INTRODUCTION

MMPs are a family of structurally and functionally related endoproteases capable of degrading most of the components of an extracellular matrix (1). They share common functional activities and activation mechanisms in which they depend on Ca^{2+} and Zn^{2+} ions and are active at neutral pH. A MMP’s expression can be rapidly stimulated by different cytokines and growth factors and other physical cellular interactions. There are many studies showing data related to a matrix MMP’s over expression in various tumour types, when compared to normal tissues (1). Several studies have provided evidence that certain MMPs in specific cancers can be useful as the indicators of tumour progression and this leads to the possibility of improving cancer treatment therapies and tumour management overall (1,2). Previously, the contribution of MMPs to tumour metastasis was attributed to their matrix degrading activity. However, recent studies demonstrated that MMPs are involved in all stages of tumour progression starting from the initial tumour development, growth, angiogenesis, invasion, metastasis as well as growth at secondary sites(3). Therefore, the study of the functional activities of MMPs is of great importance in understanding the role of MMPs and also in the development of applications where MMPs can be used as tumour biomarkers in cancer diagnostics and therapy.

The role of MMPs in cancer dissemination led to the development of MMP inhibitors (MMPIs) for tumour treatment (4,5). In order to inhibit growth and invasion of cancer cells, several hypotheses have been developed on how to block MMP activity in an extracellular environment (2). Studies, where small molecule MMPs inhibitors (MMPIs) were used for the treatment of cancer in its early stages, report that MMPIs are useful as therapeutic drugs
to treat and prevent metastasis\(^{(6)}\). It was also suggested that MMPs may be more important in early stages of cancer due to their cytotoxic rather than cytotoxic effects \((2)\). However, clinical studies of the anti-tumour application of synthetic MMP had only limited success. A possible reason for that is the lack of specificity of the inhibitors used, which can lead to unexpected and unwanted effects as in some instances where MMP inhibition may be detrimental rather than therapeutic\(^{(2)}\).

It is known that proteins can be biologically “active” only by achievement of a certain active native conformation as a three-dimensional structure \((3D)\). It is generally accepted that 3D conformations of proteins are fully predetermined by their primary structures. It is known that a protein’s biological function is encrypted within its primary structure, i.e. a sequence of amino acids. The RRM \((7-14)\) is able to determine a protein’s functional and structural information by analysis of its primary sequences using signal processing methods, Fourier and Wavelet Transforms. It is assumed that the selectivity of protein interactions is based on the resonant electromagnetic energy transfer at the frequency specific for each interaction. One of the main applications of this approach is the prediction of the location of active/binding site(s) within protein primary structures \((7,9)\).

The main disadvantage of the signal analysis based on the standard Fourier Transform is that the information about frequency characteristic along the series is hidden, and we can obtain only an averaged time and frequency content of the analysed signal. In the last 20 years, time-frequency distribution methods have become powerful alternative tools for signal analysis. A time-frequency transform presents energy distribution of a signal over the time and frequency domains.

In this study we applied the time-frequency signal processing technique to structure-function analysis of the selected proteins analysed, aiming to demonstrate how the signal’s energy is distributed over the two-dimensional time-frequency space. By incorporating the Smoothed Pseudo Wigner Ville distribution (SPWV) in the standard RRM approach we intend to overcome the problem of non-localization events currently present in the model and improve the RRM predictive capabilities and accuracy for investigation of a proteins’ physical characteristics. In particular, the RRM and SPWV methods were applied to investigate the structure-function relationships between a MMP proteins and their mutual interaction with serine protease proteins and metalloproteinase inhibitors aiming at better understanding of the MMPs’ activation and inhibition that can provide a new insight into the role of MMPs in tumour development.

**MATERIALS AND METHODS**

**Resonant Recognition Model**

The RRM, central to this study, involves a transformation of amino acid sequences into numerical sequences and then analysis of these sequences by the appropriate digital signal processing methods, Fourier and Wavelet \((7-14)\). In the RRM, a protein’s primary structure is presented as a numerical series by assigning to each amino acid a physical parameter value relevant to the protein’s biological activity. Our previous investigations \((7-14)\) as well as studies of other authors \((15)\) have shown that the best correlation is achieved with parameters related to the energy of delocalized electrons from each amino acid. These findings can be explained by the fact that electrons delocalized from a particular amino acid have the strongest impact on the electronic distribution of energy in the entire protein. In this study the energy of delocalized electrons, calculated as the Electron Ion Interaction Potential (EIIP) \(^{(10)}\) of each amino acid residue, was used. By assigning to each amino acid the EIIP value we convert the original protein sequence into a numerical sequence. This resulting numerical sequence represents then the distribution of free electron energies along the protein molecule.

