# Investigating an Efficient and Accurate Protocol for Sampling Structures from Molecular Dynamics Simulations: A Close Look by Different Wavelet Families

Mateus A. Gonçalves[a], Arismar M. G. Júnior[b], Elaine F.F. da Cunha [a,], Teodorico C. Ramalho[a*]

[a] Chemistry Department, Federal University of Lavras, 37200-000 Lavras, MG, Brazil

[b] Minas Gerais State University, Divinópolis, MG, Brazil

**Corresponding Author**

*e-mail: teo@ufla.br; teodorico.ramalho@gmail.com

Tel.:+55 35 3829-1522; Fax: +55 35 3829-1271

**Abstract**

Molecular Dynamics (MD) simulations are widely used to predict the behavior of molecular systems over time. However, one of the great challenges of MD simulations is how to treat the thousands of configurations obtained from calculations, since the number of the quantum calculations (QM) required for evaluating electronic parameters is too high and, sometimes, computationally impracticable. Thus, an efficient and accurate sampling protocol is essential for combining classical MD and QM calculations. In this article, based on the OWSCA methodology, 93 wavelet signals were analyzed in order to further refine the methodology and identify the best wavelet family for $[Fe(H_2O)_6]^{2+}$ and $[Mn(H_2O)_6]^{2+}$ complexes in solution. Our results point out that the bior1.3 was the best wavelet, values closest to the experimental data were obtained for both studied systems.

**Keywords**

Wavelet decomposition, Computational Chemistry, molecular dynamics, OWSCA, Theoretical calculations

## 1.0- Introduction

Molecular dynamics (MD) calculations are used to predict the physical movements of atoms and molecules and it is one of the most versatile computational techniques for the study of biological macromolecules [1–4]. In fact, in the rational structure-based drug design, MD simulations have exhaustively contributed to various stages of the process[1,5]. In fact, MD simulations have evolved into a mature technique that can be used effectively to understand macromolecular structure-to-function relationships [1,6,7].

By using MD simulations, it is possible to study the solvent effect employing explicit molecules to obtain time-averaged properties of the system, such as density, conductivity, and dipolar moment, as well as different thermodynamic parameters, including interactions energies and entropies [2,8].

MD calculations generally provide thousands of conformations throughout of simulation, therefore, one of the biggest challenges is how to select these conformations for QM calculations required, for example, calculations of Gibbs free energy calculations, spectroscopic properties, such as NMR or EPR[9,10].

Currently there are several methods for selecting MD structures, as the method of statistical inefficiency [11–13], Clustering[14–16], PCA (Principal component analysis)[17,18], random method (selection of 50 out of 50 structures, for example)[19] and recently our research group developed the OWSCA[9] method. It is important to mention that each method has its particularities and must be used carefully, in order not to compromise the results.

Recently, the use of the Wavelet transform has demonstrated to be a very promising in the selection of structures[9]. In this method, OWSCA, the signal of a MD

simulation is compressed without damaging the main characteristics of the original MD simulation data and, representative conformations are selected. It is noteworthy, however, that the number of selected conformations can be dependent of the wavelet family used in the compression procedure. In this perspective, to investigate MD simulations by different wavelet families may give rise to a new, efficient and inexpensive computational protocol for improving the OWSCA methodology.

It should be kept in mind that the Wavelet transform is a mathematical procedure that can convert the signal into a different form[20–22]. This conversion reveals the main features hidden within the sign and shows the original sign in a more succinct way. The wavelet transform using the Haar function has been shown to be very promising in the study of the selection of MD conformations [9].

With the selection of MD structures, several calculations can be performed to study the properties of the systems. A property that has been extensively explored by our research group, is the hyperfine coupling constant ($A_{iso}$)[23–27]. The hyperfine coupling constant is the most sensitive parameter for determining the values of relaxation rates, longitudinal ($R_1 = 1/T_1$) or transverse ($R_2 = 1/T_2$). The $T_1$ and $T_2$ values are very useful for predicting the efficiency of a contrast agent (CA), which are paramagnetic compounds used in Magnetic Resonance Imaging (RMI). MRI uses the NMR signal of water protons present in the body, whose contrast among different tissues is related to $T_1$ and $T_2$ relaxation rates. In most cases, only the natural relaxation rate of water protons is not sufficient for providing good visualization of tissue images, therefore it is necessary the use of contrast agents[28,29].

In the present work, we explored a new methodology for the selection of MD structures. This methodology is based on the OWSCA[30] already described and successfully applied, however, in this work the aim is to further optimize the

methodology. Thus, for this purpose several families of wavelets were studied and validated with experimental results of $A_{iso}$ of the $[Fe(H_2O)_6]^{2+}$ and $[Mn(H_2O)_6]^{2+}$ complexes.

## 2.0- Computational methods

### 2.1- Geometries construction, optimization and Molecular dynamics simulations

The geometries construction of the compounds $[Fe(H_2O)_6]^{2+}$ and $[Mn(H_2O)_6]^{2+}$ were performed in the GaussView program, thus, after construction, the geometries were fully optimized in the Gaussian 09 program[31] with the functional B3LYP and the basis set Lanl2dz for the $Fe^{2+}$, $Mn^{2+}$ atoms and 6-311g++ (d, p) for hydrogen and oxygen atoms.

From the optimized geometries, molecular dynamics (MD) simulations were performed using the REAXFF program, which was developed and validated by van Duin and coworkers[32] (REAX-FF), for the simulations we used the FEOCH[33] force field for $[Fe(H_2O)_6]^{2+}$ and NiCH for the compound $[Mn(H_2O)_6]^{2+}$. The MD simulations were performed at 310.65 K (physiological temperature). In fact, this temperature is suitable to simulate the behavior of compounds in biological systems. As usual, periodic boundary conditions (PBC) and a cutoff distance of 10.0 Å have been applied. The system consists of 300 water molecules in a cubic cell with a side of 20 Å. The volume of the cube was determined by the density of liquid water ($\rho$ = 0.996 g $cm^{-3}$). The constant atom number, temperature and volume (NVT) ensemble, known as the canonical ensemble, was applied for both systems. First, the initial configuration was minimized using the steepest descente and the conjugate gradient algorithm until an energy gradient of 0.01 kcal $mol^{-1}$ $Å^{-1}$ was reached. The simulations, with water molecules, consisting of a thermalization phase of 500 ps, followed by an additional period of 2.0 ns, was employed.

