Sci. Rep. Kanazawa Univ. **ARTICLES**
Vol. 67, pp. 41-56, 2024

All-Atom Molecular Dynamics Simulations and Infrared Spectral Analysis of Myosin

Lilih Siti SOLIHAT*, Kazutomo KAWAGUCHI, and Hidemi NAGAO

Graduate School of Natural Science and Technology, Kanazawa University
Kanazawa, 920-1192, Japan

(Received January 6, 2024 and accepted in revised form March 4, 2024)

Abstract Myosin is a crucial myofibrillar protein that plays a pivotal role in the structural composition of muscle. In this study, we conduct 100 ns all-atom molecular dynamics (MD) simulations of myosin proteins from three species: pig, chicken, and cattle, with atom counts before solvation being 6387, 5781, and 5569, respectively. We use the autocorrelation function of the time derivative of the dipole moment, also called the electrical flux-flux correlation function, which is based on particle velocities, to overcome the discontinuity issues posed by periodic boundary conditions. We obtain infrared (IR) spectra profiles from velocity files sampled every 20 picoseconds NVE MD simulation from the trajectories of 100 ns NPT MD simulation. The equilibrated molecular systems in the NVE ensemble allowed for the observation of classical trajectories and minute molecular vibrations, with data captured at an interval of 0.002 picoseconds. Our IR spectra analysis, in the amide I and II regions, show distinctive peaks corresponding to α-helices and β-sheets across all three species. Species-specific structural variations were evident, such as a unique peak at 1574 cm^{-1} in pig myosin, and a peak at 1555 cm^{-1} in cattle myosin, potentially indicating the presence of β-sheets. Comparative analysis with experimental data enabled the correlation of IR spectra peaks with molecular functions and amino acid sequences from Protein Data Bank (PDB) entries. Notably, the absence of a peak associated with Tryptophan vibrations in pig myosin suggests differences in local structures or conformations. The consistency of peaks across species, particularly at 1656 cm^{-1} for α-helix and 1685 cm^{-1} for β-sheets, validated the secondary structures observed in the simulations. Overall, our research underscores the integration of computational and experimental approaches to enrich the understanding of myosin dynamics and structural biology.

Keywords. Infrared spectra, myosin, molecular dynamics, protein

1 Introduction

Myosin is one of the myofibrillar proteins and a major component of muscle, functioning as a motor protein that powers various cellular activities. This motor protein is also known for its roles in muscle contraction, among other functions [1]. The structure is highly asymmetric and hexameric. Various structures of myosin have been investigated. In 2018, the crystal structure of

* *Corresponding author* E-mail: lilih@wriron1.s.kanazawa-u.ac.jp

myosin VI related to *Sus scrofa* (pig) was analyzed by F. Blanck *et al.* [2]. This crystal structure, known as the pre-transition state (PTS), has been solved at 2.2 Å resolution. This protein structure consists of 6387 atoms and is modelled with 763 residues. The structure of myosin related to another animal, specifically smooth muscle myosin from *Gallus gallus* (chicken), has also been reported [3]. This protein, with a molecular weight of 92.4 kDa, comprises 5781 atoms, excluding any hydrogens and deuterium. It is modelled with 708 residues and a resolution of 2.67 Å. On the other hand, the crystal structure of cardiac myosin in the pre-power stroke state from *Bos taurus* (cattle) has been studied to develop a potential treatment for heart failure with reduced ejection fraction [4]. This myosin structure, resolved at 3.10 Å, consists of 5569 atoms and is modelled with 710 residues. The structures of these myosins are shown in Figure 1. The purple ribbons in the structure represent the α-helix of the myosin proteins, which are common elements of secondary structure characterized by their coiled appearance. The yellow sheets depict β-strands that are aligned side by side, which together form β-sheets, another key element of secondary structure, noted for their flat, pleated ribbon appearance. These structures play a critical role in the stability and function of the protein.

In experimental studies, Fourier-transform infrared (FTIR) spectroscopy has been utilized to detect adulteration in meat and meat products [5, 6]. From a theoretical perspective, Molecular Dynamics (MD) simulations have also been employed to explore aspects of food processing and safety [7]. The use of MD simulations in food science is expected to increase to address fundamental problems that current instruments cannot adequately resolve. For instance, in our previous work, we investigated the dynamics of the Hsp90 protein. We calculated the free energy profile for the dissociation of ADP from Hsp90 and suggested that Met98 obstructs the dissociation pathway of ADP at r=1.0 nm [8].

Infrared (IR) absorption spectra arise from the vibrational and rotational energy level transitions within a molecule. Molecular covalent bonds can be likened to stiff springs, capable of stretching and bending. Absorption of specific wavelengths of IR radiation is often associated with these mechanical movements of bonds within the molecule. The IR spectrum of organic molecules is generally complex due to the myriad bond vibrations, which affect the absorption of incident radiation and produce a range of overtones and harmonics.

There have been several studies utilizing MD simulation to calculate IR spectra. Zhu and Robinson, in their 1989 work, reported an MD simulation of liquid CO_2 [9]. They derived an empirical intermolecular potential for liquid CO_2 and incorporated internal degrees of freedom to examine the molecular geometry and vibrational spectrum. Additionally, they calculated the liquid structure parameters and various time-dependent correlation functions. The results of their vibrational spectrum showed a good correlation with experimental data. In 1991, Guillot calculated the far-IR spectrum of liquid water at room temperature using MD simulation across the spectral range of 0.5–1000 cm^{-1} [10]. The simulation results were quantitatively consistent with experimental findings.

More recently, in 2011, a study explored the simulation of cellulose Iβ as it transitioned to a high-temperature (550 K) structure representing the initial stage of cellulose pyrolysis [11]. Utilizing constant pressure MD simulations with the GROMOS 45a4 united atom force field, the researchers evaluated the force field's accuracy by computing the density, thermal expansion coefficient, total dipole moment, and dielectric constant of cellulose Iβ. IR spectra of cellulose Iβ were computed across the 300–550 K range to examine hydrogen bonding. The simulated IR spectra exhibited semi-quantitative consistency with experimental data, particularly in the OH stretching region. The simulations indicated a structural transformation in cellulose Iβ beyond 450 K, which agreed with experimental IR data. The low-temperature structure (300–400 K) of cellulose Iβ is characterized by intrachain hydrogen bonds, whereas at higher temperatures (450–550 K), these bonds transition to longer, weaker interchain hydrogen bonds. At elevated temperatures, a three-dimensional hydrogen-bonding network forms due to the establishment of new interchain hydrogen bonds, which may contribute to the structural stability of cellulose under such conditions.

On the other hand, recent research on calculating IR spectra from computational simulation results has predominantly leveraged quantum effects, encompassing both quantum dynamics and ab initio MD studies. This focus on quantum effects is due to their contribution to more precise IR spectral predictions than those obtained from classical MD simulations. However, quantum dynamics simulations are highly resource-intensive, especially for large molecules composed of thousands of atoms. Considering these costs, classical MD simulations remain a practical alternative for estimating the IR spectra of large molecules, although with lesser accuracy [12].

