BIRZEIT UNIVERSITY

THz Frequency Spectrum of BLIP Protein-Water Interaction Energy

الطيف الترددي في مدى التيراهيرتز للطاقة التفاعلية بين بروتين BLIP والماء المحيط به

By
KHADEEJA MOSA ABO LEBDEH

Thesis committee:
    Dr. Wael Karain (Principal advisor)
    Dr. Esma'el Badran (Member)
    Dr. Abdallah Sayyed-Ahmad (Member)

This Thesis is submitted in partial fulfillment of the requirements for the Master's Degree in Physics from the Faculty of Graduate Studies at Birzeit University, Palestine.

21/6/2018

THz Frequency Spectrum of BLIP Protein-Water Interaction Energy

By
KHADEEJA MOSA ABO LEBDEH

Accepted by the Faculty of Graduate Studies, Birzeit University, in partial fulfillment of the requirements of the Master's Degree in Physics.

Thesis committee:
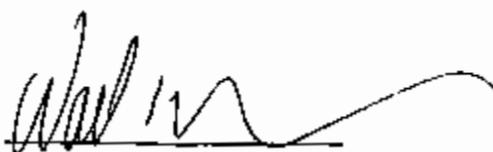
Dr. Wael Karain (Principal advisor)

Dr. Esma'el Badran (Member)
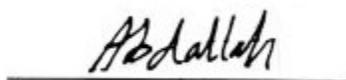
Dr. Abdallah Sayyed-Ahmad (Member)

21/6/2018

DEDICATION

Special thanks

To my friends and my family especially my soulmate "Ahmad" for providing me with unfailing support and continuous encouragement through the process of researching and writing this thesis.

To all people I ever love and respect

To all living souls seeking knowledge

God bless you all

ACKNOWLEDGMENT

Table of Contents

Abstract

The vibration spectrum of water hydrogen bonds and the large amplitude motion modes of proteins can be characterized by the THz frequency spectrum of protein-water non-bonded interaction energy. This spectrum can be computed using recurrence plots based on Wiener-Khinchin method (RPWK). In this thesis, THz frequency spectra for non-bonded interaction energy between all residues in the β-Lactamase Inhibitory Protein (BLIP), and water molecules within 10 angstroms are calculated at 150K and 310K temperatures, respectively. The results obtained are also compared to those determined using the auto-covariance based Wiener-Khinchin method (WK). Moreover all residues are grouped together according to their sidechain physico-chemical properties to compare their spectra. We found that RPWK method robustly detects frequency data points for residues especially at high frequencies. We also found that most of detected are hydrophilic which consistent with their likely high contact with the solvent.

ملخص

يمكن وصف طيف الاهتزاز للروابط الهيدروجينية في الماء وأشكال الحركة في جزيء البروتين باستخدام الطيف الترددي في مدى التيراهيرتز للطاقة التفاعلية بين البروتين والماء، ويمكن حساب هذا الطيف باستخدام طريقة تسمى ب "Recurrence plots based on Wiener–Khinchin"، وفي هذه الرسالة تم حساب الطيف الترددي في مدى التيراهيرتز للطاقة التفاعلية بين جميع الأحماض الأمينية في بروتين BLIP وجزيئات الماء ضمن مسافة 10 أنجستروم على درجتي الحرارة 150 كلفن و 310 كلفن. تم مقارنة النتائج التي حصلنا عليها مع تلك التي نتجت من استخدام طريقة أخرى تسمى " Auto–covariance based Wiener–Khinchin". إضافة إلى أنه تم تصنيف جميع الأحماض الأمينية بناء على الخصائص الفيزيائية والكيميائية لسلاسلها الجانبية للمقارنة بين أطيافها الترددية. وجدنا ان الطريقة الأولى "Recurrence plots based on Wiener–Khinchin" كشفت بفعالية عن نقاط ترددية خاصة عند الترددات العالية. كما وجدنا أن معظم الأحماض الأمينية التي تم الكشف عنها تشترك في كونها محبة للماء (Hydrophilic) وهذا أمر منطقي لأنها تتفاعل مع المذيب.

LIST OF FIGURES

LIST OF TABLES

1. Introduction

1.1 Molecular dynamics simulation

Molecular dynamics (MD) is a computer simulation method that aims to study many physical and chemical properties of an interacting system of atoms and molecules. It is widely used to study structure of proteins [1]. In particular, it gives a dynamic dimension to structural data, and hence helps interpreting many biological and chemical processes [2].

MD was developed in the middle of the twentieth century [3]. It was used to study different systems like the following examples. In 1975, computer simulations were used by Levitt & Warshel to simulate protein folding [4]. In 1982, van Gunsteren and Karplus investigated the dynamics of proteins in a solvent and a crystalline environment [5]. In 1995, two molecular dynamics simulations were used by Brunne et al. to investigate the properties of bovine pancreatic trypsin inhibitor (BPTI) in solvent [6]. Recently, the non-bonded interaction energy were calculated between the β-lactamase inhibitory protein, the residues 49ASP, 53TYR, and 142PHE in this protein, and water at four different temperatures [7].

Molecular dynamics simulations were used in the beginning to study many systems like rigid molecules, alkanes and trypsin inhibitor protein [3]. However, it had severe limitations in simulation time and simulated system size. For example, a study aimed in 1959 to calculate the solution of many-body problem using molecular dynamics and the Monte Carlo method. The number of molecules that could be handled in this study was about five hundred. This limitation was due to the speed and the memory capacity of computers back then, which roughly required half an hour for one collision per molecule [8]. Nowadays, advances in computational capabilities allow studying systems consisting of millions of atoms

A classical molecular dynamics simulation gives a numerical solution of the classical Newton's equations of motion [9]

$$m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{f}_i \qquad \text{(eq 1.1.1)}$$

$$\vec{f}_I = -\vec{\nabla}_i U \qquad \text{(eq 1.1.2)}$$

where $\vec{r}_i$ and $\vec{f}_i$ represents the coordinates of each atom and the force acting on it, respectively. The latter is the derivative of the potential energy $U(\vec{r_1}, \vec{r_2}, \ldots \vec{r_N})$ of a system of N interacting atoms [9]. In protein-water systems, potential energy includes non-bonded inter- and intra-molecular interactions like van der Waals and Coulomb interactions [10, 11]. The rest of the intramolecular interactions are given by the simplest molecular model that include terms of the following kind:

$$U_{intramolecular} = \frac{1}{2} \sum_{bonds} k_{ij}^{r} \left(r_{ij} - r_{eq}\right)^2$$

$$+ \frac{1}{2} \sum_{\substack{bend \\ angles}} k_{ijk}^{\theta} \left(\theta_{ijk} - \theta_{eq}\right)^2$$

$$+ \frac{1}{2} \sum_{\substack{torsion \\ angles}} \sum_m k_{ijkl}^{\varphi,m} \left(1 + \cos\left(m\varphi_{ijkl} - \gamma_m\right)\right) \dots \text{(eq1.1.3)}$$

where k^r is the bond stretching force constant. The "bonds" typically involves the separation $r_{ij} = |r_i - r_j|$ between adjacent pairs of atoms in a molecular framework, (eq1.1.3) assumes a harmonic form with specified equilibrium separation (r_{eq}). k^{θ} is the bending force constant. The "bend angles" θ_{ijk} are between successive bond vectors such as $r_i - r_j$ and $r_j - r_k$, and therefore involve three atom coordinate. $k^{\varphi,m}$ is the torsional (bond twisting) force constant, and m is the periodicity of the rotation. The "torsion angles" φ_{ijkl} are defined in terms of three connected bonds, hence four atomic coordinates such as r_{ij}, r_{jk} and r_{kl}. The geometry is illustrated the Fig.1 [9].
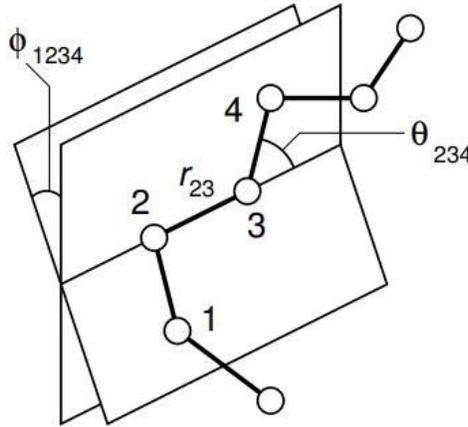
Fig. 1: Geometry of a simple chain molecule explaining the definition of intramolecular bond vector r_{23}, bend angle θ_{234}, and torsion angle φ_{1234} used in molecular dynamics simulations.

The intramolecular bonding interactions must be considered for molecules. This term "intramolecular bonding" is defined as the forces between atoms within a molecule, which is also responsible for its chemical bonding. So, these interactions are due to valence electrons that determine the chemical behavior of molecules like covalent bond [12]. The simplest form of these interactions will include terms as eq1.1.3

The intermolecular bonding interactions also must be considered for molecules. They play a critical role in biological events like binding molecules to initiate an action and then separate. Therefore, intermolecular bonding interactions help in bounding molecules for a short time. They can be classified as hydrogen bonding, charge-charge interactions and Van der Waals forces.

- The Van der Waals potential is given by

$$U(r_{ij}) = 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \dots\dots (eq1.1.4)$$

where r_{ij} is the distance between two particles (i and j), ϵ is an energy parameter determines the depth of the minimum in the potential curve and leads to more stable bonds as it become larger and σ is a length parameter that is equal to intermolecular separation at which the potential energy is equal to zero [13, 14].

Van der Waals are very weak electrostatic forces that result when an atom or group of atoms induce an opposite dipole on a non-bonded neighbor [15].

- The electrostatic potential energy of two charges (q_1, q_2) can be expressed as

$$U = k\frac{q_1 q_2}{r_{12}} \dots\dots (eq1.1.5)$$

where k=8.9875×10^9 NM ^{2}C $^{-2}$, and r$_{12}$ is the distance between the two charges [16].

The electrostatic interaction is a force between charged atoms, which may be attractive or repulsive. This type of interaction is very significant in biomolecules [17].

Hydrogen bonds are attractive interactions between a hydrogen atom bounded to another higher electronegative atom like oxygen or nitrogen. In proteins, the hydrogen bond N-H…O=C is the most dominant one. It is responsible for forming and stabilizing the beta sheets, alpha helices and turns in the secondary structure of protein [18]

Molecular dynamics simulations provide us a time series of energy values, for example. A time series is a set of observation values that are computed over some time interval [19]. These time series can be analyzed using many methods that aim to build a suitable model for the data [20]. In this thesis we utilize the Wiener-Khinchin method [21] and recurrence plot based Wiener-Khinchin Method.

1.2 Wiener-Khinchin Method
Wiener-Khinchin is used often to analyze time series resulting from molecular dynamic simulations. This method is based on three steps that determine the frequency spectra for a time series using the Fourier Transform [22].

The first step is to determine the autocorrelation function of the time series. This function is a mathematical fundamental measurement of the linearity between two points in the time series [23]. It helps us to see the change in patterns between variables when the conditions are changed [24]. For example, (Fig. 2) shows how the autocorrelation functions of lipid mixing between two proteins (protein-free liposomes and 7 µM sMunc18-1 protein) at different time points [25].
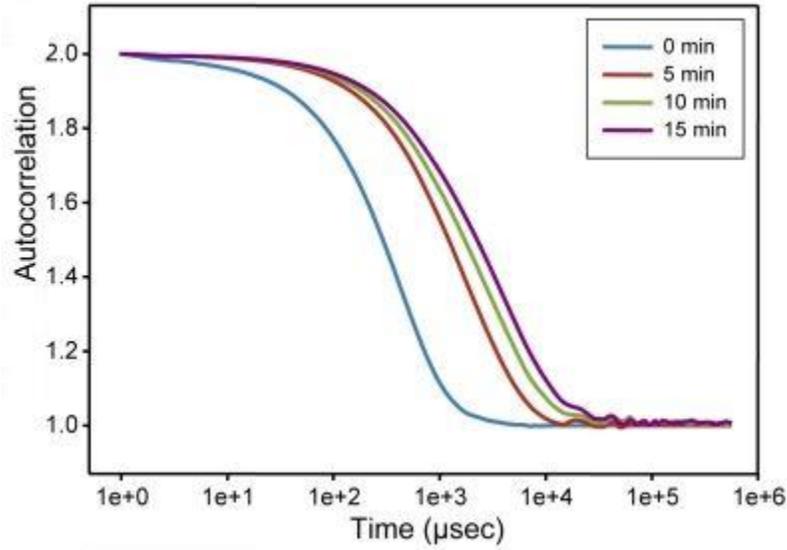
Fig. 2: Autocorrelation functions obtained at different time points after mixing protein-free liposomes with 7 µM sMunc18-1 protein [25].

Autocorrelation functions are used for signal analysis to detect their periodicity [26]. As the following (Fig.3) [27]
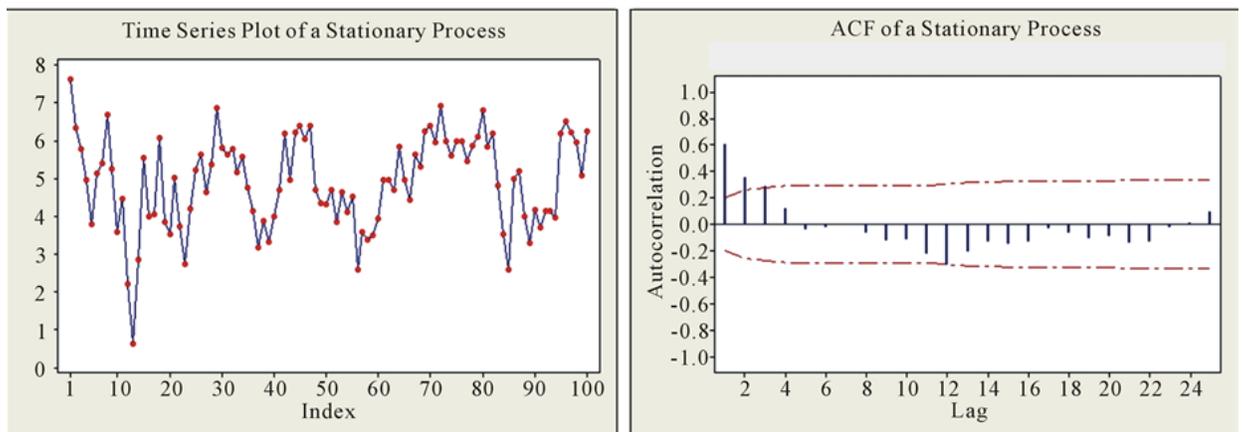
Fig. 3: A time series of a stationary [e.g. its mean and variance do not change over time] process (left) and its autocorrelation function (right).