## 2.2- Wavelet transform

The Wavelet Transform is a signal processing technique widely used in the analysis of irregular and non-periodic waveforms. Such a method provides signal information in the time and frequency domains simultaneously, being able to signal denoising, hidden features revealing or to represent the original signal in an compactly way [12].

Mathematically, the Continuous Wavelet Transform (CWT) of a signal *x(t)* can be obtained according to Equation (1), which represents the internal product of the signal with a base function $\varphi_{\tau,s}(t)$ given by Equation (2), where $\tau$ is a translation factor, $s$ is a scale factor and $\varphi$ is the mother wavelet or reference wavelet, which is the transforming function itself [20,21].

$$CWT_x^{\varphi}(\tau,s) = \int_{-\infty}^{\infty} x(t).\varphi_{\tau,s}(t)dt, \quad (1)$$

$$\varphi_{\tau,s}(t) = \frac{1}{\sqrt{s}}.\varphi\left(\frac{t-\tau}{s}\right). \quad (2)$$

In general, during the wavelet transformation, the function $\varphi_{\tau,s}(t)$ is displaced along with the signal *x(t)* at different time instants, according to the translation factor $\tau$, and the CWT is calculated for different compression/expansion values of the mother wavelet, according to the $s$ scale factor. Several wavelet families can be used to transform/decompose the signal, with the Haar, Daubechies, Bior, and Symlets wavelets being the most common types employed in the literature [21,34].

Computationally, the wavelet analysis is performed by the CWT discretization. In this way, the Discrete Wavelet Transform (DWT) constitutes an algorithm that quickly and efficiently calculates the wavelet transform using digital filtering techniques based on low-pass and high-pass filters, which are synthesized according to the characteristics of the desired mother wavelet.

Figure 1 illustrates the DWT procedure, consisting of systematic low-pass and high-pass filtering of the *x(t)* signal, with subsequent sub-sampling by 2, in such a manner the original signal is decomposed into several frequency ranges, with different resolutions. The resulted coefficients at each filtering/decomposition level, called approximation coefficients (*cA*) and detail coefficients (*cD*), for the low-pass and high-pass filters, respectively, represent the signal information in different frequency ranges.

**Figure 1**

### 2.3- Signal Compression

When applied in conjunction with an optimization strategy for selecting a truncation point for the coefficients cA and cD, the DWT enables signal compression. This procedure was called Optimal Wavelet Signal Compression (OWSC), being proposed in[30] to reduce the number of structures for calculating DM. In this work, only the Haar wavelet was analyzed, enabling an efficient MD conformation compression. Nevertheless, as different mother wavelets result in different coefficients in the DWT procedure, and consequently, indistinct reconstructed/compressed signals (through the inverse transform application), several wavelets families will be investigated in this work in the application of OWSCA, aiming to find the most appropriate mother wavelet to MD conformation compression.

To briefly summarise, the OWSCA techinique for signal compression is performed according to the following steps. For more details, see[30]:

i) DWT application in the signal decomposition, i.e., in the MD conformation data decomposition;

ii) Optimization problem solution to determine a threshold value ($\delta j$) for the coefficients *cA* and *cD*, being *j* the jth-level of the DWT decomposition. Such a step is performed employing a genetic algorithm, which searches for the optimal values of $\delta$ in order to minimize the difference between the original signal *x(t)* and the compressed signal *y(t)*, i.e., to minimize $\|x(t)-y(t)\|$;

iii) Signal compression, canceling the cA and cD indexes whose values are less than the founded optimal threshold ($\delta_j^{¿}$), i.e., $cA_j(k)=0$ if $cA_j(k)\leq\delta_j^{¿}$, where *k* is the kth-index of the jth *cA* coeficient;

iv) Application of the inverse wavelet transform to obtain the compressed signal *y (t)*.

## 3.0-Results and discussion

### 3.1- Study of the different families of wavalets

In the present work, we study several families of waveles (table 1) in order to determine an efficient and accurate protocol for selection structures from MD simulations. For the analysis, it was used the absolute Pearson's correlation coeficiente (*r*) to compare the compressed and the original signals, obtained as Equation 3, where *n* is the sample size and $\bar{x}$ and $\bar{y}$ are the statistical mean of the original and compressed signals, respectively. In this way, the closer *r* to 1, the more efficient is the signal compression since its waveform is more similar to the original case. Thus, it is possible

to determine which wavelets families are more suitable for the selection of DM structures.

$$r = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}}$$

(3)

In the OWSCA methodology, 12 levels of decomposition of the wavelet signal were used. The amplitude of the coefficients can be viewed on the vertical axis, while the horizontal axis shows the number of points. The lowest wavelet levels contain high frequency data, while the highest wavelet levels contain low frequency data, thus, each level must have a different limit. Figure 2 shows the 12 levels for the fk8 wavelet, the coefficients that had a greater contribution were 1, 2, 3, 9 and 10, the other coefficients lead to a small contribution to the sign.

