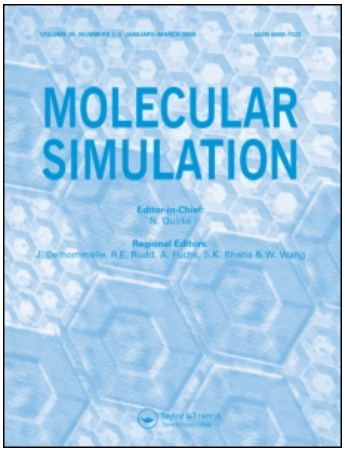

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Computational analysis of DNA photolyases using digital signal processing methods

E. Pirogova[ab]; V. Vojisavljevic[ab]; Q. Fang[ab]; I. Cosic[ab]
[a] School of Electrical and Computer Engineering, RMIT University, Melbourne, Victoria, Australia [b] Australian Centre of Radiofrequency Bioeffects Research (ACRBR), Melbourne, Australia

To cite this Article Pirogova, E. , Vojisavljevic, V. , Fang, Q. and Cosic, I.(2006) 'Computational analysis of DNA photolyases using digital signal processing methods', Molecular Simulation, 32: 14, 1195 — 1203
To link to this Article: DOI: 10.1080/08927020601052997
URL: http://dx.doi.org/10.1080/08927020601052997

Molecular Simulation, Vol. 32, No. 14, 15 December 2006, 1195–1203

Taylor & Francis
Taylor & Francis Group

Computational analysis of DNA photolyases using digital signal processing methods

E. PIROGOVA†‡*, V. VOJISAVLJEVIC†‡, Q. FANG†‡ and I. COSIC†‡

†School of Electrical and Computer Engineering, RMIT University, GPO Box 2476V, Melbourne, Victoria 3001, Australia
‡Australian Centre of Radiofrequency Bioeffects Research (ACRBR), Melbourne, Australia

(Received August 2006; in final form October 2006)

Sun light energy is used by plants to trigger their growth and development. However, an increase of UV-B light may lead to DNA damage. DNA photolyases are enzymes that repair the cyclobutane pyridine dimer (CPD) and 6–4 photoproduct lesions formed through UV irradiation of DNA. Many aspects of the repair process are under intense scientific investigation but still poorly understood. Here we have computationally analysed DNA-photolyases using the resonant recognition model (RRM), a physico-mathematical approach based on digital signal processing methods. The RRM proposes that protein interactions represent the transfer of resonant electromagnetic energy between interacting molecules at the particular frequency. Within this study we have determined photolyases characteristic frequency, "hot spots" amino acids corresponding to the functional mutations and functional active/binding sites, and designed photolyase peptide analogous. A mutual relationship between photolyase and p53 tumour suppressor protein has also been investigated. The results obtained provide new insights into the structure–function relationships of photolyase protein family.

Keywords: Active site; Characteristic frequency; Protein function; Signal processing

1. Introduction

DNA damage in cells exposed to ultraviolet (UV) radiation plays significant roles in cell-cycle arrest, activation of DNA repair, cell death and mutation, and neoplastic transformation. The major types of DNA damage induced by UVB (280–315 nm, component of sunlight) and by UVC (200–280 nm) are cyclobutane pyrimidine dimers (CPDs) and 6–4 photoproducts 6–4PPs). DNA photolyases, a type of flavoenzymes, bind to the damaged part of the DNA strand in a light-independent step and catalyse a repair of CPDs when exposed to near-UV or visible light ($300\,\text{nm} < \lambda < 500\,\text{nm}$) [1–6]. Understanding of the DNA repair process at the molecular level was greatly enhanced through the determination of the 3D X-ray structures of the DNA photolyases from two organisms, *Escherichia coli* [7] and *Anacystis nidulans* [8].

Photochemical reactions involving DNA have been linked to mutagenesis, carcinogenesis and cell death. Due to the importance of DNA photolyase as part of cancer prevention in many organisms and the unique mechanism of DNA photo-reactivation, this process has been a topic of much recent study. However, difficulties associated with the enzyme–DNA substrate complex have placed significant limitations on experimental studies. Computational studies have therefore played an important role in deciphering inconclusive and even conflicting experimental evidence, providing insights into the mechanisms of repair at an atomic level and models to guide experiment [9–11].

It is known that cells can accumulate damage to DNA. A cell will become cancerous when the right combination of genes is altered. However, some genes help to prevent cells malignant behaviour and therefore are referred to as tumour-suppressor genes. These genes can only contribute to cancer when they are inactivated or mutated. Mutant p53 can no longer bind to DNA in an effective way, and as a consequence the p21 protein is not made available to act as the "stop signal" for cell division. Thus, cells divide with no order and form tumours. 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 tumour-suppressor genes and activation of certain proto-oncogenes. Therefore an

*Corresponding author. Email: elena.pirogova@rmit.edu.au

Molecular Simulation
ISSN 0892-7022 print/ISSN 1029-0435 online © 2006 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/08927020601052997

analysis of mutual relationships between "natural defenders"—p53 tumour suppressor proteins and DNA-photolyases is off great importance in the development on new methodology and drug design for cancer treatment. In this study the RRM was employed for structure–function analysis of photolyase protein group and its interactions with p53 proteins.