The autocorrelation function can be defined for a time series $x(n)$ as follows:

$$C_x(\tau) = \varepsilon\{x(i)x^*(i+\tau)\} = \frac{1}{N}\sum_{i=0}^{N-1-\tau} x(i)x^*(i+\tau) \dots \text{(eq 1.2.1)}$$

where $\varepsilon\{.\}$ donates the expectation operator, $x(i)$ is a time series, N is the length of the time series and τ is the time lag [29].

The second step is to find the Fourier transform for the autocorrelation function. A Fourier transform will transfer a signal from the space or time domain to the frequency domain by integrating the series multiplied be by a complex exponential that is involved the frequency [30] (Fig.4).
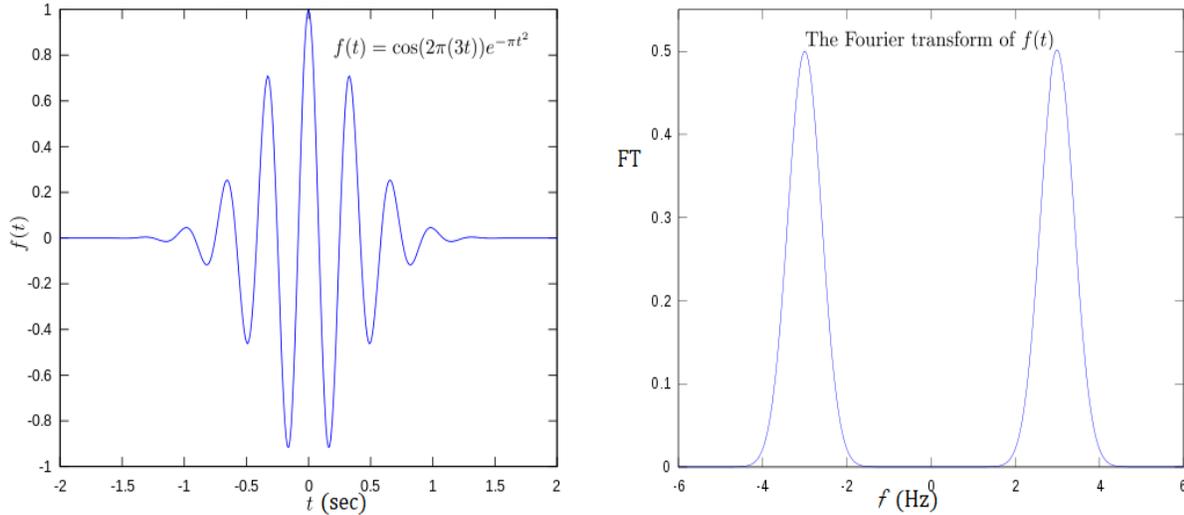
Fig. 4: The function $f(t) = cos(2\pi(3t))e^{-\pi t^2}$ (left) and its Fourier transform using Matlab [31].

The main use of it is to characterize the importance of each frequency band in the time series [32]. Fourier transform of the autocorrelation function $C_x(\tau)$ of a signal is defined as follows:

$$S_x(\omega) = \sum_{\tau=-\infty}^{\infty} C_x(\tau)e^{-j\omega\tau} \ \dots\dots \text{(eq 1.2.2)}$$

where ω is the frequency at which $S_x(\omega)$ is calculated, j is the imaginary unit and τ is the time lag.

Then the third step is to find the power spectrum which is the magnitude squared of the Fourier transform of the autocorrelation for signal as follows [29]

$$S_x(\omega) = \frac{1}{N}|X_n(\omega)|^2 \dots\dots \text{(eq 1.2.3)}$$

$$X_n(\omega) = \sum_{n=0}^{N-1} x(n)e^{-j\omega n} \quad \text{.... (eq 1.2.4)}$$

where N is the length of the time series, j is the imaginary unit and $x(n)$ is a time series.

However, WK is not suitable for all types of time series. Fourier transforms are more suitable for stationary and linear system that exhibits dynamics that do not reside in higher dimensional space [29, 33]. If one of these conditions is not met, we have to use another appropriate method to calculate the power values of the frequency spectra from the corresponding time series. One of these alternative methods is known as recurrence plot Recurrence Plot based Wiener-Khinchin RPWK method [29].

1.3 Recurrence Plot based Wiener-Khinchin Method

The phase space is dimensional space that is made of generalized or representative points for particles system. Each point corresponds to a definite state that is specified by the two independent variables or coordinates $q_1, q_2, ...q_n, p_1, p_2, ...p_n$ which called the canonical variables. Each generalized momentum (p_k) is canonically conjugated to generalized coordinate (q_k) [43]. If the initial state or conditions of motion (position and velocity or momentum) are known at time (t_i), then the Newton's equations of motion determine the future trajectory of this motion at any later time (t_f) [35].

The trajectory of any system can be built by using the method of time delays. This method depends on the dimensions of an available state of the time series such as time delay and embedding dimension to reconstruct the trajectory for the whole system [36]. The main steps while using this method is to calculate the values of time delay and embedding dimension for the time series.

So if there is a time series of a scalar variable $X(t_i)$. A vector in phase space in time t_i can be constructed as follows:

$$X(t_i) = [x(t_i), x(t_i + \tau), x(t_i + 2\tau), \ldots\ldots, x(t_i + (m-1)\tau)] \quad \text{....(eq 1.3.1)}$$

where τ is a time delay and m is an embedding dimension [37].

The time delay (τ) is specified by the first minimum in the mutual information MI function. This function measures the dependence between two variables. It can be expressed as follows

$$I^{XY} = H(X) + H(Y) - H(X,Y) \quad \text{.... (eq 1.3.2)}$$

where I^{XY} is the mutual information between time series (X and Y), H(X, Y) is the joint entropy that is measured from a joint histogram and H(X) is the entropy for time series X that can be expressed as follows

$$H = -\sum_i p_i \log_2 p_i$$

where p_i is the probability of observing symbol i and the summation is taken over all i [38].

If the mutual information value is zero, then the two variables (X, Y) are not related (dependent) to each other. As the mutual information value will be higher, the dependency between X and Y will be stronger, [39].

The embedding dimension (m) can be determined by using the False Nearest Neighbors (FNN) method. The False Nearest Neighbors (FNN) method is a tool to select the embedding dimension. Since the MI function is used to find the time delay, so FNN method is selected to determine the embedding dimension results [40]. FNN method assumes that points in phase space with small embedding dimension are far apart in the original space, so they could be closer as neighbors in the reconstructed one [41].

Recurrence plots RP are non-linear graphical tools used to display the phase space trajectories in time series in two dimensional maps. These recurrence plots provide us plenty of information that are very useful in giving us new insights in the signals' analysis field. They were introduced in 1987 by Eckmann *et al.* to analyze a nonlinear data time series by visualizing recurrences of trajectories in higher-dimensional phase space in this time series [42].

Recurrence plots depend on computing the recurrence matrix ($R_{i,j}$) as follows

$$R_{i,j} = \Theta\ (\epsilon - D_{i,j})\ \text{.....\ (eq 1.3.3)}$$

$$D_{i,j} = \|x_i - x_j\|\qquad \text{.......(eq 1.3.4)}$$

where i,j=1….N is the number of states, Θ (.) is Heaviside step function, ϵ is a threshold distance, $\| . \|$ a norm, x_i is a state at time (i), x_j is a state at time (j) and $D_{i,j}$ is the distance matrix between all state space vectors. If the two states are recurrent then $R_{ij} = 1$ and it will be represented as a black dot. If they are non-recurrent then $R_{ij} = 0$ and it will be represented as a white dot [42, 20]. The main diagonal corresponds to the recurrence of the points itself [43]. Fig.5 shows a RP example [44].
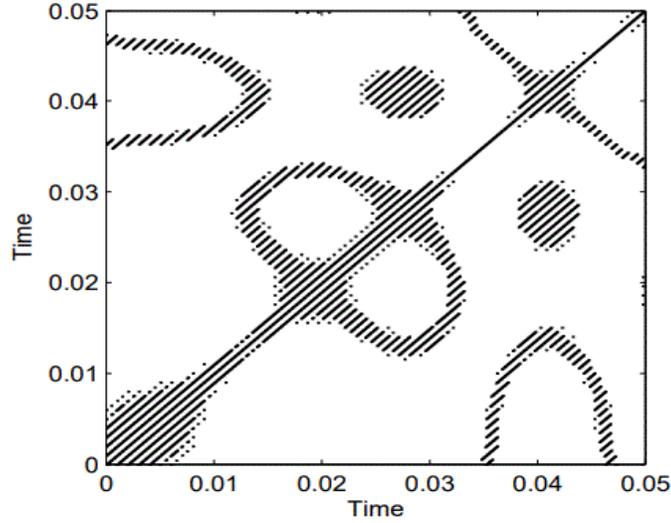
Fig. 5: Recurrence Plot (RP) of a modulated harmonic oscillation $sin(2000\pi(\pi + t) + 2\pi \, sin(88\pi t)t)$. Used RP parameters: embedding dimension m=3 and time delay τ=1.

Recurrence plot based Wiener-Khinchin method depends on choosing the proper embedding dimension (*m*) and time delay (τ) to reconstruct the phase space. Also it depends on using the Tau-recurrence rate spectrum $RR(\tau)$ function. This function $RR(\tau)$ gives the probability that a state will recur after a time delay (τ) by considering all diagonal lines that are parallel to the diagonal line whose distance is τ from the main diagonal [43]. It can be defined as follows

$$RR(\tau) = \frac{1}{N-\tau}\sum_{i=1}^{N-\tau} R(i, i + \tau) = \frac{1}{N-\tau}\sum_{i=1}^{N-\tau} R_{i,i+\tau} \ \dots\dots \text{(eq 1.3.5)}$$

where τ is the time delay between the main diagonal and the parallel diagonal line, i =0, 1.....N is the length of the time series and $R_{i,i+\tau}$ is the recurrence matrix for states whose distance is τ from each other. So $RR(\tau)$ is determined by summing the values of $R_{i,i+\tau}$ and divide them by $N - \tau$.

Then the power spectrum can be calculated by finding the magnitude squared of the Fourier transform of $RR(\tau)$. So $C_x(\tau)$ will be replaced by tau-recurrence rate spectrum $RR(\tau)$ function as follows:

$$S_x(\omega) = \sum_{\tau=-\infty}^{\infty} RR(\tau)e^{-j\omega\tau} \ \dots\dots \text{(eq 1.3.6)}$$

by replacing the expectation value of the time series by the expectation value of the recurrence matrix $(R_{i,i+\tau})$

$$C_x(\tau) = \varepsilon \{\Theta \ (\epsilon - \|x_i - x_{i+\tau}\|)\} \ \dots\dots \text{(eq 1.3.7)}$$

In brief, RPWK is a method to find the power spectrum by applying the Fourier transform on the probability of a state recurrence rate $RR(\tau)$ function. This function $RR(\tau)$ bases on a

8

recurrence plots by using the proper embedding dimension, time delay and the recurrence matrix [29].

1.4 Solvated Protein

A suitable case to compare between Wiener-Khinchin method (WK) and Recurrence plot based Wiener-Khinchin method (RPWK) in the field of molecular dynamics simulations is the interaction between a protein and its surrounding solvent. Water molecules play an important part in the structure, stability, dynamics, and function of proteins and other biomolecules [45, 46, 47]. Also protein functional properties like solubility are strongly related to the pattern in which protein and water interact with each other [48].

Water does not interact only with the protein's surface, but it also interacts with the protein interior backbone and side chains [47]. These interactions help to ensure flexible conformation conditions that are required for a functional active protein [49]. Since water has the ability to make essential hydrogen bonds to the polar parts of protein [50], this ability has an important effect on the strength of hydrogen bonds between these parts [15]. The strength and the number of hydrogen bonds also play a role in the different affinities (e.g. bonds or links) of bounded water molecules [51].

Hydrated protein undergoes a dynamical transition –also named as glass transition-at around 200 K. This transition describes protein motion that is coupled near the protein surface to water molecules [52, 53].

Above 200K, the dynamical transition is due to the motions of side chains, and motions over the whole of solvated protein [54]. So the solvated protein changes from rigid, glasslike, harmonic vibrations and no biological functions, to become a flexible, liquid-like, anharmonic dynamics and biologically active one [53, 54]. The life time of protein-water hydrogen bond decreases at dynamical transition temperature, while the number of participating water molecules increases [55].

1.5 Water Spectrum

The water vibration frequency spectrum ranges from zero up to $v<1,000$ cm^{-1} ~30THz [56]. In this range, it is very critical to any structural changes in protein [7]. THz spectroscopy shows a significant correlation between structural changes in biomolecules and changes in the water hydration dynamics, which also affect the THz absorption of solvated protein [57]. The water frequency spectrum consists of three main broad regions (Fig. 6) which are:

1) The hydrogen-bond bending (HBB) band which is due to molecule pairs moving in a transverse direction to the hydrogen bond connecting them. It is centered at~1.5THz.

2) The hydrogen-bond stretching (HBS) band which is due to water molecule pairs moving in the same direction as the hydrogen bond connecting them. It is centered at ~6THz.