**Figure 2**

**Table 1**

Table 1 shows the values of the correlation coefficients for the $[Fe(H_2O)_6]^{2+}$ [35] and $[Mn(H_2O)_6]^{2+}$ complexes, thus, the best ten wavelets were selected, then aiso calculation and methodology validation were carried out. Table 2 shows the selected wavelets, the best wavelets were the Biorthogonal (bior1.1, bior1.3 and bior1.5), Reverse Biorthogonal (rbior1.1 and rbior1.3), Daubechies (db1 and db26), Discrete

Meyer (dmey) and Fejer-Korovkin (fk18 and fk22), it is important to note that bior1.3 and rbior 3.1 wavalets reached the best correlation coeficiente, because for these wavelets, we have the values of r equal to r = 0.6969 and 0.6589 for $[Fe(H_2O)_6]^{2+}$ and r = 0.6395 and 0.7469 for the compound $[Mn(H_2O)_6]^{2+}$ for the bior1.3 and rbior3.1, respectively. Figure 1 shows the energy form MD conformations ($[Fe(H_2O)_6]^{2+}$) at each time for the selected wavelets (blue is the original signal and red is the compressed signal). In that way, we can highlight the rbio 3.1 wavelet, in this wavelet it is possible to observe that practically all DM structures were selected. Therefore, we can discard this wavelet because the large number of structures makes the $A_{iso}$ calculations unfeasible.

**Table 2**

**Figure 3**

In DM simulations, it was considered for the analysis the period between 40,000 and 10,000 fs was observed that the period of energy decrease (dynamic period) was between 0 and 40,000 fs. It was observed that the period of energy decrease (dynamic period) was between 0 and 40,000 fs. In view of this, such period was excluded from the analysis and the compression was performed only in the stationary phase, where the energy oscillates periodically. It can be observed that the original data set x(t) presents a high oscillatory (noisy) profile while the compressed data set y(t), with ny equals to 50, presents a smoother profile not equidistantly spaced. Notice that the trajectory of both curves is similar, meaning that y(t) captures the main features of x(t).

Table 2 shows the numbers of the structures selected for the group of the best wavelets, it is possible to observe that the rbior1.3 wavelet is not found in this table 2, in figure 3, it is possible to observe that practically all the conformations for the wavelet rbior 3.1 are important, so it is impracticable to perform the calculations all the points and thus we exclude this wavelets due to the great number of conformations, as already mentioned. Another important point is that the numbers of selected structures were very close for both compounds ($[Fe(H_2O)_6]^{2+}$ and $[Mn(H_2O)_6]^{2+}$), this shows that the OWSCA method can be used in a similar way for different compounds.

**Table 3**

### 3.2- Validating theoretical strategies by $A_{iso}$ calculations

Paramagnetic compounds, which have unpaired electrons, currently have a great interest, mainly, in the understanding of biological systems, for example, contrast agents for magnetic resonance imaging (MRI) [36,37]. Contrast agents (CAs) have a minimum requirement to be paramagnetic compounds, thus, through EPR spectroscopy it is possible to study these compounds, having the hyperfine coupling constant ($A_{iso}$) as a parameter to assess these properties[30].

From $A_{iso}$ calculations, it is possible to predict longitudinal ($T_1$) and transverse ($T_2$) relaxation times for water molecules, the equations 4 and 5 represent the relaxation time $T_1$ and $T_2$ induced by paramagnetic ions in aqueous solution, from this equation, it is possible to highlight the Aiso parameter, which is the most sensitive parameter in solution. In other words, any small variation in the system directly influences the values of Aiso and consequently the values of $T_1$ and $T_2$.

$$R1=\frac{1}{T_1}\cong\frac{1}{15}\frac{S(S+1)g_e^2\beta^2g^2{}_N\beta^2{}_N}{\hbar^2r^6}+¿R1=\frac{1}{T_1}\cong\frac{1}{15}\frac{S(S+1)g_e^2\beta^2g^2{}_N\beta^2{}_N}{\hbar^2r^6}+¿$$

$$R_1=\frac{1}{T_1}\cong\frac{1}{15}\frac{S(S+1)g_e{}^2\beta^2g^2{}_N\beta^2{}_N}{\hbar^2r^6}+¿ \quad (4)$$

$$R_2=\frac{1}{T_2}\cong\frac{1}{15}\frac{S(S+1)g_e^2\beta^2g_N^2\beta_N^2}{h^2r^6}+¿ \quad (5)$$

Analyzing equations 2 and 3, the relaxation time $T_1$ and $T_2$ depends on the electron spin (S), the electronic and proton g factors ($g_e$ and $g_N$, respectively), the Bohr magneton (β), the nuclear magneton ($\beta_N$), the hyperfine coupling constant ($A$), the ion-nucleus distance ($r$), and the Larmor frequencies for the proton and electron spins ($\omega_I$ $\omega_I$ and $\omega_S\omega_S$, respectively) [38] . The correlation times $\tau_c$ $\tau_c$ and$\tau_e\tau_e$ are characteristic of the rate of change of the interactions between the metallic species and neighboring protons. Each of the relaxation rates is a sum of two terms. The first term comes from the dipolar coupling and the second term from the scalar coupling. Hence, there is a dependence of the relaxation time on the τ and hyperfine coupling constants values. This dependence is valid for all types of collisions, which leads to an expression for the relaxation time (Equations 1 and 2) [39]. The paramagnetic ions most commonly found $T_1$ and $T_2$ depend on the correlation times $\tau_c$ and $\tau_e$ and hyperfine coupling constants [40].

Thus, $A_{iso}$ values were studied for two complexes $[Fe(H_2O)_6]^{2+}$ and $[Mn(H_2O)_6]^{2+}$, both coordinated with water molecules. Figure 4 shows the structures for the $[Fe(H_2O)_6]^{2+}$ complexafter the DM calculations with the first layer of solvation, this approach is important to reduce the computational cost of the system, this same

methodology has already been used for several systems without losing the important statistical information [10,41].

