



Unpolarized laser method for infrared spectrum calculation of amide I C=O bonds in proteins using molecular dynamics simulation

Viet Hoang Man^{a, **}, Xibing He^a, Phuong H. Nguyen^b, Celeste Sagui^c, Christopher Roland^c, Xiang-Qun Xie^a, Junmei Wang^{a, *}

^a Department of Pharmaceutical Sciences and Computational Chemical Genomics Screening Center, School of Pharmacy, University of Pittsburgh, Pittsburgh, PA, 15261, USA

^b CNRS, Université Paris Cité, UPR9080, Laboratoire de Biochimie Théorique, Institut de Biologie Physico-Chimique, Fondation Edmond de Rothschild, 13 Rue Pierre et Marie Curie, 75005, Paris, France

^c Department of Physics, North Carolina State University, Raleigh, NC, 27695-8202, USA

ARTICLE INFO

Keywords:

IR spectra

Amide I

Radiation absorption

External electric field

MD simulation

ABSTRACT

The investigation of the strong infrared (IR)-active amide I modes of peptides and proteins has received considerable attention because a wealth of detailed information on hydrogen bonding, dipole-dipole interactions, and the conformations of the peptide backbone can be derived from the amide I bands. The interpretation of experimental spectra typically requires substantial theoretical support, such as direct *ab-initio* molecular dynamics simulation or mixed quantum-classical description. However, considering the difficulties associated with these theoretical methods and their applications are limited in small peptides, it is highly desirable to develop a simple yet efficient approach for simulating the amide I modes of any large proteins in solution. In this work, we proposed a comprehensive computational method that extends the well-established molecular dynamics (MD) simulation method to include an unpolarized IR laser for exciting the C=O bonds of proteins. We showed the amide I frequency corresponding to the frequency of the laser pulse which resonated with the C=O bond vibration. At this frequency, the protein energy and the C=O bond length fluctuation were maximized. Overall, the amide I bands of various single proteins and amyloids agreed well with experimental data. The method has been implemented into the AMBER simulation package, making it widely available to the scientific community. Additionally, the application of the method to simulate the transient amide I bands of amyloid fibrils during the IR laser-induced disassembly process was discussed in details.

1. Introduction

Infrared (IR) spectroscopy, which is based on the absorption of IR radiation, is one of the most popular methods in molecular structure determination [1]. It has been widely used to investigate protein structures in protein folding, unfolding and misfolding [2–7]. Among the IR normal modes of the protein IR spectra, the amide I mode, which arises mainly from the C=O stretching vibration, is most commonly used for protein secondary structure analysis due to its strong signal. Additionally, the amide I vibration is hardly affected by the nature of the side-chain, but it is sensitive to hydrogen bonding, dipole-dipole interactions, and the conformation of the backbone [8,9]. The amide I IR

spectra of a protein for a given protein structure can reveal the percentages of the basic secondary structural contents including α -helix, β -sheet, turns and disordered loops in the protein structure [10,11]. However, it cannot directly show the protein tertiary structure which may be modeled from the data of X-ray or NMR experiments, or by pure theoretical modeling. In this context, one can first calculate the theoretical IR spectra of a protein tertiary structure model, and then investigate how well the calculated IR spectra are consistent with the experimental ones. In this way, the quality of the protein model can be assessed. Moreover, the combination of theoretical and experimental amide I IR spectra can reveal the transition-state structures in a protein dynamics process [12,13].

* Corresponding author. Department of Pharmaceutical Sciences and Computational Chemical Genomics Screening Center, School of Pharmacy, University of Pittsburgh, 3501 Terrace Street, Pittsburgh, PA, 15261, USA.

** Corresponding author.

E-mail addresses: vhm3@pitt.edu (V.H. Man), junmei.wang@pitt.edu (J. Wang).

<https://doi.org/10.1016/j.combiomed.2023.106902>

Received 6 December 2022; Received in revised form 7 March 2023; Accepted 9 April 2023

Available online 13 April 2023

0010-4825/© 2023 Elsevier Ltd. All rights reserved.

In the last few decades, several theoretical-computational approaches have been proposed to reproduce the amide I IR spectra of well-defined molecules [14–26]. These approaches are typically based on mixed quantum-classical strategies and can be divided into two categories. In the first category, the vibration frequencies and eigenstates for single amides are determined by employing quantum mechanical calculations and then calibrated for the whole protein [14–18]. In the second category the Hessian matrix for the whole isolated protein is calculated, and then the local, single-residue-based, vibrational frequencies are obtained via a matrix reconstruction [19–21]. In these approaches, diagonalization of the whole Hamiltonian matrix or localized amide Hamiltonian matrices provides the vibrational eigenstates and energy of the whole protein for a given protein conformation. These approaches have been successfully applied to investigate different proteins, including β -hairpin peptides, ubiquitin, and other globular proteins, as well as to characterize the IR spectral features of individual secondary-structure elements [14,16,18,27]. Conversely, given an IR spectrum, we can deduce the secondary-structure elements that contributed to it.

In this study, we simulated an “unpolarized” laser via external electric field (EEF) to probe the infrared spectrum of C=O bonds in different protein structures by using all-atom nonequilibrium molecular dynamics (MD) simulations. In our method, the IR radiation is mimicked by a time dependent EEF which has a shape of laser pulse and IR frequencies. The EEF is applied along the directions of the individual C=O bonds to get maximize EEF’s effect. The basic principle of the method is that the EEF causes C=O stretching vibration and subsequently leads to a system energy change which represents the radiation absorption. This application of an EEF along covalent bonds allows us to simulate single-wavelength radiation. Note that polarized lasers are popular in experiments as well as applications for IR spectroscopy. In an experimental IR spectrum measurement, each protein in the solution will only

experience an electric field in a given direction, but the signal is averaged over the ensemble. Therefore, to model an experiment with a polarized laser, we need to use many same proteins with different orientations in a simulation system and/or multiple scanning with different angles of the EEF. This will result in considerable computational cost. On the other hand, the simulated “unpolarized” laser, where all excitation orientations are sampled, allows us to model an experimental IR measurement with a single protein in a simulation system and without the consideration of the EEF angles. This method has been implemented in the AMBER software package [28]. Using the unpolarized radiation, we computed the amide I IR spectrum of the C=O bonds in representative protein structures including β -sheet and α -helix. The optimization of the important input parameters for the IR spectrum calculation has been investigated. Finally, we discuss the application of the laser radiation for force field development and molecular mechanism studies for radiation absorption.

2. Methods and materials

2.1. Protein structural models and equilibrium MD simulation

The protein structures selected to calculate C=O bond infrared spectra include amyloid anti-parallel β -sheet (PDB ID: 2KIB) [29], amyloid parallel β -sheet (PDB ID: 1YJP) [30], β -hairpin (Tryptophan Zipper 2, PDB ID: 1LE1) [31], α -helix (PDB ID: 6N8C) [32], and the structure of the transmembrane region of the M2 protein H⁺ channel (PDB ID: 1MP6) [33]. The representative protein secondary structures, anti-parallel β -sheet (PDB ID: 2KIB), parallel β -sheet (PDB ID: 1YJP) and α -helix (PDB ID: 6N8C), are shown in Fig. 1b. For each structure, the protein was placed at the center of an octahedron box with a minimum distance of 10 Å between box borders and any atoms of protein. The protein was solvated with explicit water using the TIP3P water model