3) The libration (LIB) bands which are due to frustrated rotations of water molecules about their center of mass due to the presence of other water molecules around them. They consist of the L_1 libration band that is centered at ~12THz, and the L_2 libration band that is centered at ~20THz [58, 7].
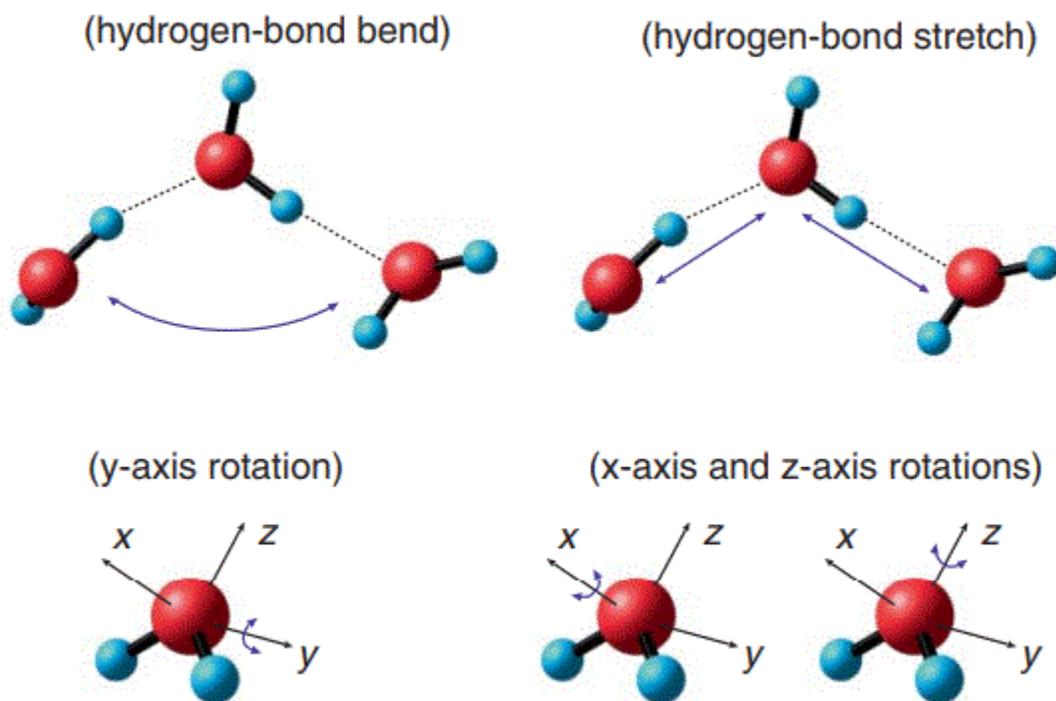
Fig. 6: Schematics of the four intermolecular motions [59].

There is also a relaxation region at frequencies less than 0.6THz [56]. Moreover, proteins are flexible and undergo large-amplitude modes that affect the hydration shell surrounding them [60]. So the THz frequency spectrum includes the frequency range of large-amplitude motions of proteins. The coupling between protein and water extends into the hydration shell surrounding the protein for at least 10Å [10] and hence solvation and collective modes can be probed by THz spectroscopy on a length scale around (0.3–1 nm) [60].

1.6 β-Lactamase Inhibitory Protein

In this thesis, the solvated protein that is used in the simulation is the β-Lactamase Inhibitory Protein BLIP [7]. It contains 165 amino acids. This protein is secreted by the soil bacterium Streptomyces clavuligerus. Its function is to inhibit β-lactam enzymes, which hydrolyze β-lactam antibiotics and nullify their effect [61, 62].

BLIP is a flat shaped protein with unique fold, which means it has a unique pattern of the polypeptide chain in space [63]. As Fig. 7 shows, each protein molecule contains two repeated domains. Each domain has a helix-loop-helix motif that packs against a four standard antiparallel β-sheet [64]. Motif and domain have two overlapped meanings. Motif is a set of residues that are functionally important in protein, while domain is a structurally unit that might contain one or motifs [65].
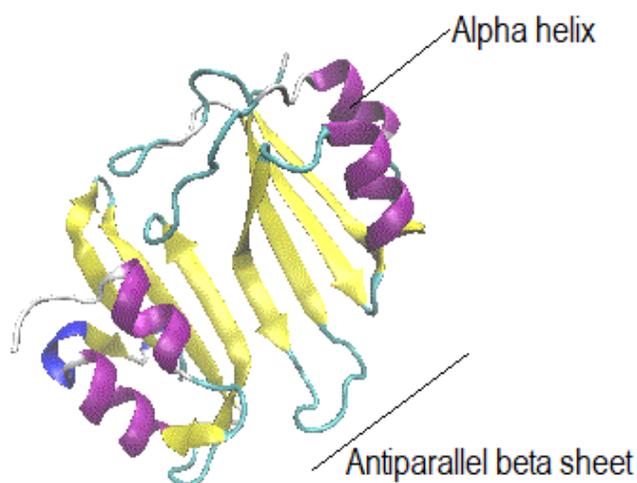
Figure 7: The shape of β-Lactamase Inhibitory Protein -using VMD software- which contains alpha helices and antiparallel beta sheets

In this thesis the non-bonded interaction energy time series are investigated between all BLIP residues and water molecules within 10Å from each residue at 150K and 310K respectively. The water-BLIP system is studied by Wiener-Khinchin method (WK) and Recurrence plot based Wiener-Khinchin method (RPWK) to calculate the THz frequency spectra for non-bonded interaction energy time series extracted from molecular dynamics simulations.

The THz frequency for the non-bonded interaction energy between the '165' residues in BLIP protein molecule and water molecules within 10Å from each residue using WK method and RPWK method are graphed versus the Relative Solvent Accessibility (RSA) for each residue. These values of RSA give an insight to figure out the properties of protein such as its function and structure since RSA helps in predicting the protein secondary structure [66]. Also it help us to

11

distinguish between buried or exposed residues. They are calculated by dividing the accessible surface area (ASA) for residues by a maximum reference theoretical value for the solvent-ASA of the corresponding standard residue by using the CSU program to determine ASA for each BLIP's residue [67].

1.7 Questions

Our goals are to compare the performance of Wiener-Khinchin method (WK) and Recurrence plot based Wiener-Khinchin method (RPWK) in detecting THz frequencies at temperatures below and above the protein dynamical transition. Also we aim to search for common characteristic frequencies in THz range for certain groups of residues when they interact with water molecules. So basically, these two questions will be addressed:

1. How well will the RPWK perform when compared to WK in detecting the frequency data points for non-bonded interaction energy between all residues in BLIP and water molecules?

2. What will be the similarities and differences between residues in solvated protein depending on their characteristic THz frequencies?

2. Methods

2.1 Molecular dynamics simulations

Molecular dynamics simulations were performed and analyzed at 150K and 310K using the computer programs NAMD [68] and VMD [69]. The starting BLIP structure (PDB entry 3gmu) was downloaded from the protein data bank [70]. Periodic boundary conditions were used in an 80Å X 80Å X 80Å box. The protein was neutralized and solvated using 20 Cl - and 22 Na + ions as well as 15264 TIP3P waters. The Particle-Mesh-Ewald method was used to do the electrostatic calculations [18]. A switching function was used for non-bonded interactions with a switch distance of 10Å and a cutoff distance of 12Å. A pair-list distance of 14Å was used. All simulations were performed at constant pressure of 1atm with an integration step of 2fs. The protein was minimized using the conjugate gradient method for 5000 steps (10ps). This was followed by a gradual heating from an initial temperature of 100K in steps of 10K, with the simulation running for 10ps at each temperature step, until reaching one of the respective two final temperatures. The equilibration period was 5ns long (depending on the solvent and solute). Each production run was 1ns long.

2.2 Non-bonded interaction energy time series

Non-bonded interaction energy time series -in Kcal/mol- between protein residues and solvent molecules within 10Å from each respective residue, are calculated using the Namdenergy plugin in VMD [69]. Each energy series have a length equal to the number of frames s in the dcd file. The time period between consecutive pair of values is 20fs, for a total time period of over 20 ps for the whole time series. This covers the THz frequency range 0.05THz-25THz.

2.3 RPWK method

The recurrence parameters of embedding dimension and time delay are calculated for each time series using the CRP toolbox [71]. A threshold parameter giving a recurrence value of around 25% using the maximum norm was found to give the best results, since it corresponds to 10% of the mean and 5% of the maximal phase space diameter [7, 72]. The RPWK frequency spectra are calculated using the RRSPEC subroutine in the CRP toolbox [71], which evaluate eq1.3.4 to find the Tau-recurrence rate spectrum $RR(\tau)$ using the time delay and embedding dimension value data. For example, the calculation procedure that was used to get the frequency spectrum for the 12ILE residue at 150K is detailed in APPENDIX A.

2.4 WK method

The autocorrelation function and Fourier transform function are calculated for each time series using the Matlab toolbox [31]. The frequency spectra up to 50THz are then calculated using the Fourier transform and its conjugate For example, the calculation procedure that was used to get the frequency spectrum for the 12ILE residue at 150K is detailed in APPENDIX B.

3. Results and discussion

3.1 Frequency power spectra at 150K and 310K

In this section, we will show the frequency data points which are detected by WK and RPWK at both temperatures 150K and 310K, in each band of power spectra (relaxation region, HBB band, HBS band, L_1 band and L_2 band), for the non-bonded interaction energy between the residues in BLIP protein and water molecules within 10Å from each residue, over a time period of 20ps. The residues shown will have power values with z-score larger than 1. It will be noisy if the residues under one z-score were shown. Each band will be shown in a separate table. The residues in brackets are detected by one method and not the other.

Table 1: The frequency data points which are detected by WK and RPWK at 150K and 310K in the relaxation region.

Residue's name	Method	At 150K	Residue's name	Method	At 310K
ALA	WK	1, 7, 25, 27, 46, 47, 52, 54, 61, 62, 63, 65, 77, 80, 86, 98, 102, 118, 123, 147	ALA	WK	1, 7, 25, 27, 46, 47, 52, 54, 61, 62, 63, 65, 77, 80, 86, 98, 102, 118, 123, 147
	RPWK	1, 7, 25, 27, 46, 47, 52, 54, 61, 62, 63, 65, 77, 80, 86, 98, 102, 118, 123, 147		RPWK	1, 7, 25, 27, 46, 47, 52, 54, 61, 62, 63, 65, 77, 80, 86, 98, 102, 118, 123, 147
VAL	WK	3, 21, 67, 91, 93, 100, 104, 125, 134, 155, 165	VAL	WK	3, 21, 67, 91, 93, 100, 104, 125, 134, 155, 165
	RPWK	3, 21, 67, 91, 93, 100, 104, 125, 134, 155, 165		RPWK	3, 21, 67, 91, 93, 100, 104, 125, 134, 155, 165
SER	WK	35, 39, 60, 69, 71, 79, 108, 113, 121, 128, 130, 138, 139, 146	SER	WK	35, 39, 60, 69, 71, 79, 108, 113, 121, 128, 130, 138, 139, 146
	RPWK	35, 39, 60, 69, 71, 79, 108, 113, 121, 128, 130, 138, 139, 146		RPWK	35, 39, 60, 69, 71, 79, 108, 113, 121, 128, 130, 138, 139, 146

THR	WK	5, 10, 17, 32, 55, 59, 82, 84, 92, 96, 103, 110, 111, 122, 126, 140, 152	THR	WK	5, 10, 17, 32, 55, 59, 82, 84, 92, 96, 103, 110, 111, 122, 126, 140, 152
	RPWK	5, 10, 17, 32, 55, 59, 82, 84, 92, 96, 103, 110, 111, 122, 126, 140, 152		RPWK	5, 10, 17, 32, 55, 59, 82, 84, 92, 96, 103, 110, 111, 122, 126, 140, 152
GLN	WK	11, 13, 19, 20, 72, 90, 99, 106, 157, 161	GLN	WK	11, 13, 19, 20, 72, 90, 99, 106, 157, 161
	RPWK	11, 13, 19, 20, 72, 90, 99, 106, 157, 161		RPWK	11, 13, 19, 20, 72, 90, 99, 106, 157, 161
LYS	WK	8, 66, 70, 74, 87, 159	LYS	WK	8, 66, 70, 74, 87, 159
	RPWK	8, 66, 70, 74, 87, 159		RPWK	8, 66, 70, 74, 87, 159
GLU	WK	28, 31, 73, 114	GLU	WK	28, 31, 73, 114
	RPWK	28, 31, 73, 114		RPWK	28, 31, 73, 114
PRO	WK	78, 81, 117, 120	PRO	WK	78, 81, 117, 120
	RPWK	78, 81, 117, 120		RPWK	78, 81, 117, 120
MET	WK	4, 16, 95	MET	WK	4, 16, 95
	RPWK	4, 16, 95		RPWK	4, 16, 95
TRP	WK	112, 150, 162	TRP	WK	112, 150, 162
	RPWK	112, 150, 162		RPWK	112, 150, 162
GLY	WK	2, 6, 15, 26, 33, 34, 37, 44, 48, 57, 94, 105, 107, 124, 136, 141, 145, 154, 158	GLY	WK	2, 6, 15, 26, 33, 34, 37, 44, 48, 57, 94, 105, 107, 124, 136, 141, 145, 154, 158
	RPWK	2, 6, 15, 26, 33, 34, 37, 44, 48, 57, 94, 105, 107, 124, 136, 141, 145, 154, 158		RPWK	2, 6, 15, 26, 33, 34, 37, 44, 48, 57, 94, 105, 107, 124, 136, 141, 145, 154, 158

