Macromolecular Bioactivity: Is It Resonant Interaction Between Macromolecules?—Theory and Applications

Irena Cosic, *Member, IEEE*

Abstract— **Biological processes in any living organism are based on selective interactions between particular biomolecules. In most cases, these interactions involve and are driven by proteins which are the main conductors of any living process within the organism. The physical nature of these interactions is still not well known. This paper represents a whole new view to biomolecular interactions, in particular protein-protein and protein-DNA interactions, based on the assumption that these interactions are electromagnetic in their nature. This new approach is incorporated in the Resonant Recognition Model (RRM), which was developed over the last 10 years. It has been shown initially that certain periodicities within the distribution of energies of delocalized electrons along a protein molecule are critical for protein biological function, i.e., interaction with its target. If protein conductivity was introduced, then a charge moving through protein backbone can produce electromagnetic irradiation or absorption with spectral characteristics corresponding to energy distribution along the protein. The RRM enables these spectral characteristics, which were found to be in the range of infrared and visible light, to be calculated. These theoretically calculated spectra were proved using experimentally obtained frequency characteristics of some light-induced biological processes. Furthermore, completely new peptides with desired spectral characteristics, and consequently corresponding biological activities, were designed.**

I. Introduction

THE ENTIRE genetic information of any living organism is written as linear information within DNA sequences and is coded by four different nucleotides. DNA molecules serve as backup for complete genetic information for the whole organism. The particular and well-defined fragments of this information, so-called coding sequences, are then translated, using complex molecular mechanisms, into other linear information then contained within protein sequences and coded with 20 different amino acids.

Proteins are the main conductors and workforce in any living process within a cell, tissue or organism. They are composed of sequentially linked amino acids but can only express their biological function when they achieve a certain active three-dimensional (3-D) structure. Their biological function as well as their active (3-D) structure is determined primarily by the amino acid sequence within the protein. Although both function and structure of a number of proteins are known, the crucial problem of understanding how the biological function and active 3-D structure are "written" within the protein sequence still remains. If biological function is considered as selective interaction of protein and its target, then a more fundamental question arises: What is the physical basis of this interaction and how is this selectivity achieved? Once this understanding has been gained, it should be possible to design peptides and even proteins *de novo*, with the chosen biological function, and thus to produce new and more efficient drugs and other biotechnological products.

There have been many attempts to discover the rules governing the coding of the biological function into the sequence of amino acids within the protein. Typical approaches deal with either homology characterization of specific features of the primary and secondary structure of proteins or molecular modeling of protein tertiary structure. Although such approaches permit a significant insight into protein structure and active site location, they still do not provide sufficient knowledge about the informational, structural and physicochemical parameters crucial to the selectivity of protein interactions which can be used for *de novo* design of peptide or protein analogs with the desired biological activity. The existing knowledge in the field of computer-aided molecular modeling and protein structure/function analysis can be classified in terms of the of primary, secondary and tertiary structure analysis of proteins.

A. Primary Structure Analysis

Primary structure analysis is concerned mainly with the search for homologies among amino acid sequences. The main concept behind this method is that proteins with the same biological function share amino acid sequence alignments and these homologous fragments carry the main information about the protein function [1]–[3]. However, similar sequences may appear occasionally in totally unrelated proteins, through convergent evolution of sequences with similar properties. On one hand, there are a number of cases of unexpected but significant resemblances between functionally dissimilar proteins, while on the other hand, there are cases of insignificant resemblance between functionally related peptides [3]. Thus, the available methods rarely reach a satisfactory level of predictive accuracy for *de novo* prediction of the biological function from sequence similarities or for rational peptide

Manuscript received July 22, 1993; revised July 15, 1994. This work was supported by the National Health and Medical Research Council of Australia, the Australian Research Council and the Monash University Faculty of Engineering Research Fund.

The author is with the Department of Electrical and Computer Systems Engineering, Monash University, Caulfield, VIC 3145 Australia.

IEEE Log Number 9405971.

design. Optimal alignment programs [2] with certain new improvements in sequence analysis are designed to distinguish between analogous and homologous sequences, but they are still based on sequence similarities. Problems remaining with optimal alignment programs include difficulties associated with the length of the sequence string and insertion of gaps in order to increase the number of matching residues. Inserting gaps too liberally and assigned gap-weights arbitrarily can lead to biologically irrelevant alignments. Combination of primary, secondary and tertiary structure homologies to overcome some of these problems has recently been attempted [4].

B. Secondary Structure Analysis and Predictions

In the absence of experimental knowledge about protein secondary structure, various empirical, computer-aided algorithms are available for its prediction from the known primary structure [5], with results often presented in the form of preferred regions of regular secondary folding for a given peptide chain [6], [7]. Most of these algorithms are based on the average probability that any particular amino acid residue will be found in an α-helical, ß-sheet or "random coil" conformation [5]–[7]. While these approaches can be relatively successful in predicting secondary structure (with probabilities in the region of 60–70% with selected examples) [6], direct relationships to protein function cannot generally be described with these methods [8].

C. Tertiary Structure Analysis

The folding of the linear, primary polymer chain of a protein into a defined 3-D structure results in a spatial relationship between the various constituent amino acids which is found to be crucial for determining the functional behavior of the protein. In particular, a widely accepted model of protein interactions proposes that selectivity of these interactions is based on the structural matching between active sites of interacting molecules. Experimentally, the tertiary structure and stability of proteins have been studied using such techniques as X-ray crystallography, circular dichroism, fluorescence spectroscopy and 2-D NMR [9]. However, these methods have been limited due to the need for relatively large amounts of protein and the inability of many techniques to detect low abundances of conformational intermediates. These methods may also be limited by structural distortions caused by particular techniques, e.g., crystallization. The increasing data base of experimentally derived protein primary structures combined with the computer algorithms for performing molecular mechanics and dynamics has the potential to establish computational algorithms as a powerful tool to study protein tertiary structure and predict peptide/protein active conformations [10]–[12]. These methods are still not able to predict protein 3-D structure solely from its sequence, and thus they usually use a number of constraints from experimental measurements, or they are based on the sequential and functional homology to the proteins with known 3-D structure.

Although all of these techniques have significant success in prediction of protein structure and/or function, they are still not sufficient for understanding protein interactions, and consequently for the design of peptides or proteins with desired bioactivity. The lack of knowledge about the informational parameters of protein sequence important to the protein biological function, as well as a lack of understanding of the physical processes which are behind the biological activity of the proteins, has limited the success of these techniques. The investigations presented in this paper are aimed at solving some of these problems.

II. Resonant Recognition Model (RRM)

The new physicomathematical approach presented here is called the Resonant Recognition Model (RRM) [13]–[22]. The RRM is based on the representation of the protein primary structure as a numerical series by assigning to each amino acid a physical parameter value relevant to the protein's biological activity. Although a number of amino acid indices (222 have been published up to now [23], [24]) have been found to correlate in some way with the biological activity of the whole protein, our investigations [13]–[22], as well as studies of other authors [25]–[27], have shown that the best correlation can be achieved with parameters which are related to the energy of delocalized electrons of each amino acid. These findings can be explained by the fact that the electrons delocalized from the particular amino acid have the strongest impact on the electronic distribution of the whole protein. In our studies, the energy of delocalized electrons (calculated as the electron-ion interaction pseudopotential, EIIP [28], [29]) of each amino acid residue was employed. The resulting numerical series then represented the distribution of the free electrons' energies along the protein. This numerical series was then converted into a discrete Fourier spectrum which carried the same information content about the arrangement of amino acids in the sequence as the original numerical sequence [30]. The initial step in the RRM, where protein sequences are represented as numerical spectra, was named Informational Spectrum Method (ISM) [13], [15]. Later, when it was found that proteins can recognize their targets on the basis of the same characteristic frequency, which is in fact resonant recognition, the whole model was renamed with a more appropriate name: The Resonant Recognition Model (RRM).

Approaches similar to the RRM, based on the Fourier transform and physical characteristics of amino acids, have been successfully applied by Mandell who has shown that the characteristic hydrophobic mass energy Fourier modes are signatures of isomorphism and immunological reactivates [31]. Viari *et al.* have used our RRM approach with scale independent coding to localize biologically relevant patterns in calcium-binding proteins [32].

A. Definition of Common Frequency Characteristics

The RRM is a physical and mathematical model which interprets protein sequence linear information using signal analysis methods. It comprises two stages: The first involves 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) [16] which describes

TABLE I
THE ELECTRON-ION INTERACTION POTENTIAL (EIIP) VALUES FOR NUCLEOTIDES AND AMINO ACIDS

nucleotide	EIIP (Ry)
A	0.1260
G	0.0806
T	0.1335
C	0.1340

amino acid	EIIP (Ry)
Leu	0.0000
Ile	0.0000
Asn	0.0036
Gly	0.0050
Val	0.0057
Glu	0.0058
Pro	0.0198
His	0.0242
Lys	0.0371
Ala	0.0373
Tyr	0.0516
Trp	0.0548
Gln	0.0761
Met	0.0823
Ser	0.0829
Cys	0.0829
Thr	0.0941
Phe	0.0946
Arg	0.0959
Asp	0.1263

the average energy states of all valence electrons, in particular amino acids. The EIIP values for each amino acid were calculated using the following general model pseudopotential [29]:

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

where q is a change of momentum k of the delocalized electron in the interaction with potential w, while

$$Z = \left(\sum Z_i\right)/N$$

where Z_i is the number of valence electrons of the i-th component of each amino acid and N is the total number of atoms in the amino acid. The EIIP values for 20 amino acids, as well as for five nucleotides (the whole procedure can be applied to the DNA and RNA, too), are shown in Table I. Each amino acid or nucleotide, irrespective of its position in a sequence, can thus be represented by a unique number.