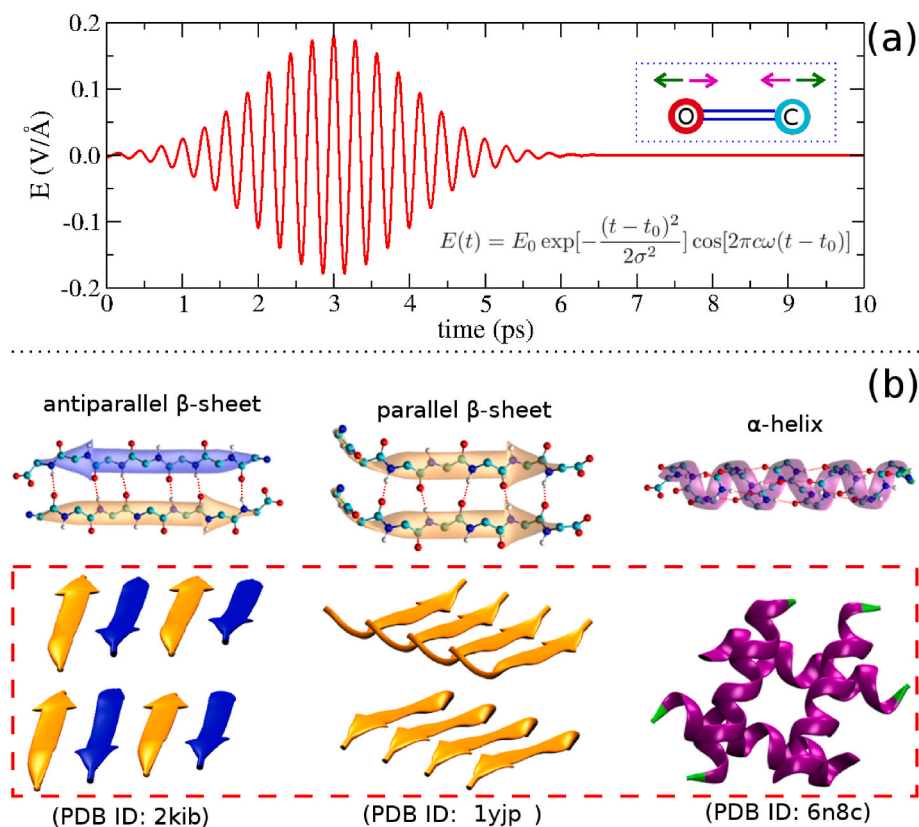


Fig. 1. Laser pulse mimicked by an external electric field and representative protein secondary structures studied in this work. (a) Cartoon of the carbonyl (C=O) bond vibration (dotted blue box) when the unpolarized laser shown in the figure is applied, (b) representative protein secondary structures.

[34]. The ff14SB force field [35] was used to describe proteins. The Na^+/Cl^- ions were properly added to neutralize the systems. Detailed information of the simulation systems is provided in the Supporting Information (Table S1). Each MD system underwent the following steps. First, the steepest descent minimization followed by a conjugate gradient minimization with the peptide atoms restrained at their initial positions, then the steepest descent minimization without restraint followed by a conjugate gradient minimization were carried out. The minimization step was followed by a short MD simulation under constant volume while the system was heated from 0 K to 300 K with weak restraints on the protein atoms. Next, Langevin dynamics at constant temperature (300 K) and constant pressure (1 atm) were run for 100 ps, after which the density of the system was found to be stable around 1.0 g/cm^3 . Finally, in the sampling phase, NPT simulations at 300 K were performed using the leap-frog algorithm with a time step of 2 fs and a duration time of 10 ns. The long-range Coulomb interaction was accounted by means of the Particle-Mesh Ewald (PME) method [36] with a $10 \text{ \AA}$ cutoff. The van der Waals interactions were calculated by means of a $10 \text{ \AA}$ atom-based nonbonded lists, with continuous corrections applied to the long-range parts. The constant pressure simulations were carried out at 1 atm via the Berendsen barostat [37] with the pressure relaxation time τ_p of 3.0 ps. The temperature was regulated using Langevin dynamics with a collision frequency of 1 ps^{-1} . The SHAKE algorithm [38] was applied to all bonds involving hydrogen atoms. The periodic boundary condition was applied to all MD simulations which were performed using the pmemd.cuda module implemented in the AMBER 19 package [28].

2.2. Simulated laser applied on the amide bond (C=O)

An “unpolarized” laser pulse was simulated by a time dependent electric field as shown in Equation (1).

$$E = E_0 * \exp\left[-\frac{(t - t_0)^2}{2\sigma^2}\right] * \cos[2\pi\omega(t - t_0)] \quad (1)$$

Here, E_0 (in $\text{V/\AA}$) represents the amplitude of the electric field, σ (ps) is the pulse temporal width, t (ps) is the time, t_0 (ps) is the time when the pulse is the maximum (E_0), c (cm/ps) is the speed of light, and ω (cm^{-1}), the space frequency is the inverse of the wavelength, as such $2\pi\omega$ is the wavenumber. This technique has been implemented in the GROMACS simulation package [39], but not in the AMBER package [28] yet. We have used this mimic laser for laser-induced simulations to study the dissociation of amyloid fibrils, poliovirus, peptide nanotubes and cellulose fibers and fast DNA melting [40–44]. The input parameters for a laser-induced simulation include E_0 , t_0 , ω and σ , in which, E_0 and σ control the laser power. In these laser simulations, the mimic laser is applied on a molecule such that the harmonic electric field excites the vibrational transitions of ions or partially charged atoms with the corresponding laser frequency, ω . In the case of the C=O amide bond, C and O atoms have positive and negative partial charges, respectively. Therefore, their vibration transitions are opposite to each other, resulting in bond stretch vibration. If the electric field frequency, ω , equals the natural frequency of the bond, the bond vibration will be resonant, and the energy of the simulated system will increase. In other words, one can state that the amide bond absorbs the laser energy at the specific bond stretch frequency, namely amide I frequency. In experiment, although the polarized laser is used, there are many molecules in the sample, and the measurement is repeated for each sample with different angles. Then, the spectrum is basically averaged over all orientations of molecules. In this context, we can think that the laser used in experiment is “unpolarized”. On the other hand, in MD simulations, the sample usually consists of a single protein. Therefore, if the mimic laser behaves like a polarized light beam for which the electric field vectors only vibrate in a specific plane, it shows maximum effect on the vibration of a bond only when the electric field vector is applied along

with the bond direction, and it has no or very little effect on the bond vibration when its electric field vector is perpendicular to the bond direction. The mimic laser can be used to measure the infrared spectra of amide I in some special peptide/protein structures such as β -sheet and single helices in which all C=O bonds are mostly in the same direction. In those cases, we can choose the direction of the electric field vector as parallel to the C=O bond to maximize the laser effect. However, in most protein structures, all C=O bonds are not ordered and take random directions. Thus, it is impossible to mimic the effect of applying a polarized laser on an ensemble of protein molecules in experiment by using a single protein in macromolecule in periodic MD simulations.

To overcome the above limitation, we applied a previously unused strategy to mimic the effect of applying an unpolarized laser to a protein: for a given C=O bond, the magnitude of the electric field is described by Equation (1), but the direction of the electric field vector is adapted and always aligned to the direction of the C=O bond. With this new strategy, the forces acting on the C and O atoms due to the electric field, $\vec{F}_C$ and $\vec{F}_O$, are now calculated by Equations (2) and (3):

$$\vec{F}_C = E * q_c * \hat{r}_{CO} \quad (2)$$

$$\vec{F}_O = E * q_o * \hat{r}_{CO} \quad (3)$$

Where, E is electric field magnitude described by Equation (1). q_c and q_o are partial charges of C and O atoms, respectively. The $\hat{r}_{CO}$ is the unit vector which points from C atom to O atom. The parameters for controlling the laser now include the C=O bond information, E_0 , t_0 , ω and σ . By scanning some parameters especially ω , we are able to study the infrared absorption of a covalent bond mimicking the effect of applying a polarized laser beam on a protein sample in experiment.

2.3. Laser-induced simulations for the infrared spectra investigation

To investigate the IR spectra of amide bonds in the protein structures, we have implemented aforementioned technique into the AMBER software package. For each considered protein structure, we performed the laser-induced simulations for 201 parameter sets with the laser frequencies, ω , varied from 1550 cm^{-1} to 1750 cm^{-1} at a step of 1 cm^{-1} and the other laser parameters being fixed at values of $E_0 = 0.18 \text{ V/\AA}$, $t_0 = 3 \text{ ps}$ and $\sigma = 1 \text{ ps}$. The total simulation time is 10 ps. For each parameter set/frequency, ten trajectories were carried out with ten different starting conformations taken from the 10-ns equilibrium simulation. We also investigated ten different simulated time steps varying from 0.01 fs to 2 fs for each trajectory. Thus, there are about 20,100 laser-induced simulations which were carried out for each protein model. Additionally, we carried out the laser-induced simulation to estimate the effect of C=O bond stretching force constant on the IR spectra (for the simulation time step of 0.1 fs only). Overall, the total number of laser-induced simulations performed in this study is more than 80,000.