**Figure 4**

In order to validate our methodology and check which wavelet is best for the selection of DM configuratoins, theoretical $A_{iso}$ calculations were performed. Table 2 shows the $A_{iso}$ values for both compounds in different wavelets, thus, initially, we will analyze the $[Fe(H_2O)_6]^{2+}$ complex. For the $[Fe(H_2O)_6]^{2+}$ complex, the experimental the $A_{iso}$ values of the $Fe^{2+}$ atom is equal to 0.500 MHz and the wavelets that came closest to that value was the bior 1.1 with the value of 0.496 MHz. Indeed, there is a very small difference between the theoretical and experimental values, this difference is 0.004 MHz. The wavelet that had a value more distant from the experimental was the Dmey with a value of 0.454 MHz, a difference of 0.046 MHz between the theoretical and experimental values. In fact, for the Dmey wavelet, there are very few structures to represent the entire thermal effect of the system. The fk18 wavelet, which had only 12 structures selected, shows also a $A_{iso}$ value (0.459MHz) far from the experimental value, a difference of 0.041 MHz between the theoretical and experimental values is observed. These two wavalets were the worst models for the $[Fe(H_2O)_6]^{2+}$ complex. This result puts in evidence that the behavior of the wavelet directly influences the system and that few selected structures cannot realistically describe the system. In fact, important configurations can be left out and this factor depends on the type of wavelet that will be used.

Turning now to the $[Mn(H_2O)_6]^{2+}$ complex, it is possible to observe that the best wavelet was again bior1.3, being that the experimental $A_{iso}$ values for this complex is 0.860 and 5.400MHz for the atoms of $^{1}H$ and $^{17}O$, respectively. Thus, the theoretical

values for the bior1.3 wavetet were 0.853 and 5,382 MHz for the $^{1}H$ and $^{17}O$ atoms, respectively, a difference between the theoretical and experimental values of 0.007 MHz for the atom of $^{1}H$ and 0.018 MHz for the atom of $^{17}O$. It is also observed that the wavelet that had the most distant result from the experimental was the dmey, with theoretical values of 0.818 and 4,923 MHz for the atoms of $^{1}H$ and $^{17}O$, respectively, a difference of 0.049 MHz for $^{1}H$ and 0.295 MHz for $^{17}O$ was obtained.

**Table 4**

From the $A_{iso}$ results, it was noticed that the wavelet that best describes the behavior of our system is the dmey, the fact, the theoretical values are very close to the experimental values for this wavelet.

There is a variety of wavelet families[21], as mentioned. The difference of the wavelet families varies according to some criteria such as: the length of the support of the mother wavelet, the number of disappearance moments, the symmetry or the regularity. Another criterion that is also important and worth considering is the existence of a corresponding scale function and the orthogonality or bi-orthogonality of the resulting analysis. When the mother wavelet produces through translation and scaling all the wavelet functions used in the transformation, thus, it is possible to determine the characteristics of the resulting Wavelet Transform. With this in mind, it is important that the details of a particular application must be taken into account and the appropriate mother wavelet must be chosen to use the wavelet transform effectively to get satisfactory results for the analysis in question. In the case of the present work, wavelets have been used for selecting representative conformations of molecular dynamics simulations. Thus, the wavelet that had a better result was the bior1.3 [34].

The superior performance of biortogonal wavelets can be related to its linear

phase property, which is fundamental for signal reconstruction since the symmetry of the filters coefficients is guaranteed [34].

## 4- Conclusions

Based on our findings, the bior 1.3 wavelet showed the best $A_{iso}$ value, compared to experimental values, of the studied compounds ($[Fe(H_2O)_6]^{2+}$ and $[Mn(H_2O)_6]^{2+}$). For the bior 1.3 wavelet, we obtain the average values of 0.496 MHz (experimental value 0.500 MHz) for the $Fe^{2+}$ atom of compound $[Fe(H_2O)_6]^{2+}$ and 0.853 and 5.382 MHz values for the atoms of $^{1}H$ and $^{17}O$, respectively (experimental values 0.860 and 5.400MHz[42]), for $[Mn(H_2O)_6]^{2+}$ compound. In this sense, this article was intended to refine the methodology already developed (OWSCA). Thus, for this purpose, 93 wavelets were analyzed being 6 families (bior, rbior, db, sym, dmey, fk, coif). Thereby, wavelet bior 1.3 can be successfully used for the selection of DM structures for coordination compounds containing transition metals. For other types of systems a thorough analysis must be carried out in order to verify which is the best wavelet for each system (in the near future we will analyze the wavelets for other systems). In this way, the use of the OWSCA methodology is able to drastically reduce the number of conformations without losing important information, introducing thermal and solvent effects in spectroscopic and thermodynamic theoretical calculations with reasonable trade-off between cost and accuracy.

## Acknowledgment

The authors wish to thank the Brazilian financial agencies Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo ao Ensino e Pesquisa de Minas Gerais (FAPEMIG) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior/Ministério da Defesa (CAPES/MD) for financial support, and the Federal University of Lavras (UFLA) and Minas Gerais State University (UEMG) for providing the physical infrastructure and work space. This work was also supported by excellence project FIM UHK.