Numerical series obtained this way are then analyzed by digital signal analysis methods in order to extract information pertinent to the biological function. The original numerical sequence is transformed to the frequency domain using the discrete Fourier transform (DFT). As the average distance between amino acid residues in a polypeptide chain is about 3.8 Å, it can be assumed that the points in the numerical sequence derived are equidistant. For further numerical analysis, the distance between points in these numerical sequences is set at an arbitrary value $d = 1$. Then, the maximum frequency in the spectrum is $F = 1/2\ d = 0.5$. The total number of points in the sequence influences the resolution of the spectrum only. Thus, for N-point sequence the resolution in the spectrum is equal to $1/N$. The n-th point in the spectral function corresponds to the frequency $f = n/N$.

In order to extract common spectral characteristics of sequences having the same or similar biological function, the following cross-spectral function was used:

$$S_n = X_n Y_n^* \qquad n = 1, 2, \ldots, N/2$$

where X_n are the DFT coefficients of the series $x(m)$ and Y_n^* are complex conjugate DFT coefficients of the series $y(m)$. Peak frequencies in the amplitude cross-spectral function define common frequency components of the two sequences analyzed. The whole procedure: protein sequence → numerical series → amplitude spectra → cross-spectra, is represented in Fig. 1 using the example of acidic and basic fibroblast growth factors.

To determine the common frequency components for a group of protein sequences, we have calculated the absolute values of multiple cross-spectral function coefficients M, which are defined as follows:

$$|M_n| = |X1n|.|X2n| \ldots |XMn| \qquad n = 1, 2, \ldots, N/2.$$

Peak frequencies in such a multiple cross-spectral function denote common frequency components for all sequences analyzed. Signal-to-noise ratio (S/N) for each peak is defined as a measure of similarity between sequences analyzed. S/N is calculated as the ratio between signal intensity at the particular peak frequency and the mean value over the whole spectrum. The extensive experience gained from previous research [13]–[22] suggests that an S/N ratio of at least 20 can be considered significant. The multiple cross-spectral function for a large group of sequences with the same biological function has been named "consensus spectrum." The presence of a peak frequency with a significant signal-to-noise ratio in a consensus spectrum implies that all of the analyzed sequences within the group have one frequency component in common. This frequency is related to the biological function provided the following criteria are met:

1) One peak only exists for a group of protein sequences sharing the same biological function.
2) No significant peak exists for biologically unrelated protein sequences.
3) 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 [13]–[22]. Multiple cross-spectral functions of four different functional groups of proteins are represented in Fig. 2. The regulatory DNA sequences were analyzed in the same way. The following fundamental conclusion was drawn from our studies: Each specific biological function of protein or regulatory DNA sequence is characterized by a single frequency.

These results are summarized in Table II where each functional group of proteins or DNA regulatory sequences is shown with its characteristic frequency and corresponding S/N ratio within the multiple cross-spectral function.

B. The Physical Meaning of the Characteristic Frequency

The correlation between the amplitude spectrum of numerical representation of genetic sequences and the corresponding

FGF basic bovine
PALPEDGGSGAFPPGHFKDPKRLYCKNGGF
FLRIHPDGRVDGVREKSDPHIKLQLQAEER
GVVSIKGVCANRYLAMKEDGRLLASKCVTD
ECFFFERLESNNYNTYRSRKYSSWYVALKR
TGQYKLGPKTGPGQKAILFLPMSAKS

FGF acidic bovine
FNLPLGNYKKPKLLYCSNGGYFLRILPDGT
VDGTKDRSDQHIQLQLCAESIGEVYIKSTE
TGQFLAMDTDGLLYGSQTPNEECLFLERLE
ENHYNTYISKKHAEKHWFVGLKKNGRSKLG
PRTHFGQKAILFLPLPVSSD

(a)

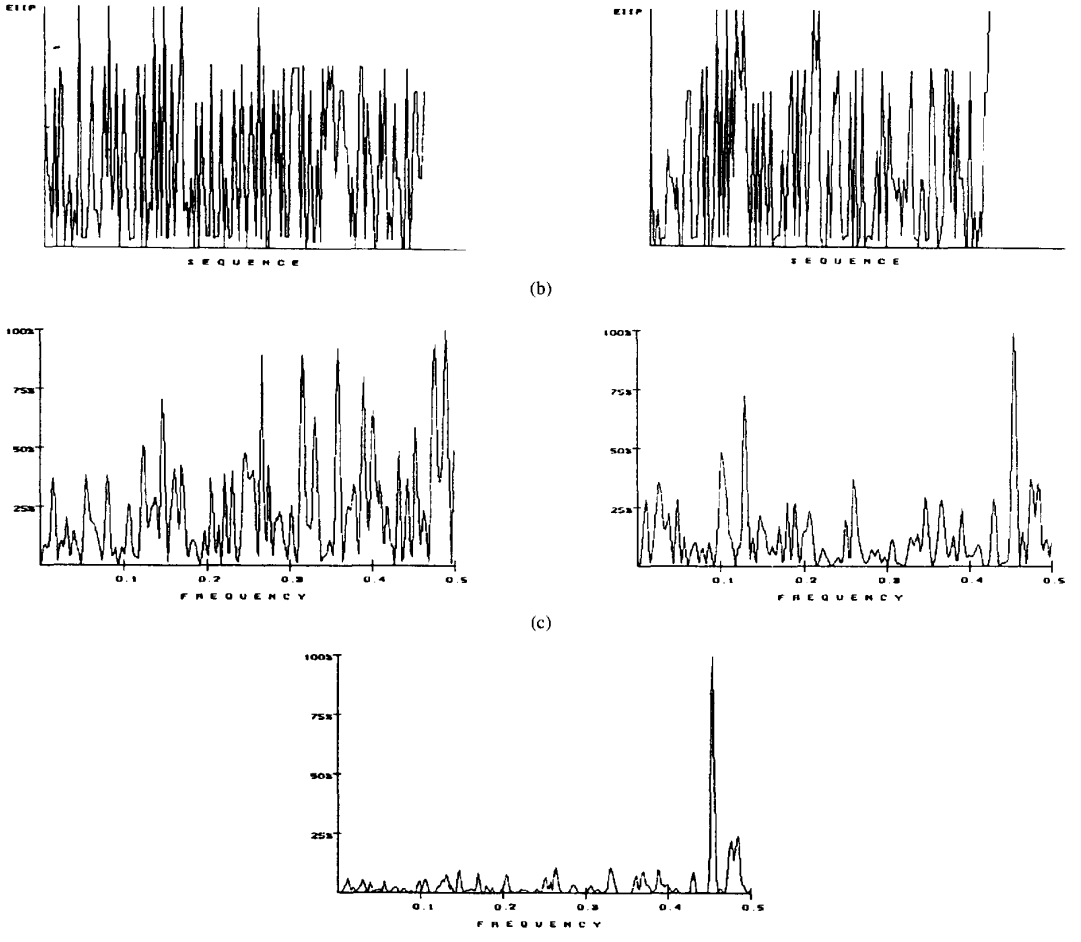


Fig. 1. The RRM procedure: (a) sequences of basic and acidic bovine FGF's, (b) graphical representation of the corresponding numerical sequences obtained by replacing every amino acid with its EIIP value; (c) spectra of both basic and acidic FGF; (d) cross-spectral function of the spectra presented in (c). The prominent peaks denote common frequency components. The abscissa represents RRM frequencies and the ordinate is normalized intensity.

biological function presented previously can lead to a completely new approach to protein dynamics. Each frequency in the RRM characterizes one biological function. To grasp the meaning of characteristic frequency, it is important first to understand what is meant by the biological function of proteins. Each biological process involves a number of interactions between proteins and their targets (other protein, DNA regulatory segment or small molecule). Each of these processes involves energy transfer between interacting molecules. These interactions are highly selective and this selectivity is defined within the protein structure. Protein and their protein or DNA targets have been analyzed to find out whether RRM characteristic frequencies denote a parameter which describes this selectivity between interacting molecules. It has been shown that proteins and their DNA or protein targets share the same characteristic frequency [13], [15], [17], [22], but of opposite phase [17], [22] for each in a pair of interacting macromolecules. Thus, it can be postulated that RRM characteristic frequencies characterize not only general functions but also provide recognition between a particular protein and its target (receptor, ligand, etc.). As this recognition arises from the matching of periodicities within the distribution of energies of free electrons along the interacting proteins, it can be regarded as resonant recognition. The whole