2.4. Data analysis

The time-dependent fluctuation of a C=O bond was calculated by the following equation: $\Delta_i(t) = |d_i(t) - 1.23|$, where $d_i(t)$ is time evolution of the i th C=O bond and $1.23 \text{ \AA}$ is its equilibrium value. The time-dependent fluctuation of the C=O bond in a protein is the ensemble average, $\langle\Delta(t)\rangle$, which was averaged over all trajectories and all C=O bonds in the protein.

To estimate the IR radiation absorption of a protein system, we calculate the time-dependent difference between the laser-induced and the equilibrium energies: $SE(t) = SE_{laser}(t) - \langle SE_{eq} \rangle$. Here, $SE_{laser}(t)$ is the energy of the system in a laser-induced simulation at a given time, and $\langle SE_{eq} \rangle$ is the averaged equilibrium energy at the same temperature as in the laser-induced simulations.

3. Results and discussions

3.1. The IR radiation absorption of C=O bonds

The basic principle of IR spectroscopy is that when IR radiation passes through a sample, the absorption of IR radiation by the sample produces a change of the radiation intensity, and the change is then detected and translated into the absorption intensity of the sample. In our laser-induced simulation, the laser mimicked by the time-dependent electric field (Fig. 1a) induces the vibration of the targeted bonds, resulting in a change of the system energy. Therefore, the IR absorption can be described by the difference between the energies of the system before and after the laser pulse applied. Fig. 2 shows the time dependence of the energy differences and the fluctuation of the C=O bonds in the different protein systems during the laser-induced simulations for the given laser frequencies. The data shows that the energy difference and the bond fluctuation strongly depend on the laser frequency, and they are highly correlated. For example, for laser frequencies of 1630 cm^{-1} and 1720 cm^{-1} , the energy of all considered systems did not change (or changed very little). Their C=O fluctuation is very small, suggesting that the C=O bonds did not absorb the laser radiation energy at those frequencies. In contrast, for laser frequencies of 1676 cm^{-1} , 1679 cm^{-1} and 1691 cm^{-1} , the C=O fluctuation is clearly shown. The energy difference was significantly changed during the simulation time. The time dependence of the energy difference can be divided into three states. In the first state, from the beginning of the simulation to 2 ps, the energy mostly did not change. The second state, from 2 ps to 4 ps, the energy difference dramatically increased. In the last state, the energy difference slightly increased from 4 ps and reached a plateau value at about 5 ps. Therefore, for each laser frequency, the energy difference at the last 10th ps of the laser-induced simulation will be used as the protein absorption energy, which is equal to the absorption intensity in IR spectra.

3.2. Important input parameters for the laser-induced simulation

Since the mimic laser absorption in the simulation was caused by the C=O vibration, the temperature coupling parameters, which directly

affect the instantaneous velocities of the atoms, are important to a laser-induced simulation. In principle, when a system is coupled to the thermostat, the velocities of all atoms in the systems will be rescaled to get the system temperature close with the desired value. This implies that the effect of laser on the atom vibrations will diminish when the temperature coupling is applied. Therefore, in an ideal laser-induced simulation, the system is only coupled to the thermostat at the beginning of the simulation for its initial thermostat temperature, and the temperature coupling should not be applied during the rest of the simulation time. To satisfy the temperature coupling condition, in our laser-induced simulations, the flag *ntt*, which is used for the temperature regulation, was set to 3 and with *gamma ln* being set to 0.00001. This setting will rescale velocities of the atoms every 100,000 ps, which is much longer than the 10 ps of a laser-induced simulation.

In the laser-induced simulations, the external forces due to the electric field applied on the atoms of the C=O bonds are calculated by Equations (2) and (3). Because the partial charge of an atom is constant, the time dependence of the external electric force is similar to that of the external electric field described by Equation (1). In any MD simulation, the force acting on an atom and the velocity of the atom is calculated and updated for every simulated time step, namely dt . Therefore, the external electric force contributed by laser is discrete. If the laser period, $T = 1/(\omega)$, is too small, the ratio T/dt (resolution) will be significant small, the effect of the discreteness will become obvious, and the external electric field will be very different from the one truly applied in the simulations. The effect of dt on a mimic laser pulse, which has the frequency ω of 6000 cm^{-1} (corresponding to the period T of 5.56 fs) was shown in Fig. 3. It shows that when $dt = 0.02$ fs, the resolution T/dt is large enough to reproduce the periodicity of the laser, while the resolution will be too low for the laser expression if $dt = 2$ fs. In general, the higher the resolution is, the better the time-dependent pattern of the laser is reproduced, and the unpolarized laser effect better achieved. Although applying a small dt can increase the resolution, it also considerably slows the simulation. Thus, it is important to determine reasonable values of dt for laser-induced simulations at a given range of laser frequencies. In this study, the scanning frequencies were in the range of 1550 cm^{-1} to 1750 cm^{-1} . We have measured the amide I spectra with 11 values of dt (2, 1, 0.5, 0.4, 0.25, 0.2, 0.1, 0.08, 0.05, 0.02

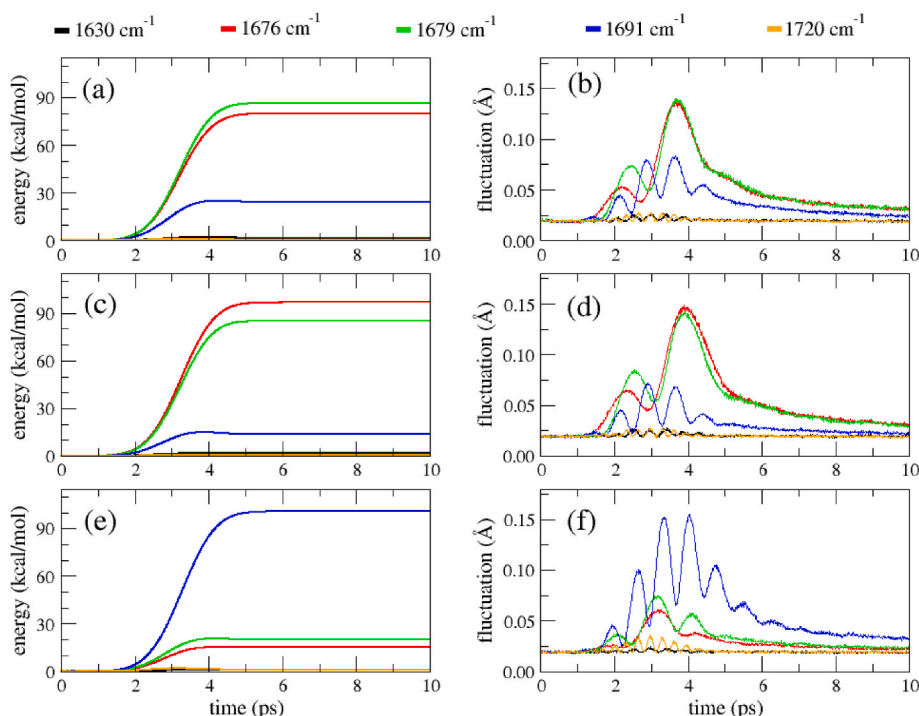


Fig. 2. The time dependence of the energy difference and the fluctuation of the C=O bond during the nonequilibrium MD simulations under an unpolarized laser for different protein structures: antiparallel β -sheet (a, b), parallel β -sheet (c, d) and α -helix (e, f). The energy difference of a system was calculated by dividing the difference between instantaneous energy and the energy at the beginning of the laser simulation by the number of targeted C=O bonds. The fluctuation was averaged over all the C=O bonds. The laser's parameters are $E_0 = 1.8$ V/nm, $t_0 = 3$ ps, $\sigma = 1$ ps and different frequencies, ω . The result with laser frequencies of 1630, 1676, 1679, 1691 and 1720 cm^{-1} are shown in black, red, green, blue and orange colors, respectively. The data was averaged over ten independent MD trajectories.