2. Method and background

2.1 Resonant Recognition Model–characteristic frequency determination

Biological processes in living organism are driven by proteins. The occurrence of such processes is attributed to selective interactions between proteins and their targets—DNA regulatory segments, other proteins or small molecules. However, the principles governing this selectivity have not been clearly understood. The RRM is a physical and mathematical model that interprets protein sequence linear information using signal processing methods [12,13]. In the RRM a 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. It has been found through 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 and interaction [12,13]. Once the characteristic frequency for a particular protein function/interaction is identified, it is possible to utilize the RRM to predict which amino acids in the sequence predominantly contribute to this frequency and consequently to the observed function. It also becomes possible to design peptides having only the desired periodicities. Initially, these amino acids were identified using Inverse Fourier Transform [12–14]. The Wavelet transform, as a new signal-processing tool for the multi-resolution analysis and local feature extraction of non-stationary signals, has been recently incorporated into the RRM. In our recent studies continuous wavelet transform (CWT) was applied successfully for the determination of functional active sites of different protein families [15,16].

It has been shown that certain periodicities (frequencies) within the distribution of energies of delocalised electrons along the protein molecule are critical for the protein's biological function (i.e. interaction with its target) [12–19]. The application of the RRM involves two stages of calculation. The first is the transformation of the amino acid sequence into a numerical sequence. Each amino acid is represented by the value of the electron–ion interaction potential (EIIP) describing the average energy states of all valence electrons in a given amino acid. The EIIP values for each amino acid are calculated using the general model of pseudo-potentials [20]. A unique number can thus represent each amino acid or nucleotide, irrespective of its position in a sequence. Numerical series obtained this way are then analysed by digital signal analysis methods in order to extract information relevant to the biological function. The original numerical sequence is transformed to the frequency domain using the Discrete Fourier Transform (DFT). To determine the common frequency components for a group of protein sequences, a multiple cross-spectral analysis is used [12,13]. Peak frequencies in such a multiple cross-spectral function denote common frequency components for all sequences analysed. Signal-to-noise ratio (*S*/*N*) for each peak is defined as a measure of similarity between sequences analysed and is calculated as the ratio between signal intensity at the particular peak frequency and the mean value over the whole spectrum. The presence of a peak frequency with significant *S*/*N* ratio in a consensus spectrum implies that all of the analysed sequences within the group have one frequency component in common [12,13]. This frequency is related to the biological function provided the following criteria are met:

- Only one peak 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.

In our previous studies, the above criteria have been tested with over 1000 proteins from 25 functional groups [12,13]. The following fundamental conclusion was drawn from our studies: one RRM peak frequency characterises one particular biological function or interaction. Therefore, those peaks are termed as the RRM characteristic frequencies [12–14].

2.2 "Hot spots" in terms of the RRM

It is known that proteins cannot express their biological function until they achieve a certain active 3D conformation. By identifying the characteristic frequency of a particular protein, it is possible to predict which amino acids in the sequence predominantly contribute to the frequency and consequently to the observed function [12–14]. Since the characteristic frequency correlates to biological function, the positions of the amino acids that are most affected by change of amplitude at the particular frequency can be defined as "hot spots" for the corresponding biological function. The strategy for this prediction includes the following steps [12,13]:

- Determine the unique characteristic frequency for the specific biological function by multiple cross-spectral analysis for the group of sequences with the corresponding biological function.

- Alter the amplitude at this characteristic frequency in the particular numerical spectrum. The criterion used for identifying the critical characteristic frequency change is the minimum number of "hot spot" amino acids that are least sensitive to further changes in the amplitude of the characteristic frequency.
- Derive a numerical sequence from the modified spectrum using IFT.

It is known that a change in amplitude at one frequency in the spectrum causes changes at each point in the numerical sequence. Thus a new numerical series is obtained where each point is different from those in the original series. Detecting the amino acids corresponding to each element of this new numerical sequence can then be achieved using tabulated values of the EIIP or other appropriate amino acid parameters. The amino acids in the new sequence that differ from the original ones reside at the points which most contribute to the frequency. These hot spots are related to this frequency and to the corresponding biological function [12,13]. The procedure described above was used in a number of examples: IL-2 [21], haemoglobins, myoglobins and lysozymes [22], repressors and transforming proteins [23], glucagons, TNFs [13,14], EGFs, FGFs [24], oncogenes [13,16], etc [13]. These examples have shown that such predicted amino acids denote residues crucial for protein functions. Consequently, these "hot spot" amino acids are found spatially clustered in the protein's 3D structure in and around the protein active site. As these specific amino acids most strongly influence the characteristic frequency, their cluster represents a site in the protein where the signal of characteristic frequency for the specific protein property is dominant. As this cluster of amino acids has been found positioned in and around the active site (figure 2), it is proposed that these specific amino acids play a crucial role in determining the structure of the active site and possibly, the active 3D structure of the whole molecule [13,14,16,21–24].

2.3 Bioactive peptide design

Knowing the RRM characteristic frequencies and corresponding phases for particular biological functions, it is possible to design amino acid sequences having those spectral characteristics only. The strategy for the design of such defined peptides is as follows [13]:

- Within the multiple cross-spectral analysis of the group of protein sequences sharing the corresponding biological function, determine the unique RRM frequency characterising this specific biological function/interaction.
- Define the characteristic phases at the characteristic frequencies for the particular protein that is chosen as the parent for agonist/antagonist peptide design.
- Derive a numerical sequence from the known characteristic frequencies and phases. This can be done by summing sinusoids of the particular frequencies, amplitudes and phases. The length of the numerical sequence is defined by the appropriate frequency resolution, and the required peptide's length.
- Determine the amino acids that corresponds to each element of the new numerical sequence, either tabulated EIIP values or other appropriate amino acid parameters [13,16,24,25].