TABLE II
CHARACTERISTIC RRM FREQUENCIES FOR DIFFERENT FUNCTIONAL GROUPS OF PROTEINS AND DNA REGULATORY SEQUENCES

MOLECULE TYPE	FREQ.	No SEQ.	S/N	ERROR
DNA REGULATORY SEQUENCES				
promoters	.34375	53	128	.016
operators	.07813	8	44	.008
SOS operators	.46875	5	13	.050
enhancers	.04883	10	467	.024
PROTEIN SEQUENCES				
oncogenes	.03130	46	468	.004
kinases	.42969	8	71	.003
fibrinogens	.44230	5	99	.001
ACH receptors	.49219	21	137	.002
phages' repressors	.10547	4	51	.005
bacterial repress.	.08398	4	56	.004
heat shock proteins	.09473	10	326	.005
interferons	.08203	18	117	.008
hemoglobins	.02340	187	119	.008
signal proteins	.14063	5	31	.016
protease inhibitors	.35550	27	203	.008
proteases	.37700	80	511	.004
restriction enzymes	.29102	3	36	.004
amylases	.41211	12	170	.002
neurotoxins	.07031	16	60	.004
growth factors	.29297	105	200	.016
ins.-like(IGF I,II)	.49220	12	72	.008
FGFs	.45120	7	121	.005
glucagons	.32030	13	71	.034
homeo box proteins	.04590	9	100	.001
cytochromes B	.05900	16	201	.004
cytochromes C	.47656	38	252	.004
myoglobins	.08200	49	128	.004
lysozymes	.32810	15	124	.004
phospholipases	.04300	29	115	.004
actins	.48000	12	163	.002
myosins	.34000	11	201	.004
RNA polymerases	.35693	10	256	.001

process can be observed as the interaction between transmitting and receiving antennae of a radio system. The RRM model assumes that characteristic frequencies are responsible for resonant recognition between macromolecules at a distance. Thus, these frequencies have to represent oscillations of some physical field which can propagate through water dipoles. One possibility is that this field is electromagnetic in nature. There is evidence that proteins and DNA have certain conducting properties [33], [34]. If so, charges would be moving through the backbone of the macromolecule and pass through different energy stages caused by different side groups of various amino acids or nucleotides. This process 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 charge transfer. The nature of this physical process is still unknown. Some models of charge transfer through the backbone of macromolecules have been accepted [31], [32]. Each of these models shows that charge transfer through the backbone of macromolecules is possible but the complexity of the system does not allow for precise calculations of charge velocity, and furthermore it is not possible to find any correlation between charge transfer and the biological

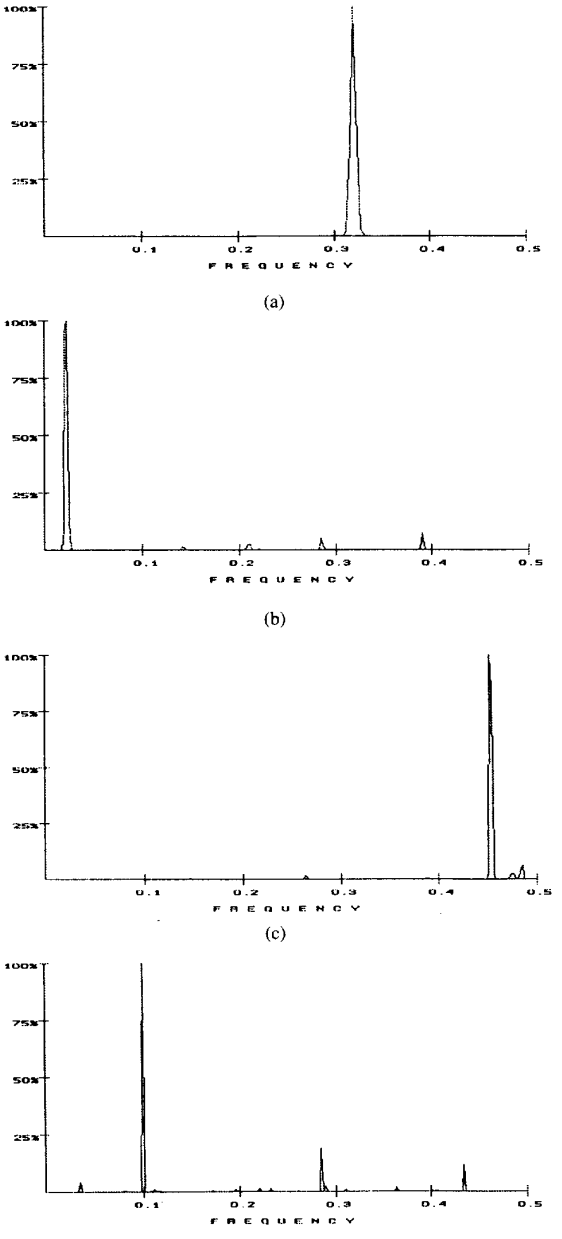


Fig. 2. Multiple cross-spectral function of four different functional groups of proteins: (a) glucagons; (b) hemoglobins; (c) FGF's; and (d) repressors. The multiple cross-spectral function of each group of proteins has a prominent peak representing the frequency characteristic for the biological function.

functioning of macromolecules. The significance of the whole process is proposed below although the author is well aware that conductive electron transfer caused by the difference of the free electron potentials PEII at the N and C terminals of the protein is beset by conditions that are firmly understood. According to pseudopotential, this potential difference is

$$W = W(\mathrm{COOH}) - W(\mathrm{NH_2}) = 0.13 \text{ Ry}.$$

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

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

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

$$V < 7.87 \times 10^5\ \text{m/s}.$$

An inherent assumption is that amino acids in the protein are equidistant and the distance is

$$d = 3.8\,\text{Å}.$$

Therefore, the maximum frequency that could be emitted during the electron transfer is

$$F_{\max} < V/2d$$
$$F_{\max} < 1 \times 10^{15}\ \text{Hz}$$

while the corresponding wavelength is

$$L_{\min} > 330\ \text{nm}.$$

The minimum frequency that could be emitted depends on the total length of the protein

$$F_{\min} = 2F_{\max}/N,$$

where N is the total number of amino acids in the protein. For example, with proteins of 200 AA in length, the minimum frequency is

$$F_{\min} < 1 \times 10^{13}\ \text{Hz}$$

and the corresponding wavelength is

$$L_{\max} < 30000\ \text{nm}.$$

The range from 30000 nm to 300 nm is very wide, starting from the very low infrared through the visible to the ultraviolet regions. This is only an estimate of the possible frequency range of the field that could be emitted by charge transfer through the backbone of proteins. A more precise estimation could only be made with biological experiments. Known experimental results have been needed to reach this aim. They are now deciphered.

C. Correlation of the RRM Characteristic Frequency with Absorption Characteristics of Light Absorbing Proteins

The RRM model is based on the concept that the biological function represents a resonant energy transfer from one biomacromolecule to another [35]. However, the function of some proteins is directly related to the absorption of visible light of defined wavelengths. Thus, correlation is expected between the absorption spectra of proteins and their RRM spectra with a predetermined frequency range. Such correlation has already been demonstrated in previous studies [16], [17] and it is now summarized.

The light is absorbed effectively on a prosthetic group bound to the protein, however, frequency selectivity in this process is defined by the protein itself. For example, rhodopsins which have the same prosthetic group within proteins of similar but different primary structures can absorb different wavelengths, e.g., there are three different variants of rhodopsins, one for the absorption of each color: red, blue and green [36]. As the biological function of these molecules is to absorb particular wavelengths (colors), their grouping was affected on this basis. Thus, the rhodopsin responsible for the absorption of blue light was compared with the bioluminescent protein aequorin which absorbs at similar wavelengths [37]. In the corresponding multiple cross-spectrum, there is only one prominent peak at frequency 0.475 ± 0.004, and this is the most likely peak to be related to the absorption of blue light. In the same manner, it is estimated that the frequency of 0.355 ± 0.004 is related to the absorption of green light, and the frequency of 0.346 ± 0.004 to the absorption of red light. The numerical frequencies obtained similarly by the RRM for various other groups of visible light-absorbing proteins are compared with their corresponding characteristic absorption frequencies in Table III. A result of considerable significance is that the scaling factor between these two sets of data is almost constant at the mean value of $K = 201$. Thus, a strong linear correlation would seem to exist between the numerical characteristic frequencies defined by the RRM and the experimentally determined frequencies corresponding to the absorption of electromagnetic radiation of such proteins. From this correlation, it can be observed that the full range of wavelengths which can be related to RRM characteristic frequencies is over 400 nm. This finding is in complete accord with the frequency range previously associated with the RRM spectra and calculated from the charge velocities through the protein backbone. It can be now inferred from both correlations that approximate wavelengths in real frequency space can be calculated from the RRM characteristic frequencies for each biologically related group of sequences. Furthermore, these calculations can be used to predict the wavelength of visible and near-infrared irradiation which may produce a biological effect.

The situation is slightly different for Cytochrome B where characteristic RRM frequency was found to be at 0.059 and for Myoglobin where characteristic frequency was found to

TABLE III
CORRELATION BETWEEN CHARACTERISTIC RRM FREQUENCIES AND VISIBLE-INFRARED ABSORPTION MAXIMA OF DIFFERENT GROUPS OF PHOTOSENSITIVE PROTEINS [16]

Protein Group	Wavelength (nm)	Frequency (cm^{-1})	Relative RRM frequency	K
Cyt. C	415±20	24,096±117	0.473±0.003	196
Blue	430±25	23,256±1278	0.475±0.004	204
Green	540±30	18,518±974	0.355±0.004	192
Red	570±30	17,544±877	0.346±0.004	197
Hemoglobin	14770±30	677±2	0.023±0.007	295
Purple	860±35	11,628±455	0.281±0.02	241
Flavodoxin	470±30	21,275±1275	0.379±0.004	178

The groups are formed according to their characteristic absorption maxima: Cyt C (cytochromes C); blue (group of proteins absorbing blue light: blue rhodopsin and bioluminescent proteins); green (green light absorbing chlorophylls and rhodopsin); red (red-light absorbing rhodopsin); Purple (purple light absorbing proteins from purple bacteria); The second column represents the wavelengths (nm) of characteristic absorption of the groups of photosensitive proteins, while the third represents the same data in cm^{-1} units. The fourth column contains characteristic RRM frequencies f of the same groups of proteins while the fifth column contains coefficient K, the scaling factor between numerical frequency space of RRM and the wavelengths λ (nm) of real frequency space: λ = K/f. The mean value of coefficient K is K=201 with a standard deviation of 15 %.

be at 0.082. According to the RRM, it is expected that they would have absorption at 3400 nm and 2451 nm, respectively. On the other hand, it is known that there is an absorption in the visible light range, for Cytochrome B at 423–427 nm and for Myoglobin at 533–582 nm. As there are not any data on absorption in Cytochrome B and Myoglobins within expected very low frequency ranges (on the order of a value in thousands of nm), we cannot conclude that these proteins are an exception or that, possibly, they have even stronger absorption at predicted low frequencies.