The numerical sequences obtained are analysed using Discrete Fourier Transform (DFT) in order to extract information pertinent to the biological function. As the average distance between amino acid residues in a protein sequence 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 \(d=1\). Peak frequencies in the amplitude cross-spectral function define common frequency components of the two sequences analysed. 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 \((7,8)\):

\[
|M_n| = |X_{1,n}| |X_{2,n}| |\alpha| X_{M,n} |\text{Kn=1,2,K N/2}
\]

(1)

Peak frequencies in such a multiple cross-spectral function denote common frequency components for all sequences analysed. 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 distinct peak frequency in a consensus spectrum implies that all of the analysed sequences within the group have one frequency component in common. This frequency is related to the biological function provided the following criteria are met:

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

In our previous research \((7-14)\) the above criteria have been implemented and the following fundamental conclusion was drawn: each specific biological function of a given protein or DNA is characterized by a single frequency. It has been found in previous research that proteins with the same biological function have a common frequency in their numerical spectra and shown that each specific biological function of protein or regulatory DNA sequence(s) is characterized by a single frequency \((7,8)\). The results of our previous studies with a number of different protein families revealed that proteins and their interacting targets (receptors, binding proteins, and inhibitors) display the same characteristic frequency in their interactions. However, it is obvious that one protein can participate in more than one biological process, i.e. revealing more than one biological function. Therefore, it has been postulated that the RRM frequency characterizes a particular biological process of interaction between selected bio-molecules. Further research in this direction has lead to the conclusion that interacting molecules have the same characteristic frequency but opposite phases at that frequency \((7,8)\). Thus, the RRM characteristic frequencies represent a proteins general functions as well as a mutual 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 the resonant recognition. Once the characteristic frequency for the particular biological function or interaction is determined, it becomes possible to identify the individual “hot spot” amino acids that contributed most to this specific characteristic frequency and thus, possibly to the observed biological behaviour of the protein.

**Time-frequency analysis**

The Wigner quasi-distribution was initially introduced to replace the classical phase-space distribution in statistical physics with corresponding quantum analogue \((17)\). Von Neumann \((18)\) established a method where two non-simultaneously measurable quantum mechanical quantities, such as the coordinate and momentum, can be measured simultaneously with a limited precision. He
also showed that all measurements, with limited accuracy, can be replaced by the absolutely accurate measurements of other quantities, which are related to their non-simultaneously measurable quantities. Although due to the uncertainty principle, the concept of phase space in quantum mechanics is somewhat problematic, various functions which bear some resemblance to true phase-space distribution functions of non-quantum world were introduced. They proved to be useful not only as calculation tools, but also provided insights into the relations between classical and quantum mechanics. The first of such functions was introduced by Wigner (17) to study quantum corrections in classical statistical mechanics. It is now known as the Wigner function. It may be shown (17,18) that the phase space distribution, which is produced in simultaneous measurements of position and momentum, can be represented as a convolution of the Wigner function of considered quantum state and the Wigner function of the filter state, which represents a measuring apparatus.

In general, Wigner-Ville distribution (WVD) describes the frequency content changes as a function of time. The distribution is the actual energy intensity of various frequency components of the signal at a given position along the protein assuming that average distance between amino acid is set at an arbitrary value d=1. In practical calculations, convolution of the signal generates the cross term that represents interference of the signals, and consequently decreases significantly the resolution of the signal. A number of methods have been developed to reduce the cross-term (19). In this investigation we replaced the WVD by the Smoothed Pseudo Wigner-Ville distribution (SPWVD), where some window functions are convolved with the WVD to restrain and decrease the effect of the interference terms.

Supposing EIIP[i], i=1,2,...,N is the numerical sequence of the Electron Ion Interaction Potentials of amino acids along the polypeptide chain, then the SPWVD of EIIP[i] is given by (19):

$$S(t,f) = \int_{-\infty}^{\infty} [h(\tau)] f(t-\tau) g(s-\tau) d\tau$$

(2)

$$W_{n,m} = \frac{1}{2N} \sum_{k=-N+1}^{N-1} \sum_{p=-M+1}^{M+1} h(k) g(p) z(n+p+k) z^{*}(n+p-k) e^{-j2\pi n m}$$

(3)

where \(h(k)\) and \(g(p)\) represent an independent frequency and time smoothing respectively. In this study as the smoothing functions we used the Gauss filters, which are defined as:

$$h(k) = e^{-k^2/2\sigma^2}$$

$$g(p) = e^{-p^2/2\sigma^2}$$

where \(s\) is standard deviation, and \(k\) and \(p\) are the average values in the frequency and distance sets. The resulting SPWVD could be shown in a t-f plane as a contour plot according to the values of \(S(t,f)\), which represents the distribution of the signal energy in the space domain. By choosing the standard deviation of the Gaussian functions \(h\) and \(g\), we are practically balancing between the resolution in frequency and space domain interferences.