In this study, we investigate the dynamics and properties of myosin using all-atom MD simulations. Subsequently, we calculate the IR spectra by employing the autocorrelation function of the time derivative of the dipole moment and its spectral density. Our goal is to predict the IR spectra of myosin from pigs, chickens, and cattle. We also assign the IR spectra peaks of three myosins from our computational findings and compare these with experimental data.

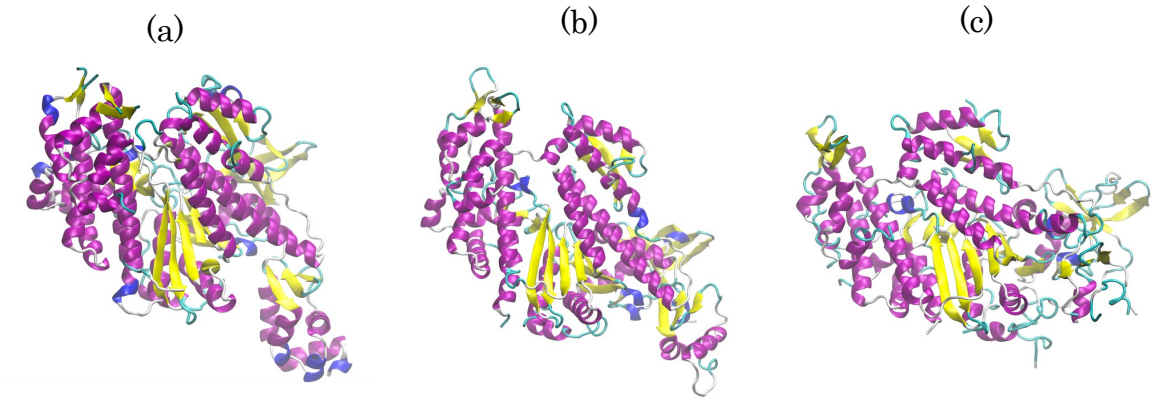

Figure 1. Structure of myosin, the purple ribbons in the structure represent the α-helix, the yellow sheets depict β-strands, that are aligned side by side, which together form β-sheets. a. pig (*Sus scrofa*), b. chicken (*Gallus gallus*), c. cattle (*Bos taurus)*

2 Method

We conduct all-atom molecular dynamics (MD) simulations using the AMBER 20 package designed for the CentOS MPI version [13]. We apply the AMBER ff19SB force field [14] for the protein molecule, and for the water molecules, we employ the OPC water model [15].

2.1 Molecular dynamics simulation

We retrieve the structures of myosin from three different species: *Sus scrofa* (PDB ID: 5O2L [2]), *Gallus gallus* (PDB ID: 5M05 [3]), and *Bos taurus* (PDB ID: 5N6A [4]) for MD simulations. Using CHARMM-GUI [16, 17, 18, 19], we prepare the initial topology, input files, and simulation boxes. We configure each myosin system to mimic physiological conditions with an ionic concentration of 150 mM and achieve a zero net charge by adding Na^+ and Cl^- ions. This solvation process yields substantial systems comprising 313354 atoms for *S. scrofa*, 284947 for *G. gallus*, and 258237 for *B. taurus*, as depicted in Figure 2.

After initial setup, we conduct energy minimization using the Particle Mesh Ewald (PME) method to calculate long-range electrostatic interactions with a 1 nm cut-off distance. The minimization process uses the steepest descent method with harmonic restraints to preserve the structure. We then heat the systems to 303.15 K over 125 ps while maintaining the restraints under an NVT (constant number of particles, volume, and temperature) ensemble. Once equilibrium is achieved, we switch to an NPT (constant number of particles, pressure, and temperature) ensemble

simulation, stabilizing the temperature with the Langevin thermostat and maintaining pressure at 1 atm with isotropic position scaling for constant pressure periodic boundary conditions. We extend these simulations to 100 ns for thorough equilibration.

With the systems equilibrated, we continue with NVE (constant number of particles, volume, and energy) simulations to run classical trajectories and enable detailed sampling at 20 ps intervals. This method allows us to collect coordinate and velocity data at each simulation step, effectively every 0.002 ps, to analyze dynamic properties, especially molecular vibrations. From these simulations, we extract five samples of velocity data files, which we use to calculate auto-correlation functions (ACFs) and to generate IR spectra for our study.

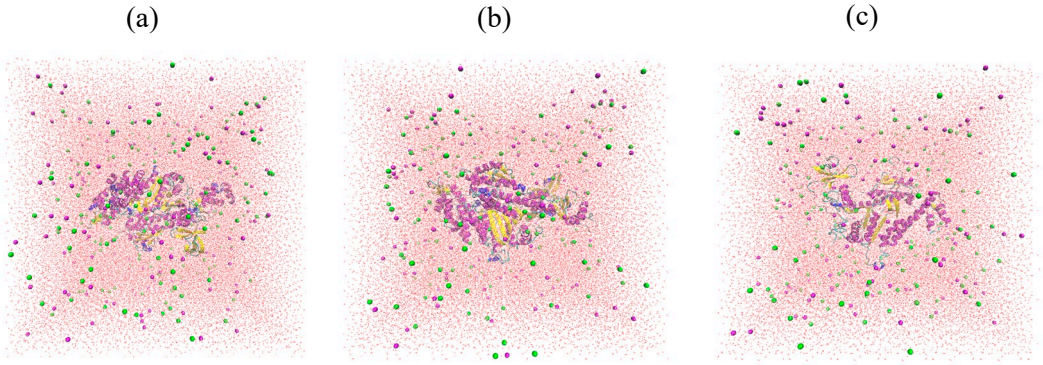

Figure 2. Initial configuration of myosins, the myosin protein at the centre of the water box, the red dots are the water molecules, the magenta beads are the Na^+ ions and the green bead are the Cl^- ions. (a) pig (*Sus scrofa*), (b) chicken (*Gallus gallus*), (c) cattle (*Bos taurus*)

2.2 Calculation of Root Mean Square Deviation (RMSD) and Root Mean Square Fluctuation (RMSF)

RMSD quantifies how much a set of target coordinates (such as a particular structure) differs from a set of reference coordinates. An RMSD of 0.0 signifies that the two sets of coordinates are perfectly superimposed. The RMSD is calculated using the following formula:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N} m_i(\boldsymbol{x_i} - \boldsymbol{y_i})^2}{M}} \tag{1}$$

In this equation, N represents the total number of atoms, m_i is the mass of the i^{th} *atom,* $\boldsymbol{x_i}$ is the coordinate vector of the target atom i, $\boldsymbol{y_i}$ the coordinate vector of the reference atom i, and M is the sum of the masses of all atoms. For a non-mass-weighted RMSD, each m_i is taken as 1, and M equals N.