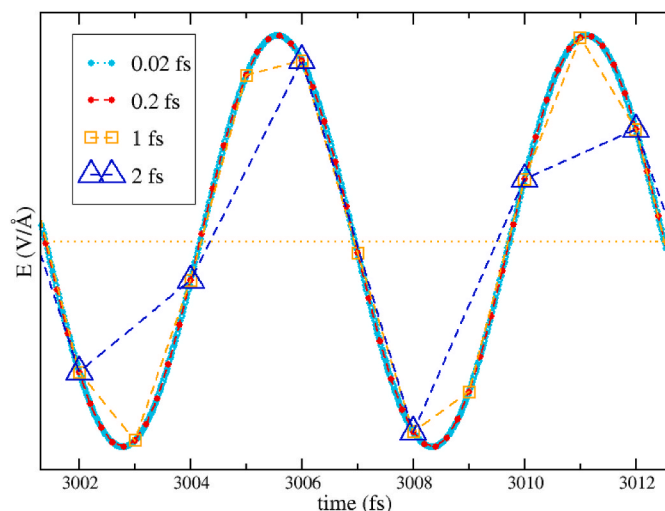


Fig. 3. View of the external electric field with frequency, $\omega = 6000 \text{ cm}^{-1}$ for different simulated time steps.

and 0.01 fs). Fig. 4a, b and 4c show the spectra of the different protein structures at $dt = 1, 0.4, 0.2, 0.1$ and 0.01 fs. The results indicate that when dt decreases, the peak of the spectrum first shifts to low frequencies, then the peak is unchanged as dt is smaller than 0.4 fs (Fig. 4d). In other words, a dt value smaller than 0.4 fs is good enough for studying the amide I spectra via laser-induced simulations. Later, we will discuss the results of simulations using $dt = 0.1$ fs.

Other tunable parameters include E_0 and σ , which control the energy of a laser pulse at a given frequency. If the laser energy is too high, the structure of the protein may dramatically change, discrepancies may also occur because the calculated spectrum is not for the true protein structure at all. If the laser energy is too low, the absorbed peak will be very low, which may not be distinguishable from the thermal noise. As the demonstration in Fig. 5, when $E_0 = 0.5 \text{ V/nm}$ and $\sigma = 1$ ps, the absorption energy is very small, while if $E_0 = 3.0 \text{ V/nm}$ and $\sigma = 3$ ps the

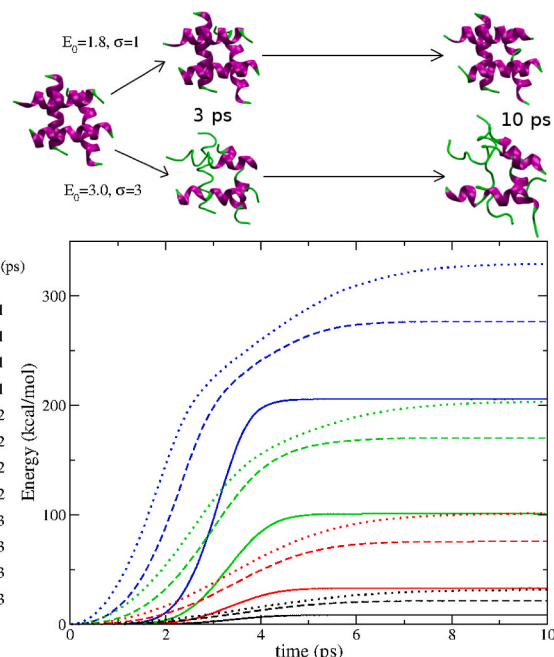


Fig. 5. The effects of the laser parameters, E_0 and σ , on the absorption energy and structure of the α -helix model (PDB ID: 6N8C). The laser frequency is 1692 cm^{-1} which gives maximum absorption energy in the model. The result is similar for the β -sheet models (data not shown).

protein structure significantly changed in comparison with the initial conformation. In this study, we used the laser parameters, $E_0 = 1.8 \text{ V/nm}$ and $\sigma = 1$ ps. It provides a good observation of the IR radiation absorption but has negligible effect on the structure of protein models.

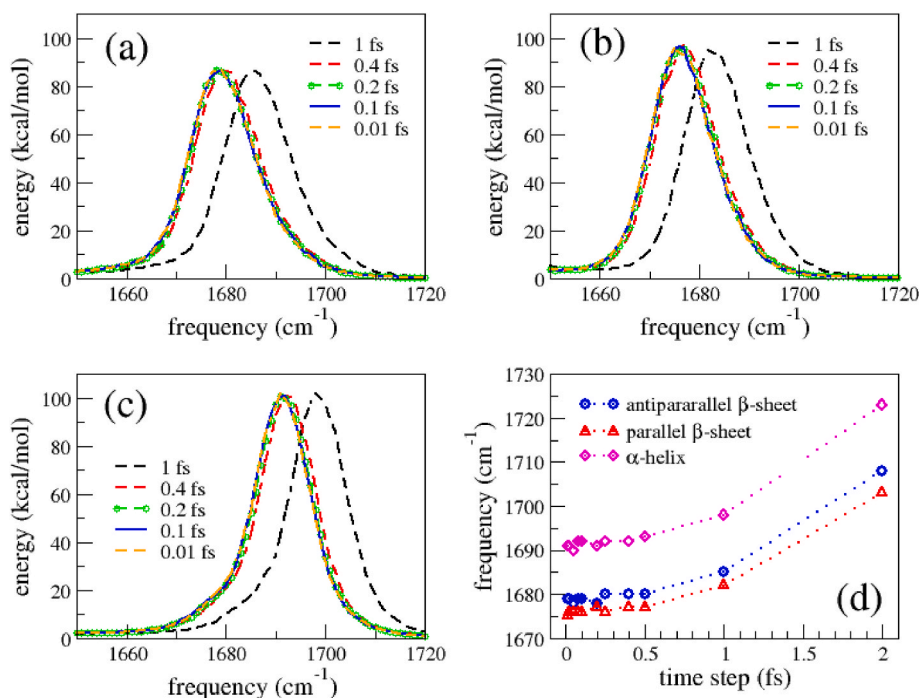


Fig. 4. The frequency dependence of carbonyl absorption energy in antiparallel β -sheet (a), parallel β -sheet (b) and α -helix (c) secondary structures at different simulated time steps. Panel (d) shows the dependence on simulation time step of the space frequency at which maximum absorption energy is observed.

3.3. Amide I spectra of proteins

3.3.1. Amide I spectra of secondary structural elements

One of the most important applications of a protein amide I IR spectra is to determine protein secondary structures. In this application, the correct identification of the absorbed peak positions, which correspond to the protein secondary structure elements, is crucial. Goormaghtigh et al. have collected and evaluated the assignment of amide I band positions to protein secondary structure based on experimental data and assignments in various publications [45,46]. Their study suggested that two IR regions, 1623–1641 cm^{-1} and 1674–1695 cm^{-1} , are assigned for β -sheet structures, and a region, 1648–1657 cm^{-1} , is assigned for α -helix structures. The average positions of those assignments are 1633 cm^{-1} and 1684 cm^{-1} for β -sheets, and 1654 cm^{-1} for α -helices. Those assignments have been widely applied in analyzing the Fourier-transform infrared (FTIR) spectroscopy [39]. In this study, we applied our specific method to calculate amide I spectra of some protein models which contain only one type of secondary structure elements, and we compared our calculated spectra with the average experimental data.

Fig. 6 shows the amide I IR spectra of the antiparallel β -sheet, parallel β -sheet and α -helix protein structures which were calculated by nonequilibrium MD simulations with the external electric field mimicking a laser pulse, with $E_0 = 1.8 \text{ V/nm}$, $\sigma = 1 \text{ ps}$ and $dt = 0.1 \text{ fs}$. For the β -sheet structures, our calculated spectra contain only one peak. The absorbed peak of the parallel β -sheet spectra is located at 1675 cm^{-1} , and the peak for the antiparallel case slightly shifted to a higher frequency, 1679 cm^{-1} . Those values are about 42–46 cm^{-1} higher than the experimental low frequency, and 5–9 cm^{-1} lower than the experimental high frequency, respectively [45,46]. Many previous studies agreed that the spectra of both the β -sheet structures contains a main peak with strong intensity at low frequency region (1617–1641 cm^{-1}), and the peak with much lower intensity at the high frequency region (above 1680 cm^{-1}) is a feature of antiparallel β -sheet structure only [47–50]. Thus, our observed peaks should belong to the low frequency region, but they were shifted 40 cm^{-1} toward the higher frequency direction in

comparison with the average experimental value. We did not observe the low intensity peak at high frequency region in our calculated spectra. For the α -helix structure, we observed a peak at 1692 cm^{-1} , which is again higher than the experimental observation about 40 cm^{-1} . Despite of the significant shift of our calculated spectrum peaks compared to the experimental values, the difference between α -helix and β -sheet peaks in our calculation (about 17 cm^{-1}) is comparable with the experimental finding (about 20 cm^{-1}). Additionally, our result, which showed a slightly difference between the spectra of parallel and antiparallel β -sheet structures, is consistent with a previous theoretical calculation [51].

