APPLICATION OF SIGNAL PROCESSING IN COMPUTATIONAL ANALYSIS OF HSPs INTERACTIONS

Elena Pirogova¹,², Irena Cosic¹,², senior member of the IEEE

¹ Australian Centre for Radiofrequency Bioeffects Research (ACRBR)
² School of Electrical and Computer Engineering RMIT University
GPO Box 2476V Melbourne VIC 3001
Australia

Abstract

Heat shock proteins (HSPs) are a group of proteins that are present in all cells. HSP immunotherapy is believed to be one of the most promising areas of developed cancer treatment technology that is characterized by a unique approach to every tumor. HSPs are induced when a cell is influenced by environmental stresses like heat, cold and oxygen deprivation. Under perfectly normal conditions HSPs act like ‘chaperones’ helping new or distorted proteins fold into shape essential for their function, shuttle proteins and transport old proteins to ‘garbage disposals’ inside the cell. HSPs also play a significant role in helping the immune system recognize diseased cells. Twenty years ago HSPs were identified as the element responsible for protecting animals from cancer, and studies towards anti-tumor vaccine development still continue today. Here we have computationally analyzed HSPs, EGF and oncogene proteins aiming to find structural similarities of EGF, EGFR, oncogene and proto-oncogene proteins that can underline possible interaction of these proteins with HSPs. The resonant recognition model (RRM) was employed in this study to perform structure-function analysis of selected proteins and determine the RRM frequencies that represent the characteristic feature of protein biological behavior.

Key Words
Molecular modeling, protein function, Heat Shock Protein

1. Introduction

In order to treat tumors purified HSP-peptide complexes are directly introduced to dendritic cells, which alert the immune system by displaying these peptides on the cell surfaces. It leads to activation of “killer” T cells, which target and destroy the diseased cells. By isolating HSPs complexes from a patient’s tumor, the vaccines are meant to be developed to capture the specific cancer’s “fingerprint” in order to program the immune system to target only cells able to bear this “fingerprint”. Similarly, by complexion HSP with common viral proteins, it is intended to target only infected cells, which bear these viral proteins. The proprietary HSP technology is designed to allow extreme specificity in cancer and infectious disease treatment, thereby preserving healthy tissue to eliminate the side effects that are common with more traditional treatments.

Here we analyze HSPs using a computational method, the so-called RRM. Biological processes are based on selective interactions between particular bio-molecules. The RRM [1,2] was invented for structure-function analysis of proteins and DNA. The RRM approach presents a physico-mathematical model that interprets protein sequence linear information using methods of digital signal analysis. The RRM main concepts are:

- biological function of a given protein is considered as a selective interaction of the protein and its target
- intermolecular interactions present the transfer of resonant energy between the interacting molecules at the frequency specific for each interaction.

Knowledge of the protein characteristic frequencies leads to the potential to determine functionally important amino acids within the protein sequence and thus, to propose effective mutations, identify protein active sites and design new peptides with the desired biological features [1-8]. Once an understanding of the nature of protein interactions and their selectivity is reached, this knowledge will benefit molecular biology, pharmacology and biotechnology.

2. Body of Paper

2.1 Methodology

The RRM presents an uncommon engineering approach to analysis of proteins and DNA. It is assumed that proteins with the same biological function or interactive activity
have the same periodic components in the distribution of delocalized electron energies along the protein molecule. This postulate is supported by the fact that electrons delocalized in the particular amino acid have the strongest impact on the electronic distribution of the whole protein sequence. The RRM is based on the findings that there is a significant correlation between spectra of the numerical presentation of amino acids and their biological activity [1,2]. The RRM involves a two-stage analysis of the protein primary structure, i.e. the sequence of linked amino acids. At the first stage the amino acid sequence is transformed into a numerical sequence where each amino acid is represented by the value of the electron-ion interaction potential (EIIP), which describes the average energy states of all valence electrons in the particular amino acid [9]. Thus, the resulting numerical series represents the distribution of the free electron energies along the protein backbone.

At the next stage the obtained numerical series are analyzed by digital signal analysis methods, Fourier and Wavelet Transform, in order to extract structural and functional information pertinent to the studied proteins [10-16]. To determine the common frequency components in the spectra for all studied protein sequences within the group the multiple cross-spectral function is used. Peaks in this function denote common frequency components for the whole protein group. The signal-to-noise ratio (S/N) for each peak is considered as a measure of similarity between the analyzed sequences.

Through an extensive study, the RRM has reached a fundamental conclusion: one RRM characteristic frequency characterizes one particular biological function or interaction [1-8]. This frequency is related to the biological function provided the following criteria are met:

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

Based on the RRM characteristic frequency it is then possible using Inverse Fourier Transform to identify the individual amino acids, so-called "hot spots" or domains, that contribute mostly to the characteristic frequency and thus, to protein's biological function [1-8,10-16]. Furthermore, based on identified protein characteristic frequency and phase we are able to design bioactive peptides having only this particular characteristic frequency and consequently expressing the desired biological function [7,8,14].