Moreover, RMSF serves to quantitatively assess the dynamic behavior of molecular structures within MD simulations. RMSF of a given atom i is calculated as:

$$RMSF = \sqrt{\langle (x_i - \langle x_i \rangle)^2 \rangle} \tag{2}$$

where x denotes atomic positions and the averages are over all input frames. If RMSF calculated by residue specifically, the mass-weighted average of atomic fluctuations of each atom for either each residue will be calculated respectively:

$$\langle Fluct \rangle = \frac{\sum AtomFluct_i \times Mass_i}{\sum Mass_i} \tag{3}$$

The MD simulation output trajectories were subjected to detailed analysis with the CPPTRAJ tool, enabling the computation of RMSD and RMSF metrics as referenced in [20].

2.3 Calculation of autocorrelation function (ACF) and IR Spectra

We use Python program to write the code for ACF and IR spectra calculation. The total dipole moment denoted as $\mu(t)$ of protein at any time t was obtained by summing the product of charges and position vectors of the individual atoms [11] as

$$\mu(t) = \sum_{i=1}^{N} q_i \boldsymbol{r_i} \tag{4}$$

where N is the total number of atoms and q_i is the charge and $\boldsymbol{r_i}$ is the position vector of the i^{th} atom. $\mu(t)$ is equal to the sum of all the individual dipole moments of the molecules in the simulation box. The autocorrelation function of the dipole moment $f_{(t)}$ is given by [21, 22]:

$$f_{(t)} = \langle \boldsymbol{\mu}(0) \cdot \boldsymbol{\mu}(t) \rangle = \langle \sum_{i=1}^{N} q_i \boldsymbol{r_i}\ (0) \cdot \sum_{i=1}^{N} q_i \boldsymbol{r_i}\,(t)\ \rangle \tag{5}$$

The angular brackets represent an average taken over all time origins. Meanwhile, the infrared spectrum for system can be computed by Fourier transforming the total dipole moment autocorrelation function as equation (6)[23]

$$I_{(\omega)\mathrm{cl}} = \frac{1}{2\pi\omega^2} \int_{-\infty}^{\infty} dt\, e^{-i\omega t} \langle \boldsymbol{\mu}(0) \cdot \boldsymbol{\mu}(t) \rangle \tag{6}$$

where $I_{(\omega)\mathrm{cl}}$ is the classical spectral density or the classical absorption line shape and ω is the angular frequency of IR radiation. Using periodic boundary conditions during MD simulations causes discontinuities in position due to leaving and entering of atoms inside the simulation box [21]. The IR spectrum can be also obtained using the autocorrelation function of the time derivative of the dipole moment (electrical flux-flux correlation function) [11, 21].

$$I_{(\omega)\mathrm{cl}} = \frac{1}{2\pi\omega^2} \int_{-\infty}^{\infty} dt\, e^{-i\omega t} \left\langle \frac{d\boldsymbol{\mu}(0)}{dt} \cdot \frac{d\boldsymbol{\mu}(t)}{dt} \right\rangle \tag{7}$$

$$I_{(\omega)\mathrm{cl}} = \frac{1}{2\pi\omega^2} \int_{-\infty}^{\infty} dt\, e^{-i\omega t} \left\langle \sum_{i=1}^{N} q_i \boldsymbol{v_i}\ (0) \cdot \sum_{i=1}^{N} q_i \boldsymbol{v_i}\,(t) \right\rangle \tag{8}$$

Charge values obtained from the topology file, while velocity ($\boldsymbol{v_i}$) values obtained from the velocity file. The expression $\langle \sum_{i=1}^{N} q_i \boldsymbol{v_i}\ (0) \cdot \sum_{i=1}^{N} q_i \boldsymbol{v_i}\,(t) \rangle$ represents the electrical flux-flux correlation function. By formulating the spectrum using particle velocities, one can avoid the complications that arise from using periodic boundary conditions.

The quantum correction factor based on the "harmonic approximation" (Q_{HA}) also was applied in this study to correct the spectral density obtained from the classical correlation function. The final spectral density is resulted as follows:

$$I(\omega) = Q_{\mathrm{HA}} I(\omega)_{\mathrm{cl}} = \left(\frac{\beta\hbar\omega}{1 - exp\,(-\beta\hbar\omega)} \right) \frac{1}{2\pi\omega^2} \int_{-\infty}^{\infty} dt\, e^{-i\omega t} \left\langle \frac{d\boldsymbol{\mu}(0)}{dt} \cdot \frac{d\boldsymbol{\mu}(t)}{dt} \right\rangle \tag{9}$$

where $\beta = \frac{1}{k_B T}$ (k_B is Boltzmann constant with the value 1.380649 $\times$ 10^{-23}J/K, T is temperature at 303.15K), while $\hbar = \frac{\mathrm{h}}{2\pi}$ ($\hbar$ is reduced Planck constant with h is Planck constant with the value 6.62607015 $\times$ 10^{-34}Js).

To evaluate the precision of our ACF and IR spectra computational tool, we perform a validation simulation using water as a reference system. This simulation adheres to the protocols outlined in the AMBER software suite's official documentation that reproduce from study by Liu and colleagues [24]. Following the established procedures for classical MD simulations, we generate a set of trajectory samples, specifically velocity files, which served as inputs for our IR

spectral analysis algorithm.

Detailed documentation of the simulation parameters and the outcomes of the IR spectral analysis are thoroughly presented in Appendix A. In brief, the validation results instil confidence in our computational tool. The IR spectrum obtained from the water simulation displayed a striking resemblance to the reference spectrum provided in the AMBER tutorial, indicating that our program performs calculations accurately. Further tests were conducted by varying the number of trajectory samples used for the IR spectrum computation—specifically, 100, 25, and 5 velocity files. The results indicated that IR spectra derived from 100 and 25 velocity files exhibited consistency in the number and intensity of spectral peaks. However, a notable change was observed when the IR spectrum was calculated with only 5 velocity files; a significant peak around 3700 cm^{-1}, typically prominent in the spectra of water, was absent (see Figure A.1(c)). This discrepancy highlights the importance of sample size in trajectory analysis: a lower number of velocity files can lead to an IR spectral profile that lacks accuracy and may omit critical absorption features. Such findings underscore the necessity for an adequate sample size to achieve a reliable representation of a system's vibrational characteristics in IR spectral simulations.

Minor adjustments in the computational parameters were employed for the IR spectra calculations of water and myosin protein, specifically concerning the sample rate used during the frequency/ wavenumber setting in spectral density calculations via Fourier transform. A sample rate of 3200 was utilized for water to adequately capture its spectral characteristics, whereas a higher sample rate of 5500 was necessary for the myosin protein to resolve the more complex vibrational modes inherent to its structure. This distinction in sample rates is essential to ensure accurate representation of the respective molecular vibrations in the computed IR spectra.