It is expected the designed peptide will exhibit the desired biological activity.

2.4 Continuous Wavelet Transform (CWT)

Using Inverse Fourier transform (IFT) we can only identify a small number of single amino acids, which predominantly contribute to the particular frequency. However, the protein active site is usually built up of domain(s) within the protein molecule. Applying the wavelet transform we observe a whole frequency/spatial distribution and thus, are able identify the domain(s) of high energy of a particular frequency along the sequence. The wavelet transform is an extension of Fourier transform which decomposes a signal *s(t)* by using a set of dilated and translated function of the original mother wavelet $\psi(t)$. The wavelet transform can be viewed as an inner product operation that measures the similarity or cross-correlation between the signal and the wavelets at different scales and translations. CWT is one of the time-frequency representations which can provide the information on how the spectral content of the signal evolves with time. The time-frequency analysis provides an ideal tool to dissect, analyse and interpret signals with transients or localised events. CWT provides the same time/space resolution for each scale therefore it may be used to localize individual events, such as the active site identification. The particular wavelet chosen here for "hot spot" amino acids identification is the Morlet, which is a locally periodic wave train:

$$cwt(a,b) = \int s(t)\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)\mathrm{d}t \qquad (1)$$

where $\omega_0 = 5$ and c is the constant used for normalization. This particular wavelet transform was chosen here to identify the location(s) of active site(s) of the selected protein molecule for its ability to help locate events simultaneously in both frequency and time [15,16,19].

2.5 The RRM physical basis

In the RRM protein interaction is considered as resonant energy transfer between interacting molecules. This energy may be transferred through oscillations of a physical field, possibly electromagnetic in nature. There is much evidence that biological processes can be induced or modulated by induction of light of particular characteristic frequencies [26–29]. This is caused directly by the light-induced

changes of energy states of macromolecules, in particular proteins. The function of some proteins is directly connected with the absorption of visible light of defined wavelengths as in the case of rhodopsins. Rhodopsin absorption properties are attributed to the presence of a colour prosthetic group bound to the protein, while the frequency selectivity of this absorption is defined by the amino acid sequence of the protein *per se*. In addition, there is evidence that light of a defined frequency can induce or enhance certain biological processes such as cell growth and proliferation [30], which are normally controlled by proteins. All these frequency selective effects of light on biological processes of protein activation imply that protein activation involves energies of the same order and nature as the electromagnetic irradiation of light.

Proteins and DNA have certain conducting or semiconducting properties [13,26–29]. Consequently, a charge moving through the molecule's backbone and passing different energy stages caused by different side groups of various amino acids or nucleotides provides sufficient conditions for the emission of electromagnetic waves. Their frequency range depends on charge velocity, which then depends on the nature of charge movement (superconductive, conductive, soliton transfer, etc.) and on the energy of the field that causes this charge transfer. The nature of this physical process is still unknown. Some accepted models denote the possibility that the complexity of the system does not allow precise calculations of charge velocity. According to the RRM postulates [12,13], despite the fact that conductive electron transfer at the N- and C-terminals of the protein is governed by poorly understood conditions, the potential energy difference is:

$$W = W(\mathrm{COOH}) - W(\mathrm{NH2}) = 0.13\mathrm{Ry} \qquad (2)$$

This energy difference allows for a maximum velocity of the electrons which is equal to:

$$V_{\max} = \sqrt{(2eW/m)} \qquad (3)$$

where e is the electron charge and m is electron mass. Therefore

$$V < 7.87 \times 10^5\ \mathrm{m/sec} \qquad (4)$$

An inherent assumption is that amino acids in the protein are equidistant and the distance between each amino acid is $d = 3.8$ Å. Thus, the maximum frequency that could be emitted during the electron transfer is $F_{\max} < V/(2d) < 1 \times 10^{15}$ Hz, while the corresponding wavelength is $L_{\min} > 330$ nm. The minimum frequency that could be emitted depends on the total length of the protein $F_{\min} = 2\ F_{\max}/N < 1 \times 10^{13}$ Hz, where N is the total number of amino acids in the protein. For example, with proteins of 200 a.a. in length, the minimum frequency is $F_{\min}$ and the corresponding wavelength is $L_{\max} < 30{,}000$ nm [13]. The range from 30,000 to 300 nm is very wide, starting from the far IR through the visible to the UV regions.

The characteristic frequency identified for a particular protein group characterises one particular biological function/interaction. It has been postulated that the protein function is directly related to the absorption of light of defined wavelength. Within the RRM it was found that a strong linear correlation exists between the predicted and experimentally determined frequencies corresponding to the absorption of electromagnetic radiation of such proteins [12,13]. It is inferred that approximate wavelengths in real frequency space can be calculated from the RRM characteristic frequencies for each biologically related group of sequences. These calculations can be used to predict the wavelength of the light irradiation, which might affect the biological activity of proteins exposed [12,13].

A number of examples including light absorbing proteins, growth factor activation, enzyme activation, and red/far-red and blue-light receptors in plants have shown that there is a definite linear correlation between RRM frequency space f_{RRM} and corresponding light wavelengths λ in nm [13]:

$$\lambda = K/f_{\mathrm{RRM}} \qquad (5)$$

All these results lead to the conclusion that the specificity of protein interactions are based on the resonant electromagnetic energy transfer on a frequency specific for each observed interaction [12,13].