ASP	WK	23, 38, 49, 64, 68, 133, 135, 153, 163	ASP	WK	23, 38, 49, 64, 68, 133, 135, 153, 163
	RPWK	23, 38, 49, 64, 68, 133, 135, 153, 163		RPWK	23, 38, 49, 64, 68, 133, 135, 153, 163
TYR	WK	50, 51, 53, 115, 116, 119, 137, 143	TYR	WK	50, 51, 53, 115, 116, 119, 137, 143
	RPWK	50, 51, 53, 115, 116, 119, 137, 143		RPWK	50, 51, 53, 115, 116, 119, 137, 143
LEU	WK	22, 75, 76, 83, 85, 101, 127, 129, 149, 156, 164	LEU	WK	22, 75, 76, 83, 85, 101, 127, 129, 149, 156, 164
	RPWK	22, 75, 76, 83, 85, 101, 127, 129, 149, 156, 164		RPWK	22, 75, 76, 83, 85, 101, 127, 129, 149, 156, 164
PHE	WK	9, 14, 36, 56, 58, 88, 132, 142, 151	PHE	WK	9, 14, 36, 56, 58, 88, 132, 142, 151
	RPWK	9, 14, 36, 56, 58, 88, 132, 142, 151		RPWK	9, 14, 36, 56, 58, 88, 132, 142, 151
ARG	WK	18, 43, 97, 144, 160	ARG	WK	18, 43, 97, 144, 160
	RPWK	18, 43, 97, 144, 160		RPWK	18, 43, 97, 144, 160
CYS	WK	30, 42, 109, 131	CYS	WK	30, 42, 109, 131
	RPWK	30, 42, 109, 131		RPWK	30, 42, 109, 131
ILE	WK	12, 24, 40	ILE	WK	12, 24, 40
	RPWK	12, 24, 40		RPWK	12, 24, 40
HIS	WK	41, 45, 148	HIS	WK	41, 45, 148
	RPWK	41, 45, 148		RPWK	41, 45, 148
ASN	WK	29, 89	ASN	WK	29, 89
	RPWK	29, 89		RPWK	29, 89

From this table we can notice that all 165-BLIP residues are detected inside the relaxation region (<0.6THz) by WK. And all of them are also detected by RPWK in the same region (<0.6THz). As a result, WK and RPWK perform equally in detecting frequency data points for BLIP residues in the relaxation region at 150K and 310K.

Table 2: The frequency data points which are detected by WK and RPWK at 150K and 310K in HBB band.

Residue's name	Method	At 150K	Residue's name	Method	At 310K
ALA	WK	(1), 7, 25, 27, 46, 47, 52, (54), 61, (62), (63), 65, (77), 80, 86, 98, (102), 118, 123, 147	ALA	WK	1, 25, 27, (47), 52, 54, 63, (65), 77, 80, 86, 98, (102), 123, (147)
	RPWK	7, 25, 27, 46, 47, 52, 61, 65, 80, 86, 98, 118, 123, 147		RPWK	1, 25, 27, 52, 54, 63, 77, 80, 86, 98, 123
VAL	WK	(3), 21, 67, 91, 93, 100, 104, 125, 134, 155, 165	VAL	WK	3, 21, 67, 91, 100, 104, (134), 155, 165
	RPWK	21, 67, 91, 93, 100, 104, 125, 134, 155, 165		RPWK	3, 21, 67, 91, 100, 104, (125), 155, 165
SER	WK	35, (39), 60, 69, 71, 79, 108, (113), 121, 128, 130, (138), (146)	SER	WK	35, (60), (69), 79, 108, 113, 130, 146
	RPWK	35, 60, 69, 71, 79, 108, 121, 128, 130		RPWK	35, 79, 108, 113, (128), 130, 146
THR	WK	5, 10, 17, 32, 59, 82, (84), 92, (103),(110), 111, 122, (126), 140, 152	THR	WK	(5), 10, 17, 32, 82, 96, 110, 111, 122, 126, (140)
	RPWK	5, 10, 17, 32, 59, 84, 92, 111, 122, 140, 152		RPWK	10, 17, 32, 82, 96, 110, 111, 122, 126

GLN	WK	11, 13, 19, 20, 72, 99, 106, 157, 161	GLN	WK	(11), 13, 19, 20, 72, (157)
	RPWK	11, 13, 19, 20, 72, 99, 106, 157, 161		RPWK	13, 19, 20, 72
LYS	WK	(8), 66, 70, 87	LYS	WK	8, 66, (70), 74, (87), 159
	RPWK	66, 70, 87		RPWK	8, 66, 74, 159
GLU	WK	28, 73, 114	GLU	WK	73, 114
	RPWK	28, (31), 73, 114		RPWK	73, 114
PRO	WK	78, (81), 117, 120	PRO	WK	81, 117
	RPWK	78, 117, 120		RPWK	(78), 81, 117
MET	WK	4, (16), (95)	MET	WK	4, 16, 95
	RPWK	4		RPWK	4, 16, 95
TRP	WK	112, 150, (162)	TRP	WK	162
	RPWK	112, 150		RPWK	(112), 162
GLY	WK	2, 6, 15, 26, 33, (34), (37), 44, 48, 57, 94, 105, 107, 124, 141, 145, 154, 158	GLY	WK	2, 6, 15, (26), (33), 34, 37, 48, 94, (105), 107, (136),(145), 154, 158
	RPWK	2, 6, 15, 26, 33, 44, 48, 57, 94, 105, 107, 124, 141, 145, 154, 158		RPWK	2, 6, 15, 34, 37, 48, (57), 94, 107, (124), 154, 158
ASP	WK	23, 38, 49, 64, 68, (133), (135), 153, 163	ASP	WK	23, 38, 49, (64), 68, 133, 135, 153, 163
	RPWK	23, 38, 49, 64, 68, 153, 163		RPWK	23, 38, 49, 68, 133, 135, 153, 163
TYR	WK	50, (51), 53, 115, 116, 119, 137, 143	TYR	WK	50, 51, 53, (115), 119

	RPWK	50, 53, 115, 116, 119, 137, 143		RPWK	50, 51, 53, 119, (143)
LEU	WK	22, 75, 76, 83, 85, 127, 129, 149, 156, 164	LEU	WK	22, 75, 83, (101), 129, 149, 156, 164
	RPWK	22, 75, 76, 83, 85, (101), 127, 129, 149, 156, 164		RPWK	22, 75, 83, (127), 129, 149, 156, 164
PHE	WK	9, 14, 36, 56, 58, 88, 132, 142, 151	PHE	WK	9, 14, 56, 58, (88), 132, 151
	RPWK	9, 14, 36, 56, 58, 88, 132, 142, 151		RPWK	9, 14, 56, 58, 132, 151
ARG	WK	18, (97), 144, 160	ARG	WK	18, 43, 97, 144, 160
	RPWK	18, 144, 160		RPWK	18, 43, 97, 144, 160
CYS	WK	30, 42, 109, 131	CYS	WK	42, 131
	RPWK	30, 42, 109, 131		RPWK	(30), 42, 131
ILE	WK	12, 24, 40	ILE	WK	(12), 24, 40
	RPWK	12, 24, 40		RPWK	24, 40
HIS	WK	41, 45, 148	HIS	WK	(41), 148
	RPWK	41, 45, 148		RPWK	148
ASN	WK	29, 89	ASN	WK	29, 89
	RPWK	29, 89		RPWK	29, 89

At 150K, WK is able to detect more data points for ALA residues than RPWK. In WK, all ALA residues are present in the HBB band. In RPWK, six ALA residues (1ALA, 54ALA, 62ALA, 63ALA, 77ALA and 102ALA) are not present in the HBB band. For GLY residues, WK fails to detect only one of them (136GLY) in this band and RPWK fails to detect three of them (34GLY, 37GLY and 136GLY). Moreover, WK detects all VAL residues in HBB band, but RPWK fails to detect one of them (3VAL).

WK detects all ASP residues in HBB band, while RPWK fails to detect two of them (133ASP and 135ASP). Also, in HBB band, WK fails to detect one SER residue (139SER), while

RPWK fails to detect five SER residues (39SER, 113SER, 138SER, 139SER and 146SER) in the same band. Moreover, WK detects all TYR residues but RPWK fails to detect one of them (51TYR) in this band.

WK fails to detect two THR residues (55THR and 96THR) but RPWK fails to detect six of them (55THR, 82THR, 96THR, 103THR, 110THR and 126THR) in HBB band. And RPWK detects all LEU residues in this band, while WK fails to detect one of them (101LEU) in the same band. Moreover, both methods fail to detect 90GLN residue in HBB band.

WK detects four LYS residues (8LYS, 66LYS, 70LYS and 87LYS) in HBB band. Three of them (66LYS, 70LYS and 87LYS) are detected in the same band by RPWK. Also, WK detects four ARG residues (18ARG, 97ARG, 144ARG and 160ARG) in HBB band. Three of them (18ARG, 144ARG and 160ARG) are detected in the same band by RPWK. Moreover, RPWK detects all GLU residues in HBB band, but WK fails to detect one of them (31GLU).

WK and RPWK detect frequency data points inside HBB band for all CYS, PHE ILE, HIS and ASN residues. And WK detects all PRO residues in HBB band, while RPWK fails to detect one of them (81PRO). Also WK detects all MET residues in HBB band, but RPWK detects only one of them (4MET). Moreover, WK detects all TRP residues in HBB band, but RPWK fails to detect one of them (162TRP).

In general, at 150K, WK detects more data points than RPWK for (12) residues. Both methods perform equally in detecting data points for (6) residues. While there are only two residues (101LEU and 31GLU) that are detected by RPWK but WK fails to detect them.

At 310K, in the HBB band, WK fails to detect five of the ALA residues (7ALA, 46ALA, 61ALA, 62ALA and 118ALA), while RPWK fails to detect these residues in addition to another four ALA residues (47ALA, 65ALA, 102ALA and 147ALA). Also, WK fails to detect four GLY residues (44GLY, 57GLY, 124GLY and 141GLY) in HBB band, while RPWK fails to detect seven GLY residues (26GLY, 33GLY, 44GLY, 105GLY, 136GLY, 141GLY and 145GLY) in the same band. Moreover, both methods fails to detect 93VAL residue. Also WK fails to detect 125VAL residue, and RPWK fails to detect 134VAL residue. So both methods are able to detect nine VAL residues in HBB band.

WK detects all ASP residues in HBB band, while RPWK fails to detect the 64ASP residue in the same band. Also, RPWK detects seven SER residues (35SER, 79SER, 108SER, 113SER, 128SER. 130SER and 146SER) in HBB band. These SER residues (except 128SER) are detected by WK, in addition to 60SER residue and 69SER residue, in the same band. Moreover, WK detects five TYR residues (50TYR, 51TYR, 53TYR, 115TYR and 119TYR) in HBB band. These residues (except 115TYR) are detected, in addition to 143TYR residue, in the same band by RPWK.

WK detects eleven THR residues (5THR, 10THR, 17THR, 32THR, 82THR, 96THR, 110THR, 111THR, 122THR, 126THR and 140THR) in HBB band, while RPWK fails to detect

two of them (5THR and 140THR) in the same band. Also, WK detects eight LEU residues (22LEU, 75LEU, 83LEU, 101LEU, 129LEU, 149LEU, 156LEU and 164LEU) in HBB band. These residues (except 101LEU) are detected, in addition to 127LEU residue, in the same band by RPWK. Moreover, RPWK detects four GLN residues (13GLN, 19GLN, 20GLN and 72GLN) in HBB band. These residues in addition to 11GLN residue and 157GLN residue are detected in the same band by WK method.

WK fails to detect two PHE residues (36PHE and 142PHE) in HBB band. RPWK also fails to detect these two residues in the same band, in addition to 88PHE residue. Also, WK detects all LYS residues in HBB band, while RPWK fails to detect two of them (70LYS and 87LYS). Moreover, both methods also detect 73GLU residue and 114GLU residue in HBB band.

WK detects 42CYS residue and 131CYS residue in HBB band. These two residues and 30CYS residue are also detected by RPWK in the same band. Also, WK detects 81PRO residue and 117PRO residue in HBB band. These two residues and 78PRO residue are also detected by RPWK in the same band. Moreover, WK detects all ILE residues but RPWK fails to detect 12ILE residue.

Both methods detect frequency data points inside HBB band for all MET, ASN and ARG residues. Also, WK detects 41HIS residue and 148HIS residue in HBB band, while RPWK detects only 148HIS residue in the same band. Moreover, RPWK detects 112TRP residue and 162TRP residue in HBB band, while WK detects only 162TRP residue in the same band.

In general, at 310K, both methods are close in their performance in detecting data points for (7) residues. WK detects more data points than RPWK for (10) residues, while RPWK detects more data points than WK for only three residues (CYS, PRO and TRP).

Table 3: The frequency data points which are detected by WK and RPWK at 150K and 310K in HBS band.