For water, which has relatively simple vibrational modes due to its small molecular size, a lower sample rate can suffice to accurately capture the necessary spectral information. For proteins like myosin, which have complex structures and consequently more intricate vibrational modes, a higher sample rate is often required. This ensures that the details of these vibrations are not lost and that the spectral representation is as accurate as possible. The statement is consistent with the Nyquist-Shannon sampling theorem and best practices in spectroscopic computational analysis [25]. The theorem posits that to accurately sample a signal, the sampling frequency must be at least twice the highest frequency present in the signal. Selecting an appropriate sample rate is therefore crucial to ensure the resulting spectra are a true representation of the molecular vibrations within the sample.

Furthermore, the horizontal axis was converted from frequency (ν) to wavenumber ($\tilde{\nu}$), by unit conversion 3.3356405 which obtained from the equation:

$$\nu = \frac{c}{\lambda} \tag{10}$$

where c is the wave's propagation velocity. In a vacuum, the velocity of electromagnetic radiation is represented by $c = 3 \times 10^8$ m/s. Frequency and wavelength serve as key descriptors for electromagnetic radiation. Therefore, the wavenumber measured in reciprocal centimeters, is the inverse of the wavelength, given by the formula in equation (9)

$$\tilde{\nu} = \frac{1}{\lambda} \tag{11}$$

In the IR spectra computation analysis for myosin also, certain components of the system, specifically water molecules and the ions Na^+ and Cl^-, were deliberately excluded from the calculations. This exclusion is a standard procedure in spectral analysis when the focus is on the vibrational modes of the macromolecule itself, without interference from the solvent or ions that are typically present in a biological context. By omitting these entities, we ensure that the resulting IR spectrum is representative solely of the myosin's molecular vibrations, thus providing a clearer understanding of the protein's intrinsic structural dynamics. This approach simplifies the

computational model and enhances the clarity of the spectral features attributed to the protein, allowing for a more precise investigation into the vibrational properties of myosin.

3 Results and Discussion

3.1 RMSD and RMSF

After completing the 100 ns production runs in MD simulations, we calculate the RMSD to assess the structural fidelity of the proteins against their initial configurations. Figure 3 shows the evolution of RMSD over time for myosin proteins from three distinct species, presenting their respective RMSD values with line colour green, blue, and pink correspond to myosin from *Sus scrofa, Gallus gallus,* and *Bos taurus* respectively.

Blanck and co-workers [2] analyzed the crystal structure of myosin VI from *Sus scrofa* and resolved it at 2.2 Å. This high-resolution structure consisted of 6387 atoms and was modelled with 763 residues, indicating a detailed depiction of the myosin's architecture. The RMSD in the green plot showed fluctuations for this protein ranging from 2 to 4 Å. These findings align with Blanck and co-workers' detailed structural analysis, suggesting that the RMSD fluctuations accurately reflect the protein's dynamic nature within the simulation, which is crucial for its functional activities.

Regarding the smooth muscle myosin from *Gallus gallus*, Sirigu *et al.* [3] reported a 2.67 Å resolution structure comprising 5781 atoms and 708 residues. The RMSD fluctuations observed in the blue plot ranged from 3 to 4 Å. This indicates a degree of stability and conformational flexibility, which may relate to the smooth muscle myosin's functional role, as reported by Sirigu *et al.*

The cardiac myosin from *Bos taurus*, studied by Planelles-Herrero *et al.* [4], was crystallized at a lower resolution of 3.10 Å and modelled with 5569 atoms across 710 residues. This simulation's RMSD values (pink line) displayed a broader fluctuation, ranging from 5 to 6 Å. Given the lower resolution reported by Planelles-Herrero *et al.,* the RMSD values may not be as precise as those for the other two proteins. However, they still provide valuable insights into the conformational dynamics pertinent to cardiac myosin's role in heart function and its potential therapeutic applications in heart failure.

In light of Carugo's findings [26] on the impact of resolution on RMSD values, this analysis underscores the importance of considering the resolution of the crystal structures when interpreting RMSD data from molecular dynamics simulations. The differences in RMSD values among the three species of myosin proteins emphasize the necessity of considering the resolution at which the crystal structures were determined to ensure the accuracy and reliability of simulation-based structural comparisons.

The part of myosin from cattle (*Bos taurus*) which is fluctuated more greatly than the other myosins elucidated by RMSF calculation result as shown in Figure 3 (b). At the initial peaks, we can see that the first notable peak for *Bos taurus* (pink line) occurs very early in the residue sequence, near the start, and it is significantly higher than the other two species. The first residue of myosin from cattle has highest RMSF value as 15.2434 Å and the 58^{th} residue has value 10.1587 Å. In correlating the RMSF data with the crystal structure of *Bos taurus* myosin (in Figure 1 (c)), we observe that regions with high RMSF values, indicative of increased flexibility, often do not coincide with the protein's more rigid α-helix (depicted as purple ribbons) or β-sheets (shown as yellow strands). The rigidity of α-helix and β-sheets in the protein structures is due to their specific hydrogen bonding patterns which confer stability to the protein [27]. The pronounced peaks in the RMSF plot, particularly at this beginning of the protein sequence, suggest that the initial residues, are likely to be in loop regions or turns that connect the structured α-helix and β-sheets.

Throughout the middle of the sequence, there are several peaks where myosin from cattle shows greater fluctuation than the others, with some particularly higher spikes which are noticeable, among others are 201^{th} residue with RMSF value 5.4491 Å and 561^{th} residue with value 6.936 Å.

Similarly, elevated RMSF values in this middle of the sequence, are characteristic of either loop regions or flexible linkers between domains, which facilitate movement and conformational changes essential for myosin's activity. Toward the C-terminus, the significant fluctuations, especially noted in the last quarter of the RMSF plot, likely correspond to a less structured C-terminal tail, as evidenced by the 719th residue with an RMSF of 10.0832 Å.