3.3.2. Correction on the C=O bond force-constant of the FF14SB force field

Although our IR spectra were calculated by targeting only the vibration of the C=O bond, we still observed the distinguished peaks for different protein structures as experiments and other theoretical calculations did. However, the peak positions in our spectrum calculations were different from the ones assigned by experimental data. We have discussed how the IR spectra are affected by the laser and simulation parameters. We showed that the peak positions only depend on the simulation time step dt when dt is larger than 0.4 fs, and this dependency disappear when dt is smaller than 0.4 fs. Thus, we wondered whether the difference was caused by force field parameters. In order to investigate the factors that affect the absorbed frequencies of the C=O bonds in different protein conformations, let us consider a simple two-atom oscillator, which has the natural frequency (ω_0) given by the equation:

$$\omega_0 = (k_b/m_r)^{0.5} / (2\pi c) \quad (4)$$

$$\frac{1}{m_r} = \frac{1}{m_1} + \frac{1}{m_2} \quad (5)$$

where, the natural frequency ω_0 is in cm^{-1} , c is speed of light, k_b is the force constant between the two atoms, m_r is the reduced mass, m_1 is mass of the first atom, and m_2 is mass of the second atom. Therefore, two factors influencing the frequency are the force constant and the mass of the vibrating atoms. The frequency rises when the force constant increases and/or the reduced mass decreases, and vice versa. Applying this oscillator to the case of our spectrum calculation with all-atom MD simulation, k_b is the force constant of C=O bond, which was defined by the force field and did not change during the simulations, and m_1 and m_2 are effective masses of C and O atoms. Those effective masses mostly depend on the masses of the atoms, intermolecular interactions, intramolecular interaction and conformation of the protein, resulting in the dependence of IR absorbed peak on the protein structure. Note that, like the force constant of a bond, the mass of an atom was also constant and defined by force field. In the AMBER FF14SB [35] we used, the atomic mass was experimentally measured, while the force constant was usually calculated using quantum chemical methods. Therefore, we only consider the impact of the bond force constant on the protein spectra. As seen in Fig. 7, the absorbed peaks of both antiparallel β -sheet and α -helix spectra shifted to lower frequency when the C=O bond force constant was reduced. Interestingly, when $k_b = 515 \text{ kcal/mol/\AA}^2$ the absorbed peaks of the β -sheet and α -helix spectra are 1633 cm^{-1} and 1654 cm^{-1} , which are the same as the experimental averaged values assigned to the β -sheet and α -helix secondary structure. In other words, with the C=O bond stretching force constant, k_b being set to 515 $\text{kcal/mol/\AA}^2$, not the original one (570 $\text{kcal/mol/\AA}^2$) in the AMBER ff14SB force field, our method can exactly reproduce the amide I IR spectra of β -sheet and α -helix protein structures. Therefore, our method can also be applied to adjust the bond force constant parameters of a force field, which may not be ideal for all-atom MD simulations of biomolecules as those force field parameters were usually developed through the quantum mechanical calculation of prototype molecules in vacuum. Note that most of AMBER biomolecular force fields share the same bond force-constants of the

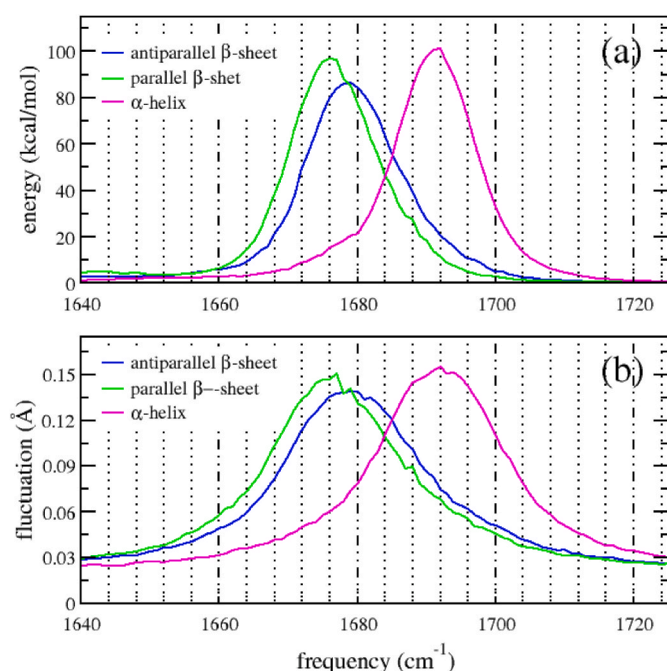


Fig. 6. The frequency dependence of the C=O bond absorption energy (a) and maximum fluctuation (b) in different protein secondary structures (PDB IDs: 2KIB, 1YJP, 6N8C). The data was obtained from the laser simulations with parameters, $E_0 = 1.8 \text{ V/nm}$, $\sigma = 1 \text{ ps}$ and simulation time step $dt = 0.1 \text{ fs}$.

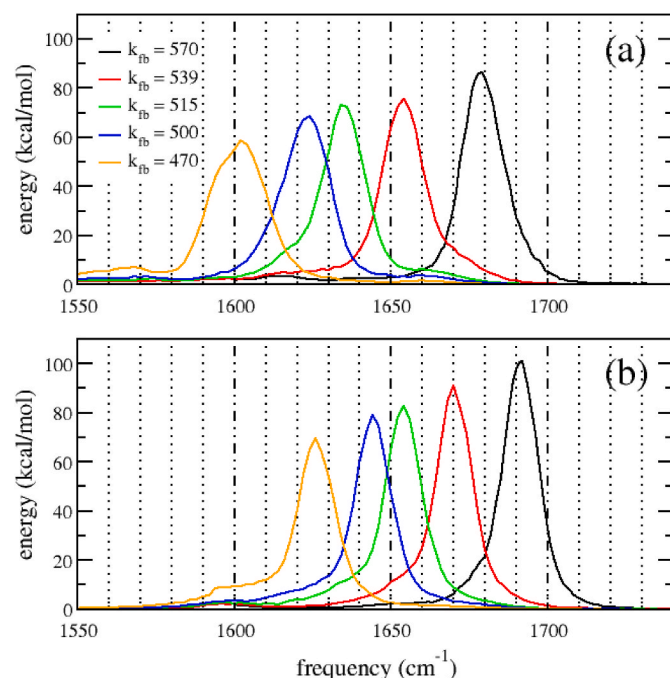


Fig. 7. The frequency dependence of the C=O bond absorption energy anti-parallel β -sheet (a) and α -helix (b) protein structures at different C=O force bond constants, k_{fb} . The $k_{fb} = 570$ is the original value of C=O force bond constant defined in the AMBER force field. The unit of the force bond constant is kcal/mol/Å².

bonds in proteins. However, other force fields such as GROMOS [52] or OPLS-AA [53] do not share the bond-force constants with AMBER force fields. Therefore, the correction of the bond-force constants based on IR spectra should be done for an individual force field or for a force field family.

3.3.3. Application of unpolarized laser all-atom MD simulations to calculate amide I spectra for representative protein secondary structures

In this section, we applied the mimic laser to calculate the IR spectra of additional proteins and compared the calculated spectra to the experimental data. It is noted that the force constant of C=O bond stretching parameter, k_{fb} , was set to 515 kcal/mol/Å². The first system considered was Tryptophan Zipper2 (Trpzip2), a β -hairpin structure. The experimental amide I spectrum of Trpzip2 was produced by Smith and Tokmakoff [18]. Our calculated spectrum agreed with the experiment in terms of the peak location that was around 1636 cm⁻¹ (Fig. 8). However, the calculated spectrum peak width was narrower than the experimental one, probably due to insufficient conformational sampling.