**Reference**

1. A. M. Namba, V. B. Silva and C. H. T. P. Silva, *Eclética Química*, **2008**, 33, 13.
2. H. Alonso, A. A. Bliznyuk and J. E. Gready, *Med Res Rev*, **2006**, 26, 531.
3. N. L. A. U. Burkert, *ACS Monogr*, **1982**, 64.
4. K. Lipkowitz, *J Comput Chem*, **1983**, 4, 605.
5. S. A. Adcock and J. A. McCammon, *Chem Rev*, **2006**, 106, 1589.
6. J. Gelpi, A. Hospital, R. Goñi and M. Orozco, *Adv Appl Bioinforma Chem*, **2015**, 8, 37.
7. A. Persidis, *Nat Biotechnol*, **1998**, 16, 393.
8. C. D. Snow, E. J. Sorin, Y. M. Rhee and V. S. Pande, *Annu Rev Biophys Biomol Struct*, **2005**, 34, 43.
9. M. A. Gonçalves, L. S. Santos, D. M. Prata, F. C. Peixoto, E. F. F. da Cunha and T. C. Ramalho, *Theor Chem Acc*, **2017**, 136, 15.
10. M. A. Gonçalves, F. C. Peixoto, E. F. F. Da Cunha and T. C. Ramalho, *Chem Phys Lett*, **2014**, 609, 88.
11. K. Coutinho, H. C. C. Georg, T. L. L. Fonseca, V. Ludwig and S. Canuto, *Chem Phys Lett*, **2007**, 437, 148.
12. K. Coutinho and S. Canuto, *Adv. Quantum Chem*, **1997**, 28, 89.
13. K. Coutinho, S. Canuto and M. C. Zerner, *J Chem Phys*, **2000**, 112, 9874.
14. X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. Van Gunsteren and A. E. Mark, *Angew Chemie - Int Ed*, **1999**, 38, 236.
15. J. Shao, S. W. Tanner, N. Thompson and T. E. Cheatham, *J Chem Theory Comput*, **2007**, 3, 2312.
16. R. De Paris, C. V. Quevedo, D. D. Ruiz, O. Norberto De Souza and R. C. Barros, *Comput Intell Neurosci*, **2015**, 2015, 1.
17. R. Susnow, C. Schutt and H. Rabitz, *J Comput Chem*, **1994**, 15, 963.

18. F. Sittel, A. Jain and G. Stock, *J Chem Phys*, **2014**, 141, 014111.

19. D. T. Mancini, E. F. Souza, M. S. Caetano and T. C. Ramalho, *Magn Reson Chem*, **2014**, 52, 129.

20. G. Oppenheim, *Wavelets and Their Applications*, **2007**.

21. S. Mallat, *A Wavelet Tour of Signal Processing*, **1999**.

22. M. Kamada, M. Toda, M. Sekijima, M. Takata and K. Joe, *Chem Phys Lett*, **2011**, 502, 241.

23. M. A. Gonçalves, L. S. Santos, F. C. Peixoto, E. F. F. da Cunha, T. C. Silva and T. C. Ramalho, *ChemistrySelect*, **2017**, 2, 10136.

24. M. a. Gonçalves, F. C. Peixoto, E. F. F. da Cunha and T. C. Ramalho, *Chem Phys Lett*, **2014**, 609, 88.

25. M. A. Gonçalves, T. C. Ramalho, *Eclet Quim*, **2020**, 45, 12.

26. B. T. L. Pereira, É. F. Silva, M. A. Gonçalves, D. T. Mancini, T. C. Ramalho, *J Chem*, **2017**, 2017,1.

27. M. A. Gonçalves, E. F. F. da Cunha, F. C. Peixoto, T. C. Ramalho, *Comput Theor Chem*, **2015**, 1069, 96.

28. C. De Angelis, R. F. Brizzi and R. Pellicano, *J Gastrointest Oncol*, **2013**, 4, 220.

29. D. Esteban-Gómez, A. de Blas, T. Rodríguez-Blas, L. Helm and C. Platas-Iglesias, *Chemphyschem*, **2012**, 13, 3640.

30. M. A. Gonçalves, L. S. Santos, D. M. Prata, F. C. Peixoto, E. F. F. da Cunha and T. C. Ramalho, *Theor Chem Acc*, **2017**, 136, 1.

31. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Farkas, J. B. Foresman, J. V Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09, Revis B01, Gaussian, Inc, Wallingford CT*, **2009**.