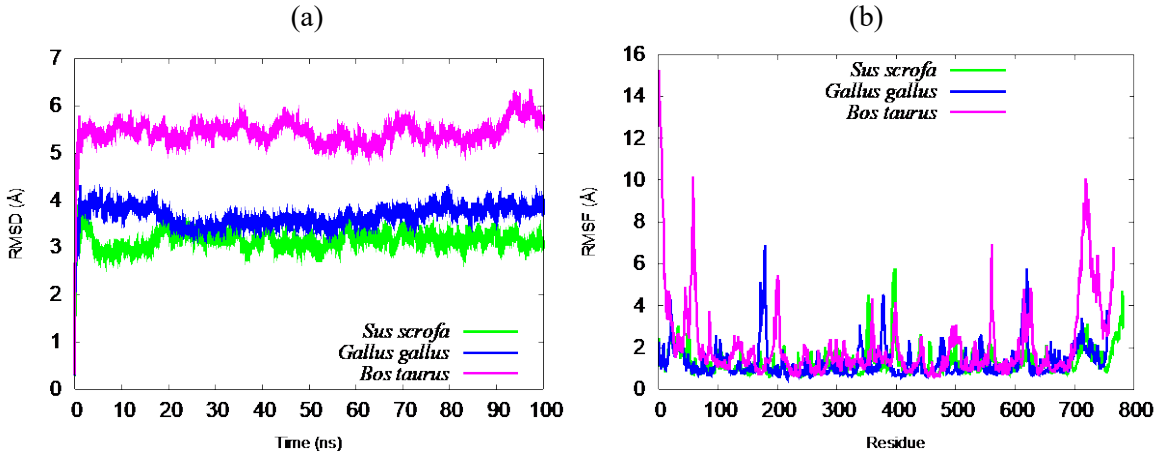

Figure 3. Dependences of RMSD on time (a) and RMSF by residue (b) of three myosins of MD simulation. The green line corresponds to myosin from *Sus scrofa*, the blue line represents myosin from *Gallus gallus*, and the pink line refers to myosin from *Bos taurus*.

3.2 IR spectra calculation

After the equilibrated systems are in hand, we process them with NVE simulations to propagate classical trajectories and facilitate detailed sampling at 20-ps intervals. This approach enabled us to capture coordinate and velocity data at each simulation step, effectively every 0.002 ps, to elucidate dynamic properties with a particular emphasis on molecular vibrations. We extract five samples of velocity data files from these simulations to serve as the basis for calculating ACFs and generating IR spectra for our study.

The analysis of the IR spectra from molecular dynamics of myosin proteins from three different species simulations as shown in Figure 4 provides a window into the protein's secondary structure and conformational dynamics. In Figure 4, we sign the five or six highest peak of IR intensity absorption. The spectra cover the amide I which ranges from 1600 to 1700 cm^{-1} and amide II regions ranging from 1500 to 1600 cm^{-1}. Both of these ranges are significant for understanding protein backbone arrangements. In the amide I region which is primarily associated with the C=O stretching vibration of the protein backbone, distinct features are observed across the three species. For the pig myosin, the highest peaks at 1613, 1656, and 1685 cm^{-1} suggest a mixture of secondary structures, with the peak at 1656 cm^{-1} indicating α-helix content, which is common in proteins, while the peaks at 1613 and 1685 cm^{-1} could be associated with β-sheet structures and β-turns or unordered structures, respectively. The chicken myosin exhibits peaks at 1541, 1570, and 1670 cm^{-1}, with the peak at 1670 cm^{-1} aligning with the presence of α-helix, similar to the pig myosin. In contrast, the cattle myosin shows peaks at 1555, 1656, and 1670 cm^{-1}, indicating a strong α-helix presence similar to the other two species, with the peak at 1555 cm^{-1} potentially pointing to β-sheet structures [28].

The amide II region corresponds to N-H bending vibrations coupled with C-N stretching. The pig myosin have some peaks in this region, which the intensity not so high as in the area above 1600 cm^{-1}. In contrast, a peak at 1574 cm^{-1} is only observed in pig myosin, indicating a structure that may be unique to the pig myosin or not present in the myosin of chicken and cattle. On the other hand, the chicken myosin's peak at 1541 cm^{-1} and the cattle myosin's peak at 1555

cm^{-1} in the amide II region both support the presence of β-sheet structures in these proteins. Peaks observed outside the typical amide I and II regions, such as those near 1714, 1800, and 1944 cm^{-1}, may represent other types of molecular vibrations, artifacts from the simulation, or data processing, and could possibly be overtones or combination bands, which are generally weaker and less structurally informative.

Overall, the spectra reveal conserved elements of secondary structure among the myosin proteins, with α-helix being a predominant feature in all three species, which is an expected finding considering the similar functional roles of myosin. The variation in the spectral signatures, however, suggests differences in tertiary structure, sequence, or interactions with other molecules within the cells of each species. These disparities may have functional implications, influencing the mechanics and activity of myosin, which plays a critical role in muscle contraction and mobility across diverse biological systems.

On the other hand, we also meticulously compare the IR spectra peaks derived from normal MD simulations for myosin proteins of three distinct species with corresponding experimental data [29], as shown in Table 1. The experimental data provides a general overview of the infrared bands associated with amino acid side chains found in proteins, without specificity to particular protein types or species. The analysis focused on identifying the wavenumber positions of IR spectral peaks, assigning their molecular functionalities, and correlating these findings with the amino acid composition detailed in the respective PDB files. The observed IR spectra peaks highlighted in our MD simulations reflect the characteristic vibrational modes inherent to specific amino acids within the myosin proteins. Notably, the presence and absence of specific peaks varied across the species, providing insights into the unique structural nuances of each protein. For example, the peak at 1512 cm^{-1}, attributed to C-N stretching, C-H bending, and N-H bending vibrations in Tryptophan (Trp), was identified in *Gallus gallus* and *Bos taurus* but was not present in the *Sus scrofa* simulation. This absence could suggest variations in the local environment or conformational state of the Trp residues in the *Sus scrofa* myosin.

Furthermore, the amino acid count from the PDB files revealed differences in myosin composition across species, which can influence the IR spectrum. For instance, the peak at 1555 cm^{-1} corresponds to the symmetric stretching of COO^- groups from Glutamic acid (Glu), closely matching the experimental data. This is significant due to the high abundance of Glu residues in myosin proteins. These counts are crucial for understanding the influence of specific amino acids on the protein's IR spectral signature.

Peaks associated with secondary structures, such as the α-helix at 1656 cm^{-1} and the β-sheet at 1685 cm^{-1}, were consistently observed across all species, indicating the preservation of these structural elements within the simulation. It is noteworthy that the absence of specific peaks in the MD simulation, when compared to experimental data, could point to the limitations of the simulation methods or differences in the experimental conditions, underscoring the need for careful consideration when interpreting these results.

This comparative analysis elucidates the correlation between the MD simulations and experimental IR spectra, shedding light on the dynamic structural behaviour of myosin proteins. It underscores the importance of integrating computational and experimental approaches to gain a comprehensive understanding of protein dynamics.