The second protein system we considered was the three peptides encompassing the transmembrane domain of Influenza A virus M2 protein H⁺ channel (Ser22-Leu46, SSDPLVVAASIIGILHLILWILDRL). The first peptide was not labeled (wild type peptide), the second one contains a ¹³C=¹⁶O isotopic labeled alanine residue at position 29 (¹³C=¹⁶O A29 peptide), and the third peptide contains a ¹³C=¹⁸O isotopic labeled glycine residue at position 34 (¹³C=¹⁸O G34 peptide). The amide spectra of those peptides were experimentally examined by Torres et al. for studying the vital role of the isotopic labels in relaying site-specific secondary structure and orientational information of proteins [54]. To calculate the peptide spectra, the initial structure of the wild-type peptide was taken from PDB databank (PDB ID: 1MP6), while the isotopic labeled residues were obtained by properly changing the mass of the related atom in the AMBER topological file of the wild-type peptide. All systems of the peptides were created and underwent the simulated process as described in the method section. As illustrated in

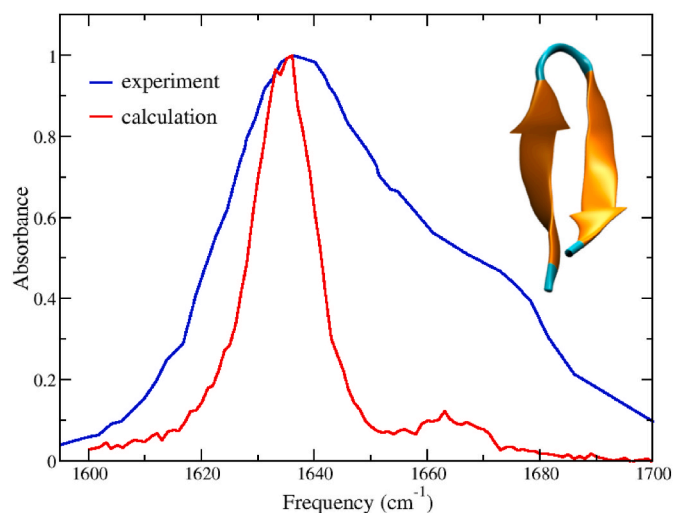


Fig. 8. IR spectra of the Trpzip2 peptide (PDB ID: 1LE1). The experimental data is from Ref.18. The calculated spectrum was normalized to the experimental value by matching the maximum absorption intensities of calculated and experimental spectra.

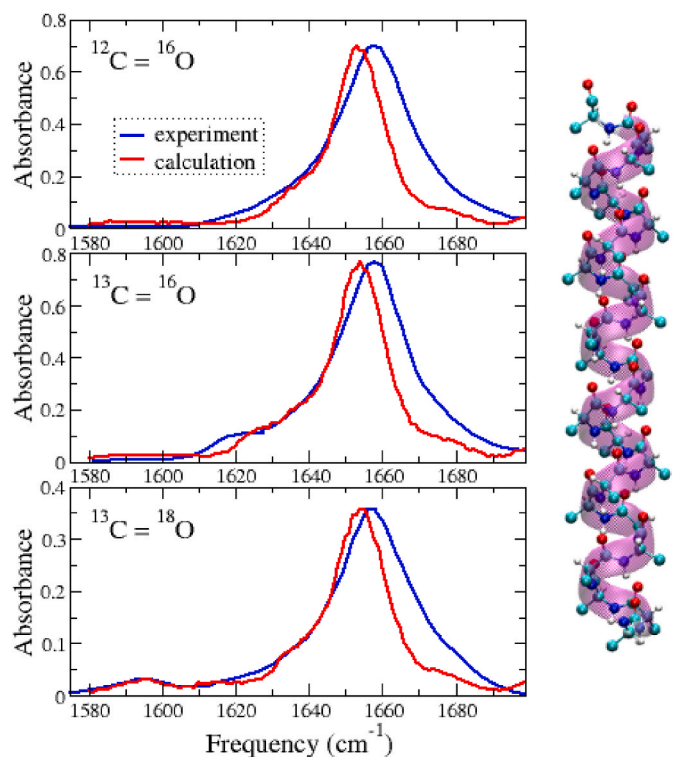


Fig. 9. IR spectra of the three different peptides encompassing the transmembrane domain of influenza A virus M2 H⁺ channel (PDB ID: 1MP6). The top, middle, and bottom panels are spectra from peptides that contain no isotopic label, a ¹³C=¹⁶O alanine residue at position 29 and ¹³C=¹⁸O glycine residue at position 34, respectively. The experimental data is from Ref.54. The calculated spectra were normalized to the experimental value by matching the maximum absorption intensities of calculated and experimental spectra.

Fig. 9, the calculated spectra were close to experimental ones. The calculated spectrum peaks, which related to ¹²C=¹⁶O bond absorbance, was at 1654 cm⁻¹, differing from experiment (1659 cm⁻¹) by only 5 cm⁻¹. The experimental weak peak at 1619 cm⁻¹ associated with the isotopic ¹³C=¹⁶O labeled A29 was not clearly observed in our calculation, while the experimental weak peak at 1595 cm⁻¹ in ¹³C=¹⁸O G34

peptide system was reproduced in our theoretical spectrum. This red-shifted peak can be qualitatively explained by the increasing of masses of labeled atoms (Equation.2). In other words, our method was able to describe the experimental red-shifted spectra caused by isotopic labeling if the effect is significant, such as the $^{13}\text{C}=^{18}\text{O}$ isotopic label.

Last, we evaluated our method using a much larger protein system, the receptor binding domain of the spike protein of SARS-CoV2. We calculated the spectrum maps of all the residues of the protein in two different states, the apo state and the binding state in complex with the ACE2 protein as shown in Fig. 10. The initial structure of the RBD-SARS-CoV2/ACE2 complex was taken from PDB databank (PDB ID: 6M0J) [55]. Two MD systems, the RBD-SARS-CoV2/ACE2 complex and the apo-structure of RBD-SARS-CoV2 (extracted from the complex), were prepared and simulated following the same protocols as described in the Method section. The residue-spectrum maps of RBD-SARS2 protein in the binding and unbound state were constructed and compared (Fig. 10). The difference of the maps reveals the residues which are possibly keys to the RBD-SARS2 and ACE2 binding. Interestingly, many residues, such as 73D, 74 E, 114G, 143 A, 144G, 161Q, 162S and 165F, which have the largest differences in vibrational frequencies between the two states are located in the Spike and ACE2 binding interface. Those residues could be hotspot residues for the protein-protein interactions. Interestingly, many of those residues are also recognized as hotspot residues through free energy decomposition analysis including 143 A and 161Q. Still the neighbors of some residues like 144G, 162S, 165F are hotspot residues in free energy decomposition analysis [56].