32. Duin V (2002) ReaxFF user manual.

33. M. Aryanpour, A. C. T. van Duin and J. D. Kubicki, *J Phys Chem A*, **2010**, 114, 6298.

34. C. Stolojescu-crisan, S. Moga and A. Isar, *EEE Optim-Aegean*.**2010,** 929.

35. D. Harris, G. H. Loew and A. Komornicki, *J Phys Chem A*, **1997**, 101, 3959.

36. M. Lepage and J. C. Gore, *J Phys Conf Ser*, **2004**, 3, 78.

37. E. D. Hedegård, J. Kongsted and S. P. A. Sauer, *J Chem Theory Comput*, **2011**, 7, 4077.

38. E. D. Hedegård, J. Kongsted and S. P. a. Sauer, *Phys Chem Chem Phys*, **2012**, 14, 10669.

39. O. V Yazyev, L. Helm, V. G. Malkin and O. L. Malkina, *J Phys Chem A*, **2005**, 109, 10997.

40. G. A. Rolla, M. Botta and C. Platas-iglesias, *Inorg. Chem*, **2013**, 52, 3268.

41. M. A. Gonçalves, E. F. F. da Cunha, F. C. Peixoto and T. C. Ramalho, *Comput Theor Chem*, **2015**, 1069.

42. D. Esteban-Gómez, C. Cassino, M. Botta and C. Platas-Iglesias, *RSC Adv*, **2014**, 4, 7094.

Table 1. Correlation Coefficient Values for the studied compounds

| | | $[Fe(H_2O)_6]^{2+}$ | $[Mn(H_2O)_6]^{2+}$ |
|---|---|---|---|
| | **Wavalet** | **Correlation coefficient** | **Correlation coefficient** |
| 1 | bior1.1 | 0.5989 | 0.5268 |
| 2 | bior1.3 | 0.6969 | 0.6395 |
| 3 | bior1.5 | 0.5071 | 0.5307 |
| 4 | bior2.2 | 0.0341 | 0.1345 |
| 5 | bior2.6 | 0.1378 | 0.3110 |
| 6 | bior2.8 | 0.0644 | 0.3579 |
| 7 | bior3.1 | 0.5111 | 0.0095 |
| 8 | bior3.3 | 0.2299 | 0.4155 |
| 9 | bior3.5 | 0.2321 | 0.3437 |
| 10 | bior3.7 | 0.0816 | 0.4356 |
| 11 | bior3.9 | 0.1379 | 0.3710 |
| 12 | bior4.4 | 0.1370 | 0.4761 |
| 13 | bior5.5 | 0.0172 | 0.4745 |
| 14 | bior6.8 | 0.1986 | 0.3282 |
| 15 | rbio1.1 | 0.5807 | 0.5268 |
| 16 | rbio1.3 | 0.1622 | 0.4680 |
| 17 | rbio1.5 | 0.1929 | 0.4587 |
| 18 | rbio2.2 | 0.0102 | 0.2112 |
| 19 | rbio2.4 | 0.0785 | 0.4327 |
| 20 | rbio2.6 | 0.0280 | 0.4769 |
| 21 | rbio2.8 | 0.0509 | 0.4831 |
| 22 | rbio3.1 | 0.6589 | 0.7469 |
| 23 | rbio3.3 | 0.3896 | 0.3434 |
| 24 | rbio3.5 | 0.2294 | 0.3127 |
| 25 | rbio3.7 | 0.3078 | 0.3706 |

| | | | |
|---|---|---|---|
| 26 | rbio3.9 | 0.2825 | 0.4694 |
| 27 | rbio4.4 | 0.0936 | 0.4409 |
| 28 | rbio5.5 | 0.0195 | 0.3381 |
| 29 | rbio6.8 | 0.0420 | 0.3972 |
| 30 | db1 (haar) | 0.5655 | 0.5001 |
| 31 | db2 | 0.0374 | 0.2359 |
| 32 | db3 | 0.0615 | 0.4139 |
| 33 | db4 | 0.0576 | 0.4244 |
| 34 | db5 | 0.1099 | 0.4924 |
| 35 | db6 | 0.0108 | 0.4370 |
| 36 | db7 | 0.2540 | 0.4129 |
| 37 | db8 | 0.1056 | 0.4403 |
| 38 | db9 | 0.1324 | 0.4839 |
| 39 | db10 | 0.1104 | 0.5130 |
| 40 | db11 | 0.2414 | 0.4926 |
| 41 | db12 | 0.1262 | 0.5520 |
| 42 | db13 | 0.2166 | 0.5273 |
| 43 | db14 | 0.0084 | 0.5244 |
| 44 | db15 | 0.1051 | 0.4915 |
| 45 | db16 | 0.2166 | 0.4871 |
| 46 | db17 | 0.0656 | 0.4773 |
| 47 | db18 | 0.0125 | 0.4088 |
| 48 | db19 | 0.0744 | 0.5137 |
| 49 | db20 | 0.1171 | 0.4925 |
| 50 | db21 | 0.1814 | 0.5233 |
| 51 | db22 | 0.0121 | 0.4501 |
| 52 | db23 | 0.0818 | 0.4965 |
| 53 | db24 | 0.1005 | 0.5309 |
| 54 | db25 | 0.0350 | 0.4873 |
| 55 | db26 | 0.2749 | 0.4809 |
| 56 | db27 | 0.0417 | 0.4827 |
| 57 | db28 | 0.1109 | 0.5608 |
| 58 | db29 | 0.0513 | 0.5235 |
| 59 | db30 | 0.2277 | 0.4908 |
| 60 | db31 | 0.0019 | 0.4836 |
| 61 | db32 | 0.0186 | 0.4458 |
| 62 | db33 | 0.0420 | 0.5581 |
| 63 | db34 | 0.1422 | 0.5152 |
| 64 | db35 | 0.0560 | 0.5198 |
| 65 | db36 | 0.0971 | 0.4982 |
| 66 | db37 | 0.1711 | 0.4898 |
| 67 | db38 | 0.0211 | 0.4257 |
| 68 | db39 | 0.2473 | 0.4800 |
| 69 | db40 | 0.1840 | 0.5189 |
| 70 | db41 | 0.0348 | 0.4545 |
| 71 | db42 | 0.1661 | 0.4335 |
| 72 | db43 | 0.0523 | 0.4724 |
| 73 | db44 | 0.0561 | 0.4774 |

| | | | |
|---|---|---|---|
| 74 | db45 | 0.1518 | 0.4755 |
| 75 | sym2 | 0.0374 | 0.2359 |
| 76 | sym3 | 0.0615 | 0.4139 |
| 77 | sym4 | 0.0285 | 0.2972 |
| 78 | sym5 | 0.0099 | 0.4043 |
| 79 | sym6 | 0.2481 | 0.4004 |
| 80 | sym7 | 0.0590 | 0.4906 |
| 81 | sym8 | 0.2301 | 0.3504 |
| 82 | dmey | 0.4373 | 0.5949 |
| 83 | fk4 | 0.3230 | 0.4917 |
| 84 | fk6 | 0.0499 | 0.3728 |
| 85 | fk8 | 0.1277 | 0.5296 |
| 86 | fk14 | 0.1013 | 0.4629 |
| 87 | fk18 | 0.4288 | 0.5766 |
| 88 | fk22 | 0.4867 | 0.6161 |
| 89 | coif1 | 0.0523 | 0.2568 |
| 90 | coif2 | 0.1532 | 0.3600 |
| 91 | coif3 | 0.2874 | 0.3438 |
| 92 | coif4 | 0.0194 | 0.3445 |
| 93 | coif5 | 0.0139 | 0.3620 |

Table 2. 10 best wavelet models

| | $[Fe(H_2O)_6]^{2+}$ | $[Mn(H_2O)_6]^{2+}$ |
|---|---|---|
| **Wavalet** | **Correlation coefficient** | **Correlation coefficient** |
| bior1.1 | 0.5870 | 0.5268 |
| bior1.3 | 0.6969 | 0.6395 |
| bior1.5 | 0.5071 | 0.5307 |
| rbio1.1 | 0.5807 | 0.5268 |
| rbio3.1 | 0.6589 | 0.7469 |
| db1 (haar) | 0.5655 | 0.5001 |
| db26 | 0.2749 | 0.4809 |
| Dmey | 0.4373 | 0.5949 |
| fk18 | 0.4288 | 0.5766 |
| fk22 | 0.4867 | 0.6161 |