2.2 Results

Recent findings in cancer research have established a connection between epidermal growth factors (EGF), epidermal growth factor receptors (EGFR) and c-erbB-2 oncoprotein and formation of pancreatic tumor in humans. [17]. A significant role in the specific anti-tumor vaccine development is attributed to HSP-based immunotherapy. In this study HSPs, EGF, EGFR, oncogene and proto-oncogene proteins were investigated. The RRM was employed here aiming to assist in analysis of interaction between HSPs and other studied proteins. A multiple cross-spectral analysis was performed for each selected protein group as well as for the mutual combination of HSPs with EGF, EGFR, oncogene and proto-oncogene using the EIIP values (Fig.1-Fig.9). In those multiple cross-spectral functions the prominent peaks denote common frequency components. The abscissa represents RRM frequencies, and the ordinate is the normalized intensity. The RRM characteristic frequencies and S/N ratio values for each analyzed protein group were calculated and are shown in Table1.

Oncogenes are a specific group of growth factors that promote uncontrolled cell growth and proliferation. These proteins are derived from normal cellular growth factors, so-called proto-oncogenes, via a limited number of modifications: mutations, insertions or deletions. Because proto-oncogenes control the cell cycle, it is obvious that should a proto-oncogene be mutated the potential for an unregulated cell cycle results. An unregulated cell cycle is the essence of cancer. Cells lose their control and begin to divide uncontrollably, forming tissue masses, tumors, and the disease known as cancer. The expressions of EGF, EGFR, and the c-erbB-2 oncoprotein were immunohistochemically examined in human pancreatic carcinoma tissues [17]. The expressions of these 3 products correlated significantly with tumor invasion into the anterior and posterior areas surrounding the pancreas. It is known that EGF and its receptor EGFR are key drivers in the process of cell growth and replication. Increased activity at the EGFR receptor has been implicated in a variety of solid tumors: lung, colon, breast, prostate, gastric and ovarian cancer, and head and neck tumors. Excessive activation of the EGFR is associated with advanced stages of the cancer disease. Thus, research into EGFR activation leads to a better understanding of how cancer cells divide providing a targeted approach to cancer treatment.

HSPs are attractive and novel immunogens against cancer [18]. HSPs trigger immune response through a variety of intracellular and extracellular activities. Because HSPs are normally found inside cells, the existence of extracellular HSPs can indicate that a sick cell has died. Thus, extracellular HSPs help the immune system to generate a response that can eliminate infection or disease. Studies in cancer research suggested that HSP-based vaccines are able to decrease the growth of the primary cancer, the metastatic load and prolong the life-span [18,19]. Immunotherapy with such groups of HSP associated proteins, unique to every tumor, is believed to be the most suitable specific therapy tailored to each
individual cancer even without prior knowledge of the identity of such molecular repertoire [19].

Following the aim of investigation of structural similarities of EGF, EGFR, oncogene and proto-oncogene proteins that can underline possible interaction of these proteins with HSPs, the RRM analysis was performed under selected protein groups. Initially the structure-function analysis was applied to a group of 30-HSP sequences, and two peak frequencies were identified in their multiple cross-spectrum: \( f_1 = 0.0195 \pm 0.033 \) characterizing HSPs anti-tumor and immunoregulatory activity, and \( f_2 = 0.4248 \pm 0.033 \) (the same frequency was identified for FGF proteins in our previous studies [2,8]) that characterize the bioactivity of fibroblast growth factors (FGF) – ability to regulate cell growth and differentiation (Fig. 1).

The same procedure was repeated with EGF, EGFR, oncogene and proto-oncogene proteins and their mutual combination with HSPs. Multiple cross-spectral spectra of these proteins are shown in Fig.2-Fig.9.

Table 1. Peak frequency and Signal-to-Noise Values of proteins analyzed

<table>
<thead>
<tr>
<th>Protein group</th>
<th>( f )</th>
<th>S/N</th>
<th>No seq.</th>
<th>Standard Error, 1/No seq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSP</td>
<td>0.0195</td>
<td>277.24</td>
<td>30</td>
<td>0.033</td>
</tr>
<tr>
<td>EGF</td>
<td>0.0586</td>
<td>98.58</td>
<td>60</td>
<td>0.017</td>
</tr>
<tr>
<td>EGF, HSP</td>
<td>0.0596</td>
<td>292.20</td>
<td>90</td>
<td>0.011</td>
</tr>
<tr>
<td>EGFR</td>
<td>0.0576</td>
<td>151.10</td>
<td>10</td>
<td>0.100</td>
</tr>
<tr>
<td>EGFR, HSP</td>
<td>0.0195</td>
<td>473.24</td>
<td>40</td>
<td>0.025</td>
</tr>
<tr>
<td>Oncogene</td>
<td>0.0322</td>
<td>468.66</td>
<td>46</td>
<td>0.022</td>
</tr>
<tr>
<td>Oncogene, HSP</td>
<td>0.0322</td>
<td>483.84</td>
<td>76</td>
<td>0.013</td>
</tr>
<tr>
<td>Proto-oncogene</td>
<td>0.0576</td>
<td>222.39</td>
<td>15</td>
<td>0.067</td>
</tr>
<tr>
<td>Proto-oncogene, HSP</td>
<td>0.0186</td>
<td>396.41</td>
<td>45</td>
<td>0.022</td>
</tr>
</tbody>
</table>

Cross-spectral analysis of EGF proteins [2] revealed one prominent peak frequency at \( f = 0.0586 \pm 0.017, \) \( S/N = 98.58 \) (Fig. 2).
biological function, and consequently EGF and HSP can be involved in the same interactive process at this identified frequency.