Residue's name	Method	At 150K	Residue's name	Method	At 310K
ALA	WK	25, (65), 147	ALA	WK	52
	RPWK	25, (61), (118), 147		RPWK	52
VAL	WK	(67), 125, (134)	VAL	WK	3, (21), (155)
	RPWK	(91), (93), 125		RPWK	3

SER	WK	(60), 79, (108), (113), (121), 130, (138)	SER	WK	(113), 130
	RPWK	79, 130		RPWK	130
THR	WK	32, 111, (122), (126), (152)	THR	(111)	(111)
	RPWK	32, 111, (140)		(32)	(32)
GLN	WK	(13), 72, 99, 157	GLN	WK	(11)
	RPWK	72, 99, 157, (161)		RPWK	(13)
LYS	WK	70	LYS	WK	8, (66), (74)
	RPWK	70		RPWK	8
GLU	WK	-	GLU	WK	-
	RPWK	-		RPWK	-
PRO	WK	(120)	PRO	WK	(81)
	RPWK	-		RPWK	-
MET	WK	-	MET	WK	(95)
	RPWK	-		RPWK	-
TRP	WK	-	TRP	WK	-
	RPWK	-		RPWK	-
GLY	WK	(6), (37), (94), 107, 154	GLY	WK	(2), (6)
	RPWK	(26), (48), 107, 154		RPWK	-
ASP	WK	(135)	ASP	WK	-
	RPWK	-		RPWK	(133), (153)
TYR	WK	(53), (115), 116	TYR	WK	-
	RPWK	116		RPWK	-

LEU	WK	(75), (129), (156)	LEU	WK	(149)
	RPWK	-		RPWK	-
PHE	WK	(56)	PHE	WK	(88), (132)
	RPWK	(88)		RPWK	-
ARG	WK	-	ARG	WK	(97)
	RPWK	-		RPWK	-
CYS	WK	(42)	CYS	WK	-
	RPWK	(109)		RPWK	-
ILE	WK	-	ILE	WK	-
	RPWK	(24)		RPWK	-
HIS	WK	(45)	HIS	WK	-
	RPWK	-		RPWK	-
ASN	WK	(29), (89)	ASN	WK	(89)
	RPWK	-		RPWK	-

At 150K, we can notice that WK is able to detect three ALA residues (25ALA, 65ALA and 147ALA) in HBS band, while RPWK is able to detect four ALA residues (25ALA, 61ALA, 118ALA and 147ALA) in the same band. Also, WK is able to detect five GLY residues (6GLY, 37GLY, 94GLY, 107GLY and 154GLY) in HBS band but RPWK detects four GLY residues (26GLY, 48GLY, 107GLY and 154GLY) in the same band. Moreover, both methods detect three VAL residues in HBS band. WK detects 67VAL, 125VAL and 134VAL. RPWK detects 91VAL, 93VAL and 125VAL.

For ASP residues, WK detects only one residue (135ASP) but RPWK fails to detect any data points. Also, WK detects seven SER residues (60SER, 79SER, 108SER, 113SER, 121SER, 130SER and 13SER) in HBS band, but RPWK detects only two SER residues (79SER and 130SER) in it. Moreover, WK is able to detect three TYR residues (53TYR, 115TY and 116TYR) in HBS band, while RPWK detects one of them (116TYR).

For THR residues, WK detects five of them (32THR, 111THR, 122THR, 126THR and 152THR) in HBS band but RPWK detects three of them (32THR, 111THR and 140THR) in the

same band. Moreover, WK detects three LEU residues (75LEU, 129LEU and 156LEU) in HBS band, but RPWK fails to detect any residue. Also, WK detects four GLN residues (13GLN, 72GLN, 99GLN and 157GLN) in HBS band. Three of them (72GLN, 99GLN and 157GLN), in addition to 161GLN residue, are detected by RPWK in HBS band.

WK detects 56PHE residue in HBS band, but RPWK detects 88PHE residue in the same band. And both methods detect only 70LYS residue in HBS band. While they fail to detect any residue for ARG, GLU, MET and TRP. Also, WK detects 42CYS residue in HBS band and RPWK detects 109CYS residue in the same band.

For PRO residues, only one PRO residue (120PRO) is detected in HBS band by WK. And only one ILE residue (24ILE) is detected in the same band by RPWK. Also only one HIS residue (45HIS) is detected in the same band by WK. Moreover, WK detects all ASN residues in HBS band, but RPWK fails to detect any of them in the same band.

In general, at 150K, WK detects more data points than RPWK for (9) residues. Both methods are similar in their performance in detecting data points for (5) residues. While RPWK detects more data points than WK for only two residues (ALA and ILE). Moreover, there are four residues (ARG, GLU, MET and TRP) that are not detected by either WK or RPWK.

From this table, at 310K, we can see that, in HBS band, WK and RPWK detects only one ALA residue (52ALA) and only WK is able to detect two GLY residues (2GLY and 6GLY). Also, RPWK detects only one VAL residue (3VAL) in HBS band. This residue, in addition to 21VAL and 155VAL, are also detected in HBS band by WK method. Moreover, RPWK detects 133ASP residue and 153ASP residue in HBS band, but WK fails to detect any residue in the same band.

WK detects two SER residues (113SER and 130SER) in HBS band, while RPWK detects one of them (130SER). For THR, WK detects 111THR residue and RPWK detects 32THR in the same band. Also, only one LEU residue (149LEU) is detected in HBS band by WK method. Moreover, WK detects 11GLN residue in this band but RPWK detects 13GLN residue in the same band.

Only WK is able to detect two PHE residues (88PHE and 132PHE) in HBS band. And WK detects three LYS residues (8LYS, 66LYS and 74LYS), but RPWK detects only one of them (8LYS). Also, the residues 97ARG, 81PRO, 95MET and 89ASN are only detected by WK method. Finally, both methods fail to detect any residue in HBS band for TYR, GLU, CYS, ILE, HIS and TRP.

In general, at 310K, WK detects more data points than RPWK for (3) residues. Also WK detects data points for (7) residues that are not detected by RPWK at all. Both methods perform almost equally in detecting data points for (3) residues. While there is only one residue (ASP) that is detected by RPWK but WK fails to detect it. Moreover, there are six residues (TYR, GLU, CYS, ILE, HIS and TRP) that are not detected by WK neither by RPWK.

Table 4: The frequency data points which are detected by WK and RPWK at 150K and 310K in L_1 band.

Residue's name	Method	At 150K	Residue's name	Method	At 310K
ALA	WK	-	ALA	WK	-
	RPWK	(118)		RPWK	-
SER	WK	-	SER	WK	-
	RPWK	-		RPWK	(130)
GLN	WK	-	GLN	WK	-
	RPWK	(99)		RPWK	-
ASP	WK	-	ASP	WK	-
	RPWK	-		RPWK	(153)
TYR	WK	-	TYR	WK	-
	RPWK	(116)		RPWK	-
LEU	WK	-	LEU	WK	-
	RPWK	(75)		RPWK	-

This table shows that at 150K only RPWK detects frequency data points in L_1 band. These data points for 118ALA, 116TYR, 75LEU and 99GLN. WK fails to detect any data points in this band.

Also, from this table we can notice that at 310K only two residues (153ASP and 130SER) are detected by RPWK in L_1 band; WK fails to detect them.

Table 5: The frequency data points which are detected by WK and RPWK at 150K in L_2 band.

Residue's name	Method	L_2 band >15THz	Residue's name	Method	L_2 band >15THz
ALA	WK		VAL	WK	125
	RPWK	(61), (118)		RPWK	125
TYR	WK		THR	WK	(152)
	RPWK	(116)		RPWK	
GLN	WK	99			
	RPWK	99			

At 150 K, this table shows that RPWK detects data points for 61ALA and 118ALA in L_2 band. While WK fails to detect any data points for ALA residue in the same band. However, both methods detects 125VAL residue and 99GLN residue in L_2 band. For TYR, only one residue (116TYR) is detected in this band by RPWK. On the other hand, WK detects one residue for THR residue (152THR) in L_2 band.

Finally, the two methods (WK and RPWK) fail to detect any data points in L_2 band of power spectra for the non-bonded interaction energy between the residues in BLIP protein molecule and water molecules at 310K.

The results above show that RPWK can play a complementary role to WK in detecting significant frequency spectra, especially in the higher frequency bands.

3.2 The effect of the RSA value on detected spectra at 150K and 310K

The THz frequency for the non-bonded interaction energy between the '165' residues in BLIP protein and water molecules within 10Å from each residue at 150K and 310K, over a time period of 20ps using WK and RPWK are graphed versus the Relative Solvent Accessibility (RSA) for each residue to distinguish between their properties. The power values are shown in each graph for z-score larger than 1 on the z-axis. The z-score values are shown by using the color bar for each graph.

Most of detected data points by WK at 150K are in the frequency range up to 4THz (Fig. 8) with no clear pattern regarding the effect of RSA. Some of them are in the region between 4THz and 5THz. These residues are 72GLN residue at (0.0225, 4.19THz /4.64THz /4.79THz /4.99THz),

70LYS residue at (0.0377, 4.19THz /4.64THz /4.79THz /4.99THz), 29ASN residue at (0.147, 4.05THz), 116TYR residue at (0.152, 4.74THz), 138SER residue at (0.194, 4.19THz), 126THR residue at (0.337, 4.84THz), 152THR residue at (0.342, 4.54THz /4.74THz), 154GLY residue at (0.41, 4.34THz /4.54THz /4.79THz), 45HIS residue at (0.44, 4.15THz /4.35THz), 157GLN residue at (0.522, 4.09THz /4.45THz), 32THR residue at (0.597, 4.84THz) and 122THR residue at (0.666, 4.59THz). All of them are polar hydrophilic residues. In addition there are three more residues, which are 156LEU residue at (0, 4.3THz /4.55THz /4.6THz), 125VAL residue at (0.168, 4.39THz), 134VAL residue at (0.719, 4.74THz) and 120PRO residue at (0.724, 4.05THz). These three residues are nonpolar and hydrophobic.

WK detects several data points for different residues in the frequency region (5THz-10THz). These data points are for 156LEU residue at (0, 5.25THz /5.4THz /6.05THz /7.05THz), 152THR residue at (0.342, 5.89THz), 116TYR residue at (0.152, 6.44THz), 70LYS residue at (0.0377, 5.79THz), 72GLN residue at (0.0225, 5.79THz) and 126THR residue at (0.337, 7.89THz), (0.337, 6.69THz) and (0.337, 6.24THz). All of them are polar hydrophilic residues except the first one (156LEU) [73].

Moreover, there are three data points in the high frequency region (10THz-25THz). The first one at (0.342, 15.98THz) is a polar hydrophilic residue (152THR). The second one also is a polar hydrophilic residue (99GLN) at (0.493, 24.2THz). However, the third one (0.168, 16.6THz) is 125VAL residue, which is a nonpolar hydrophobic residue.

The z-score values larger than 10 are detected only in the frequency region up to 2THz. These z-score values indicate that the power of these residues are larger than 10 standard deviations from the mean. Also there are other values between 1 and 10, which are spread out in all the frequency range (up to 25THz).
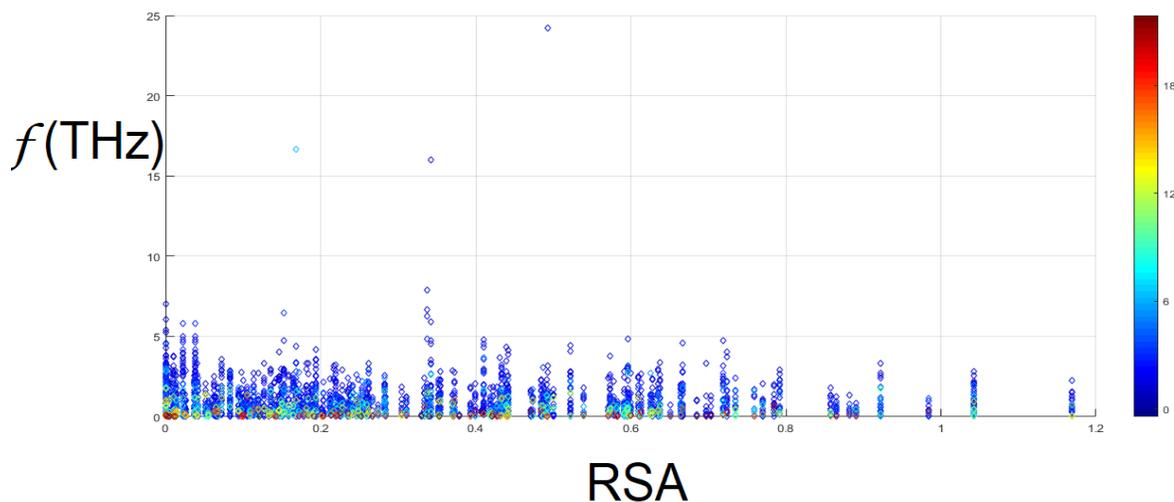
Fig. 8: The frequency for the non-bonded interaction energy between the '165' residues in BLIP protein and water molecules versus RSA by using WK method at 150K.

In Fig. 9, all detected values by WK at 310K are in the frequency range up to 5THz. Only one value is detected at 5.19THz for a polar hydrophilic residue (50TYR). However, WK detects many residues which are also close to 5THz. These residues are 89ASN residue at (0.638, 4.84THz), 11GLN residue at (0.352, 3.99THz) and two data points for 113SER residue at (0.261, 4.79THz) and (0.261, 4.34THz). All these residues are polar and hydrophilic. However, there are also three data points for nonpolar hydrophobic residues, which are 21VAL residue at (0, 4.40THz), 155VAL residue at (0.333, 4.54THz) and 149LEU residue at (0.0142, 4.15THz).

The z-score values larger than 10 are detected only in the frequency region up to 1.5THz. It also contains other values between 1 and 10, which are also spread out in the frequency range up to 5.2THz.
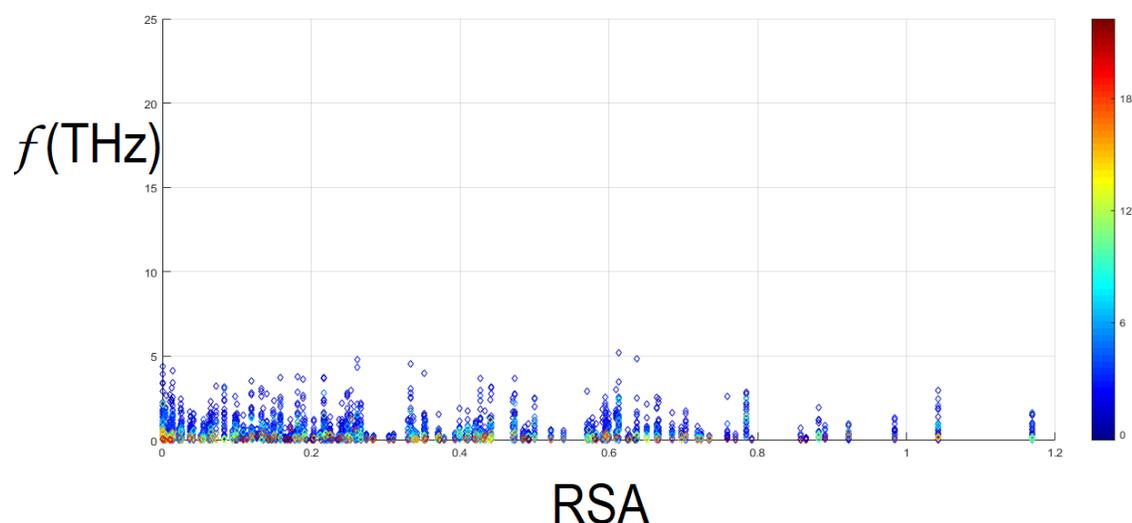
Fig. 9: The frequency for the non-bonded interaction energy between the '165' residues in BLIP protein and water molecules versus RSA by using WK method at 310K.