**RESULTS AND DISCUSSION**

To investigate the structure-function relationship between MMP proteins, we collected thirty seven MMP primary sequences (National Centre for Biotechnology Information (NCBI) Database) and analysed them using the RRM approach. The cross-spectral analysis was performed and the characteristic frequency was identified at \(f_2=0.1143\) (Figure 1). There are also other, less significant peaks at the frequencies \(f_3\) and \(f_4\) as can be observed in Figure 1.
frequency corresponding to their common biological activity, e.g. MMP inhibition. The multiple cross-spectral function of the selected MMPs is presented in Figure 2. The prominent characteristic frequency was identified at $f_{\text{characteristic}}=0.4326$. As can be seen from the multiple cross-spectral function of the analysed MMPs, the same frequency ($f=0.4326$) was identified as a less significant characteristic frequency for MMPs proteins (Figure 1). It was proposed that a RRM frequency represents a unique parameter of the interacting proteins and thus, characterises their mutual interaction. This characteristic frequency is a relevant parameter for mutual recognition between biomolecules and is significant in describing the interaction between proteins and their substrates or targets. This result reveals that MMPs can recognise and interact with MPIs on the basis that they share the same characteristic frequency at $f=0.4336$.

We were also interested to investigate if it is possible to predict the activation of MMPs. It was reported that most MMPs can be activated by other MMPs or by serine proteases $(5, 20)$. For this analysis we selected twelve serine protease sequences and applied the RRM procedure to determine their characteristic frequency. The multiple cross-spectral function of the analysed serine proteases is shown in Figure 3. The prominent characteristic frequency was identified at $f_{\text{characteristic}}=0.4609$. There is also a less significant frequency located at $f=0.4326$ as can be observed from Figure 3. This is the characteristic that MMPs and serine protease have in common and thus we can suggest a possible interaction between these groups of proteins that can occur at this frequency.

Bioinformatics methods were used to compare the primary sequences of the MMPs and the following evolutionary groupings of the MMPs were suggested: (i) MMP-19; (ii) MMPs 11, 14, 15, 16, 17; (iii) MMP-2 and MMP-9 and (iv) all the other MMPs. Gene expression studies, using microarrays have identified five new molecular subtypes of breast cancer, i.e. luminal A, luminal B, human epidermal growth factor receptor-2 (HER-2), normal breast like and basal type $(5)$. Unlike the other subtypes, targeted therapy does not exist for the basal type. In the study reported in $(5)$, expressions of the different MMPs in basal and non-basal tumours were compared in order to identify potential new targets for treatment of the basal subtype. As shown in Table 1, expressions of MMP-1, -7, -9, -12 and -15 were significantly elevated in the basal type compared with all the other subtypes combined $(5)$.

MMP-9 is one of the two MMPs known as gelatinases. MMP-9 is absent from most normal adult tissues including the intestinal epithelial cells. MMP-9 proteins are associated with tumour invasion and metastasis due to their capacity for tissue remodelling via extracellular matrix. They also are involved with basement membrane degradation and induction of angiogenesis. The study $(21)$ showed that MMP-9 expression in carcinoma cells favours survival. Evaluation of MMP-9 expression seems to add valuable information on breast cancer prognosis. In this study we analysed eight MMP-9 protein sequences using the RRM. The characteristic frequency of the MMP-9 sequences was identified at $f=0.0186$, which characterises their common biological activity. The multiple cross-spectral function is shown in Figure 4.

Another interesting example of a cancer biomarker is the MMP-1 (Figure 1). Here, we analysed the MMP-1 bovine protein sequence aiming to predict the location of active/binding sites.