D. Correlation of the RRM Characteristic Frequency with Low Intensity Light Effects on Cell Proliferation

The frequency selectivity of many light induced biological processes is directly caused by the quantum-mechanical energetic states of photosensitive molecules. There is considerable evidence to suggest that the induced change of the energy states of biomacromolecules (such as the effect of ultraviolet and visible light on primary photoacceptors) leads to the modulation of some biological processes in cells [38]. There is also considerable evidence that low intensity light irradiation at precisely defined wavelengths (frequencies) can produce defined, frequency-dependent effects on living systems *in vivo* or *in vitro*. The correlation between the frequency selectivity of light induced biological processes and RRM characteristics of biomacromolecules involved in these processes has been described for the case of cell growth and proliferation [16]. These results are summarized in Table IV, where light irradiation frequencies which have been shown to effect cell growth are compared with characteristic RRM frequencies for different groups of growth factors normally involved in these cell proliferation processes. The linear correlation was again obtained with the same regression factor as that found for light absorbing molecules. The most illustrative comparison is the one between expected wavelengths, calculated from RRM frequencies using correlation obtained with light absorbing molecules and measured wavelengths which can produce certain biological effects.

The same graph (Fig. 3) shows results of the comparisons between RRM characteristic frequencies for light absorbing proteins and their absorbing characteristics on the one hand, and between RRM frequencies of different groups of growth factors and frequencies of light irradiation which can produce effects on cell growth on the other. It can be observed that in both cases there is the same linear correlation. Strikingly, this correlation was already exhibited in the same coefficient (scaling factor) of $K = 201$ between RRM frequency space and corresponding light wavelengths in nm [14], [15]. It can be represented as follows:

$$\lambda = \mathbf{K}/\mathbf{f_{rrm}}$$

where λ is the wavelength of the light irradiation in nm which can influence a particular biological process (cell growth or light absorption), f_{rrm} is the numerical frequency obtained by the RRM, and $\mathbf{K}$ is the estimated coefficient of the linear correlation obtained. All these results lead to the conclusion that specificity of protein interactions is based on the resonant electromagnetic energy transfer at a frequency specific to the interaction observed.

TABLE IV
CORRELATION BETWEEN CHARACTERISTIC RRM FREQUENCIES OF FIVE GROUPS OF GROWTH REGULATING FACTORS AND THE LOW-INTENSITY LIGHT IRRADIATION FREQUENCIES WHICH PRODUCE SIGNIFICANT EFFECTS ON CELL GROWTH [16]

Growth factor group	Effect observed observed	Characteristic RRM frequency f_c	Expected wavelength λ_c (nm)	Wavelength of maximum effect (nm)
IGFs	DNA synthesis	0.492±0.008	406±6.5	400
FGFs	DNA synthesis Therapeutic effects	0.453±0.004	441.5	441.6 441.6
insulins	DNA synthesis	0.344±0.16	581.4±10	552
growth factors (EGF, CSF, GH, PLF)	DNA synthesis ATP synthesis Therapeutic effect	0.293±0.016	682.6±35.4	633 650 633
PDGFs	Therapeutic effect	0.242±0.008	826±26	830

Second column represents biological effect observed under laser irradiation of particular frequency presented in the fifth column. The third column contains RRM characteristic frequencies of different growth factor groups listed in the first column. The fourth column list the real frequency space wavelengths corresponding according to the relation: λ=201/f, to the calculated numerical RRM frequencies. It can be observed that expected wavelengths calculated using the RRM and above relation, are in a complete agreement with the wavelengths which can produce maximum biological effect.

E. Correlation Between Predicted and Measured Characteristic Frequency of Chymotrypsin Activation

As discussed above, there is evidence that biological processes can be induced or modulated by irradiation with characteristic frequencies [38], [39]. This is directly caused by light-induced changes of the energy states of molecules and, in particular, of proteins. Consequently, one may infer that biological processes can be influenced by an external radiation of defined frequency. The strongest support for this idea emanates from the experiment described by Biscar [39] where it is shown that protease activity of α-chymotrypsin is significantly increased in the presence of a near infrared beam of defined wavelength. These phenomena are discussed here in terms of the RRM which suggests that protein activities, i.e., protein interactions) are based on the resonant electromagnetic energy transfer within a range of infrared and visible light [16], [17].

The RRM was applied to the example of chymotrypsins which belong to the group of proteases [40]. Their common biological function is hydrolysis of proteins. Chymotrypsin is synthesized as one chain, chymotrypsinogen, which after cleavage results in a fully active enzyme, α-chymotrypsin. Cross-spectral analysis of five active chymotrypsins from different origins (bovine, human, rat, vop, and oriental horne) revealed a common frequency component at $f = 0.2363 \pm 0.004$ with $S/N = 35$. This frequency component can be considered a characteristic specifically associated with chymotrypsin activation rather than with its protease activity. In comparison with the other proteases, a different frequency ($f = 0.377 \pm 0.004$) was obtained as general protease characteristic [14], [17]. Using the relation between RRM

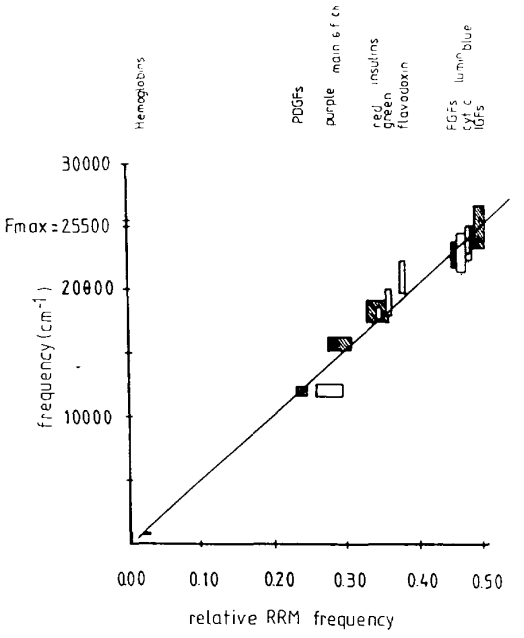


Fig. 3. Graphical presentation of the correlation between RRM frequency space and real frequency space. The abscissa represents the dimensionless RRM frequencies (in the range 0–0.5), while the ordinate represents real frequencies (cm^{-1}) in the visible and near-infrared region of the electromagnetic spectrum. Each functional group of proteins is denoted by a rectangle whose sides represent calculational uncertainties in the RRM frequency space and experimental errors in the real frequency space. The unshaded rectangles represent results from the relation between the light absorption characteristics of photosensitive proteins, while the shaded rectangles represent the relation between frequencies of low-intensity laser irradiation causing effects on cell growth and proliferation (DNA formation) and characteristic RRM frequencies of different groups of growth factors. A linear correlation between RRM frequency space and real frequency space is evident.

frequency and light irradiation wavelength: $\lambda = 201/f_{\text{rrm}}$ obtained previously, the expected light wavelength which can influence chymotrypsin activation was calculated to be 851 nm. However, Biscar [39] has reported an experimental increase in chymotrypsin activity by a factor greater than two under infrared irradiation of wavelength 855 nm. The activity of the enzyme was unaffected (equal to the control in the absence of irradiation) outside the range 850–860 nm. As can be observed from our results, the light wavelength predicted by the RRM as characteristic of chymotrypsin activation is exactly within the measured value for chymotrypsin activation. This result explicitly supports the idea that the RRM characteristic frequencies represent specific oscillations of electromagnetic field within the infrared and visible range which are crucial for protein activity and interactions. This result also reinforces linear correlation between the RRM frequency space and wavelengths of biologically effective light irradiation with a correlation coefficient of $K = 201$. With this coefficient in mind, it is now possible to calculate wavelengths of light irradiation which, it is proposed, will affect other biological processes. The triggering or control of some biological processes in the cell by irradiating them with light of a defined wavelength is then also possible, and indeed would seem to occur.

III. Applications of the RRM

Once there is an understanding of the nature of protein interactions and their selectivity, one can use the knowledge to benefit molecular biology and consequently, medicine, pharmacology and agriculture. In general, the following possibilities emerge: a) to predict functionally important amino acids, so-called "hot spots," within the protein sequence, and thus propose effective mutations; and b) to design peptides with the desired spectral and consequently, functional characteristics. These applications of the RRM are described now in a number of examples.

A. Prediction of "Hot Spots"

Knowing the characteristic frequency of a particular protein function, it is possible to predict which amino acids in the sequence predominantly contribute to this frequency and, consequently, to the observed function. Since the characteristic frequency $f(\mathbf{x})$ correlates with the biological function x, the positions of the amino acids that are most affected by the change of amplitude at the frequency $f(\mathbf{x})$ can be defined as "hot spots" for the corresponding biological function x. The strategy for this prediction includes the following steps:

1) Determination of the unique characteristic frequency for the specific biological function analyzed by multiple cross-spectral analysis for the group of protein sequences with the corresponding biological function.
2) Alteration of the amplitude at this characteristic frequency in the particular numerical spectrum. The criterion used for the identification of the critical characteristic frequency change is the minimum number of "hot spot" amino acids which are least sensitive to further changes in the amplitude of the characteristic frequency.
3) Derivation with DFT of a numerical sequence from the modified spectrum.