As Fig. 10 shows, most of detected data points by RPWK at 150K are in the frequency range up to 4THz, but there are also many detected data points in the region between 4THz and 25THz. RPWK detects five residues in the region between 4THz and 5THz. These residues are 91VAL residue at (0.0137, 4.65THz), 88PHE residue at (0.0832, 4.66THz), 154GLY residue at (0.409, 4.54THz), 157GLN at (0.522, 4.09THz /4.38THz /4.45THz), 93VAL residue at (0.582, 4.29THz /4.34THz) and 32THR residue at (0.597, 4.83THz). Two of them (VAL and PHE) are nonpolar hydrophobic residues, while the rest (GLY, GLN and THR) are polar and hydrophilic residues.

RPWK detects many data points for different residues in the high frequency region between 5THz and 25THz. These residues are 75LEU residue at (0.127, 10.66THz), 26GLY residue at

(0.135, 5.46THz), 116TYR residue at (0.152, 12.62THz /16.94THz), 125VAL residue at (0.1683, 16.64THz), 161GLN residue at (0.304, 4.99THz /5.43THz), 118ALA residue at (0.484, 4.16THz/ 4.39THz/ 4.84THz /4.87THz /4.99THz /5.97THz /13.4THz /13.8THz /14.25THz /22.07THz/24.72THz), 61ALA residue at (0.628, 6.37THz/ 7.59THz/ 21.54THz/ 22.19THz) and 99GLN residue at (0.493, 5.02THz /5.60THz /5.65THz), also RPWK detects (14) data points in the range between 10THz to 13THz for 99GLN residue and (12) data points for the same residue in the region between 19THz-25THz. Some of these residues are hydrophobic like ALA, LEU and VAL. While the rest of them (GLY, TYR and GLN) are polar hydrophilic residues.

The z-score values larger than 10 are detected only in the frequency region up to 1THz. It also contains other values between 1 and 10, which are also spread out in the frequency range up to 25THz.
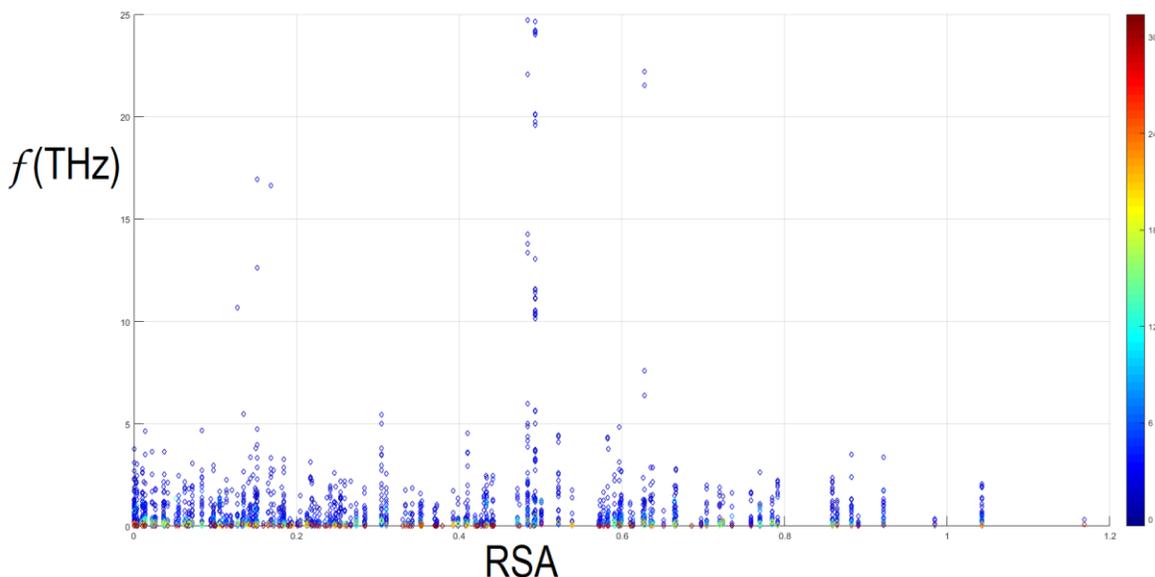
Fig. 10: The frequency for the non-bonded interaction energy between the '165' residues in BLIP protein and water molecules versus RSA by using RPWK method at 150K.

In Fig. 11, most of detected residues by RPWK at 310K are in the frequency range up to 4THz, but there are also many detected data points in the region between 4THz and 15THz. RPWK detects five data points for different residues in the region between 4THz and 5THz. These residues are 130SER residue at (0.0720, 4.35THz), 8LYS residue at (0.119, 4.67THz), 13GLN residue at (0.433, 4.17THz), 32THR residue at (0.597, 4.30THz) and 52ALA residue at (0, 4.48THz). All of them are polar hydrophilic residues except the last one (52ALA), which is nonpolar and hydrophobic.

Moreover, RPWK detects many data points in the region 5THz-15THz. These data points for one nonpolar hydrophobic residue which is 52ALA residue at (0, 5.73THz /8.39THz /8.62THz). And two polar hydrophilic residues, which are 130SER residue at (0.0720, 8.14THz /8.75 THz /8.85 THz /11.6 THz /13.45 THz) and 153ASP residue at (0.610, 5.58 THz /5.76 THz /14.1THz).

The z-score values larger than 10 are detected only in the frequency region up to 1THz. It also contains other values between 1 and 10, which are also spread out in the frequency range up to 15THz.
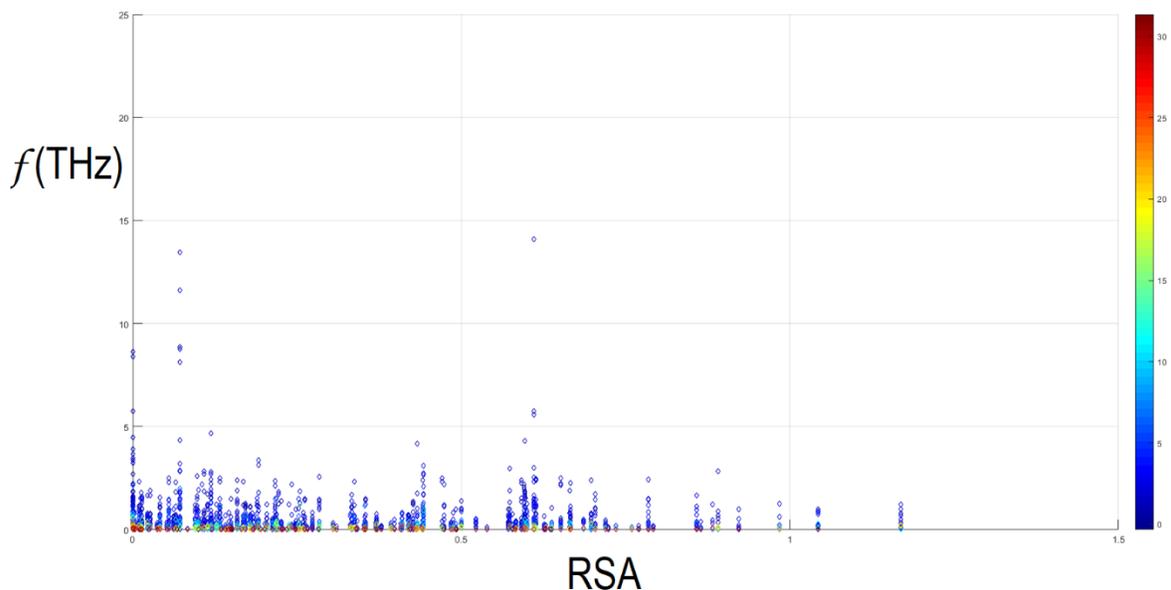
Fig. 11: The frequency for the non-bonded interaction energy between the '165' residues in BLIP protein and water molecules versus RSA by using RPWK method at 310K.

4. Conclusion

We have used Wiener-Khinchin (WK) and plot based Wiener-Khinchin (RPWK) methods to analyze the protein-water interactions obtained from molecular dynamics simulations. In particular, the two methods are used to determine the THz frequency spectra for the non-bonded interaction energy series between BLIP protein residues and surrounding water molecules at 150K and 310K.

We found that, at frequencies less than 4THz, WK is able to detect more residues than RPWK and sometimes they acts equally in the number of detected residues. However, we found that RPWK detects 16 different residues at 310K with frequencies more than 4THz, while WK detects only 7 such residues. At 150K, we found that RPWK detects 60 residues with frequencies larger than 4THz, while WK detects only 43 residues. This suggests that the number of detected residues with high frequency range by RPWK is larger than obtained by WK at both temperatures.

This research suggests that both methods are useful in finding the frequency spectra for the non-bonded interactions time series in water-protein system. However, RPWK is a more robust analysis method to use in investigating molecular dynamics trajectories. This is because it depends on recurrence plots that are useful for displaying the trajectories of phase space into two dimensional maps through specifying only two parameters for the time series, which are the time delay and the embedding dimension. Conversely, WK method is made of two functions (autocorrelation and Fourier transform) with a lot of parameters that make WK usage more complicated than RPWK.

5. References

1. Karplus, M. "Molecular Dynamics Simulations of Biomolecules" *Accounts of Chemical Research* 35, no. 6 (2002): 321-23.

2. Hansson, T., C. Oostenbrink, and W. V. Gunsteren. "Molecular dynamics simulations" *Current Opinion in Structural Biology* 12, no. 2 (2002): 190-96.

3. Gunsteren, V., W. F. Berendsen, and J. C. Herman "Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry" *Angewandte Chemie International Edition*. December 22, 2003

4. Levitt, M., and A. Warshel. "Computer simulation of protein folding." *Nature* 253, no. 5494 (1975): 694-98.

5. Gunsteren, W. F. V., and M. Karplus. "Protein dynamics in solution and in a crystalline environment: a molecular dynamics study." *Biochemistry* 21, no. 10 (1982): 2259-274.

6. Brunne, R. M., K. D. Berndt, P. Güntert, K. Wüthrich, and W. F. V. Gunsteren. "Structure and internal dynamics of the bovine pancreatic trypsin inhibitor in aqueous solution from long-time molecular dynamics simulations." *Proteins: Structure, Function, and Genetics* 23, no. 1 (1995): 49-62.

7. Karain, W. THz frequency spectrum of protein-solvent interaction energy using a recurrence plot-based Wiener-Khinchin method." *Proteins: Structure, Function, and Bioinformatics* 84, no. 10 (2016): 1549-557.

8. Alder, B. J., and T. E. Wainwright. "Studies in Molecular Dynamics. I. General Method." *The Journal of Chemical Physics* 31, no. 2 (1959): 459-66.

9. Allen, M. P. "Introduction to Molecular Dynamics Simulation." *Computational Soft Matter: From Synthetic Polymers to Proteins*, NIC Series, 23 (2004): 1-28.

10. Prabhu, N., and K. Sharp. "Protein−Solvent Interactions." *Chem. Rev.* 106, no. 5 (2006): 1616-623.

11. Spassov, V. Z., A. D. Karshikoff, and R. Ladenstein. "The optimization of protein-solvent interactions: Thermostability and the role of hydrophobic and electrostatic interactions." *Protein Science* 4, no. 8 (1995): 1516-527.

12. Myers, R. L. *The Basics of Chemistry*. Westport, CT: Greenwood Press, 2003.

13. Griebel, M., S. Knapek, and G. Zumbusch. *Numerical Simulation in Molecular Dynamics: Numerics, Algorithms, Parallelization, Applications*. Berlin: Springer, 2010.

14. Mortimer, R. G. *Physical Chemistry*. Amsterdam: Elsevier Academic Press, 2008.

15. Petsko, G. A. and D. Ringe. *Protein structure and function*. London: New Science Press, 2009.

16. Kumar, N. *Comprehensive physics for class XII*. New Delhi: Laxmi Publications, 2004.

17. Fry, M. and E. Page. *Catch up chemistry: for the life and medical sciences*. Banbury: Scion, 2012.

18. Jeffrey, G. A., and W. Saenger. *Hydrogen bonding in biological structures*. Berlin: Springer, 1994.

19. Brockwell, P. J. and R. A. Davis. *Time Series: Theory and Methods*. New York: Springer-Verlag, 1991.

20. Dufour, J. M. *Introduction to Time Series Analysis*, Canada, Montreal, QC:Univ. of Montreal Press, 1998.

21. Lu, W., and N. Vaswani. "The Wiener-Khinchin Theorem for Non-wide Sense stationary Random Processes." April 03, 2009. https://arxiv.org/abs/0904.0602v1.

22. Braiman, M. S., O. Bousche, and K. J. Rothschild. "Protein dynamics in the bacteriorhodopsin photocycle: submillisecond Fourier transform infrared spectra of the L, M, and N photointermediates." *Proceedings of the National Academy of Sciences* 88, no. 6 (1991): 2388-392. doi:10.1073/pnas.88.6.2388.