Table 3. Number of structures selected from each wavelet

| | $[Fe(H_2O)_6]^{2+}$ | $[Mn(H_2O)_6]^{2+}$ |
|---|---|---|
| **Wavalet** | **Número de estruturas selecionadas** | **Número de estruturas selecionadas** |
| bior1.1 | 46 | 42 |
| bior1.3 | 34 | 35 |
| bior1.5 | 49 | 47 |
| rbio1.1 | 43 | 44 |
| db1 (haar) | 47 | 45 |
| db26 | 16 | 23 |
| dmey | 14 | 18 |
| fk18 | 12 | 14 |
| fk22 | 22 | 21 |

Table 4. $A_{iso}$ values for different Wavelets

| | $[Fe(H_2O)_6]^{2+}$ | $[Mn(H_2O)_6]^{2+}$ | |
|---|---|---|---|
| **Wavalet** | **$A_{iso}$ (MHz)** | **$A_{iso}$ (MHz)** | |
| | **$Fe^{2+}$** | **$^{1}H$** | **$^{17}O$** |
| bior1.1 | 0.473 | 0.823 | 5.175 |
| bior1.3 | 0.496 | 0.853 | 5.382 |
| bior1.5 | 0.491 | 0.831 | 5.375 |
| rbio1.1 | 0.482 | 0.821 | 5.112 |
| db1 (haar) | 0.488 | 0.915 | 5.275 |
| db26 | 0.460 | 0.824 | 5.021 |
| dmey | 0.454 | 0.811 | 5.105 |
| fk18 | 0.459 | 0.815 | 4.923 |
| fk22 | 0.519 | 0.825 | 5.435 |
| *Experimental* | 0.500 | 0.860[42] | 5.400[42] |

**Figure Captions:**

**Figure 1:** Filter bank for DWT signal decomposition.

**Figure 2:** Optimal thresholds δj* (vertical axis: threshold value; horizontal axis: individual threshold), $[Fe(H_2O)_6]^{2+}$ of fk8 wavelet.

**Figure 3:** Energy of MD $[Fe(H_2O)_6]^{2+}$ conformations (original and compressed) at each time (fs) for the ten selected wavelets.

**Figure 4:** Structure of the $[Fe(H_2O)_6]^{2+}$ complex.

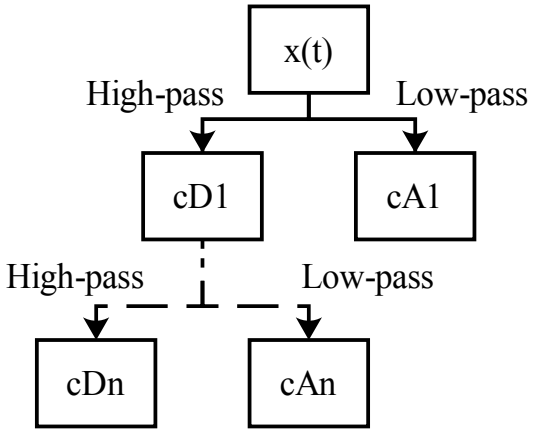



**Figure 1.**

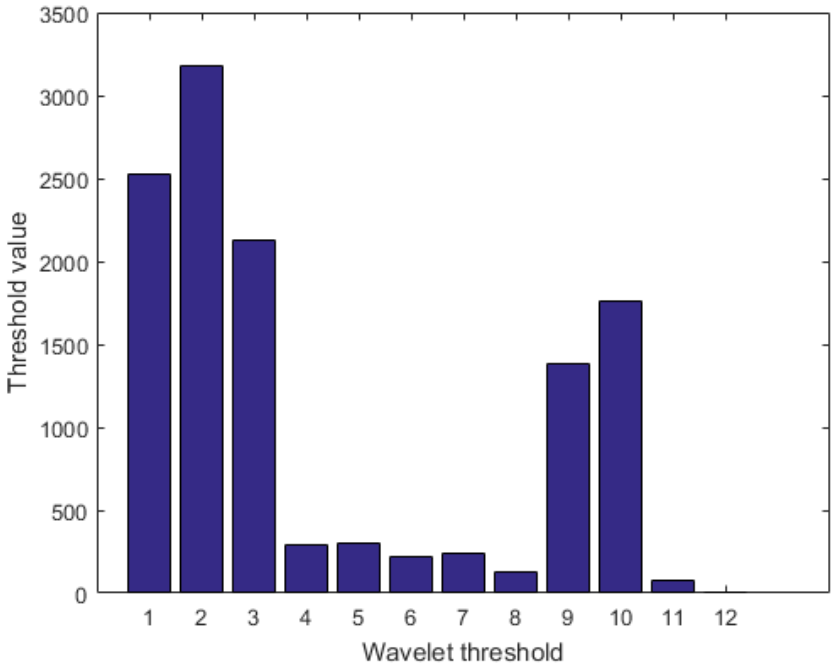
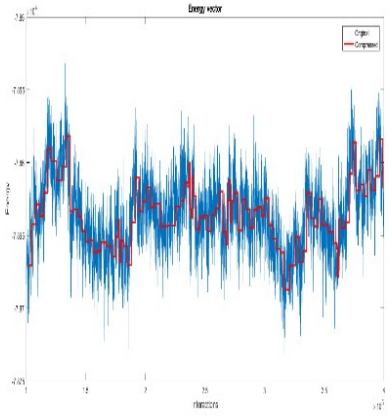
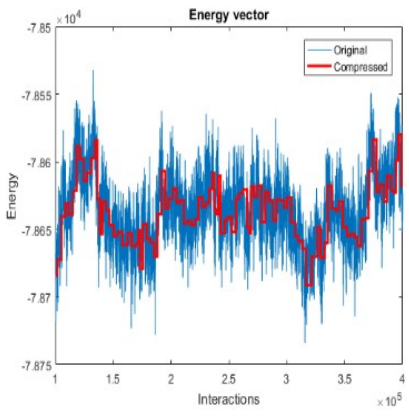