It is known [30] that the amplitude change 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 the original series. Determination of the amino acid corresponding to each element of this new numerical sequence can then be achieved from tabulated values of EIIP. The amino acids in the new sequence, which differ from the original ones, reside at the points most important for the frequency $f(\mathbf{x})$. These "hot spots" are related to this frequency and to the corresponding biological function. The procedures described have already been applied to a number of examples: previous studies with interleukin-2 [15], SV40 enhancer [14], tumor necrosis factors (TNF's) [41], Ha-ras p21 oncogene product [19], [41], glucagons [Cosic *et al.*, unpublished results], hemoglobins, myoglobins and lysozymes [18], all have documented evidence that such predicted amino acids denote residues crucial for protein function. Furthermore, in the examples studied to date [18]–[20], these amino acids are found to be spatially clustered in the protein tertiary structure, and to be positioned in and around the protein active site.

The prediction of "hot spots" with typical results is now described for the case of Cytochrome C proteins. Their com-

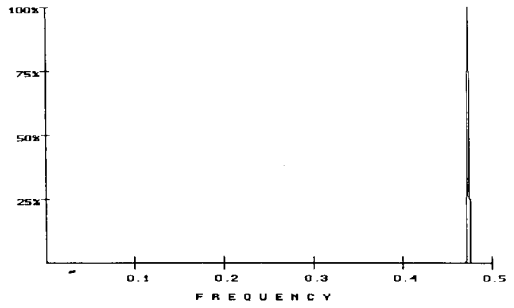


Fig. 4. Multiple cross-spectral function of 38 Cytochrome C proteins. The prominent peak is located at a frequency of 0.476 ± 0.004 with $S/N = 252$.

mon biological function is electron transfer through excitation of the heme group which is positioned in the active site cleft within the protein 3-D structure [42]. Cross-spectral analysis of 38 Cytochrome C proteins from different origins revealed the common frequency component at the $f = 0.476 \pm 0.004$ with $S/N = 252$ (Fig. 4). Consequently, the "hot spot" amino acids in the tuna heart Cytochrome C were determined at the positions: 41, 45, 56, 77. Although these amino acids are not sequentially linked, they form spatial cluster in the protein 3-D structure which is positioned around the protein active site cleft (Fig. 5). In numerous other examples, it was found that "hot spot" amino acids were clustered around the active site of the protein as well [18]–[20]. Since "hot spots" are the amino acids where the resonant characteristic frequency signal is dominant, the cluster of "hot spots" may define specific sites in the 3-D structure of a protein which act as a resonator for the characteristic frequency. This resonator is located just around the active site cleft, and thus it can be considered a resonant box which provides the optimum conditions for the resonant energy transfer at the frequency characteristic for the protein function. As a consequence, the active protein conformation can be predicted to provide the optimal resonant conditions for a particular frequency characteristic, thus dictating the specificity for protein biorecognition with its target and the subsequent energy transfer, which defines the functional selectivity of the biomolecular interaction. At this stage, it is still not clear what the relationship is between the RRM and protein 3-D structure. The result that protein is folded in such a way that amino acids are clustered together and positioned around the active site where characteristic frequency is strongest can have some implication on the protein overall structure but not necessarily to the local structure of the active site.

B. Peptide Design

Following the determination of the characteristic RRM frequency and corresponding phases for particular biological functions, it is possible to design amino acid sequences which will have only those spectral characteristics. It is then expected that these peptides will have the desired biological activity. The strategy for the design of such defined peptides is as follows:

1) Determination of the characteristic unique frequencies for the specific biological function or interaction analyzed by multiple cross-spectral analysis for the group of protein sequences with the corresponding biological functions, as described above.
2) Definition of the characteristic phases at the characteristic frequencies for the particular protein which is chosen as the parent for agonist/antagonist peptide design.
3) Derivation of a numerical sequence from the knowledge of characteristic frequencies and phases. This is done by summation of sinusoids of particular frequencies, amplitudes and phases. The length of the numerical sequence is defined first, by appropriate frequency resolution which allows different characteristic frequencies to be separated and second, by the peptide length desired.
4) Determination of the amino acid corresponding to each element of this new numerical sequence can then be achieved from tabulated values of EIIP. Peptides obtained in this way will have the desired spectral characteristics and are purported to have the corresponding biological activity. The applications of the above strategy are now presented.

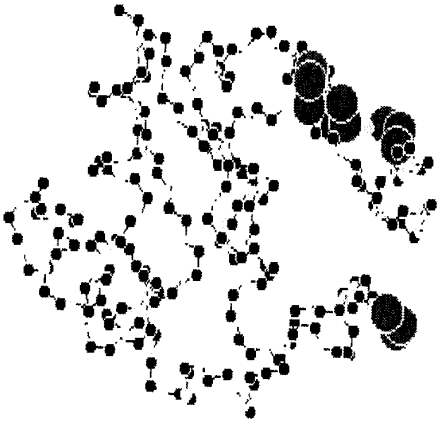


Fig. 5. Cytochrome C (tuna heart) 3-D structure in backbone presentation. Predicted "hot spot" amino acids are highlighted with shaded VDW surfaces. Predicted "hot spot" amino acids are clustered in a space around the active site cleft.

C. Design of Glucagon Agonists

Design procedures are demonstrated here in an example with human glucagon. The aim was to design shorter peptides with the same biological function. Application of the RRM analysis to 13 different glucagons from 18 species and seven different glucagon-like peptides revealed a single prominent peak in the cross-spectral function at frequency $f = 0.3242 \pm 0.034$ ($S/N = 70.54$). The frequency component at $f = 0.3242$ was common to all analyzed glucagon-related polypeptides, and therefore could be considered the consensus frequency characteristic of these polypeptides for a common biological property, for example, the potential to stimulate glycogenolysis.

The next step was to define the active site of human glucagon. With the characteristic frequency of glucagons defined, it was possible to identify amino acids in human glucagon sequence which are the major contributors to the

characteristic frequency. The locations of these predicted amino acids are within the known functionally important regions of the human glucagon sequence. Three of these predicted amino acid residues (Gly 4, Thr 5 and Thr 7) are located within the N-terminal region, known to be critical for the transduction of the biological message. The other two amino acid residues (Phe 22 and Thr 29) are located at the C-terminal of human glucagon, previously shown [43]–[45] to be primarily involved in receptor recognition. This correlation between the RRM predictions and experimentally derived data provides additional evidence that the concepts implicit in the RRM model can be employed as an aid to the prediction of the contribution which a specific amino acid residue may make to the global arrangement of a biorecognition site of a polypeptide.

In order to examine the spatial disposition and topology of these predicted amino acid residues, an appropriate solution structure for human glucagon was needed. A currently accepted model for human glucagon in solution is the Korn and Ottensmeyer model [46] which is based primarily on solution measurements, i.e., CD, fluorescence quenching, and NMR, as well as Chou-Fasman [7] secondary structure predictions. Based on our investigations, a modified model of human glucagon structure in solution was suggested. The model resulted from conformational simulations using molecular mechanics and dynamics [10]–[12]. This stable conformation represents local energy minimum which can be achieved under the 300°K conditions, i.e., energy conditions which may approximate those in the living cell. This simulated human glucagon conformation agreed well with previous experimental observations, as well as with the Korn-Ottensmeyer model, in terms of the compact overall structure with the N- and C-terminus crossing each other and the general shape of the backbone structure [46].

Consequently, this simulated structure was used as a model of the tertiary structure of human glucagon in solution in order to investigate the spatial disposition of the functional key amino acid residues predicted by the RRM procedure. In the case of human glucagon, these predicted amino acids were found to be clustered along one face of the structure of the globular polypeptide, forming an arch across the N- and C-terminal regions and the hydrophobic cleft (Fig. 6). This result strongly supports earlier conclusions reached with a number of other polypeptides and proteins, that such key amino acid residues predicted by the RRM procedures are located in spatially clustered regions in the tertiary structure of the polypeptide or protein and denote a specific biorecognition or interaction site.

With a glucagon characteristic frequency, it was then possible to design a variety of polypeptide analogs having this spectral characteristic only [17], [22]. In particular, several polypeptides were designed to satisfy the criterion that their RRM spectra exhibited only the frequency $f = 0.3242$. It is important to note that no significant homology exists between these derived peptides and human glucagon, while these peptides all show very high mutual sequence homology. The structures of these peptides were then subjected to molecular mechanics and dynamics simulations using the same procedures as for human glucagon. Simulated structures of these peptides were then compared, in terms of their backbone superimposition, onto the structure of the human glucagon interaction site, encompassing the cluster of key amino acid residues (Gly 4, Thr 5, Thr 7, Phe 22, and Thr 29). The goodness of fit was then determined from the root mean square deviation (RMS) of the peptide backbone from that of the corresponding interactive region of the human glucagon.

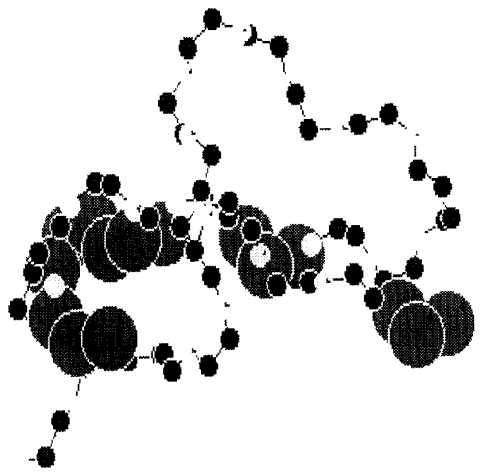


Fig. 6. Predicted 3-D structure of human glucagon in backbone presentation. Predicted "hot spot" amino acids are highlighted with shaded VDW surfaces. Predicted "hot spot" amino acids are clustered in a space around the active site cleft.