23. Davier, A. *Statistical models for test equating, scaling, and linking*. New York: Springer, 2011.

24. Taylor, D. "Time-Series Analysis." *Western Journal of Nursing Research* 12, no. 2 (1990): 254-61.

25. Xu, Y., A. B. Seven, L. Su, Q. Jiang, and J. Rizo. "Membrane Bridging and Hemifusion by Denaturated Munc18." *PLoS ONE* 6, no. 7 (2011). doi:10.1371/journal.pone.0022012.

26. Rabiner, L. R., and R. W. Schafer. *Introduction to digital speech processing*. Boston, MA: Now, 2007.

27. Davies, R., T. Coole, and D. Osipyw. "The Application of Time Series Modelling and Monte Carlo Simulation: Forecasting Volatile Inventory Requirements." *Applied Mathematics* 05, no. 08 (2014): 1152-168. doi:10.4236/am.2014.58108.

28. cancelled

29. Zbilut, J. P., and N. Marwan. "The Wiener–Khinchin theorem and recurrence quantification." *Physics Letters A* 372, no. 44 (2008): 6622-626. doi:10.1016/j.physleta.2008.09.027.

30. Gray, R. M., and J. W. Goodman. *Fourier transforms: an introduction for engineers*. New York: Springer , 2013.

31. MATLAB version 7.2.0.232. Natick, Massachusetts: The MathWorks Inc., 2006

32. Masset, P. "Analysis of Financial Time-Series Using Fourier and Wavelet Methods." SSRN Electronic Journal, 2008. doi:10.2139/ssrn.1289420.

33. Matei, F., and N. Alde. "Fourier Series and Fourier Transform with Applications in Nanomaterials Structure." *Fourier Transform - Materials Analysis*, 2012. doi:10.5772/37890.

34. Chaichian, M., I. Merches, and A. Tureanu. *Mechanics An Intensive Course*. Berlin: Springer Berlin, 2014.

35. Müller-Kirsten, H. J. W. *Classical Mechanics and Relativity*. Hackensack, N.J: World Scientific, 2008.

36. Tao, T. "Phase Space." 2008 http://math.ucla.edu/~tao/preprints/phase_space.pdf.

37. Klikov´a, B., and A. Raidl "Reconstruction of phase space of dynamical systems using method of time delay." *WD,* 2011.

38. March, T.k., S.c. Chapman, and R.o. Dendy. "Recurrence Plot Statistics and the Effect of Embedding." *Physica D: Nonlinear Phenomena* 200, no. 1-2 (2005): 171-84. doi:10.1016/j.physd.2004.11.002.

39. Li, W. "Mutual Information Functions versus Correlation Functions." *Journal of Statistical Physics* 60, no. 5-6 (1990): 823-37. doi:10.1007/bf01025996.

40. Jiang, X., and H. Adeli. "Fuzzy Clustering Approach for Accurate Embedding Dimension Identification in Chaotic Time Series." *Integrated Computer-Aided Engineering* 10, no. 3 (2003): 287-302.

41. Krakovská, A., K. Mezeiová, and H. Budáčová. "Use of False Nearest Neighbours for Selecting Variables and Embedding Parameters for State Space Reconstruction." *Journal of Complex Systems* 2015 (2015): 1-12. doi:10.1155/2015/932750.

42. Webber, C. L., C. Ioana, and N. Marwan. *Recurrence Plots and Their Quantifications: Expanding Horizons*. Grenoble: Springer, 2016.

43. Mukherjee, S., S. K. Palit, S. Banerjee, M. Ariffin, and D. Bhattacharya. "Phase Synchronization of Instrumental Music Signals." *The European Physical Journal Special Topics* 223, no. 8 (2014): 1561-577. doi:10.1140/epjst/e2014-02145-7.

44. Marwan, N. "How To Avoid Potential Pitfalls In Recurrence Plot Based Data Analysis." *International Journal of Bifurcation and Chaos* 21, no. 04 (2011): 1003-017. doi:10.1142/s0218127411029008.

45. Bellissent F., M. Claire, A. Hassanali, M. Havenith, R. Henchman, P. Pohl, F. Sterpone, D. V. D. Spoel, Y. Xu, and A. E. Garcia. "Water Determines the Structure and Dynamics of Proteins." *Chemical Reviews* 116, no. 13 (2016): 7673-697. doi:10.1021/acs.chemrev.5b00664.

46. Qin, Y., L. Wang, and D. Zhong. "Dynamics and Mechanism of Ultrafast Water–protein Interactions." *Proceedings of the National Academy of Sciences* 113, no. 30 (2016): 8424-429. doi:10.1073/pnas.1602916113.

47. Levy, Y., and J. N. Onuchic. "Water and Proteins: A Love-hate Relationship." *Proceedings of the National Academy of Sciences* 101, no. 10 (2004): 3325-326. doi:10.1073/pnas.0400157101.

48. Chou, D. H., and C. V. Morr. "Protein-water Interactions and Functional Properties." *Journal of the American Oil Chemists Society* 56, no. 1 (1979). doi:10.1007/bf02671785.

49. Martini, S., C. Bonechi, A. Foletti, and C. Rossi. "Water-Protein Interactions: The Secret of Protein Dynamics." *The Scientific World Journal* 2013 (2013): 1-6. doi:10.1155/2013/138916.

50. Huggins, D. J., and B. Tidor. "Systematic Placement of Structural Water Molecules for Improved Scoring of Protein-ligand Interactions." *Protein Engineering Design and Selection* 24, no. 10 (2011): 777-89. doi:10.1093/protein/gzr036.

51. Chung, E., D. Henriques, D. Renzoni, M. Zvelebil, J. M. Bradshaw, G. Waksman, C.V. Robinson, and J. E. Ladbury. "Mass Spectrometric and Thermodynamic Studies Reveal the Role of Water Molecules in Complexes Formed between SH2 Domains and Tyrosyl Phosphopeptides." *Structure* 6, no. 9 (1998): 1141-151. doi:10.1016/s0969-2126(98)00115-4.

52. Doster, W. "The Dynamical Transition of Proteins, Concepts and Misconceptions." *European Biophysics Journal* 37, no. 5 (2008): 591-602. doi:10.1007/s00249-008-0274-3.

53. Liu, Z, J. Huang, M. Tyagi, H. O'Neill, Q. Zhang, E. Mamonto, N. Jain, Y. Wang, J. Zhang, J. C. Smith, and L. Hong. "Dynamical Transition of Collective Motions in Dry Proteins." *Physical Review Letters* 119, no. 4 (2017). doi:10.1103/physrevlett.119.048101.

54. Kim, C. U., M. W. Tate, and S. M. Gruner. "Protein Dynamical Transition at 110 K." *Proceedings of the National Academy of Sciences* 108, no. 52 (2011): 20897-0901. doi:10.1073/pnas.1110840108.

55. Tarek, M., and D. J. Tobias. "Role of Protein-Water Hydrogen Bond Dynamics in the Protein Dynamical Transition." *Physical Review Letters* 88, no. 13 (2002). doi:10.1103/physrevlett.88.138101.

56. Taschin, A., P. Bartolini, R. Eramo, R. Righini, and R. Torre. "Evidence of Two Distinct Local Structures of Water from Ambient to Supercooled Conditions." *Nature Communications* 4 (2013). doi:10.1038/ncomms3401.

57. Luong, T. Q. *Terahertz and Infrared Spectroscopy of Confined Water*. Master's thesis, Bochum.

58. Heyden, M., J. Sun, S. Funkner, G. Mathias, H. Forbert, M. Havenith, and D. Marx. "Dissecting the THz Spectrum of Liquid Water from First Principles via Correlations in Time and Space." *Proceedings of the National Academy of Sciences* 107, no. 27 (2010): 12068-2073. doi:10.1073/pnas.0914885107.

59. Yagasaki, T., and S. Saito. "A Novel Method for Analyzing Energy Relaxation in Condensed Phases Using Nonequilibrium Molecular Dynamics Simulations: Application to the Energy Relaxation of Intermolecular Motions in Liquid Water." *The Journal of Chemical Physics* 134, no. 18 (2011): 184503. doi:10.1063/1.3587105.

60. Leitner, D. M., M. Havenith, and M. Gruebele. "Biomolecule Large-amplitude Motion and Solvation Dynamics: Modelling and Probes from THz to X-rays." *International Reviews in Physical Chemistry* 25, no. 4 (2006): 553-82. doi:10.1080/01442350600862117.

61. Doran, J. L., B. K. Leskiw, S. Aippersbach, and S. E. Jensen. "Isolation and Characterization of a Beta-lactamase-inhibitory Protein from Streptomyces Clavuligerus and Cloning and Analysis of the Corresponding Gene." *Journal of Bacteriology* 172, no. 9 (1990): 4909-918. doi:10.1128/jb.172.9.4909-4918.1990.

62. Rudgers, G.W., W. Huang, and T. Palzkill. "Binding Properties of a Peptide Derived from Beta -Lactamase Inhibitory Protein." *Antimicrobial Agents and Chemotherapy* 45, no. 12 (2001): 3279-286. doi:10.1128/aac.45.12.3279-3286.2001.

63. Chasman, D. I. *Protein Structure: Determination, Analysis and Applications for Drug Discovery*. New York: Marcel Dekker, 2003.

64. Strynadka, N. C. J., S. E. Jensen, K. Johns, H. Blanchard, M. Page, A. Matagne, J. M. Frère, and M. N. G. James. "Structural and Kinetic Characterization of a β-lactamase-inhibitor Protein." *Nature* 368, no. 6472 (1994): 657-60. doi:10.1038/368657a0.

65. Koonin, E. V., and M. Y. Galperin. *Sequence - Evolution - Function: Computational Approaches in Comparative Genomics*. New York: Springer, 2011.

66. Wu, W., Z. Wang, P. Cong, and T. Li. "Accurate Prediction of Protein Relative Solvent Accessibility Using a Balanced Model." *BioData Mining* 10, no. 1 (2017). doi:10.1186/s13040-016-0121-5.

67. Karain, W. I., and N. I. Qaraeen. "The Adaptive Nature of Protein Residue Networks." *Proteins: Structure, Function, and Bioinformatics* 85, no. 5 (2017): 917-23. doi:10.1002/prot.25261.

68. Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, and K. Schulten. "Scalable Molecular Dynamics with NAMD." *Journal of Computational Chemistry* 26, no. 16 (2005): 1781-802. doi:10.1002/jcc.20289.

69. Humphrey, W., A. Dalke, and K. Schulten. "VMD: Visual Molecular Dynamics." *Journal of Molecular Graphics* 14, no. 1 (1996): 33-38. doi:10.1016/0263-7855(96)00018-5.

70. Gretes, M., D. C. Lim, L. De Castro, S. E. Jensen, S. G. Kang, K. J. Lee, and N. C. Strynadka. "Insights into Positive and Negative Requirements for Protein–Protein Interactions by Crystallographic Analysis of the β-Lactamase Inhibitory Proteins BLIP, BLIP-I, and BLP." *Journal of Molecular Biology* 389, no. 2 (2009): 289-305. doi:10.1016/j.jmb.2009.03.058.

71. Marwan N., Cross Recurrence Plot Toolbox for Matlab, *Reference Manual, Version 5.15, Release 28.6, 2010,* http://tocsy.pik-potsdam.de.

72. Schinkel, S., O. Dimigen, and N. Marwan. "Selection of Recurrence Threshold for Signal Detection." *The European Physical Journal Special Topics* 164, no. 1 (2008): 45-53.

73. Polanski, A, and M. Kimmel. *Bioinformatics*. Berlin: Springer, 2011.

74. Warner, R. M. *Spectral Analysis of Time-series Data*. New York: Guilford Press, 1999.

APPENDIX A

The calculation procedure that was used to get the frequency spectrum for the 12ILE residue at 150K using RPWK method

x=Nonbond12150;

where Nonbond12150 is the time series for non-bonded energy of the 12ILE residue at 150K

The time delay is found by using mutual information MI function [41] as follows:

mi(x, k, m)

where m is a scalar in order to compute the MI up to (m) values, and k is number of bins (default k=10).

The time delay is the time that corresponds to the first minimum, so time delay τ =7 (Fig. 12)

Then the embedding dimension is measured by using the calculated time delay and the false nearest neighbor function FNN as follows:

fnn(x, m, τ)

Where m is a scalar in order to compute the FNN up to (m) values (default m=10).

The embedding dimension is the value that corresponding to FNN=zero, so the embedding dimension (m) =6 in this example (Fig. 13)

After that the Tau-recurrence rate spectrum function is used to find the frequency spectrum as follows:

y=rrspec(x, m , τ, 0.25, 'rr', [ ], 50).

where 0.25 is the recurrence threshold e ('rr': fixed recurrence rate), [ ] maximal lag or delay at which maximal recurrence is observed for tau-recurrence and 50THz is the sampling frequency (fs) that equals the inverse of the time period between each pair of energy series values: 1/20fs = 50THz [74].

