Proceedings of the 25th Annual International Conference of the IEEE EMBS
Cancun, Mexico • September 17-21, 2003

Application of the Resonant Recognition Model to Analysis of Interaction between Viral and Tumor Suppressor Proteins

I. Cosic[1], E. Pirogova[1], M.Akay[2]
[1]School of Electrical and Computer Engineering, RMIT University, Melbourne, Australia
[2]Thayer School of Engineering, Dartmouth College, NH, USA

Abstract—**Recent findings in cancer research has established a connection between a T-antigen - common virus – and a brain tumor in children. The studies suggested the T-antigen, the viral component of a specific virus, called the JC virus, plays a significant role in the development of the most frequent type of malignant brain tumors by blocking the functionality of tumor suppressor proteins such as p53 and pRb. Here we have investigated the structure and function relationships of T-antigen, p53 and pRb proteins using the Resonant Recognition Model (RRM), a physico-mathematical approach based on digital signal processing methods.**

Keywords—**Signal processing, modeling, protein function, and characteristic frequency.**

I. Introduction

JC virus (JCV) is a neurotropic polyomavirus infecting greater than 70% of the human population worldwide during early childhood. JCV possesses an oncogenic potential and induces development of various neuroectodermal origin tumors including medulloblastomas and glioblastomas. The most important role in this process is attributed to T-antigen, which has the ability to associate with and functionally inactivate well-studied tumor suppressor proteins p53 and pRB. The immunohistochemistry analysis studies revealed expression of JCV T-antigen in the nuclei of tumor cells [1-3].

The RRM [4,5] is a novel engineering approach invented for the analysis of protein-protein and protein-DNA interactions. The RRM suggests that these bio-molecular interactions present the transfer of the resonant energy between the interacting molecules at the frequency specific for each observed function/interaction. This model is aimed to find possible patterns of significance and thus assists in structure-function studies of different protein families.

The RRM is a physico-mathematical approach that interprets protein sequence linear information using digital signal processing. In the RRM the protein primary structure is represented as a numerical series by assigning to each amino acid in the sequence a physical parameter value relevant to the protein's biological activity. The RRM concept is based on the finding that there is a significant correlation between spectra of the numerical presentation of amino acids and their biological activity [4,5].

It has been found through an extensive research that proteins with the same biological function have a common frequency in their numerical spectra. This frequency was found then to be a characteristic feature for protein biological function or interaction. Once the characteristic frequency for a particular protein function/interaction is identified, it is possible then to utilize the RRM to predict the amino acids in the sequence predominantly contributed to this frequency and consequently to the observed function. Also it becomes possible to design peptides having the desired periodicities [4,5].

II. Methodology

The RRM, used in this study, involves a transformation of the amino acid sequence into a numerical sequence and then analysis of this sequence by appropriate digital signal processing methods (Fourier Transform, wavelets etc.) [7-18]. To determine the common frequency components in the spectra for a group of proteins, the multiple cross-spectral function is used. Peaks in this function denote common frequency components for all protein sequences analyzed. Through an extensive study the RRM has reached a fundamental conclusion: *one RRM characteristic frequency characterizes one particular biological function or interaction* [4,5]. This frequency is related to the biological function provided the following criteria are met:

- One peak only exists for a group of protein sequences sharing the same biological function.
- No significant peak exists for biologically unrelated protein sequences.
- Peak frequencies are different for different biological functions.

The assignment of a particular number for each amino acid in the protein sequence is a crucial step in all these calculations. This set of numbers should have a physical meaning related to the protein's biological function. In this study, the energy of delocalized electrons (calculated as the electron-ion interaction pseudo-potential, EIIP [6]) of each amino acid residue is used. Thus, the resulting numerical series represents the distribution of the free electrons energies along the protein. Then, these numerical series are analyzed by digital signal analysis methods, in particular by Discrete Fourier Transform, in order to extract information pertinent to the protein biological function.

A multiple cross-spectral function is defined and calculated to obtain the common frequency components from the spectra of a group of proteins. Peaks in such function denote common frequency components for all sequences analyzed. Signal-to-noise ratio (S/N) for each peak, described a similarity between the analyzed

0-7803-7789-3/03/$17.00 ©2003 IEEE

sequences, is calculated as the ratio between the signal intensity at the particular peak frequency and the mean value over the whole spectrum [4,5].