The following cross-spectral analysis of the combined protein group of EGFR and HSP (Fig.5) revealed a single prominent frequency at $f=0.0195 \pm 0.025$, $S/N=473.24$, the same frequency as was found for HSP proteins (Table 1). Therefore we can conclude that this informational frequency is considered to be a relevant parameter for biological recognition between HSPs and EGFR and underlines the possibility for these two proteins to be involved in the same biological process.

RRM approach was applied for analysis of 28 viral and 18 cellular oncogenic proteins. As is evident from Fig.6, there is one prominent frequency component at $f=0.0322 \pm 0.022$, $S/N=468.66$ in the cross-spectral function, common to all analyzed protein sequences and characterizing a common biological behavior of this group of oncogene proteins, i.e their ability to transform cells [2,14]. Interesting results were obtained in analysis of structure-function relationships of the combined group of HSPs and oncogene proteins (Fig.7). We already pointed out that the protein and its target have different biological roles but they can participate in the same biological process [1,2].

This means they should interact at the same frequency, which is proposed to be a crucial condition for their mutual recognition. Observing consensus spectra of HSP and oncogene proteins (Fig.7) we found one prominent peak at $f=0.0322 \pm 0.013$, $S/N=483.84$. The presence of this peak with the significant $S/N$ implies that all of the analyzed sequences within this group have one frequency component in common, implying that oncogene and HSP may participate in interactive processes.

Also RRM analysis was used to analyze 15 proto-oncogene sequences and one dominant peak (Fig.8) was identified at frequency $f=0.0576 \pm 0.067$, $S/N=222.39$ (same frequency as for EGF proteins). We propose that this frequency presents the main characteristic feature of proto-oncogene bioactivity – normal cell growth control. However, from multiple cross-spectrum of proto-oncogene we observe several less significant peaks corresponding to the ability of these proteins to participate in more than one biological process. Examining this
spectrum we can see another frequency at \( f=0.0322\pm0.067 \), which is a characteristic feature of oncogene proteins, and is common frequency for both oncogene and proto-oncogene proteins [2,4,14]. Based on this frequency oncogenes and proto-oncogenes can be involved in the same biological process. Also from Fig.8 we observe another peak at \( f=0.4248\pm0.067 \), which was found also in multiple cross-spectral function of HSPs (Fig.1) as a second, less significant, peak frequency. This spectral similarity implies that proto-oncogene and HSP can recognize each other and then interact based on this identified frequency \( f=0.4248 \). Cross-spectral analysis of combined group of HSP and proto-oncogene proteins revealed one dominant peak at frequency \( f=0.0186\pm0.022 \), S/N=396.41 (Fig.9). As was mentioned above the same (within the calculation error) frequency at \( f=0.0195\pm0.033 \) was found for HSPs and for combined group of HSP and EGFR proteins as well (Table 1).

Therefore, we can conclude that this identified frequency can be considered as the characteristic feature of the mutual recognition and participation in the same biological process between HSP and proto-oncogene proteins. Finally, the results of our computational analysis revealed the possibility of involvement of HSP in the interactive process with EGF, EGFR oncogene and proto-oncogene respectively. As a consequence of this interactive process it is suggested that HSPs may cause tumor growth arrest and degradation of oncogenic proteins.

3. 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 about protein biological activity [1,2,4-14]. This study extends the application of the RRM procedures to investigation of the role HSP proteins in affecting the biological functionality of EGF, EGFR, oncogene and proto-oncogene proteins. Based on results of our computational analysis of structure-function relationships of HSP with EGF, EGFR, oncogene and proteo-oncogene proteins we conclude that HSPs can play a key role in tumor growth arrest and affect the degradation of oncogenic proteins. Therefore, our computational predictions confirmed the experimental studies of other authors under HSPs, basic compounds in HSP immunotherapy of cancer. Thus, the RRM approach is able to determine the protein characteristic frequencies crucial for biological activity/interaction of analyzed proteins. It is concluded that the RRM characteristic frequency may dictate the specificity of the protein interactions because they characterize interactive process between macromolecules. No doubt the RRM approach presents a novel engineering tool, which enables researchers to look differently on selective process of interaction between macromolecules.

4. Acknowledgement

The authors are thankful for support of Australian Centre for Radiofrequency Bioeffects Research (ACRBR) funded by the NHMRC.

References