The best fit, RMS = 1.98 Å, for the peptides in the range of 15-mer to 20-mer was found for the 18-mer peptide with the following primary sequence: LCRIQDGQDEWDPYDHKD. The predicted 3-D structure of the designed 18-mer peptide is comparable in backbone topography to the structure of glucagon active site, while it is not similar in primary structure to any sequentially linked fragment of human glucagon. Since biorecognition sites of a protein are usually composed of amino acids which are not contiguously linked, the folded structures of predicted peptidic analogs should mimic the overall shape of the binding site, rather than have primary structures which are simply homologous to sequentially linked fragments of the protein under examination.

Finally, the additional studies are necessary in order to characterize the biological properties of synthetic peptide analogs developed from the composite computational approach described above. Such biological evaluations have been successfully applied on a few examples up to now, including design of HIV envelope protein immunological analogs [47], as well as fibroblast growth factor (FGF) antagonist [21].

D. Design of FGF Antagonists and Some Experimental Results

The proliferative and angiogenic potential of FGF-ß has been documented during embryogenesis, vascularization of the ovary and related reproductive tissues, in pathological states such as wound healing, tissue repair and tumor development and expansion [48]. The angiogenesis of solid tumors is one of the FGF functions which is of great interest. By preventing this

angiogenesis activity, it would be possible to prevent tumor growth as well. The design and synthesis of highly specific peptide analogs capable of interacting with FGF receptors may provide a therapeutic mode for the prevention and treatment of solid tumor angiogenesis. In these studies, RRM was used to define frequency characteristics of FGF's responsible for mitogenic activity and receptor recognition.

Multiple cross-spectral function of the three acidic and four basic FGF's (Fig. 7(a)) reveals one prominent peak at the frequency $f1 = 0.4512 \pm 0.005$ ($S/N = 120$). In order to determine if this frequency is relevant for FGF receptor recognition, further comparison with two FGF receptors, flg and bek, was established. The resulting multiple cross-spectral function shows that frequency $f1$ is even more prominent ($S/N = 215$), while the other frequency components were diminished. From these data, it can be concluded that frequency $f1$ is relevant for the recognition between FGF's and their receptors. The other frequency component $f2 = 0.255 \pm 0.002$ was found to characterize the biological activity of competence growth factors including FGF's (Fig. 7(b)). Thus, in the case of the FGF's, it was possible to distinguish receptor recognition from growth promoting activity in terms of different characteristic frequencies. Consequently, peptides designed to have receptor binding characteristics without growth promoting activity incorporate a frequency component at $f1 = 0.4521$ omitting a frequency component at $f2 = 0.256$. Using the peptide design module of RRM, a number of peptides satisfying the frequency conditions were designed with lengths of 10 to 24 amino acids. The procedure was the same as described for glucagons.

Molecular modeling techniques were then applied to the designed peptides with the aim of identifying their stable structures. The proposed structures were then compared with the 3-D structure of the FGF 97–120 fragment, which is known to be a receptor binding domain. Although no sequence homology exists between the designed peptides and the FGF 97–120 fragment, a high structural similarity was detected. On the basis of RMS deviations between the predicted backbone structures, a peptide 16 amino acids in length (FGFRRM16) was found to be the most similar, RMS = 1.37 Å, to the FGF 97–120 fragment. The predicted 3-D structures of FGF fragments 97–120 and FGFRRM16 are presented in Fig. 8. The proposed amino acid sequence of FGFRRM16 is: **MWYRPDLDERKQQKRE**. The designed peptide FGFRRM16 was synthesized and its function was tested experimentally.

Fibroblast cultures were used for testing the bioactivity of FGFß peptide antagonists since FGFß exerts stimulatory effects on DNA synthesis and proliferation of these cells *in vitro* [21]. These observations were confirmed in both ^{3}H-thymidine incorporation and cell proliferation bioassays with a dose-dependent stimulation of tracer uptake and proliferation of fibroblasts in culture. The peptide itself exerted no effect on ^{3}H-thymidine uptake or the proliferation of fibroblasts in culture, however, in the presence of a stimulatory dose of FGFß (25 ng/mL), peptide FGFRRM16 antagonized the actions of FGFß, inhibiting tracer incorporation and the proliferation of fibroblasts *in vitro*. Cell proliferation bioassays were consistently more sensitive than DNA synthesis bioassays with

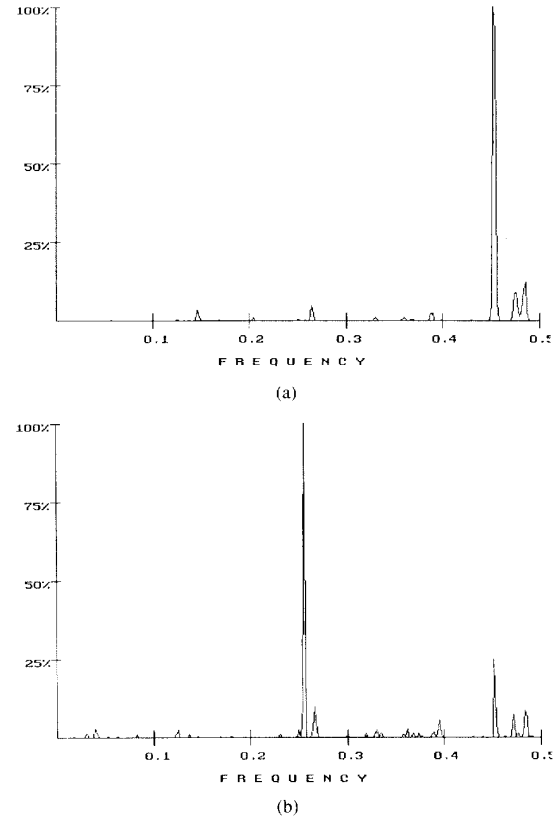


Fig. 7. Characteristic RRM frequencies of fibroblast growth factors (FGF's). (a) Cross-spectral function of five FGF's (three acidic: human, bovine, rat, and two basic: human and bovine). The prominent peak is at a frequency $f = 0.4512 \pm 0.005$ with $S/N = 121$. The same frequency peak is prominent in the multiple cross-spectral function of five FGF's plus two FGF receptors (flg and bek) but with higher signal-to-noise ratio $S/N = 215$. (b) Cross-spectral function for five FGF's and four Interleukins 1. Both groups of proteins belong to the wider functional group of competence growth factors. The common frequency component was found at a frequency $f = 0.255 \pm 0.002$.

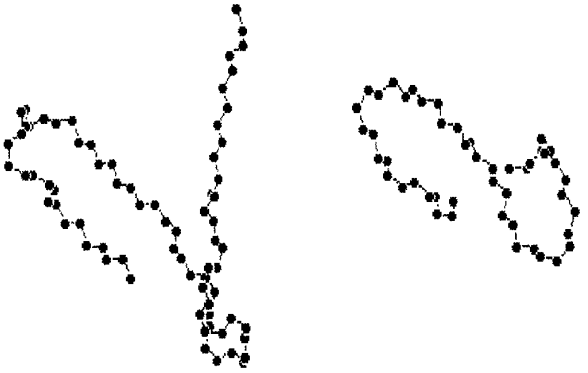


Fig. 8. Backbone structures of (left) FGF active site (97–120 fragment) and (right) FGFRRM16 peptide. Although there is no sequence homology between them, their 3-D backbone shapes are comparable.

doses of 0.5 mg/mL FGFRRM16 antagonizing and 2mg/mL FGFRRM16 totally blocking FGFß stimulated proliferation of fibroblasts. It is important to note that while DNA synthesis

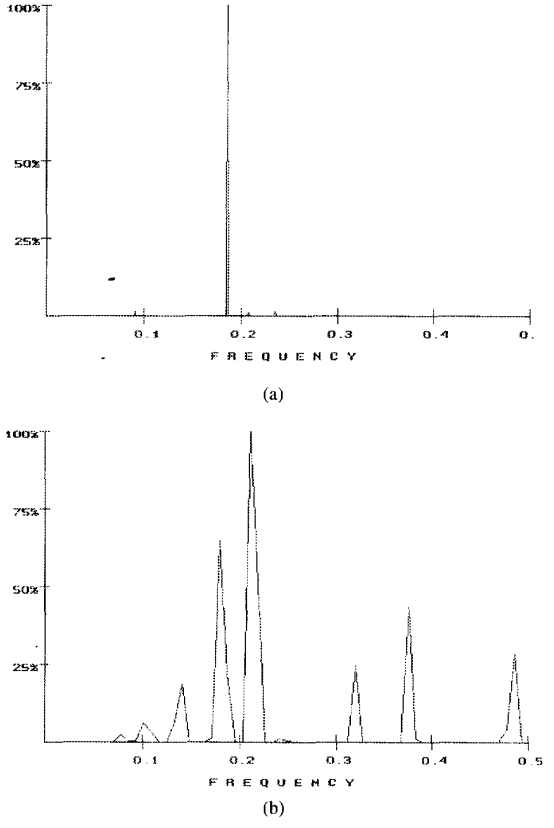


Fig. 9. HIV envelope characteristic frequencies. (a) Multiple cross-spectral function of envelope proteins gp 160 from 21 HIV isolates (lav1/bru, hxb2, nl43, sf2, sc, mn, rf, wmj2, cdc451, ny5, jh3, brva, eli, mal, z6, z2z6, z3, z321, jy1, ndk, oyi). Prominent peak was obtained at frequency $f = 1855 + 0.001$ with $S/N = 484$. (b) Multiple cross-spectral function of regions from gp 120, which are crucial for binding to CD4, 44 amino acid in length, from 11 HIV isolates (lav, arv2, eli, ny5, wmj2, mal, z3, 3cg, cdc451, hxb2, hat). Cross-spectral analysis reveals that those regions share the same characteristic frequency as gp160 env protein but also have their own prominent characteristic at a frequency $f2 = 0.2188 \pm 0.022$.

continued at 2 mg/mL FGFRRM16, this replication of DNA was not translated in terms of proliferation of fibroblasts. These experiments were performed by Dr. A. Drummond (Monash University, Australia) and detailed experimental procedures and results are presented in [21].