Table 1. Assignment of IR spectra

NO.	IR SPECTRA PEAKS FROM THE NORMAL MD SIMULATION			ASSIGNMENT/ FUNCTIONALITY	WAVENUMBER POSITION IN EXPERIMENTAL DATA	NUMBER AMINO ACID OF EACH MYOSIN (FROM PDB FILE)		
	Sus scrofa	*Gallus gallus*	*Bos taurus*			*Sus scrofa*	*Gallus gallus*	*Bos taurus*
1	-	1512	1512	ν(CN), δ(CH), δ(NH) from Tryptophan (Trp)	1509	Trp: 2	7	7
2	1555	1555	1555	$\nu_s(COO^-)$ from Glutamic acid (Glu)	1556-1560	Glu:53	63	64
3	1570	1570	1570	1. $\nu_s(COO^-)$ from Aspartic acid (Asp)	1. 1574–1579	Asp:50	45	42
4	1574	-	-	2. ν(CN) from Histidine (His)	2. 1575	His:25	15	18
5	-	1584	-	1. $\delta(NH_2)$ from Glutamine (Gln)	1. 1586-1610	Gln:30	42	35
6	-	1598	1598	2. ν(C=C) from His	2. 1594	His:25	15	18
7	1613	1613	1613	3. $\delta(NH_2)$ from Asparagine (Asn)	3. 1612-1622	Asn:46	44	44
8	1627	1627	1627	1. ν(CC), ν(C=C) from Trp	1. 1622	Trp: 2	7	7
9	-	1642	-	2. ν(C=C) from His Secondary structure: Disordered	2. 1631 Average: 1654, extremes:1642-1657	His:25	15	18
10	1656	1656	1656	Secondary structure: α-helix	Average:1654, extremes:1648-1657			
11	1670	1670	1670	1. ν(C=O) from Asn 2. ν(C=O) from Gln 3. Secondary structure: Turns	1. 1677-1678 2. 1668-1687 3. Average: 1672, extremes: 1662-1686	Asn:46 Gln:30	44 42	44 35
12	1685	1685	1685	1. ν(C=O) from Gln 2. Secondary structure: β-sheet	1. 1668-1687 2. average: 1684, extremes:1674-1695	Gln:30	42	35
13	1714	1714	-	1. ν(C=O) from Glu 2. ν(C=O) from Asp	1. 1712 2. 1716	Glu:53 Asp:50	63 45	64 42

Its attribution to amino acid based on experimental data [29] and its correlation with number of amino acids focused on 1500-1800 region. – there is no peak, v: stretching vibration, v_s: symmetric stretching vibration, δ: in plane bending vibration.

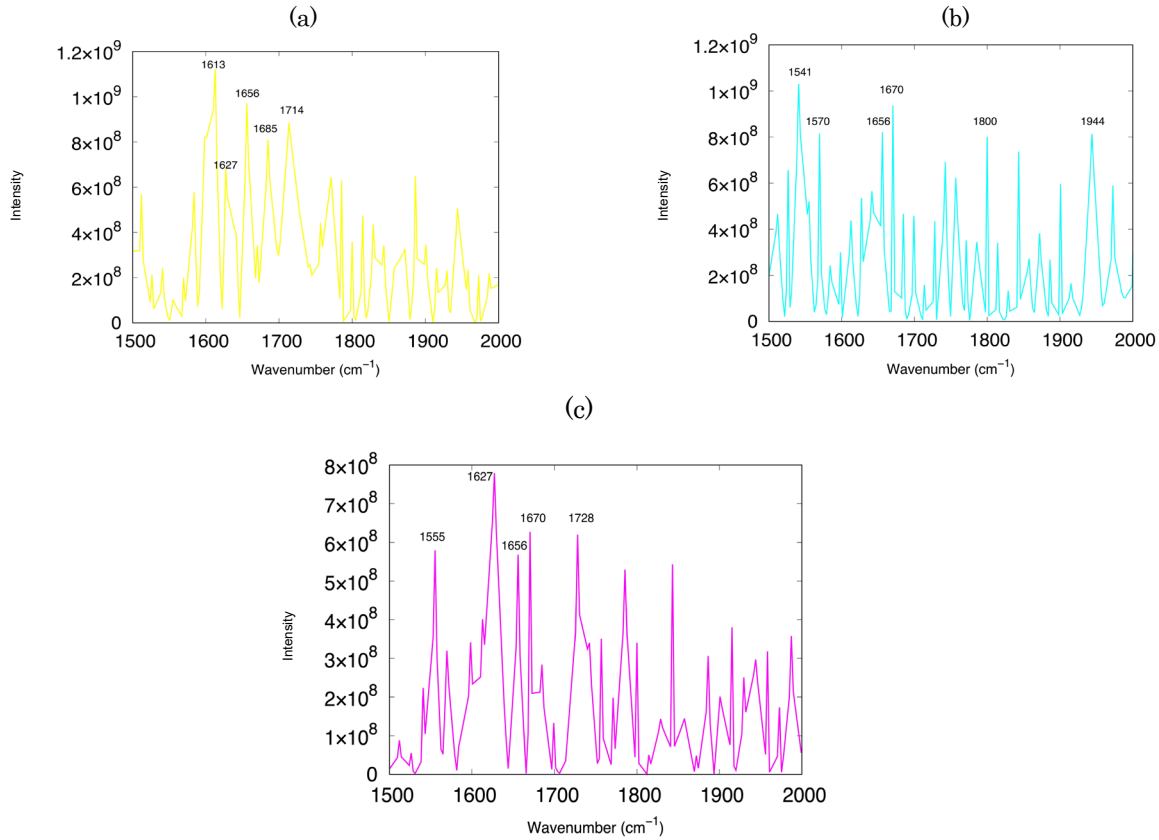

Figure 4. The IR spectra peaks of each myosin in amide I and II region range for three kind simulations; a. Myosin of pig (*Sus scrofa*), b. myosin of chicken (*Gallus gallus*), c. myosin of cattle (*Bos taurus*)

4 Summary

We perform a 100 ns all-atom MD simulation of three myosins: pig, chicken, and cattle, with the number of atoms each myosin before solvation, 6387, 5781, and 5569, respectively. In MD simulations, using periodic boundary conditions may cause discontinuities due to atoms moving across the simulation box boundaries. To avoid this, the infrared spectrum is computed using the autocorrelation function of the dipole moment's time derivative, also called the electrical flux-flux correlation function. This approach, which is based on particle velocities, helps to maintain continuity despite the periodic boundaries.

After equilibrating the molecular systems, NVE ensemble simulations were conducted to monitor classical trajectories and enable detailed data sampling every 20 picoseconds, capturing minute changes in coordinate and velocity data every 0.002 picoseconds. This provided insight into the dynamic properties and molecular vibrations of the myosin proteins. From this data, five samples of velocity files were extracted to calculate the autocorrelation functions and generate the IR spectra for the study.

In the amide I region, pig myosin shows peaks at 1613, 1656, and 1685 cm^{-1}, indicating a mix of structures, with 1656 cm^{-1} suggesting α-helix and 1613 and 1685 cm^{-1} suggesting β-sheets and disordered structures respectively. Chicken myosin has peaks at 1541, 1570, and 1670 cm^{-1}, and cattle myosin at 1555, 1656, and 1670 cm^{-1}, both indicating α-helix content, with cattle myosin's peak at 1555 cm^{-1} potentially indicating β-sheets. The amide II region involves N-H

bending and C-N stretching vibrations. Pig myosin has peaks in this region, with a unique peak at 1574 cm^{-1} possibly indicating a structure unique to pig myosin. Chicken and cattle myosin have peaks at 1541 and 1555 cm^{-1}, respectively, which may suggest β-sheet presence.