3.4. Limits of unpolarized laser all-atom MD simulations in calculating IR spectra

A number of methods for IR spectra calculation, remarkably

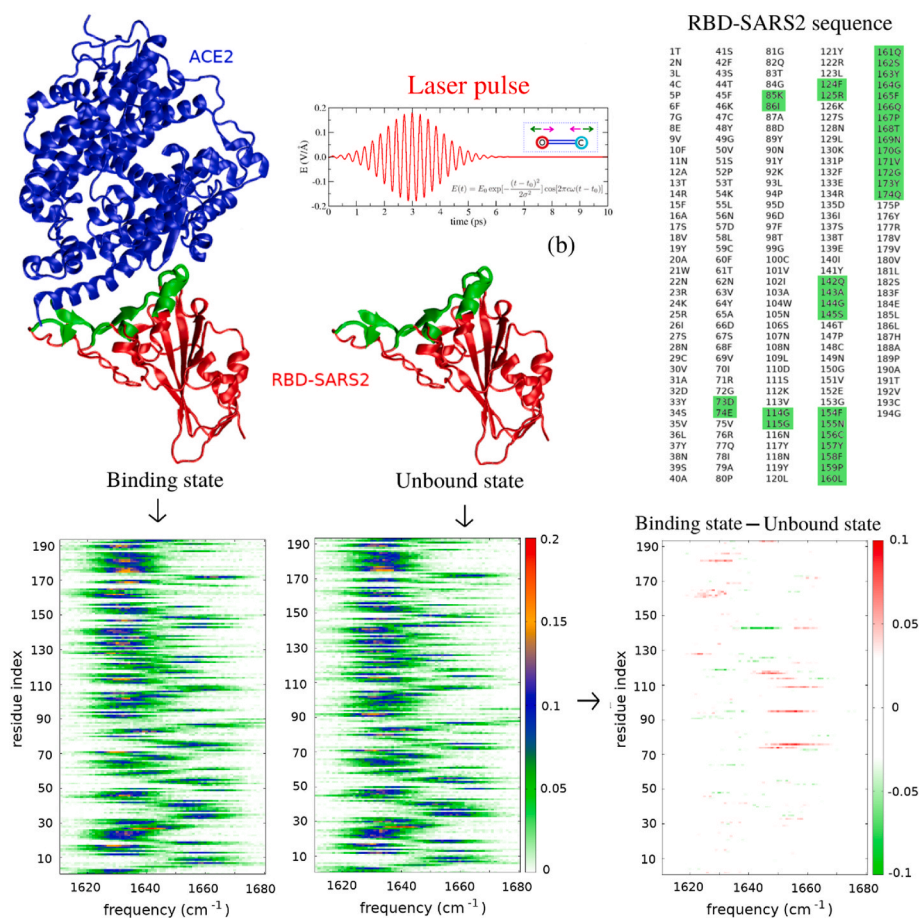


Fig. 10. Two conformational states of the SARS-CoV2 spike receptor-binding domain (RBD-SARS-CoV2), bound and unbound to ACE2 receptors. Top Panel: the RBD-SARS-CoV2 sequence and initial structures of the simulations. The RBD-SARS-CoV2 residues that are close and directly interact with ACE2 receptors are colored in green. Bottom Panels: the residue-spectrum maps of RBD-SARS-CoV2 in different states and the difference map. The color bars indicate the intensity of C=O bond fluctuation (in Å). The residues that showed largest spectrum differences include 73D, 74 E, 114G, 143 A, 144G, 161Q, 162S and 165F at the RBD-SARS-CoV2/ACE2 binding interface.

approach complements other methods. Nevertheless, the current theory which only considers C=O bond stretching vibration has limitations. First, although our method can reproduce the locations of the experimental absorption peaks well, the distribution widths, however, may be narrower than those found for the Tryptophan Zipper2 protein. Second, the anharmonicity effects were not completely considered because only implicit coupling between carbonyls was counted, while the contribution of the C–N bond absorption was neglected. We could observe the delocalization due to the natural coupling between C=O bond vibrations (Fig. S1). Of note, the carbonyls were linked by the backbone and NH–CO hydrogen bonds in the MD simulation, resulting in a coupling effect between the C=O bond vibrations. Our computational algorithm can be further improved to take the full anharmonicity effect into consideration.

3.5. Applications of unpolarized laser all-atom MD simulations in phototherapies

Besides calculating IR spectra for a biomolecular system, our method may also be applied to find optimal conditions of phototherapies. The radiation absorption of biomolecular systems has been studied and widely applied for many health therapies such as infrared and visible light therapies. The phototherapies have been medically used as the treatments for skin condition, cancer, sleep disorder, etc. [58–60]. Although the effectiveness of the phototherapies is evidently confirmed, the risks still exist [58]. Therefore, understanding the atomistic mechanism of the radiation absorption is vital for the therapy usage. It is still a grand challenge to fully reveal the atomistic picture of the phototherapies experimentally. Although computational approaches can describe biomolecular systems at the atomic level, a model/tool to mimic an unpolarized radiation in MD simulation is not available yet. Thus, the approach and implementation shown in this study will strongly support future studies to investigate the optimal conditions of phototherapies. Along this direction, we will focus on the development of the IR laser as a non-drug approach for the treatment of Alzheimer's disease (AD). This is important because despite intensive studies in both academia and the pharmaceutical industry for many decades, there is still a lack of drugs, including small molecules, vaccines and antibodies, that could arrest or reverse the cognitive impairment of AD. A structural feature of amyloid fibrils in AD is the cross β -sheet stacking conformation stabilized by the hydrogen bonding network involved in the C=O bonds. In our previous studies, we have used the polarized IR laser to excite the amide I modes of the fibrils, resulting in the dissociation of the fibril structures [40]. We find that the primary step in the dissociation process is due to the strong resonance between the fibril amide I vibrations and the tuned laser frequencies, and not just the deposited infrared thermal energy. Through a combination of experiments and simulations on disassembling of amyloid fibril of the 7-residue peptide GNNQQNY from yeast prion protein, we revealed the dissociation mechanism in an unprecedented detail [61]. Although the results are encouraging, the use of the polarized IR laser to target the amide I modes required high laser intensity in simulations compared to that used in experiments, such that the dissociation can be observed within reasonable simulation timescales. With the new unpolarized IR laser simulation method developed in this work, we will be able to directly target the C=O bonds. As a result, the required laser intensity is much lower, making it possible to make meaningful comparison between the simulation results and experimental observations.

4. Conclusions

In conclusion, we have developed a comprehensive method for calculating the IR spectrum of the amide I modes of a single protein or amyloid fibril via all-atom MD simulations. The method involves mimicking an unpolarized laser that applies the energy only along the C=O bonds, equivalent to the effect of applying an IR polarized laser on

an ensemble of identical proteins or a single molecule at a certain direction. The method has been implemented in the GPU version of the molecular dynamics simulation engine (pmemd.cuda) in AMBER simulation package. We have tested the performance of the method with various single proteins which have both β -sheet and α -helix structures and amyloid fibrils. The simulated IR spectra are roughly about 40 cm^{-1} blueshifted compared to experimental data. This is likely due to the inaccuracy of classical force fields which are not well-parametrized for high frequency bond stretching modes such as that for C=O bond stretching. We proposed a simple remediation by rescaling the force constant of the C=O bond taken from the employed force field. With this correction, we can produce not only the experimental spectra but also the difference between the peak positions of the α -helix and β -sheet structures. More encouragingly, the simulation method can also reveal small differences between the IR spectra of the parallel and anti-parallel β -sheet structures of proteins, consistent with the previous high-level *ab initio* calculations. We believe that the method will be useful for obtaining the amide I IR spectra for various types of proteins, given its simplicity in theory, efficiency in computation, and accuracy in the result.

Our simulation method can also be combined with the IR pump-probe experimental technique [62] to track conformational changes in proteins and amyloid fibrils. In this experimental technique, an ultra-short IR laser pulse is used to excite the system to generate a non-equilibrium state, and another pulse is used to monitor the time evolution of the IR spectrum. In our simulation, the unpolarized laser excites the C=O bonds, resulting in conformational changes in the molecule. By calculating the IR spectrum for each conformation during the simulation, we can obtain the time evolution of the spectra. Comparing the simulated spectra with experimental counterparts allows us to identify the conformations from the simulation trajectory whose IR spectra closely match those obtained from the experiment. We plan to develop this simulation/experimental approach to study the molecular mechanism of the IR laser-induced dissociation of amyloid oligomers and fibrils. This could contribute to the development of non-drug approaches for treating Alzheimer's disease.

Associated content

Supporting information (SI) shows the details of the simulated systems (Table S1) and the spectra of dimer anti-parallel β -sheet when the laser was employed to two residues and individual residue (Fig. S1).

Author contributions

V.H.M and J.W designed the project. V.H.M performed the simulations and analyzed data. V.H.M, X. H, P.H.N, C.S, C.R, X.Q.X, and J.W discussed and wrote the paper.

Notes

The authors declare that they have no competing financial interests.

Acknowledgements

This work was supported by the funds from the National Institutes of Health (NIH) NIH R01GM147673, NIH K25AG070277, and the National Science Foundation (NSF) 1955260. The authors thank the computing resources provided by the Center for Research Computing (CRC) at University of Pittsburgh.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.combiomed.2023.106902>.

