. The accuracy of the two methods is the same. Better Recall, F-Measure, and MCC values have been obtained with S-transform but at the cost of increased computational load. The computational complexity of MGWT based hot spot detection is relatively less than the S-transform method as it does not require multiplication by consensus spectrum in each time instant and time-frequency filtering. Thus in this work, a novel DSP-based method for hot spots identification with satisfactory prediction performance and less computational load has been developed.

REFERENCES

Alberts, B., Bray, D., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (1998). *Essential Cell Biology*. New York: Garland Publishing.

Anastassiou, D. (2001). Genomic signal processing. *IEEE Signal Processing Magazine*, *18*(4), 8-20.

Baldi, P., Brunak, S., Chauvin, Y., Andersen, F. A. C., & Nielsen, H. (2000). Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics*, *16*(5), 412–424.

Bogan, A., & Thorn, K. S. (1998). Anatomy of hot spots in protein interfaces. *Journal of Molecular Biology*, *280*, 1-9.

Cosic, I. (1994). Macromolecular bioactivity: is it resonant interaction between macromolecules? – Theory and applications. *IEEE Transaction on Biomedical Engineering, 41*(12), 1101-1114.

Cosic, I. (2001). Analysis of HIV proteins using DSP techniques. *Proceedings of the 23rd Annual EMBS International Conference,* Istanbul, Turkey*,* 2886-2889.

Kortemme, T., & Baker, D. (2002). A simple physical model for binding energy hot spots in protein-protein complexes. *Proceeding of Natural Academic Science*, *U.S.A., 99*, 14116-14121.

Kortemme, T., Kim, D. E., & Baker, D. (2004). Computational alanine scanning of protein-protein interfaces. *Science STKE*, *219*, 12.

Lazovic, J. (1996). Selection of amino acid parameters for Fourier transform-based analysis of proteins. *Computer Application Bioscience, 12*(6), 553-562.

Mena-Chalco, J. P., Carrer, H., Zana, Y., & Cesar, JR. R. M. (2008). Identification of protein coding regions using the modified Gabor-wavelet transform. *IEEE/ACM Transaction on Computational Biology and Bioinformatics, 5*(2), 198-206.

Ofran, Y., & Rost, B. (2007a). Protein-Protein interaction hotspots carved into sequences. *PLOS Computational Biology, 3,* 1169-1176.

Ofran, Y., & Rost, B. (2007b). ISIS: Interaction sites identified from sequences. *Bioinformatics, 23,* 13-16.

Pirogova, E., Fang, Q., Akay, M., & Cosic, I. (2002). Investigation of the structural and functional relationships of oncogene proteins. *Proceeding of the IEEE, 90*(12), 1859-1867.

Ramchandran, P., Antoniou, A., & Vaidyanathan, P. P. (2004). Identification and location of hot spots in proteins using the short-time discrete Fourier transform. In *Proceeding 38th Asilomer Conference Signals, Systems, Computers*, Pacific Grove, CA*,* 1656-1660.

Ramchandran, P., & Antoniou, A. (2008). Identification of hot-spot locations in proteins using digital filters. *IEEE Journal of Selected Topics in Signal Processing*, *2*(3), 378-207.

Rao, K. D., & Swamy, M. N. S. (2008). Analysis of genomics and proteomics using DSP techniques, *IEEE Transaction on Circuits and System –I, 55*(1), 370-378.

Sahu, S. S., & Panda, G. (2011). Efficient localization of hot spots in proteins using a novel S-transform based filtering approach. *IEEE Transaction on Computational Biology and Bioinformatics, 8*(5), 1235-1246.

Thorn, K. S., & Bogan, A. (2001). ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics, 17*(3), 284-285.

Tuncbag, N., Kar, G., Keskin, O., Gursay, A., & Nussinov, R. (2009). A survey of available tools and web servers for analysis of protein-protein interactions and interfaces. *Briefings in Bioinformatics*, *10*(3), 217-232.

Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., & Judson, R. S. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature*, *403*, 623-627.

Vaidyanathan, P. P., & Yoon, B. J. (2004). The role of signal processing concepts in genomics and proteomics. *Journal of the Franklin Institute, 341*, 111-135.

Veljkovic, V., Cosic, I., Dimitrijevic, B., & Lalovic, D. (1985). Is it possible to analyze DNA and protein sequences by the methods of digital signal processing. *IEEE Transaction Biomedical Engineering*, *32*(5), 337-341.