In addition, we also compare IR spectra peaks of myosins from our MD simulations against experimental data. The analysis pinpoints the wavenumber positions of IR peaks, associates them with molecular functions, and relates these to amino acid sequences from Protein Data Bank (PDB) entries. IR spectra from the MD simulations exhibit vibrational modes indicative of amino acids in the myosin proteins. Species-specific differences in peaks suggest structural variations; for example, a peak at 1512 cm^{-1} linked to Tryptophan vibrations is missing in pig myosin, hinting at different local structures or conformations.

The amino acid composition, such as the prevalence of Glutamic acid, reflected in a peak at 1555 cm^{-1}, affects the IR spectra, aligning with experimental observations. Peaks at 1656 cm^{-1} and 1685 cm^{-1}, associated with α-helix and β-sheets, respectively, were consistent across species, confirming the retention of these secondary structures in the simulations. The research highlights the value of combining computational and experimental methods for a more thorough understanding of myosin dynamics.

A Appendix

In this segment, classical molecular dynamics simulations are utilized to determine the diffusion constant and the infrared (IR) spectrum for a system of 216 water molecules contained within a cubic simulation box under periodic boundary conditions. Initially, an isothermal-isobaric (NPT) simulation is conducted to ascertain the density of liquid water at standard conditions of 298.15 K and 1 atmosphere. This is followed by a microcanonical (NVT) simulation aimed at acquiring the canonical equilibrium properties. Post-equilibration, the coordinates and velocities garnered from the NVT simulation are used to set up the initial conditions for the subsequent constant energy (NVE) ensemble trajectories. The findings from these analyses are presented towards the conclusion of this section.

The 100 velocity files of the corresponding trajectories are used to calculate the autocorrelation function of the time derivative of the dipole moment (electrical flux-flux autocorrelation function) then Fourier transform it to obtain the infrared spectra intensities. Our python code is used to this calculation. The results are plotted below:

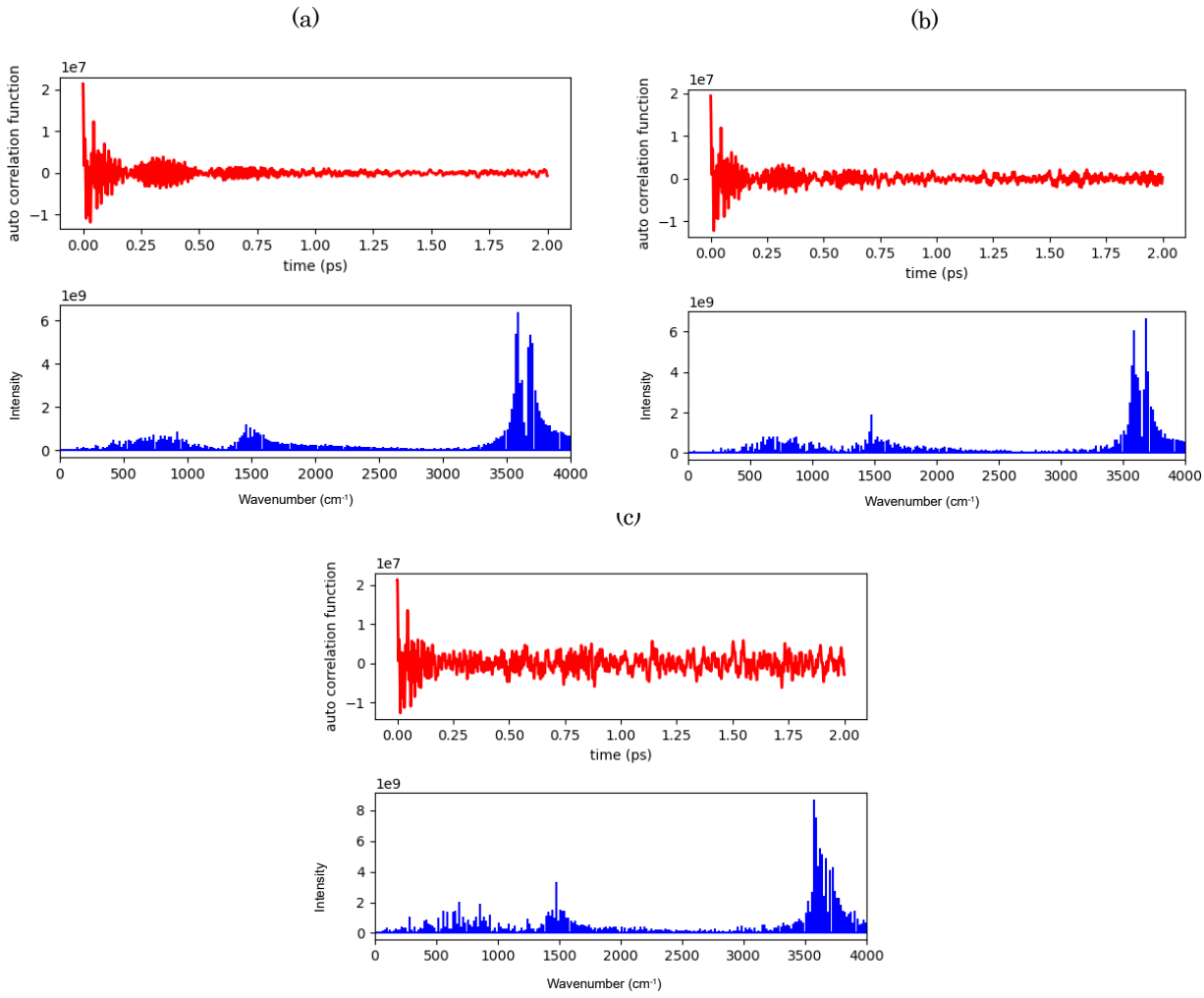

Figure A.1 The ACF and IR spectra result calculation of water with different sample number of velocity files . (a) 100, (b) 25, and (c) 5.

Observations from figure A.1 indicate that the ACF and the graphical representation of the third dataset differ markedly from the others. Additionally, the IR spectra peaks for the datasets utilizing 100 and 25 velocity files data points share similarities in their wavenumber locations. However, the IR spectrum calculated with only 5 velocity files data points shows a significant deviation, with the absence of one of the peaks that were present in the other datasets.

Acknowledgements

A part computation was performed using Research Center for Computational Science, Okazaki, Japan (Project: 23-IMS-C987).