3. Results

In this study DNA-photolyase sequences were investigated concerning the understanding of the structure-function relationship within this protein group. All sequences are taken from the Expert Protein Analysis System (ExPASy) proteomics server of the Swiss Institute of Bioinformatics (SIB).

Here we have determined the RRM characteristic frequencies of analysed photolyases as a whole functional group (16 sequences) at $f_1 = 0.2715 \pm 0.0625$ and $f_2 = 0.4053 \pm 0.0625$, as well as the specific characteristics of different subgroups: 3 sequences from archaea, 9 sequences from bacteria and 4 eukaryotic sequences were analysed. A multiple cross-spectral analysis was performed for each selected protein group using the EIIP [13,20] values (figures 1–4).

The RRM analysis was also carried out on 13 p53 tumour suppressor proteins and their combination with photolyase proteins, with the aim of determining the characteristic frequency which would correspond for their mutual interactions (figures 5 and 6). The characteristic frequency and *S*/*N* ratio values of the combined group of p53 and photolyase sequences are shown in table 1 and figure 6.

Initially the structure–function analysis was applied to a group of 16 photolyase sequences, and two prominent peak frequencies were identified in their multiple cross-spectrum at $f_1 = 0.2715 \pm 0.0625$ and $f_2 = 0.4053 \pm 0.0625$

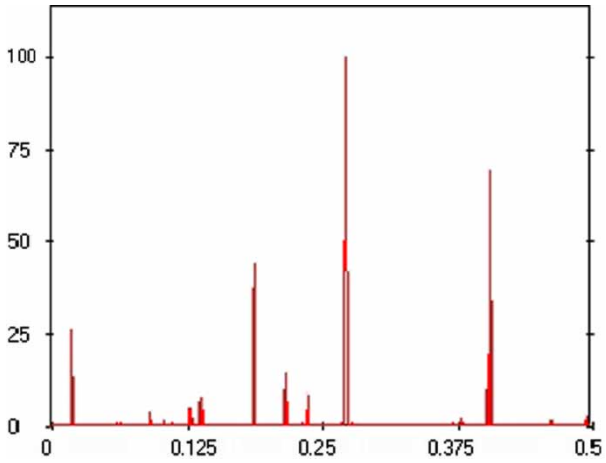

Figure 1. Multiple cross-spectral function of 16-DNA-photolyase sequences.

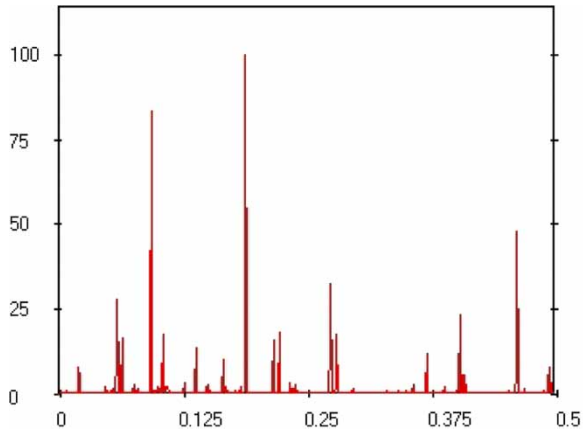

Figure 3. Multiple cross-spectral function of 9-DNA-photolyase sequences from bacteria.

(figure 1, table 1). It should be noted that a computationally predicted less prominent activation frequency $f_2 = 0.4053$ (495 nm) (figure 7) is located within the blue/UV-A region of light energy ($300\,\text{nm} < \lambda < 500\,\text{nm}$) determined experimentally [1–4], and used by DNA-photolyases for catalysis of the reactivation of irradiated/damaged DNA. Furthermore, from multiple cross-spectral function of 16 photolyases (figure 1) we can observe the existence of another more significant peak at $f_1 = 0.2715$ (740 nm) which is of great importance as it implies that photolyases can participate in more than one biological process (interact with other proteins).

The RRM procedure was repeated with photolyase subgroups where characteristic frequencies were identified at $f = 0.4502$ (for 3 sequences from archaea), $f = 0.1875$ (for 9 sequences from bacteria) and $f = 0.4043$ (for 4 eukaryotic sequences). Following the aim of understanding the structure–function relationship between photolyase and p53 tumour suppressor protein a multiple cross-spectral analysis was carried out. Consensus spectrum of 13 p53 protein sequences is shown in figure 5 while the consensus spectrum of the mutual combination pf photolyase and p53 proteins is presented in figure 6. Peak frequencies and *S*/*N* values of the analysed proteins are shown in table 1. The frequency $f = 0.4316$ identified within the RRM analysis is suggested to be considered a characteristic feature of the specific biological activity of p53 proteins—its activity stops the formation of tumours.