<table>
<thead>
<tr>
<th>MMP</th>
<th>Basal</th>
<th>Other subgroups</th>
<th>$P$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMP-1</td>
<td>1.233</td>
<td>0.473</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>MMP-2</td>
<td>0.726</td>
<td>0.834</td>
<td>NS</td>
</tr>
<tr>
<td>MMP-3</td>
<td>0.993</td>
<td>0.782</td>
<td>NS</td>
</tr>
<tr>
<td>MMP-7</td>
<td>1.349</td>
<td>0.359</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>MMP-8</td>
<td>1.074</td>
<td>1.03</td>
<td>NS</td>
</tr>
<tr>
<td>MMP-9</td>
<td>0.729</td>
<td>0.461</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>MMP-10</td>
<td>0.459</td>
<td>0.647</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>MMP-11</td>
<td>0.439</td>
<td>0.809</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>MMP-12</td>
<td>1.337</td>
<td>0.611</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>MMP-13</td>
<td>0.469</td>
<td>0.773</td>
<td>0.004</td>
</tr>
<tr>
<td>MMP-14</td>
<td>0.962</td>
<td>0.873</td>
<td>NS</td>
</tr>
<tr>
<td>MMP-15</td>
<td>1.039</td>
<td>0.897</td>
<td>0.004</td>
</tr>
<tr>
<td>MMP-16</td>
<td>1.033</td>
<td>1</td>
<td>NS</td>
</tr>
<tr>
<td>MMP-17</td>
<td>1.153</td>
<td>1.064</td>
<td>NS</td>
</tr>
<tr>
<td>MMP-19</td>
<td>1.047</td>
<td>1.038</td>
<td>NS</td>
</tr>
<tr>
<td>MMP-20</td>
<td>0.953</td>
<td>1.026</td>
<td>NS</td>
</tr>
<tr>
<td>MMP-24</td>
<td>1.122</td>
<td>1.054</td>
<td>NS</td>
</tr>
</tbody>
</table>

Table 1. Relationship between MMP mRNA expression and breast cancer subgroups $(5)$

Because expression of MMP-1 serves as a biomarker for cancer invasion, the allocation of binding sites is particularly important for the development of anti-tumour drugs which can bind the MMP-1 and prevent further tumour development. The EIIP was used as a physico-chemical parameter characterising amino acids in the selected MMP-1 bovine protein sequence. Instead of the standard Fourier Discrete Transformation usually used in the RRM model, we applied the Smoothed Pseudo Wigner Ville Distribution (using time-space analogy) for analysis of structural and functional content of the MMP-1 protein sequence. Thus, “spatial component” or the position of specific amino acids along the protein chain was represented in the time-space plane (Figure 5). As can be observed from Figure 5, in the time-space plane, there are high-energy areas that correspond to the MMPs characteristic frequency identified at $f_{\text{characteristic}}=0.1143$. This area is located at 155-175 amino acid positions in the MMP-1 bovine protein sequence. These computational predictions were compared with the available structural data obtained using X-ray diffraction (http://www.ncbi.nlm.nih.gov/protein/NP_776537.1).

It was found that the amino acids 160, 172, 179 are regarded as tissue inhibitor of metalloproteinase (TIMP) binding surface in this selected protein.
CONCLUSION

The development of effective anti-MMP therapies would greatly increase with improved knowledge of the contribution of MMPs to the progression of specific cancer types and stages with the appropriate tools for evaluating MMP inhibitory activity at both the molecular and clinical levels. There is a need to better assess and investigate natural agents with tumoricidal activity, e.g., MMPi for cancer therapy. These natural bio-drugs may offer several advantages as anti-cancer products because these diet-derived compounds are nontoxic, widely available and inexpensive. In this study we demonstrated that our computational models, the RRM and SPWVD, can be used efficiently for structure-function analysis of MMP proteins, for modelling MMPs recognition and inhibition with MMPIs and serine proteases. We also showed, using the MMP-1 bovine as an example, that by analysing a single primary sequence of the selected MMP-1 protein we are able to find/predict the specific regions of the protein sequence responsible for binding to the MMPIs. Thus, the results of this study revealed that the presented computational methods based on signal processing present valuable approaches to the analysis of the biological behaviour of MMPs.

Apstrakt

Matrix Metalloproteinase (MMPs) su proteolitički enzimi, sposobni da razgrade međučelijjski matriks. Oni su uključeni u različite fiziološke čeljske procese i takođe se mogu povezati sa razvojem i invazijom tumora. MMPs su poznati prognostički biomarkeri kod različitih tipova kancera i samim tim predstavljaju obećavajući cilj za terapiju kancera. U ovom radu su predstavljena i prodiskutovana dva slična koncepta Resonant Recognition Model (RRM) i Smoothed Pseudo Wigner Ville raspodela (SPWV), primenjena na analizi struktura-funkcija relacije između različitih MMPs. Dodatno smo studirali aktivaciju i inhibiciju MMPs analizirajući njihove interakcije sa serinskim proteazama i metalproteaznim inhibitorima. Rezultati takve analize su pokazali da RRM pristup predstavlja efikasan alat za analizu MMPs funkcionalne aktivnosti. Dobijeni rezultati jasno pokazuju da SPWV može da bude korišćena uspešno za predikciju aktivnih/vezujućih mesta unutar selektovane MMP sekvence.

REFERENCES


/The paper was accepted on 30.01.2011.