In this study the RRM approach has been used to investigate the mutual relationships of brain tumor associated JC viral oncoprotein T-antigen and tumor suppressor proteins p53 and pRb.

III. Results

Over time, cells can accumulate damage to DNA. If the right combination of genes becomes altered, the cell will become cancerous. Some genes help to prevent malignant behavior and are therefore referred to as tumor-suppressor genes. These genes contribute to cancer when they are inactivated or lost as a result of DNA damage (mutations). Other genes, known as proto-oncogenes, can promote cancer if they acquire new properties as a result of mutations at which point they are called oncogenes. Most common cancers involve both inactivation of specific tumor-suppressor genes and activation of certain proto-oncogenes. The human neurotropic polyomavirus, JCV, produces a regulatory protein named T-antigen, which is expressed at the early phase of viral lytic infection and plays a critical role in completion of the viral life cycle. Furthermore, this protein has the ability to transform neural cells *in vitro* and its expression has been detected in several human neural-origin tumors.

JC virus T-antigen most likely infects humans through the upper respiratory tract and remains in most people throughout their lives, and, in some cases, causes minor sub-clinical problems. However, in people whose immune systems are depressed, either through chemotherapy given to organ transplant recipients or an illness such as AIDS, JCV can become active and may contribute to cancer in the brain. As it was established the interaction of viral T-antigen oncoprotein with p53 and pRb tumor suppressor proteins leads to induction of cancer [1-3].

In this study 8-JC viral T-antigen protein sequences, 13-p53 protein sequences and 9-pRb protein sequences were investigated concerning the understanding of the structure-function relationship within these proteins.

TABLE 1.
PEAK FREQUNCY AND SIGNAL-TO-NOISE RATIO OF PROTEIN GROUPS

Protein Group	Freq.	No seq.	S/N	Error
T-antigen	0.2061	8	129.58	0.125
p53	0.4326	13	159.97	0.077
pRb	0.4316	9	164.28	0.111
T-antigen, p53	0.2021	21	312.36	0.048
T-antigen, pRb	0.2041	17	167.12	0.059
T-antigen, p53, pRb	0.2021	30	506.28	0.033

A multiple cross-spectral analysis was performed for each selected protein group as well as for their mutual combination using the EIIP values (Fig.1-Fig.6). As a result, characteristic frequencies of analyzed protein groups were obtained and are shown in Table1. The RRM analysis was applied to a group of 8 T-antigen proteins, and the common feature in terms of characteristic frequency was identified at f=0.2061, S/N=129.58. This frequency component is common to all analyzed sequences and therefore can be considered as the consensus characteristic of their common biological activity of all protein sequences in this functional group.

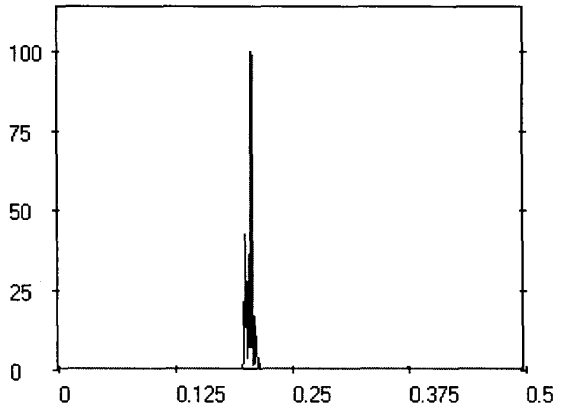

Fig. 1. Multiple cross-spectral function of JC viral T-antigen proteins. The prominent peak(s) denote common frequency components. The abscissa represents RRM frequencies, and the ordinate is the normalized intensity.