"Hot spot" analysis using the IFT was performed on *E. coli* DNA photolyase (1dnp). Predicted "hot spot" amino acids that correspond to functional mutations were found to be within the regions of: FAD binding domain 201–469: "hot spots" are at Gly219, Phe229, Gly234, Gly245, Gly246, Phe281, Gly382, Ser388 and Gly390; N-terminal domain 1–200: "hot spots" are at Gly138, Gly179; and DNA binding 312–331 domain: "hot spots" are at Gly320, Gly323. Similar analysis was undertaken with *Thermus thermophilus* photolyase (1iqr). The "hot spot" amino acid which most contribute to the characteristic frequency $f = 0.2715 \pm 0.0625$ and thus, to the biological function were identified. The computationally predicted "hot spots" were found to be positioned within the FAD domain 176–418 regions identified experimentally:

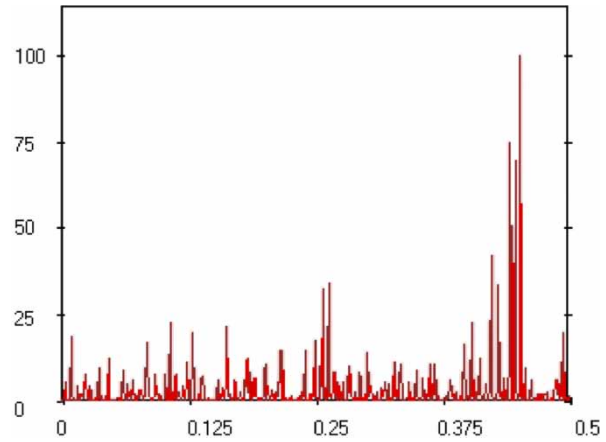

Figure 2. Multiple cross-spectral function of 3-DNA-photolyase sequences from archaea.

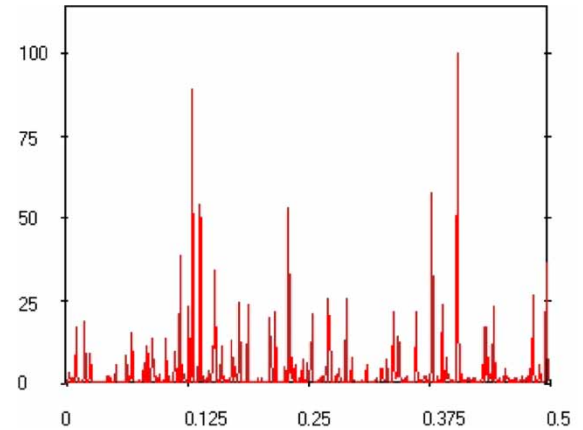

Figure 4. Multiple cross-spectral function of 4-photolyase sequences form eukaryotes.

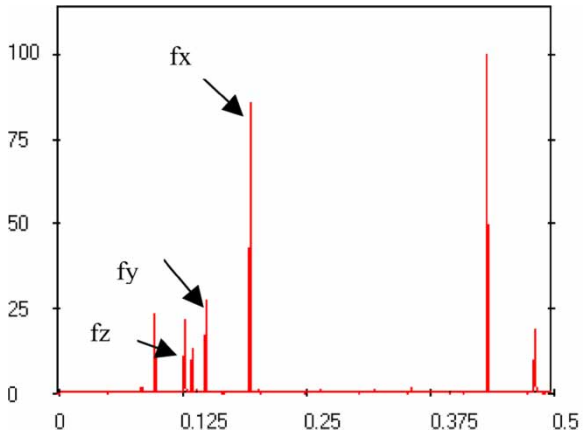

Figure 5. Multiple cross-spectral function of 13-p53 sequences.

Gly159, Gly163, Gly185, Phe216, Phe282, Gly288, Gly306, Phe319 and Gly358.

It should be mentioned that using IFT it is possible to identify only a number of single amino acids that are the greatest contributors to the particular frequency. However, the protein active site is usually built up of domain(s) within the protein sequence. Applying the wavelet transform leads to the possibility of observing a whole frequency/spatial distribution along the sequence and thus, identifying domains of high energy of particular frequency for this protein molecule. The results obtained within the study are considered to be useful input towards the prediction procedure of protein active sites. Here we have studied the performance of Morlet wavelet function. In our previous work [16,18,19] we have shown that Morlet wavelets were the most suitable for the identification of active sites of EGF, FGF and other protein sequence. The continuous scalograms of *E. coli* photolyase (1dnp) using Morlet wavelet function is shown in figure 9. The scalogram generated here has the maximum scale at 10. As mentioned above, the common frequency components of photolyase proteins corresponding to the common biological function (ability to reactivate the irradiated/damaged DNA) were found at $f_1 = 0.2715 \pm 0.0625$ and $f_2 = 0.4053 \pm 0.0625$. It can be observed that at lower frequencies (upper part of the scalogram) there are a definite areas of high energy between about the 100th and the 150th amino acid, as well as between the 225th and 240th. These domains correspond to the last part of amino acid's domains, experimentally predicted to be *E. coli* photolyase FAD binding domain (201–469) and N-terminal domain (1–200).

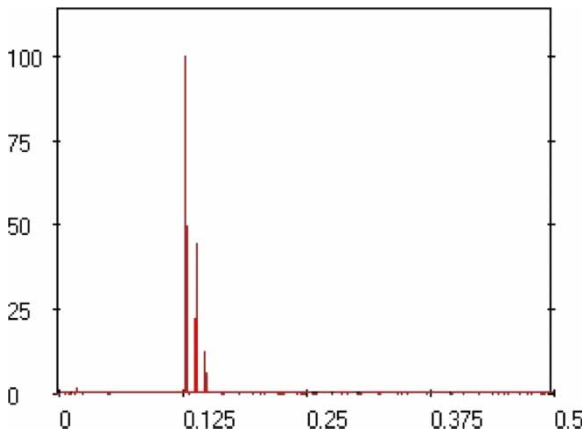

Figure 6. Multiple cross-spectral function of photolyase and p53 (29 sequences).