References

[1] Harrington, W. F. and Rodgers, M. E. : Myosin. Ann. Rev. Biochem 53 (1984), 35-73.
[2] Blanck, F., Isabet, T., Benisty, H., Sweeney, H. L., Cecchini, M., and Houdusse, A. : An intermediate along the recovery stroke of myosin VI revealed by X-ray crystallography and molecular dynamics. Proc. Natl. Acad. Sci. U. S. A 115 (2018), 6213-6218
[3] Sirigu, S., Hartman, J. J., Planelles-Herrero, V. J., Ropars, V., Clancy, S., Wang, X., Chuang, G., Qian, X., Lu, P. P., Barrett, E., Rudolph, K., Royer, C., Morgan, B. P., Stura, E. A., Malik, F. I., and Houdusse, A. M. : Highly selective inhibition of myosin motors provides the basis of potential therapeutic application. Proc. Natl. Acad. Sci. U. S. A. 113 (2016), E7448-E7455
[4] Planelles-Herrero, V. J., Hartman, J. J., Robert-Paganin, J., Malik, F. I., and Houdusse, A. : Mechanistic and structural basis for activation of cardiac myosin force production by omecamtiv mecarbil. Nat. Comm 8 (2017), 190
[5] Deniz, E., Altuntas, E. G., Ayhan, B., Igci, N., Demiralp, D. O., and Candogan, K. : Differentiation of beef mixtures adulterated with chicken or turkey meat using FTIR spectroscopy. J. Food Process Preserv 42 (2018), E13767
[6] Rohman, A., Sismindari, Erwanto, Y., and Che-Man YB. : Analysis of pork adulteration in beef meatball using Fourier transform infrared (FTIR) spectroscopy. Meat Sci 88 (2011), 91-95
[7] Chen, G., Huang, K., Miao, M., Feng, B., and Campanella, O. H. : Molecular Dynamics Simulation for Mechanism Elucidation of Food Processing and Safety: State of the Art. Compr. Rev. Food Sci. F. (2019), 18.
[8] Kawaguchi, K., Saito, H., Okazaki, S., and Nagao, H. : Molecular dynamics study on the free energy profile for dissociation of ADP from N-terminal domain of Hsp90. Chem. Phys. Lett. 588 (2013), 226-230.
[9] Zhu, S.B. and Robinson, G.W. : Molecular dynamics study of liquid carbon monoxide, Computer Physics Communications 52 (3) (1989).
[10] Guillot, B. : A molecular dynamics study of the far infrared spectrum of liquid water. *J. Chem. Phys* 95 (3) (1991), 1543–1551.
[11] Agarwal, V., Huber, G. W., Conner, W.C. Jr, Auerbach, S. M. : Simulating infrared spectra and hydrogen bonding in cellulose Iβ at elevated temperatures. J Chem Phys. 2011 Oct 7;135(13):134506.
[12] Leach, A. R. : Molecular Modelling Principles and Applications 2nd edition, Pearson Education, (2001).
[13] Case, D.A., Belfon, K., Ben-Shalom, I.Y., Brozell, S.R., Cerutti, D.S., Cheatham III, T.E., Cruzeiro, V.W.D., Darden, T.A., Duke, R.E., Giambasu, G., Gilson, M.K., Gohlke, H., Goetz, A.W., Harris, R., Izadi, S., Iz- mailov, S.A., Kasavajhala, K., Kovalenko, A., Krasny, R., Kurtzman, T., Lee, T.S., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R. , Man, V. , Merz, K.M., Miao, Y., Mikhailovskii, O., Monard, G., Nguyen, H., Onufriev, A., Pan, F., Pantano, S., Qi, R., Roe, D.R., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shen, J., Simmerling, C.L., Skrynnikov, N.R., Smith, J., Swails, J., Walker, R.C., Wang, J., Wilson, L., Wolf, R.M., Wu, X., Xiong, Y., Xue, Y., York, D.M., and Kollman, P.A. AMBER 2020, University of California, San Francisco, (2020).
[14] Tian, C., Kasavajhala, K., Belfon, K. A. A., Raguett, L., Huang, H., Migues, A. N., Bickel, J., Wang, Y., Pincay, J., Wu Q., Simmerling, C. : ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. J. Chem. Theory Comput. 16 (1) (2020), 528–552.
[15] Izadi, S., Anandakrishnan, R., Onufriev, A. V. : Building Water Models: A Different Approach. *J. Phys. Chem. Lett. 5* (2014), 3863–3871.
[16] Jo, S., Kim, T., Iyer, V.G., and Im, W. : CHARMM-GUI: a web-based graphical user interface for CHARMM. J. Comput. Chem. 29 (2008), 1859-1865.
[17] Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr, Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M.,

Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Karplus, M. : CHARMM: the biomolecular simulation program. Journal of computational chemistry 30(10) (2009), 1545–1614.
[18] Lee, J., Cheng X., Swails, J.M. Yeom, M. S., Eastman, P. K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J. C., Qi, Y., Jo, S., Pande, V. S., Case, D. A., Brooks III, C. L., MacKerell Jr., A.D., Klauda, J. B., and Im, W. : CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. Journal of Chemical Theory and Computation 12 (1) (2016), 405-413.
[19] Lee, J., Hitzenberger, M., Rieger, M., Kern, N.R., Zacharias, M., and Im, W. : CHARMM-GUI supports the Amber force fields. The Journal of chemical physics 153(3) (2020), 035103.
[20] Roe, D. R. and Cheatham III, T. E. : PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. Journal of Chemical Theory and Computation 9 (7**)** **(**2013), 3084-3095.
[21] Bornhauser, P. and Bougeard, D. : Intensities of the Vibrational Spectra of Siliceous Zeolites by Molecular Dynamics Calculations. I. Infrared Spectra. The Journal of Physical Chemistry B 105 (1) **(**2001), 36-41.
[22] Praprotnik, M. and Janezic, D. : Molecular dynamics integration and molecular vibrational theory. III. The infrared spectrum of water. J. Chem. Phys. 122 (2005), 174103.
[23] Noid, D. W., Koszykowski, M. L., and Marcus, R. A. : A spectral analysis method of obtaining molecular spectra from classical trajectories. J. Chem. Phys. 67 (1977). 404.
[24] Liu, J., Miller, W. H., Paesani, F., Zhang, W., and Case. D. A. : Quantum dynamical effects in liquid water: A semiclassical study on the diffusion and the infrared absorption spectrum . J. Chem. Phys. 131 (2009), 164509.
[25] Oppenheim, A. V. and Schafer, R. W. : Discrete-time signal processing. Pearson Education, (2009).
[26] Carugo, O. : How root-mean-square distance (r.m.s.d.) values depend on the resolution of protein structures that are compared. J. Appl. Cryst. 36 (2003), 125-128.
[27] Clark, J: The Structure of Proteins. (2022, July 5). Truro School in Cornwall. https://chem.libretexts.org/@go/page/3962
[28] De Meutter, J., & Goormaghtigh, E. : Evaluation of protein secondary structure from FTIR spectra improved after partial deuteration. European biophysics journal : EBJ, 50(3-4) (2021), 613–628.
[29] Barth, A. : Infrared spectroscopy of proteins. Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1767 (9) (2007), 1073-1101.