Then the spectra up to 25THz frequency range are graphed by using the standard score (z-score) of the power, and they are normalized to the total power for easy comparison as follows:

n=length(y); f = 25*(0:n-1)/n;

where n is the length of y and f is the frequency.

plot(f, zscore(y/sum(y)))

Z-score indicates how many standard deviations a value is from the mean of the values, so it allows us to check the differences or the similarities of two scores or residues at different frequency bands. Its formula is as follows:

Z-score $=(y- \text{mean}(y))/$ standard deviation of (y)
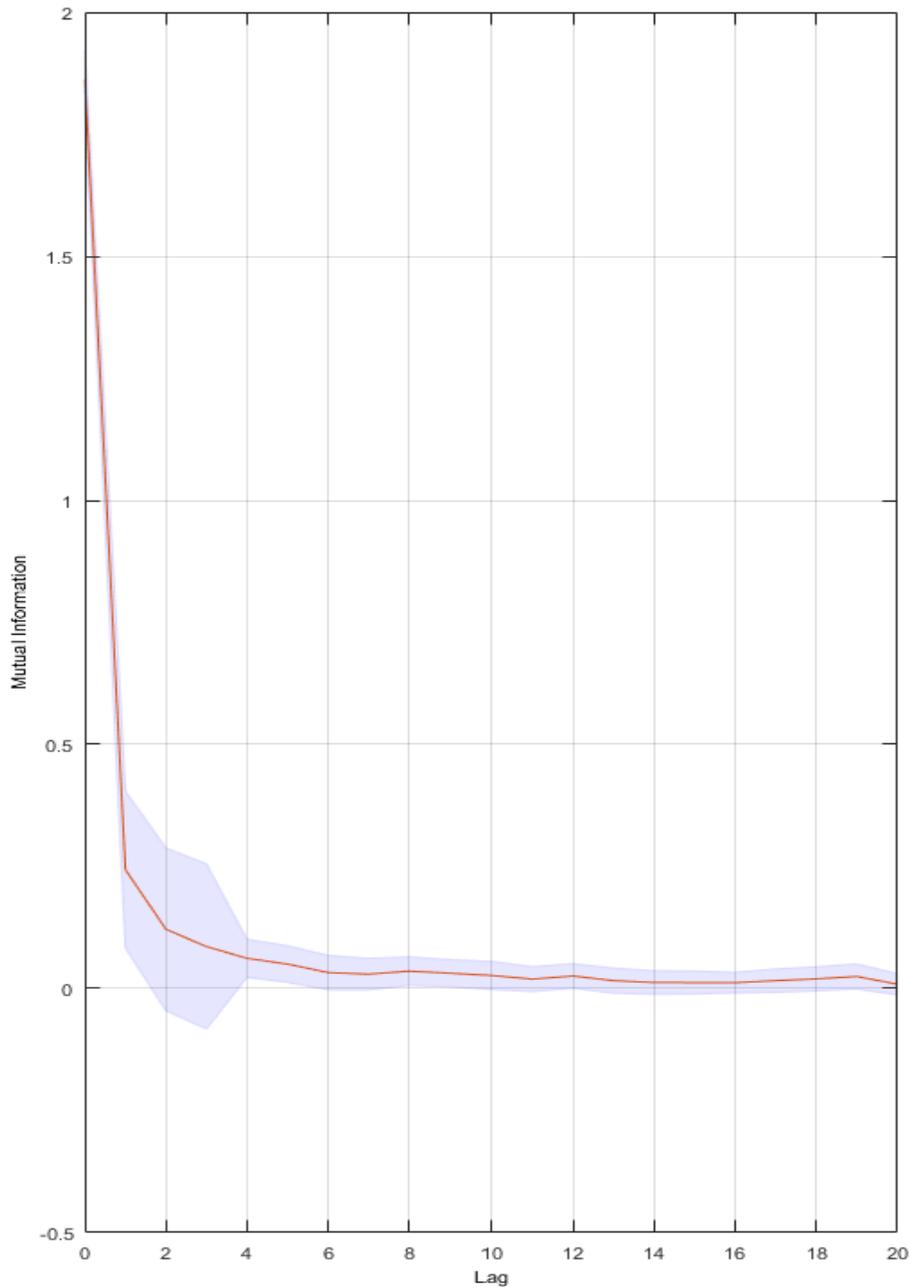
Fig. 12: MI function as it appear in Matlab. The time delay is the Lag that corresponds to the first minimum. Using zoom function, we can see that the time delay $\tau = 7$.
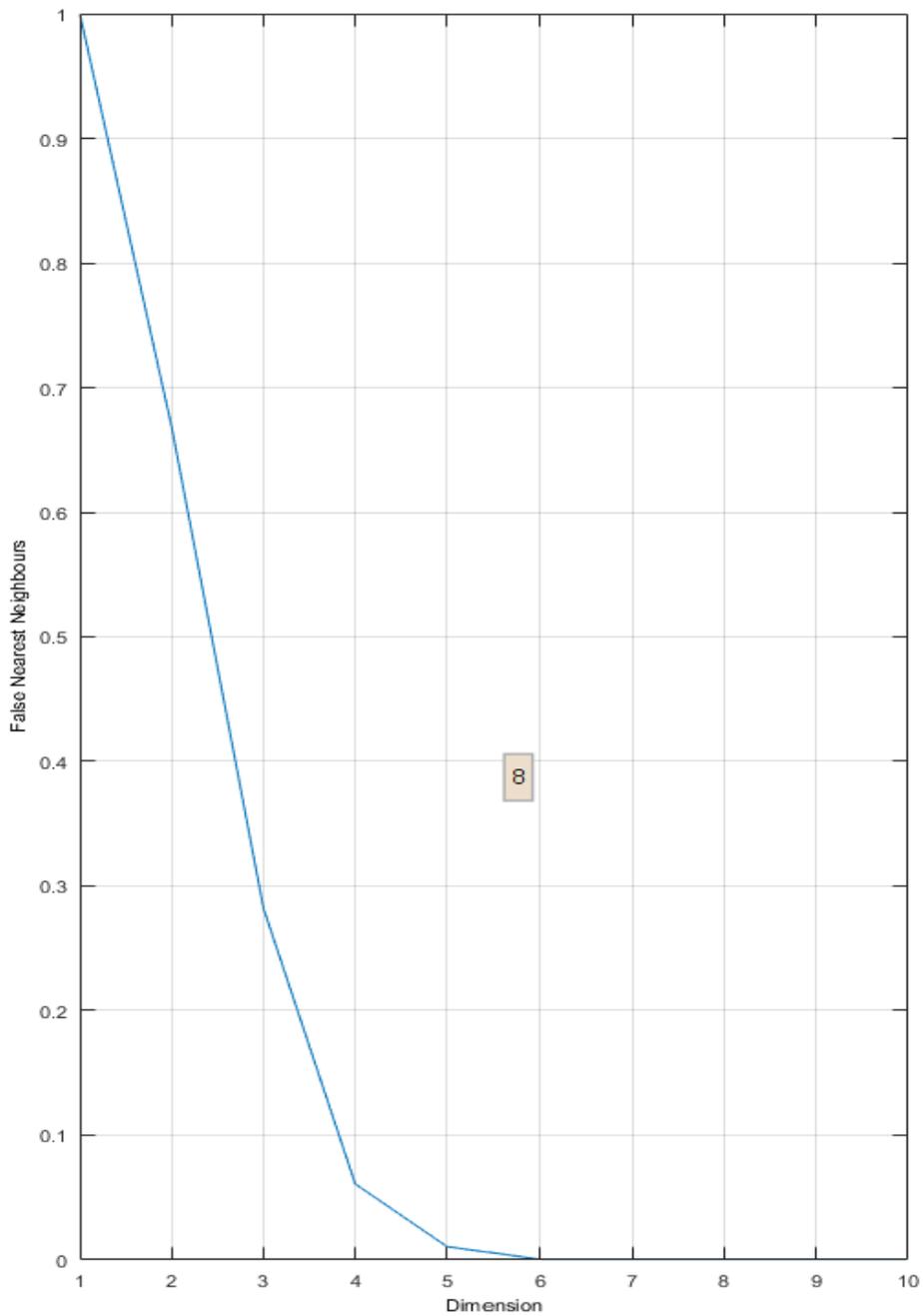
Fig. 13: FNN function as it appear in Matlab. The embedding dimension is the dimension with FNN equal zero. In this example m=6.

Finally, the power spectra of all ILE residues (12ILE, 24ILE and 40ILE) are gathered into one graph for comparison (Fig. 14)
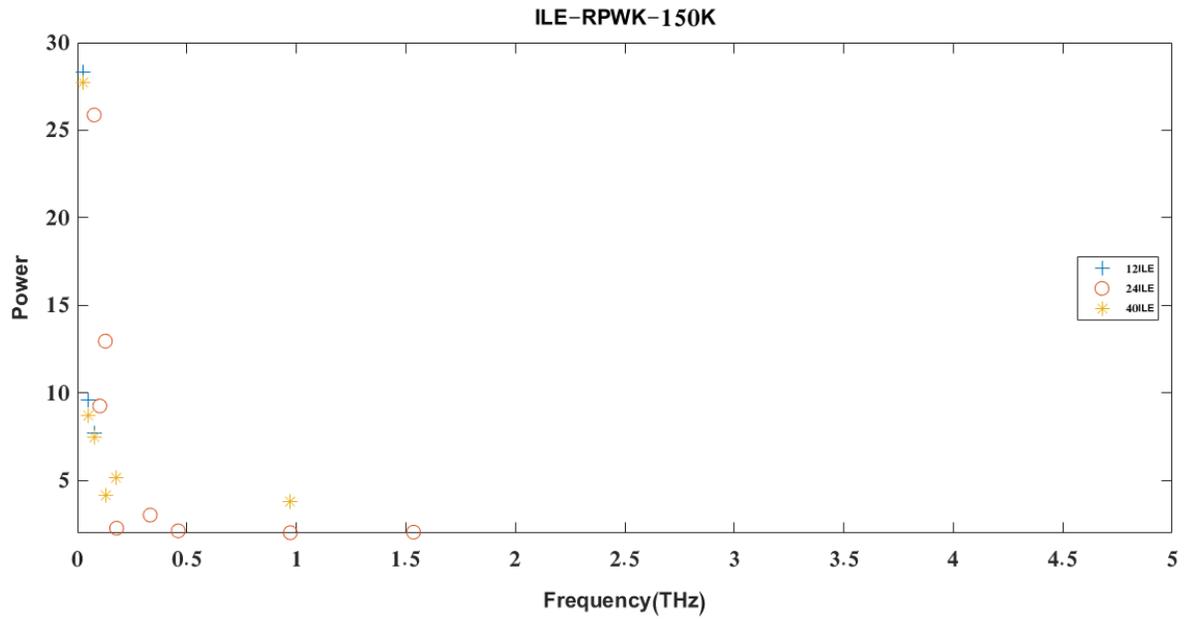
Fig. 14: The power spectra of the three ILE residues (12ILE, 24ILE and 40ILE) in the frequency range up to 5THz by using RPWK at 150K.

APPEDIX B

The calculation procedure that was used to get the frequency spectrum for the 12ILE residue at 150K using WK method

x=Nonbond12150;

where *Nonbond12150* is the non-bonded energy of the 12ILE residue at 150K.

Then autocorrelation function for x is found as follows:

 a= autocorr(x, 'NumLags' ,'NumMA', 'NumSTD')

Where 'NumLags' is the number of lags in y or the length of energy series (1000), 'NumMA' is the Number of lags in a theoretical Matlab model of y (default NumMA=0) and 'NumSTD' is the number of standard errors or standard deviation (default NumSTD=2).

The Fourier transform function of the autocorrelation function (a) is then calculated as follows:

fa = fft(a, n, dim)

Where n is the transform length (default [ ] which means the transform for all x) and dim is a dimension to return the Fourier transform of each column (default dim=1) or to return the Fourier transform of each row (default dim=2).

After that the Fourier transform results and its conjugate are used as following to find the power spectra for the time series:

y=pfa= fa.*conj(fa)/1000;

The power spectra up to 50THz frequency range are graphed by using the z-score of the power as follows:

n = 1001; f = 50*(0:n-1)/n;

Where n is the length of y and f is the frequency.

plot(f, zscore(y/sum(y)))

As RPWK, the power spectrum of the three ILE residues (12ILE, 24ILE and 40ILE) are graphed together in one graph for comparison (Fig. 15).
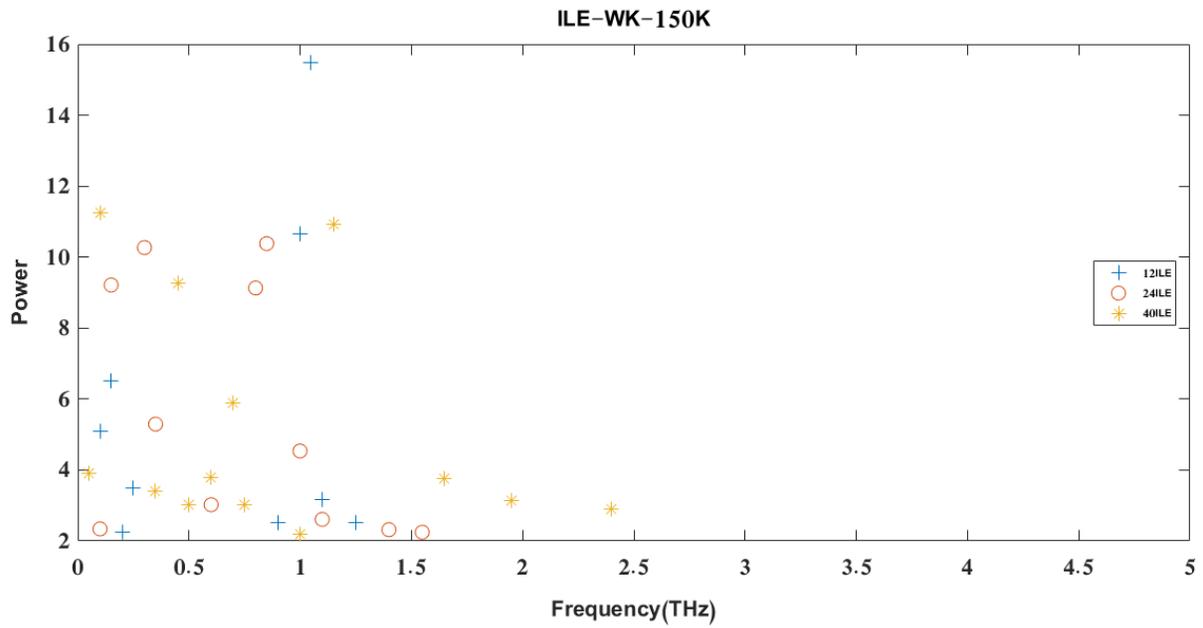
Fig. 15: The power spectra of the three ILE residues (12ILE, 24ILE and 40ILE) in the frequency range up to 5THz by using WK at 150K.