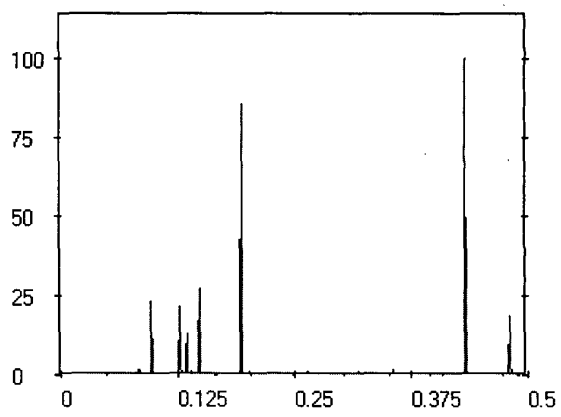

Fig. 2. Multiple cross-spectral function of p53 proteins

The same procedure was repeated with p53 and pRb tumor suppressor proteins (Fig.2, Fig.3). The p53 tumor suppressor gene has proven to be one of the genes most often mutated in human cancers. It involves mainly point mutations leading to amino acid substitutions in the central region of

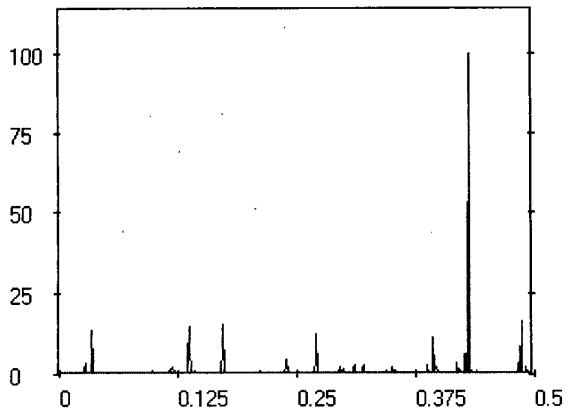

Fig. 3. Multiple cross-spectral function of pRb proteins

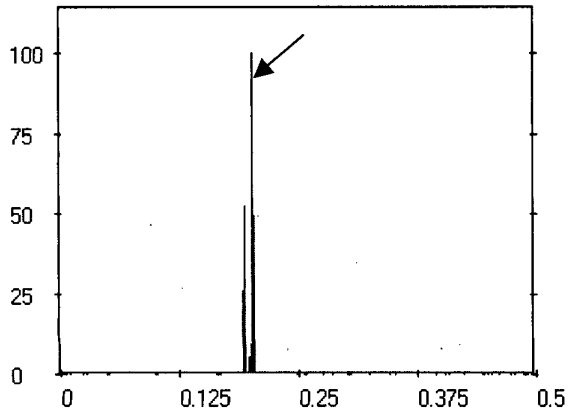

Fig. 4. Multiple cross-spectral function of JC viral T-antigen and p53 proteins

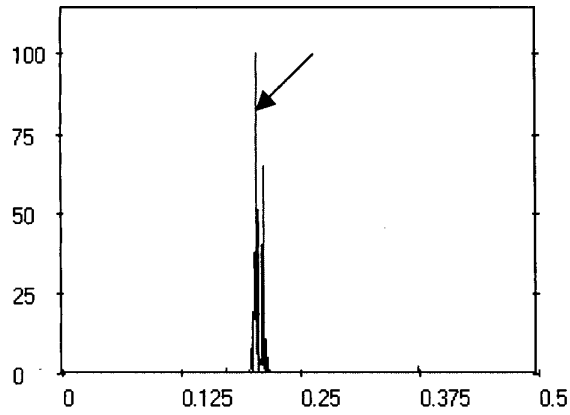

Fig. 5. Multiple cross-spectral function of JC viral T-antigen and pRb proteins

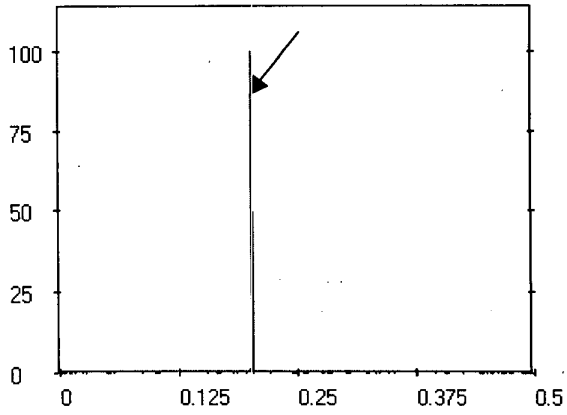

Fig. 6. Multiple cross-spectral function of T-antigen, p53 and pRb proteins

the protein and thus, causes to its abnormal functions. The prominent characteristic frequencies of p53 and pRb tumor suppressor proteins were identified at f=0.4326, S/N=159.97 and at f=0.4316, S/N=164.28 respectively. As was mentioned above each specific biological function of the protein is characterized by a single frequency. The similarity of characteristic frequencies of p53 and pRb proteins (Table 1) is expected as both p53 and pRb proteins are tumor suppressors sharing the same biological function. Thus, the frequency f=0.4326 identified within the RRM analysis is considered as a characteristic feature of the specific biological activity of p53 and pRB proteins – ability to stop the formation of tumors. After careful examining of the corresponding consensus spectrums of p53 and pRb proteins (Fig. 2, Fig. 3) we observe more than one (less significant) peaks corresponding to other biological functions determined within the RRM analysis.

IV. Discussion

It is proposed that the RRM characteristic frequencies present the common feature of the interacting sequences and thus, a common interaction. In our previous work it was also proposed that this characteristic frequency could represent the oscillations of a physical field, which are responsible for information transfer between the interacting bio-molecules [7-19]. As a consequence, it is postulated that RRM characteristic frequency is a relevant parameter for mutual recognition between bio-molecules and is significant in describing the interaction between proteins and their substrates or targets. Therefore, it is concluded that the RRM characteristic frequency may dictate the specificity of the protein interactions.

From Fig.4 we can observe only one dominant peak corresponding to the common frequency component for the

combined group of T-antigen and p53 proteins at f=0.2021, S/N=312.36. Analogous results were obtained in the analysis of interactions between T-antigen and pRb proteins. A single peak corresponding to the protein's biological activity was identified at f=0.2041, S/N=167.12. These characteristic frequencies are very close to each other (Table 1) and they overlap with each other within the calculation error, indicating their mutual recognition. Therefore, we can conclude that this identified frequency can be considered as characteristic feature of the mutual interactions between the analyzed proteins T-antigen and p53, and T-antigen and pRb respectively causing the cell damage and induction of tumor formation in brain.

