XI CONGRESO LATINOAMERICANO
DE BIOLOGÍA MATEMÁTICA
SoLaBIMa 2019
Universidad Católica del Maule
22-25 de Octubre de 2019, Talca, Chile

RECOGNITION OF PROTEIN INTERACTION REGIONS THROUGHT TIME-FREQUENCY ANALYSIS

Nicolás Arango Plaza*
Universidad del Quindío, Armenia, Colombia

Resumen

Protein-protein interactions govern all molecular processes for living organisms, even those involved in pathogen infection. Pathogens, like virus, bacteria, and parasites contain proteins that help the pathogen to attach, penetrate, and settle inside the target cell. Thus, it is necessary to know the regions in pathogenic proteins that interact with host cell receptors. Currently, powerful pathogen databases are available and many pathogenic proteins have been recognized, but many pathogenic proteins have not been characterized. This work applied Time-Frequency Analysis (TFA) to recognize important regions in proteins. TFA shows the highest local variances in a protein string from 3 different time-frequency distributions.

More specifically, taking advantage of the fact that the distance between amino acid residues in a protein sequence is about $3.8\mathring{A}$, most of corresponding numeric representations can be analyzed as an equidistance realization (or *time series*) from some stochastic process that can be stationary or not. Since several studies have reported the non-stationarity feature of genomic and biomolecular sequences [7], time-dependent spectra are an useful tool to identify localized characteristics of a protein. For instance, hot spot aminoacids or motifs that most contribute to an important frequency that describes either a biological function or an interaction [1].

For a real-valued signal $x(t),\ t \in \mathbb{R}$, Ville [2] in 1948 introduced the analytical signal concept and a quadratic transform previously studied by Wigner [3] in 1932 on quantum thermodynamic and rediscovered by Cohen [4] in 1966 for applications in statistical mechanics and signal processing of light waves. The Wigner-Ville (WV) transform of $x(t),\ t \in \mathbb{R}$ is given by

$$\begin{aligned} W_x(t,\omega) &= \frac{1}{2\pi}\int x^*\left(u-\frac{\tau}{2}\right)x\left(u+\frac{\tau}{2}\right)e^{-i\tau\omega}d\tau \\ &= \frac{1}{2\pi}\int R_x(u,\tau)e^{-i\tau\omega}d\tau, \end{aligned} \tag{1}$$

where $x^*(t)$ is the analytic signal associated to $x(t)$. Function $R_x(u,\tau) = x^*(u-\frac{\tau}{2})x(u+\frac{\tau}{2})$ represents a form of local autocovariance and measures the covariance between values at time points separated by an interval τ and symmetrically placed about the time $t = u$. In this

*Parcialmente financiado por COLCIENCIAS grant:111374455483, e-mail: narangop@uqvirtual.edu.co

transform time and frequency have a symmetric role, so by applying the Parseval formula [5] this time-frequency distribution can also be rewritten as a frequency integration, i.e.

$$W_x(t,\omega) = \frac{1}{2\pi}\int f_x^*\left(\omega - \frac{\lambda}{2}\right) f_x\left(\omega + \frac{\lambda}{2}\right) e^{i\lambda t} d\lambda.$$

Even though WV transform looks like a powerful tool to analyze the time-frequency structures of a signal, this is not the case due to the interferences created by the cross terms in (1). These interferences can be attenuated by smoothing the WV transform as proposed by Cohen [4], however the consequence of this is a decrease of the time and frequency resolutions, and more generally a loss of theoretical properties. The general family of Cohen's quadratic time-frequency distributions is

$$C_x(t,\omega) = \frac{1}{4\pi^2}\int\int\int \phi(\theta,\tau) R_x(u,\tau) e^{-i\theta t - i\tau\omega + i\theta u}\, du\; d\tau\; d\theta,$$

where $\phi(\theta,\tau)$ is a function independent of time and frequency that acts as a smoothing kernel. By choosing different kernels we obtain different distributions as well, and the mathematical properties of $C_x(t,\omega)$ depend on kernel choosen. If $\phi(\theta,\tau) = 1$ we obtain the WV distribution which satisfies many desirable properties as time and frequency marginals, convolution, real valued, time and frequency shifts, group delay, among others. In order to balance both properties and resolutions, in this work we used the *spectrogram, Choi-Williams* and *reduced interference* distributions [6].

The aforementioned three TFDs were applied for some study cases from pathogenic co-crystallized structures in order to search the frequencies that characterize interaction regions in pathogenic proteins and with this information tried to identify new interaction regions in either paralogs or orthologs. We found that this approach was able to detect under several descriptors important regions in proteins even those related to interaction. We performed a benchmark for interaction regions from 127 proteins obtained acceptable sensibility and specificity for some descriptors (sensibility/specificity around 0.85). We also identified a peptide from a mouse protein IRGb2-b1 which showed the highest local variance in TFA model and this peptide was assessed in a growth assay in Toxoplasma gondii model. The peptide was able to delayed Toxoplasma growing.

Trabajo realizado en conjunto con:
Gladys Elena Salcedo E.[1], Universidad del Quindío, Armenia, Colombia.
Ailan Farid Arenas Soto[2], Universidad del Quindío, Armenia, Colombia.

Referencias

[1] Arenas, A. F., Salcedo, G. E., Montoya, A. M., and Gomez-Marin, J. E. Msca : a spectral comparison algorithm between time series to identify protein-protein interactions. *BMC bioinformatics 16*, 1 (2015), 152.

[2] Ville, J. Theorie et application dela notion de signal analytique. *Câbles et transmissions 2*, 1 (1948), 61–74.

[3] Wigner, E. P. On the quantum correction for thermodynamic equilibrium. In *Part I: Physical Chemistry. Part II: Solid State Physics*. Springer, 1932, pp. 110–120.

[1] e-mail: gsalcedo@uniquindio.edu.co
[2] e-mail: aylanfarid@yahoo.com

[4] Cohen, L. Generalized phase-space distribution functions. *Journal of Mathematical Physics 7*, 5 (1966), 781–786.

[5] Mallat, S. *A wavelet tour of signal processing: the sparse way.* Academic press, 2008.

[6] Flandrin, Patrick. *Time-frequency/time-scale analysis.* (Vol. 10), Academic press, 1998.

[7] Saadi, H. H., Sameni, R., and Zollanvari, A. Interpretive time-frequency analysis of genomic sequences. *BMC bioinformatics 18*, 4 (2017), 154.