The RRM procedure has been applied to 1dnp and 1iqr photolyase sequences, aiming to design peptide analogues that would exhibit photolyase-like activity, i.e. ability to repair DNA damaged by irradiation. The design is based on the particular values of the characteristic frequency and phase determined earlier for each analysed protein: photolyase *E. coli* (1dnp) $f = 0.2715$, $\phi = 0.361$, and photolyase *T. thermophilus* (1iqr) $f = 0.2715$, $\phi = -2.900$. The designed sequences are shown below:

DHPDRGYDWLMDHPRREA (**parent protein is 1dnp**),
NTDPHDCLWDYLTDPPDT (**parent protein is 1iqr**).

These designed peptides have only the computationally identified spectral characteristic.

4. Discussion

It can be observed from the corresponding multiple cross-spectra (figures 1 and 5) that photolyase and p53 tumour suppressor proteins have the same frequencies identified at f_x, f_y and f_z. This agrees with the main RRM postulate that proteins and their targets recognize and interact with each other on the basis of the same (similar within the calculation error) characteristic frequency, which implies a possibility for photolyase and p53 proteins to be involved in the same biological process (interacting with each other). Analysing the mutual interactions between photolyases and p53 we have obtained one prominent frequency at $f = 0.1279 \pm 0.0345$ (table 1, figure 6). Accordingly to the RRM criteria the existence of one dominant peak in the cross-spectral function indicates that all analysed sequences within the group have this frequency component in common (share the common biological activity). Therefore, we can conclude that the identified frequency at $f = 0.1279$ can be considered a characteristic feature of the two-component mutual interactions between photolyase and p53 proteins revealing a potential ability of these proteins to defend cells against cancer by suppressing the tumour formation. Consequently, this might lead to the possibility of the development of the anti-tumour vaccines or other technology for counteracting many types of cancer.

In this study using the RRM approach we have performed the "hot spot" analysis of *E. coli* DNA photolyase (1dnp) and *T. thermophilus* photolyase (1iqr), which were used as examples for the prediction of the

Table 1. Peak frequency and *S/N* values of analysed proteins.

Protein group	*Frequency*	*Signal-to-Noise*	*Standard error, $1/N_o$*
Photolyase (16 seq)	0.2715	111.84	0.0625
Photolyase (3 seq from archaea)	0.4502	31.04	0.3333
Photolyase (9 seq from bacteria)	0.1875	82.90	0.1111
Photolyase (4 seq eukaryote)	0.4043	30.72	0.2500
p53 (13 sequences)	0.4326	159.97	0.0769
Photolyase and p53 (29 seq)	0.1279	321.9	0.0345

amino acids that would contribute mostly to the identified characteristic frequency and thus, to the observed protein function. The choice of these photolyases for our computational analysis is attributed to their well-known structural and functional characteristic obtained experimentally. *E. coli* photolyase repairs CPD in UV irradiated DNA by a reaction in which light energy drives electron transport from a catalytic chromophore, reduced flavin adenine dinucleotide (FADH), to the pyrimidine dimer, leading to its photolysis [1–4]. The enzyme is a monomeric protein of 471 amino acids. It includes two non-covalently attached cofactors, the blue light harvest cofactor methenyltetrahydrofolate (MTHF) and the catalytic cofactor FADH. Photolyase selectively binds to the CPD in a light-independent step. The MTHF cofactor absorbs a blue light photon and then excites FADH- by energy transfer. The excited FADH, *FADH, transfers an electron to split the CPD to regenerate the pyrimidines. The electron is transferred back to the photolyase, and the intact DNA dissociates (figure 8). Upon finding the characteristic frequency of photolyase sequences we were able to identify the so-called "hot spot" amino acids using IFT, in order to locate the amino acids related to the frequency $f_1 = 0.2715 \pm 0.0625$ identified previously. Predicted "hot spots" for the analysed sequence have been compared with other biological and crystallographic findings [31,32] and found to be positioned within the regions of FAD binding domain 201–469, N-terminal domain 1–200 and DNA binding 312–331 domain.

Similar analysis was performed on *T. thermophilus* photolyase (1iqr). The "hot spot" amino acid which most contribute to the characteristic frequency $f = 0.2715 \pm 0.0625$ and thus, to the biological function were identified. Again, these predictions have been compared with other biological and crystallographic findings relevant to the functional and binding sites of the analysed photolyase. The computationally predicted "hot spots" were found to be positioned within the FAD domain 176–418 regions identified experimentally. Results obtained reveal a close correlation between the predicted amino acids and active site identified via experimental research. Our previous research found that "hot spot" amino acids are found spatially clustered in the protein's 3D structure, in and around the protein active site forming the so-called "resonant boxes". As these specific amino acids most strongly influence the characteristic frequency, this clustering represents a site in the protein where the signal of characteristic frequency for the specific protein property is dominant. Due to this specific position of the amino acid cluster inside and around the protein active site, it is proposed that these specific amino acids play a crucial role in determining the structure of the

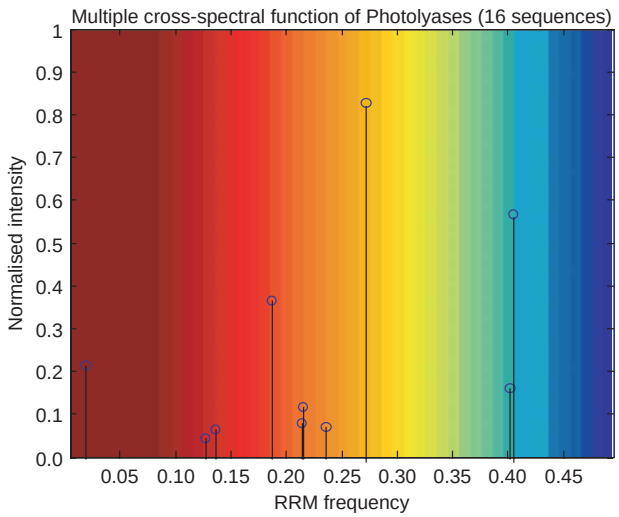

Figure 7. Distribution of photolyases characteristic frequencies along the electromagnetic spectrum.