3500
3000
2500
2000
1500
1000
500
0
Threshold value
1
2
3
4
5
6
7
8
9
10
11
12
Wavelet threshold

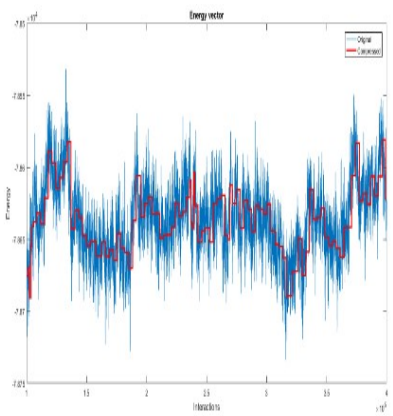
**Figure 2**



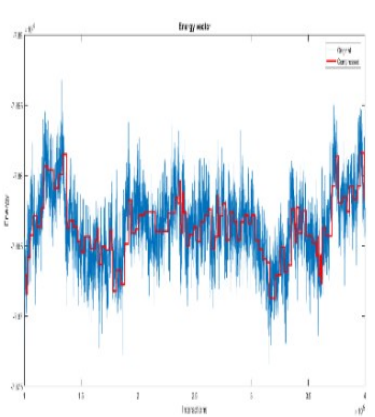
bior1.1



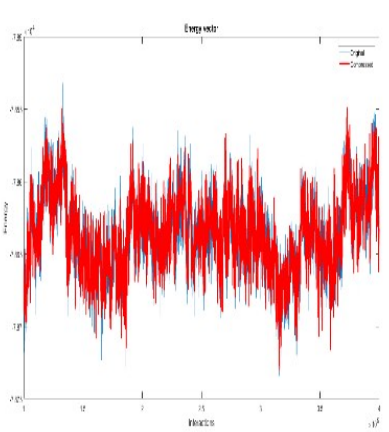

bior1.3



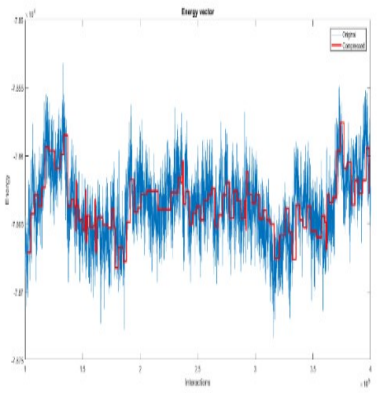
bior1.5



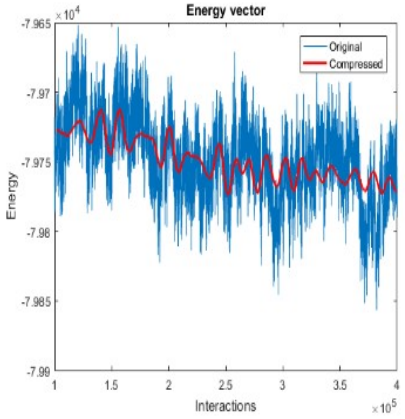
rbior1.1



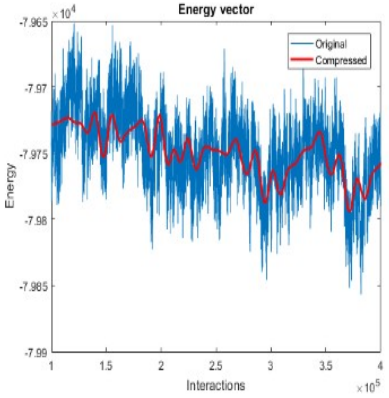
rbior1.3



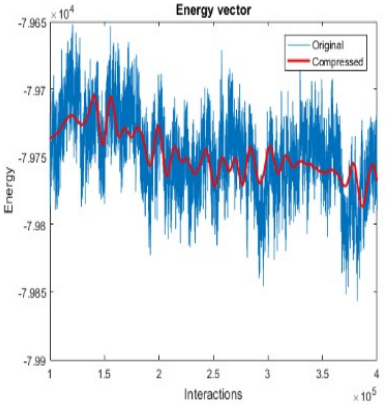
db1 (Haar)



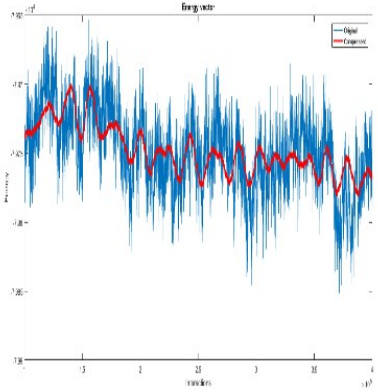

db26




dmey




fk18



fk22

**Figure 3**

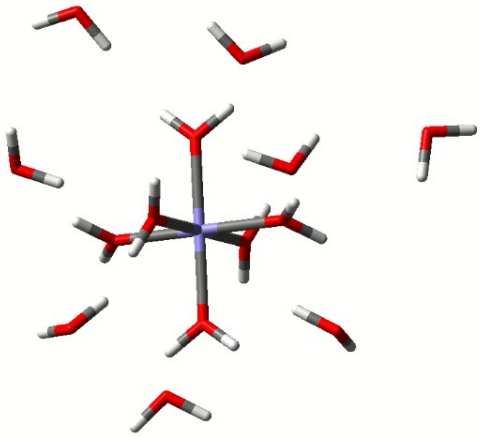


**Figure 4**