Finally, we have explored the possibility of mutual 3-components interaction between T-antigen, p53 and pRb proteins by applying the RRM cross-spectral analysis to all three functional groups of proteins. Interestingly, we have found that there is a very prominent frequency component (Fig. 6) at f=0.2021, S/N=506.28 (Table 1) shared by all analyzed sequences. It should be noted that this frequency is the same (within the calculation error) as the characteristic frequency of T-antigen and pRb proteins (Fig. 5) and of T-antigen and p53 proteins respectively. Thus, the results obtained indicate the common characteristic frequency for all three interacting proteins: T-antigen, p53 and pRb proteins at f=0.2021, which is a characteristic feature of the tumor formation. This confirms the RRM main concept that proteins and their targets recognize/interact with each other based on the same (similar) characteristic frequency. Consequently, we conclude that 3-components interaction between T-antigen, p53 and pRb proteins can be considered as a crucial condition in the process of the brain tumor formation.

V. Conclusion

We have shown previously that the digital signal processing methods can be used to analyze linear sequences of amino acids to reveal the informational content of proteins [7-19]. This study extends the utility of the RRM procedures, to viral T-antigen oncoprotein and tumor suppressor p53, pRb proteins.

The results of our computational analysis clearly indicate that the RRM can determine the protein characteristic frequencies crucial for biological activity/interaction of analyzed proteins. Moreover, our successful investigations [7-18] have proved that the RRM presents an engineering tool based on digital signal processing that is able to identify the protein characteristic patterns in protein sequences related to the common biological function/interaction of studied proteins. And consequently, knowing the protein characteristic frequency allow us to allocate the protein's biological active site(s), and design new peptides with the desired biological function [12,13]. This novel prediction scheme can be used to facilitate the structure-function studies under different protein families and thus, save the experimental cost greatly.

References

[1] S. Croul, F. D. Lublin, L. Del Valle, R. J. Oshinsky, A. Giordano, K. Khalili, and C. K. Ritchie. (2000). The cellular response of JC virus T-antigen induced brain tumor implants to a murine intra-ocular model. J. Neuroimmunol. 106, pp. 181-188.

[2] B. Krynska, L. Del Valle, J. Gordon, J. Otte, S. Croul, and K. Khalili,. (2000). Identification of a novel p53 mutation in JCV-induced mouse medulloblastoma. Virology 274, pp. :65-74.