E. Design of Peptides Able to Mimic HIV Immunogenicity

The interaction between HIV virus envelope proteins and CD4 cell surface antigen has a central role in the process of the virus entry into the host cell. Thus, blocking the interaction between the envelop glycoproteins and CD4 surface antigen, known to be the HIV receptor, should inhibit infection. We have applied the RRM model to analyze these interactions [17] and, consequently, to design peptide capable of mimicking the HIV envelope gp160 protein immunorecognition function. Prior to designing peptide analogs capable of functioning as HIV-CD4 interaction inhibitors, it was necessary to define the characteristics of the gp160 HIV envelop proteins crucial for their interaction with CD4. Despite the high sequence variability between different isolates, it can be said that at least one specific characteristic has to be common for all different isolates.

Envelop proteins from 21 HIV isolates (lav1/bru, hxb2, nl43, sf2, sc, mn, rf, wmj2, cdc451, ny5, jh3, brva, eli, mal, z6, z2z6, z3, z321, jy1, ndk, oyi) were analyzed using the RRM procedure with the aim of defining a common frequency characteristic. Only one prominent peak was obtained in the cross-spectral function at frequency $f1 = 0.1855 \pm 0.001$ with $S/N = 484$ (Fig. 9(a)). This frequency component was common for gp160 glycoproteins for all HIV isolates analyzed. According to RRM concepts, it can be suggested that this frequency characterizes the common biological behavior of all analyzed proteins, i.e., recognition and binding to CD4. To validate this idea, the CD4 molecules (human and mouse) were compared with the previously obtained cross-spectral function of 25 gp160 HIV envelop proteins. The prominent peak at the same frequency $f1 = 0.1855$ occurs but the signal-to-noise S/N ratio is significantly higher, implying that CD4 molecules also share the same frequency characteristic. This is completely in accordance with our previous findings, [13], [15], [20] in a number of examples, that protein ligands and their receptors (irrespective of whether they are protein or DNA targets) share the same characteristic frequency which defines their mutual recognition.

Regions from gp 120 which are crucial for binding to CD4 [50], 44 amino acid in length, from 11 HIV isolates (lav, arv2, eli, ny5, wmj2, mal, z3, 3cg, cdc451, hxb2, hat) were also analyzed. Cross-spectral analysis reveals that those regions share the same characteristic frequency as the gp160 env protein, but they also have their own prominent characteristic at frequency $f2 = 0.2188 \pm 0.022$ (Fig. 9(b)). To provide confirmation as to whether one or both of the characteristic frequencies characterize interaction between HIV envelope and CD4 antigen six peptides, each of 20 amino acids in length with the following spectral characteristics, were designed:

1) Peptide A: KQQYYWYAWCQPPQDQLIMD: Spectrum consists of two frequencies $f1 = 0.1855$ and $f2 = 0.2188$, with the phases of these frequencies opposite to the phases characteristic for the LAV1 gp120 region.

2) Peptide B: LKRDQEPMDFHIWDDYLKRD: Spectrum consists of the one frequency only, $f1$, with its phase opposite to the phase of the LAV1 gp120 region.

3) Peptide C: TPPTDWLADRHEMDQNKDDK: Spectrum consists of the one frequency only, $f2$, with its phase opposite to the phase of the LAV1 gp120 region.

4) Peptide D: DDALYDDKNWDRAPQRCYYQ: Spectrum consists of both characteristic frequencies, $f1$ and $f2$, with the same phases as in the LAV1 gp120 region.

5) Peptide E: DFHIWDDYLKRDQEPMDFHI: Spectrum consists of the frequency $f1$ with the same phase as in the LAV1 gp120 region.

6) Peptide F: TDWIYDRHEMDQNKDDKNQD: Spectrum consists of the frequency $f2$ with the same phase as in the LAV1 gp120 region.

It is important to note that the frequency $f1$ was also found to be significant in the spectra of p55gag and gp41env HIV proteins, while the frequency $f2$ was found to be significant in the spectra of p18gag HIV proteins. Thus, it can be anticipated

that the designed peptides, besides their reactivity relative to the LAV1 gp120 env protein, may have a reactivity relative to the other HIV proteins listed.

To validate the RRM predictions, the designed peptides were chemically synthesized and their immunoreactivity experimentally tested. Biological experiments with immunogenicity of synthetic peptides were conducted by Dr. V. Krsmanovic and Dr. J. M. Bicquard (CNRS, France). These investigations were performed initially by evaluating the reactivity and cross-reactivity of all six designed peptides with the corresponding antibodies. These results have shown that significant cross-reactivity to the polyclonal antibodies raised against peptides which share at least one characteristic frequency and phase at this frequency can be observed. In contrast, antibodies raised against peptides which do not share a common frequency and phase at this frequency do not show significant cross-reactivity (see [47]). These results give experimental confirmation of the concept that RRM frequency characteristics reflect important parameters associated with biomacromolecular recognition and, in particular, antibody-antigen recognition.

Furthermore, to ascertain whether any of the proposed frequency characteristics are important in determining the interaction between HIV envelope proteins and CD4 surface antigen, the reactivity of antibodies raised against the designed peptides and LAV HIV viral proteins was investigated [47]. These results show clearly that the peptide D is able to induce polyclonal antibodies in rabbits which can be captured by all gag-related env-related proteins of the HIV-1, except the p15-gag protein. This cross-reactivity of the anti-D serum is compatible with the existence of the common RRM spectral characteristic(s) which the peptide D shares with the HIV-1 proteins. If we consider that peptide A has the same frequency components as peptide D but opposite phases at those frequencies, and that anti-A antisera does not react with any of the HIV proteins, it can be concluded that phases at characteristic frequencies are also important parameters in biomolecular recognition. Furthermore, peptide E has only one frequency component, $f1$, and the phase at $f1$ is the same as that for peptide D. Since anti-E antisera does not react with HIV proteins, we can deduce that $f1$ is not a critical frequency for recognition between CD4 antigens and HIV proteins. Nevertheless, at this stage, it is possible to indirectly propose that frequency $f2 = 0.2188 \pm 0.022$ characterizes the recognition between CD4 antigens and HIV proteins. More generally, we can conclude that RRM frequencies, together with their corresponding phases, are important physicochemical parameters that characterize specific biorecognition processes.

IV. Conclusion

This review paper presents a complete model of selective protein interactions relevant to their biological function. The physical nature of the protein biological function is its ability to selectively interact with a particular target. The selectivity of these interactions is explained in terms of the resonant energy transfer between interacting molecules. It is predicted that these energies are electromagnetic in nature. Consequently, the characteristic resonant frequencies for a number of different interactions, i.e., biological functions, were theoretically calculated. Initially, these calculations were based on the following key finding: **Proteins with the same biological functions have common periodicities in the distribution of energies of delocalized electrons along the protein.** With this in mind and taking into account the conductive properties of the protein backbone, the theoretical model of biologically relevant protein resonances was established. These resonances were calculated to be in a very wide frequency range, including infrared and visible light. In order to justify these propositions, calculated resonant frequencies for the particular protein function were compared with the characteristic frequencies of the light absorbing and light activating proteins. Complete agreement was reached. The model presented generally describes a new view of biologically relevant intermolecular interactions, which are suggested to consist in a resonant energy transfer between interacting molecules. Although it is possible to calculate characteristic frequencies for particular interactions or biological functions, the complete physical basis of these interactions is still not completely understood. Nonetheless, with the knowledge of characteristic frequencies for a particular interaction/function, it is possible to predict active site and functional mutations, and to design bioactive peptides with the desired function. This could have great impact on molecular biology and, consequently, for medicine, pharmacology, and agriculture. The fact that completely new, nonhomologous, but biologically active, peptides were designed solely from a knowledge of characteristic frequencies is further evidence that these theoretically calculated frequencies are, in fact, the critical parameter for protein biological functions, i.e., selective interactions. This also provides indirect proof of the whole concept of resonant energy transfer between interacting biological molecules.

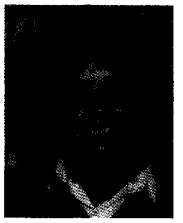
Acknowledgment

I would like to thank Asst. Prof. P. Voumard for reading and reviewing this article.

References

[1] M. O. Dayhoff, *Atlas of Protein Sequence and Structures*, 5th ed. Washington: Nat. Biomed. Res. Fund, 1978, suppl. 3, pp. 345–352.
[2] J. Devereux, P. Haeberly and O. Smithies, "A comprehensive set of sequence analysis programs for the VAX," *Nucl. Acids Res.*, vol. 12, pp. 387–395, 1984.
[3] R. Doolittle, "Similar amino acid sequences: Chance or common ancestry?" *Science*, vol. 214, pp. 149–159, 1981.
[4] J. N. Breg, J. H. J. van Opheusden, M. J. M. Burgering and R. Captein, "Structure of arc repressor in solution: Evidence for a family of ß-sheet DNA-binding proteins," *Nature*, vol. 346, pp. 586–589, 1990.
[5] J. King, "Deciphering the rules of protein folding," *Chem. & Eng. News*, vol. 32, 1989.
[6] G. D. Fasman, Ed., *Prediction of Protein Structure and the Principles of Protein Conformation*. New York: Plenum Press, 1989.
[7] Y. Chou and G. D. Fasman, "The conformation of glucagon: Predictions and consequences," *Biochem.*, vol. 14, pp. 2536–2540, 1975.
[8] C. DeLisi and J. A. Berzofsky, "T-cell antigenic sites tend to be amphipatic structures," in *Proc. Nat. Acad. Sci. USA*, vol. 82, pp. 7048–7052, 1985.
[9] E. D. Schmid, F. W. Schneider and F. Siebert, *Spectroscopy of Biological Molecules—New Advances*. New York: Wiley, 1988.