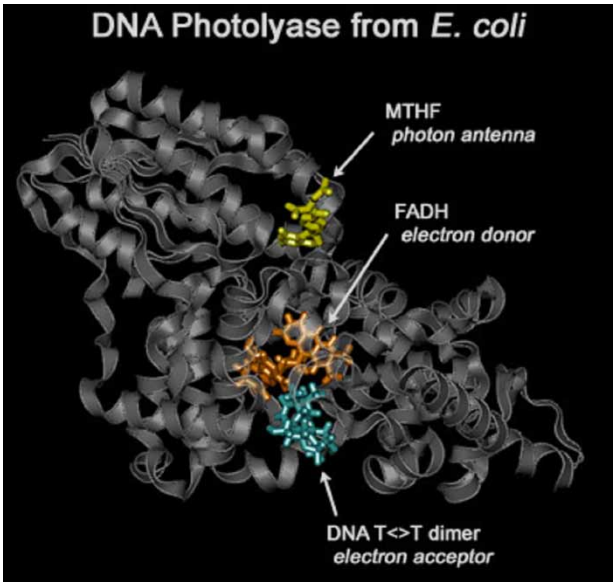

Figure 8. DNA photolyase from *E. coli* [31]. Ligand MHF, 5,10-methenyl-6-hydrofolic acid; Ligand FAD, flavin-adenine dinucleotide.

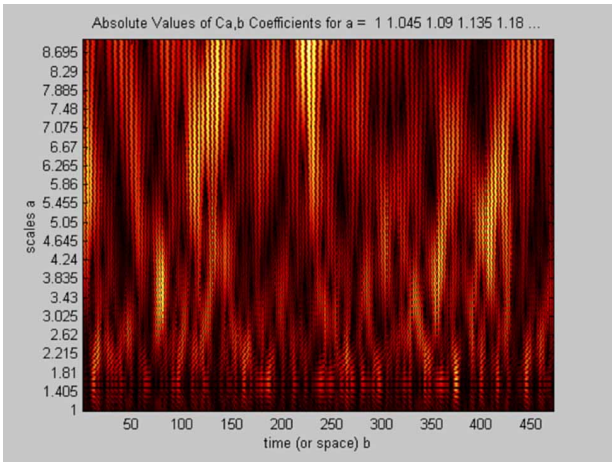

Figure 9. Scallogram of *E. coli* photolyase (scale 10).

active site and possibly, the active structure of the entire molecule [13].

Knowing the RRM characteristic frequencies and corresponding phases for particular biological functions, we have designed amino acid sequences having those spectral characteristics only and exhibiting photolyase-like activity, i.e. ability to repair irradiated/damaged DNA. The design is based on the particular values of the characteristic frequency and phase determined earlier for each analysed protein: photolyase *E. coli* (1dnp) $f = 0.2715$, $\phi = 0.361$ and photolyase *T. thermophilus* (1iqr) $f = 0.2715$, $\phi = -2.900$.

The designed peptide analogous DHPDRGYDWLMD-HPRREA (parental protein is 1dnp) and NTDPHDCLW-DYLTDPPDT (parental protein is 1iqr) have only the desired biological function related to the chosen characteristic frequency and purported to have the corresponding biological activity.

5. Conclusions

This study extends the application of the RRM approach to analysis of structure-function relationship of photolyase proteins. As was outlined in our aims above here we have identified frequencies of light that can activate photolyase proteins at $f_1 = 0.2715 \pm 0.0625$ and $f_2 = 0.4053 \pm 0.0625$. These computationally identified frequencies correspond, according to the RRM principles described above, to actual light wavelengths determined to be at 740 and 495 nm, respectively. We also predicted functional amino acids which most contributing to the proteins' biological activity, and designed photolyase peptide analogues having the same functionality as their parent proteins. The postulated interaction between photolyase and p53 tumour suppressor proteins has been examined as well. The common frequency identified for photolyase and p53 proteins at $f = 0.1279 \pm 0.0345$ implies that according to the RRM concepts photolyases maybe involved in the interactive biological process with tumour-suppressor proteins. Thus, photolyases may play a key role in the process of tumour growth arrest and consequently, be a crucial component in anti-tumour vaccines development. The results of our computational analysis presented could be useful in predicting photolyase active mutations, designing active peptide analogues as well as understanding DNA repair mechanisms.

Acknowledgements

We wish to acknowledge the financial support of our research provided by Australian Centre of Radiofrequency Bioeffects Research (ACRBR) and RMIT University's Research Investment Fund.