[3] L. Del Valle, S. A. Azizi, B. Krynska, S. Enam, S. E. Croul, and K. Khalili, (2000). Reactivation of human neurotropic JC virus expressing oncogenic protein in a recurrent glioblastoma multiforme. Ann. Neurol 48, pp. :932-936.

[4] I. Cosic, (1994). Macromolecular bioactivity: Is it Resonant Interaction between Molecules? - Theory and Applications. *IEEE Trans. On BME.* 41, pp. 1101-1114.

[5] I. Cosic, (1997) *The Resonant Recognition Model of Macromolecular Bioactivity*, Birkhauser Verlag.

[6] L. Veljkovic, M. Slavic, (1972). General Model of Pseudopotentials. *Physical Review Let.*, 29, pp. 105-108.

[7] I. Cosic, M. Pavlovic, V. Vojisavljevic, (1989). Prediction of "hot spots" in Il-2 cased on information spectrum characterictics of growth regaling factors. *Biochemie*, Vol. 71, pp. 333-342.

[8] I. Cosic,M. T. W. Hearn, (1991). "Hot spot" amino acid distribution in Haras oncogene product p21: Relationship to guanine binding site. *J. Molecular Recognition*, Vol.4, pp. 57-62.

[9] I. Cosic, A. N. Hodder, M. A. Aguilar, M. T. W. Hearn, (1991) Resonant Recognition Model and Protein Topography: Model Studies with Myoglobin, Hemoglobin and Lysozyme, *Eur. J. Biochem*, 198, pp. 113-119.

[10] I. Cosic, (1995). *Bio/Technology*, 13, pp. 236-238.

[11] V. Krsmanovic, J. M. Biquard, M. Sikorska-Walker, I. Cosic, C. Desgranges, M. A. Trabaud, J. F. Whitfield, J. P. Durkin, A. Achour, M.T.W. Hearn, (1998). Investigation Into the Cross-reactivity of Rabbit Antibodies Raised against Nonhomologous Pairs of Synthetic Peptides Derived from HIV-1 gp120 proteins. *J.Peptide Res.*, 52(5), 410-4120.

[12] I. Cosic, A. E. Drummond, J. R. Underwood and M. T. W. Hearn, (1994). In vitro inhibition of the actions of basic FGF by a novel 16 amino cid peptide. *Molecular and Cellular Biochemistry*, 130, pp. 1-9.

[13] I. Cosic, C. H. De Trad, Q. Fang, Q., M. Akay, (2000). Protein Sequences Analysis Using the RRM Model and Wavelet Transform Methods: A Comparative Study Analysis. Proc. of the IEEE-EMBS Asia-Pacific Conference on Biomedical Engineering, pp. 405-406.

[14] I. Cosic, Q. Fang, Q., (2002). Evaluation of Different Wavelet Constructions (Designs) for Analysis of Protein Sequences. Proc. of the 14th Int. Conference on DSP, Santorini, Greece.

[15] C. H. De Trad, Q. Fang, I. Cosic, (2000). The Resonant Recognition Model (RRM) Predicts Amino Acid Residues in Highly Conservative Regions of the Hormone Prolactin (PRL). *Biophysical Chemistry*, 84/2, pp. 149-157.

[16] C. H. De Trad, Q. Fang, I. Cosic, (2002). Protein sequences comparison based on the wavelet transform approach. *Protein Engineering*, Vol.15, no.3, pp 193-203.

[17] E. Pirogova, Q. Fang, M. Akay, I. Cosic, (2002). Investigation of the structure and function relationships of Oncogene proteins. *Proceeding of the IEEE*, Vol. 90, No. 12, pp. 1859-1867.

[18] E. Pirogova, I. Cosic, (2002). The use of Ionisation Constants of amino acids for protein signal analysis within the Resonant Recognition Model – Application to Proteases. *Molecular Simulation*, Vol.28, No.8-9, pp.845-851.

[19] E. Pirogova, G. P. Simon, I. Cosic, (2003). Investigation of the Applicability of Dielectric Relaxation Properties of Amino Acid Solutions within the Resonant Recognition Model. *IEEE Transactions on NanoBioscience* (in press).