[10] D. H. J. Mackay, A. J. Cross and A. T. Hagler, *Prediction of Protein Structure and the Principles of Protein Conformation*, G. D. Fasman, Ed. New York: Plenum Press, 1989, pp. 317–358.
[11] J. Maple, U. Dinur and A. T. Hagler, "Derivation of force field for molecular mechanics and dynamics from *ab initio* energy surfaces," in *Proc. Nat. Acad. USA*, vol. 85, pp. 5350–5354, 1988.
[12] M. Karplus and J. A. McCammon, "Dynamics of proteins: Elements and function," *Ann. Rev. Biochem.*, vol. 53, pp. 263–300, 1983.
[13] I. Cosic, D. Nesic, M. Pavlovic and R. Williams, "Enhancer binding proteins predicted by informational spectrum method," *Biophys. Biochem. Res. Comm.*, vol. 141, pp. 831–839, 1986.
[14] I. Cosic and D. Nesic, "Prediction of 'hot spots' in SV40 enhancer and relation with experimental data," *Eur. J. Biochem.* vol. 170, pp. 247–252, 1988.
[15] I. Cosic, M. Pavlovic and V. Vojisavljevic, "Prediction of 'hot spots' in interleukin-2 based on informational spectrum characteristics of growth regulating factors," *Biochimie*, vol. 71, pp. 333–342, 1989.
[16] ——, "The relationship of the resonant recognition model to effects of low-intensity light on cell growth," *Int. J. Radiat. Biol.* vol. 56, pp. 179–191, 1989.
[17] I. Cosic, "Resonant recognition model of protein-protein and protein-DNA interaction," in *Bioinstrumentation and Biosensors*, D. Wise, Ed. New York: Marcel Dekker, Inc., 1990, pp. 475–510.
[18] I. Cosic, A. Hodder, M. Aguilar and M. T. W. Hearn, "Resonant recognition model and protein topography: Model studies with myoglobin, hemoglobin and lysozyme," *Eur. J. Biochem.*, vol. 198, pp. 113–119, 1991.
[19] I. Cosic and M. T. W. Hearn, "'Hot spot' amino acid distribution in Ha-ras oncogene product p21: Relationship to guanine binding site," *J. Mol. Recognition*, vol. 4, pp. 57–62, 1991.
[20] ——, "Studies on protein-DNA interactions using the resonant recognition model: Application to repressors and transforming proteins," *Eur. J. Biochem.*, vol. 205, pp. 613–619, 1992.
[21] I. Cosic, A. E. Drummond, J. R. Underwood and M. T. W. Hearn, "A new approach to growth factor analogue design: Modeling of FGF analogues," *Molecular and Cellular Biochem.*, vol. 130, pp. 1–9, 1994.
[22] I. Cosic, "Protein design by methods of digital signal analysis," in *Proc. Annu. Conf. IEEE Biomed. Eng.*, vol. 2. New York: IEEE Publ. Service, 1988, pp. 901–905.
[23] M. Kanehisa, "A multivariate analysis method for discriminating protein secondary structural segments," *Protein Eng.*, vol. 2, pp. 87–92, 1988.
[24] K. Nakai, A. Kidera and M. Kanehisa, "Cluster analysis of amino acid indices for prediction of protein structure and function," *Protein Eng.*, vol. 2, pp. 93–100, 1988.
[25] V. Veljkovic, *A Theoretical Approach to Preselection of Cancerogens and Chemical Carcinogenesis*. New York: Gordon & Breach, 1980.
[26] M. Charton, "Electrical effect substituent constants for correlation analysis," *Prog. Phys. Organic Chem.*, vol. 13, pp. 119–251, 1981.
[27] M. Charton and B. I. Charton, "The structural dependence of amino acid hydrophobicity parameters," *J. Theor. Biol.*, vol. 99, pp. 629–644, 1982.
[28] V. Heine, M. Koen and D. Uir, *Theory of Pseudopotentials*. Moscow: MIR, 1973.
[29] V. Veljkovic and I. Slavic, "General model of pseudopotentials," *Physical Review Lett.*, vol. 29, pp. 105–108, 1972.
[30] L. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*. New York: Prentice-Hall, 1975.
[31] A. Mandell, *Chaos and Order in the Natural Science*. Berlin: Springer-Verlag, 1988.
[32] A. Viari, H. Soldano and E. Ollivier, "A scale-independent signal processing method for sequence analysis," *Comput. Appl. Biosci.*, vol. 6, pp. 71–80, 1990.
[33] W. A. Little, "Possibility of synthesizing an organic superconductor," *Phys. Rev.*, vol. 134, pp. A1416–A1424, 1964.
[34] A. S. Davidov, *Biology and Quantum Mechanics*. Oxford: Pergamon Press, 1982.
[35] M. D. Williams and J. L. Fox, "Coupled oscillator theory of enzyme action," in *Chemical Recognition in Biology*. New York: Springer-Verlag, 1980, pp. 86–96.
[36] J. Halhaus, D. Thomas and D. Hagness, "Molecular genetics of human color vision: The genes encoding blue, green and red pigments," *Science*, vol. 232, pp. 193–201, 1986.
[37] S. Inouye, M. Noguchi, Y Sakaki, Y. Takagi, T. Miyata, S. Iwanaga, T. Miyata and F. Tsuji, "Cloning and sequence analysis of cDNA for the luminescent protein aequorin," in *Proc. Nat. Acad. of Sci. USA*, vol. 82, pp. 3154–3158, 1985.
[38] T. J. Karu, "Photobiological fundamentals of low-power laser therapy," *IEEE J. Quantum Electron.*, vol. QE-23, pp. 1703–1717, 1987.
[39] J. P. Biscar, "Photon enzyme activation," *Bullet. Math., Biology*, vol. 38, pp. 29–38, 1976.
[40] I. Cosic, "Correlation between predicted and measured characteristic frequency of chymotrypsin activation," in *Proc. 15th Annu. Conf. IEEE Eng. Med. Biolog. Soc.*, in press.
[41] V. Veljkovic and I. Cosic, "A novel method of protein analysis for prediction of biological function: Application to tumor toxins," *Canc. Biochem. Biophys.*, vol. 9, pp. 139–148, 1987.
[42] P. Nicholls, *Cytochromes and Biological Oxidation*. London: Oxford, 1975.
[43] V. J. Hruby, "Structure-conformation-activity studies of glucagon and semi-synthetic glucagon analogues," *Mol. Cell Biochem.*, vol. 44, pp. 49–64, 1982.
[44] J. L. Krstenasky, C. Zechel, D. Trivedi and V. J. Hruby, "Importance of the C-terminal-helical structure for glucagon's biological activity," *Int. J. Peptide Protein Res.*, vol. 32, pp. 468–475, 1988.
[45] C. G. Unson, E. M. Gurzendia, K. Iwasak and R. B. Merrifield, "Glucagon antagonists: Contribution to binding and activity of the amino-terminal sequence 1–5, position 12 and the putative α-helical segment 19–27," *J. Biol. Chem.*, vol. 264, pp. 789–794, 1989.
[46] A. P. Korn and F. P. Ottensmeyer, "A model for the three-dimensional structure of glucagon," *J. Theor. Biol.*, vol. 105, pp. 403–425, 1983.
[47] V. Krsmanovic, I. Cosic, J. M. Biquard and M. T. W. Hearn, "Artificial peptides and induced antibodies," Patent Appl. PCT/FR 93/00171, Feb. 1993.
[48] A. S. Goustin, E. B. Leof, G. D. Shipley and H. L. Moses, "Growth factors and cancer," *Cancer Res.*, vol. 46, pp. 1015–1029, 1986.
[49] R. Anderson, "AIDS: Trends, predictions, controversy," *Nature*, vol. 363, pp. 393–394, 1993.
[50] L. A. Lasky, G. Nakamura, D. A. Smith, C. Fennie, C. Shimasaki, E. Patzer, P. Berman, T. Gregory and D. J. Capon, "Delineation of a region of the human immunodeficiency virus type 1 gp120 glycoprotein critical for interaction with the CD4 receptor," *Cell*, vol. 50, pp. 975–985, 1987.



Irena Cosic (M'86) received the B.Eng. degree in electrical engineering in 1976, and the M.S. and Ph.D. in biomedical engineering in 1982 and 1985, respectively, at the University of Belgrade.

From 1977 to 1989, she was employed at the Institute Vinca, Belgrade, where she was working on a variety of projects in biomedical engineering, digital signal analysis and telecommunication. In 1980, she commenced her research in digital signal processing applications on linear macromolecules. This research resulted in the Resonant Recognition Model (RRM) of protein and DNA interactions. From 1989 until the beginning of 1993, she was employed as a Research Fellow in the Biochemistry Department, Monash University, where she was able to test some practical applications of the RRM model. Since 1993, she has been a Senior Lecturer in the Department of Electrical and Computer Systems Engineering at Monash University, Caulfield, where she is continuing her research in the field of biomolecular electronics.