References

[1] J. Deisenhofer. DNA photolyases and cryptochromes. *Mutat. Res.*, **460**, 143 (2000).
[2] P.F. Heelis, R.F. Hartman, S.D. Rose. Photoenzymic repair of UV-damaged DNA: a chemist's perspective. *Chem. Soc. Rev.*, **24**, 289 (1995).
[3] P.F. Heelis, R.F. Hartman, S.D. Rose. Energy and electron transfer processes in flavoprotein-mediated DNA repair. *J. Photochem. Photobiol. A Chem.*, **95**, 89 (1996).
[4] S.-T. Kim, A. Sancar. Photochemistry, photophysics, and mechanism of pyrimidine dimer repair by DNA photolyase. *Photochem. Photobiol.*, **57**, 895 (1993).
[5] A. Sancar. No "end of history" for photolyases. *Science*, **272**, 48 (1996).
[6] T. Todo. Functional diversity of the DNA photolyase/blue light receptor family. *Mutat. Res.*, **434**, 89 (1999).
[7] H.W. Park, S.T. Kim, A. Sancar, J. Deisenhofer. Crystal structure of DNA photolyase from *Escherichia coli*. *Science*, **268**(52), 1866 (1999).
[8] T.K. Tamada, Y. Kitadokoro, K. Higuchi, A. Inaka, P.E. Yasui, A.P.M. de Ruiter, P.M. Eker, K. Miki. Crystal structure of DNA photolyase from *Anacystis nidulans*. *Nat. Struct. Biol.*, **4**, 887 (1997).
[9] L.E. Eriksson, A. Ehrenberg, J.S. Hyde. Comparative electron-nuclear double resonance study of two flavoproteins. *Eur. J. Biochem.*, **17**, 539 (1970).
[10] J.M. Hahn, M.-E. Michel-Beyerle, N. Rösch. Binding of pyrimidine model dimers to the photolyase enzyme: a molecular dynamics study. *J. Phys. Chem. B.*, **103**, 2001 (1999).
[11] C.B. Harrison, L.L. O'Neil, O. Wies. Computational studies of DNA photolyase. *Phys. Chem. A.*, **109**(32), 7001 (2005).
[12] I. Cosic. Macromolecular bioactivity: is it resonant interaction between macromolecules?—theory and applications. *IEEE Trans. Biomed. Eng.*, **41**, 1101 (1994).
[13] I. Cosic. *The Resonant Recognition Model of Macromolecular Bioactivity: Theory and Applications*, Birkhauser Verlag, Basel (1997).
[14] I. Cosic. Virtual spectroscopy for fun and profit. *Biotechnology*, **13**, 236 (1995).
[15] C.H. de Trad, Q. Fang, I. Cosic. Protein sequences comparison based on the wavelet transform approach. *Protein Eng.*, **15**(5), 193 (2002).
[16] E. Pirogova, Q. Fang, M. Akay, I. Cosic. Investigation of the structure and function relationships of oncogene proteins. *Proc. IEEE*, **90**(12), 1859 (2002).
[17] I. Cosic. The resonant recognition model of bio-molecular interactions: possibility of electromagnetic resonance. *Pol. J. Med. Phys. Eng.*, **7**(1), 73 (2001).
[18] C.H. de Trad, Q. Fang, I. Cosic. The Resonant Recognition Model (RRM) predicts amino acid residues in highly conservative regions of the hormone prolactin (PRL). *Biophys. Chem.*, (84/2), 149 (2000).

[19] Q. Fang, I. Cosic. Protein structure analysis using the resonant recognition model and wavelet transform. *APESM*, **21**(4), 179 (1998).

[20] V. Veljkovic, M. Slavic. General model of pseudopotentials. *Phys. Rev. Lett.*, **29**, 105 (1972).

[21] I. Cosic, M. Pavlovic, V. Vojisavljevic. Prediction of "Hot Spots" in Interleukin2, based on informational spectrum, characteristics of growth regulating factors. *Biochimie*, **3**, 333 (1989).

[22] I. Cosic, A.N. Hodder, M.I. Aguilar, M.T.W. Hearn. Resonant recognition model and protein topography: model studies with myoglobin, hemoglobin and lysozyme. *Eur. J. Biochem.*, **198**, 113 (1991).

[23] I. Cosic, M.T.W. Hearn. Studies on protein-DNA interactions using the resonant recognition model: application to repressors and transforming proteins. *Eur. J. Biochem.*, **205**, 613 (1992).

[24] I. Cosic, A.E. Drummond, J.R. Underwood, M.T.W. Hearn. A new approach to growth factor analogue design: modelling of FGF analogues. *Mol. Cell. Biochem.*, **130**, 1 (1994).

[25] 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. 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 (1998).

[26] G. Biscar. Photon enzyme activation. *Bull. Math. Biol.*, **38**, 29 (1976).

[27] T. Karu. Photobiological fundamentals of low-power laser therapy. *IEEE J. Quantum Electron.*, **QE-23**, 1703 (1987).

[28] T. Karu. Primary and secondary mechanisms of actions of visible to near-IR radiation on cells. *J. Photochem. Photobiol.*, **49**, 1 (1999).

[29] S. Naturheilpraxis mit Naturmedizin Kiontke. *Natural Radiation and Its Effects on Biological Systems*, Pflaum Verlag, (2000).

[30] I. Cosic, V. Vojisavljevic, M. Pavlovic. The relationship of the resonant recognition model to effects of low intensity light on cell growth. *Int. J. Radiat. Biol.*, **56**, 179 (1989).

[31] S.P. Jordan, M.S. Jorns. Evidence for a singlet intermediate in catalysis by *Escherichia coli* DNA photolyase and evaluation of substrate binding determinants. *Biochemistry*, **27**, 8915 (1988).

[32] D.B. Sanders, O. Wiest. A model for the enzyme–substrate complex of DNA photolyase and photodamaged DNA. *J. Am. Chem. Soc.*, **121**, 5127 (1999).