References

- [1] A. Barth, Infrared spectroscopy of proteins, *Biochim. Biophys. Acta* 1767 (2007) 1073–1101.
- [2] E. Kauffmann, N.C. Darnton, R.H. Austin, C. Batt, K. Gerwert, Lifetimes of intermediates in the β -sheet to α -helix transitions of β -lactoglobulin by using a diffusional IR mixer, *Proc. Natl. Acad. Sci. USA* 98 (2001) 6646–6649.
- [3] H. Fabian, H.H. Mantsch, C.P. Schultz, Two-dimensional IR correlation spectroscopy: sequential events in the unfolding process of the lambda Cro-V55C repressor protein, *Proc. Natl. Acad. Sci. U.S.A.* 96 (1999) 13153–13158.
- [4] D. Reinstadler, H. Fabian, D. Nauman, New structural insights into the refolding of ribonuclease T1 as seen by time-resolved Fourier-transform infrared spectroscopy, *Proteins* 34 (1999) 303–316.
- [5] C.P. Schultz, Illuminating folding intermediates, *Nat. Struct. Biol.* 7 (2000) 7–10.
- [6] A. Troullier, D. Reinsatdler, Y. Dupont, D. Naumann, V. Forge, Transient non-native secondary structures during the refolding of α -lactalbumin detected by infrared spectroscopy, *Nat. Struct. Biol.* 7 (2000) 78–86.
- [7] A.P. Ramajo, S.A. Petty, A. Starzyk, M. Volk, The α -helix folds more rapidly at the C-terminus than at the N-terminus, *J. Am. Chem. Soc.* 127 (2005), 3784–13785.
- [8] S. Krimm, J. Bandekar, Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins, *Protein Chem* 38 (1986) 181–364.
- [9] H. Henry, D.C. Mantsch, *Infrared Spectroscopy of Biomolecules*, Wiley, New York, 1996.
- [10] D.M. Byler, H. Susi, Examination of the secondary structure of proteins by deconvolved FTIR spectra, *Biopolymers* 25 (1986) 469–487.
- [11] A. Barth, C. Zscherp, What vibrations tell us about proteins, *Q. Rev. Biophys.* 35 (2002) 369–430.
- [12] M. Reppert, A. Tokmakoff, Computational amide I 2D IR spectroscopy as a probe of protein structure and dynamics, *Annu. Rev. Phys. Chem.* 67 (2016) 359–386.
- [13] C.M. Davis, L. Zanetti-Polzi, M. Gruebele, A. Amadei, R.B. Dyer, I. Daidone, A quantitative connection of experimental and simulated folding landscapes by vibrational spectroscopy, *Chem. Sci.* 9 (2018) 9002.
- [14] H. Torii, M. Tasumi, Model calculations on the amide-I infrared bands of globular proteins, *J. Chem. Phys.* 96 (1992) 3379–3387.
- [15] S. Lee, S. Krimm, Ab initio-based vibrational analysis of α -poly(L-alanine), *Biopolymers* 46 (1998) 283–317.
- [16] J.W. Brauner, C.R. Flach, R. Mendelsohn, A quantitative reconstruction of the amide I contour in the IR spectra of globular proteins: from structure to spectrum, *J. Am. Chem. Soc.* 127 (2005) 100–109.
- [17] T. la Cour Jansen, A.G. Dijkstra, Modeling the amide I bands of small peptides, *J. Chem. Phys.* 125 (2006) 44312–44319.
- [18] A.W. Smith, A. Tokmakoff, Amide I two-dimensional infrared spectroscopy of β -hairpin peptides, *J. Chem. Phys.* 126 (2007), 045109.
- [19] S. Ham, S. Cha, J.-H. Choi, M. Cho, Amide I modes of tripeptides: Hessian matrix reconstruction and isotope effects, *J. Chem. Phys.* 119 (2003) 1451.
- [20] S. Hahn, S. Ham, M. Cho, Simulation studies of amide I IR absorption and two-dimensional IR spectra of β hairpins in liquid water, *J. Phys. Chem. B* 109 (2005) 11789–11801.
- [21] J.-H. Choi, H. Lee, K.-K. Lee, S. Hahn, M. Cho, Computational spectroscopy of ubiquitin: comparison between theory and experiments, *J. Chem. Phys.* 126 (2007), 045102.
- [22] L. Wang, C.T. Middleton, M.T. Zanni, J.L. Skinner, Development and validation of transferable amide I vibrational frequency maps for peptides, *J. Phys. Chem. B* 115 (2011) 3713–3724.
- [23] L.Z. Polzi, I. Daidone, M. Anselmi, G. Carchini, A.D. Nola, A. Amadei, Analysis of infrared spectra of β -hairpin peptides as derived from molecular dynamics simulations, *J. Phys. Chem. B* 115 (2011) 11872–11878.
- [24] R.-J. Xue, A. Grofe, H. Yin, Z. Qu, J. Gao, H. Li, Perturbation approach for computing infrared spectra of the local mode of probe molecules, *J. Chem. Theor. Comput.* 13 (2017) 191–201.
- [25] C.M. Olson, A. Grofe, C.J. Huber, I.C. Spector, J. Gao, A.M. Massari, Enhanced vibrational solvatochromism and spectral diffusion by electron rich substituents on small molecule silanes, *J. Chem. Phys.* 147 (2017), 124302.
- [26] H. Yin, H. Li, A. Grofe, J. Gao, Active-site heterogeneity of lactate dehydrogenase, *ACS Catal.* 9 (2019) 4236–4246.
- [27] Z. Ganim, H.S. Chung, A.W. Smith, L.P. DeFlores, K.C. Jones, A. Tokmakoff, Amide I two-dimensional infrared spectroscopy of proteins, *Acc. Chem. Res.* 41 (2008) 432–441.
- [28] D.A. Case, et al., AMBER 19, University of California, San Francisco, 2019.
- [29] J. Nielsen, M. Bjerring, M. Jeppesen, R. Pedersen, J. Pedersen, K. Hein, T. Vosegaard, T. Skrydstrup, D. Otzen, N. Nielsen, Unique identification of supramolecular structures in amyloid fibrils by solid-state NMR spectroscopy, *Angew. Chem., Int. Ed. Engl.* 48 (2009) 2118–2121.
- [30] R. Nelson, M.R. Sawaya, M. Balbirnie, A.O. Madsen, C. Riek, R. Grothe, D. Eisenberg, Structure of the cross- β spine of amyloid-like fibrils, *Nature* 435 (2005) 773–778.
- [31] A.G. Cochran, N.J. Skelton, M.A. Starovasnik, Tryptophan zippers: stable, monomeric β -hairpins, *Proc. Natl. Acad. Sci. USA* 98 (2001) 5578–5583.
- [32] S.A. Kotler, V. Tugarinov, T. Schmidt, A. Ceccon, D.S. Libich, R. Ghirlando, C. D. Schwieters, G.M. Clore, Probing initial transient oligomerization events facilitating Huntingtin fibril nucleation at atomic resolution by relaxation-based NMR, *Proc. Natl. Acad. Sci. USA* 116 (2019), 3562–357.
- [33] J. Wang, S. Kim, F. Kovacs, T.A. Cross, Structure of the transmembrane region of the M2 protein H⁺ channel, *Protein Sci.* 10 (2001) 2241–2250.
- [34] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* 779 (1983) 926–935.
- [35] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling, ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB, *J. Chem. Theor. Comput.* 11 (2015) 3696–3713.
- [36] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, *J. Chem. Phys.* 103 (1995) 8577–8593.
- [37] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. Dinola, J.R. Haak, Molecular dynamics with coupling to an external bath, *J. Chem. Phys.* 81 (1984) 3684–3690.
- [38] T.R. Forester, W. Smith, SHAKE, rattle, and roll: efficient constraint algorithms for linked rigid bodies, *J. Comput. Chem.* 19 (1998) 102–111.
- [39] E. Lindahl, B. Hess, D. van der Spoel, Gromacs 3.0: a package for molecular simulation and trajectory analysis, *Molecular modeling annual* 7 (2001) 306–317.
- [40] V.H. Man, P. Derreumaux, M.S. Li, C. Roland, C. Sagui, P.H. Nguyen, Picosecond dissociation of amyloid fibrils with infrared laser: a nonequilibrium simulation study, *J. Chem. Phys.* 143 (2015), 155101.
- [41] V.H. Man, N.-T. Van-Oanh, P. Derreumaux, M.S. Li, C. Roland, C. Sagui, P. H. Nguyen, Picosecond infrared laser-induced all-atom nonequilibrium molecular dynamics simulation of dissociation of viruses, *Phys. Chem. Chem. Phys.* 18 (2016) 11951–11958.
- [42] V.H. Man, P.M. Truong, P. Derreumaux, M.S. Li, C. Roland, C. Sagui, P.H. Nguyen, Picosecond melting of peptide nanotubes with infrared laser: a nonequilibrium simulation study, *Phys. Chem. Chem. Phys.* 17 (2015) 27275–27280.
- [43] V.H. Man, F. Pan, C. Sagui, C. Roland, Comparative melting and healing of B-DNA and Z-DNA by an infrared laser pulse, *J. Chem. Phys.* 144 (2016), 145101.
- [44] D. Domin, V.H. Man, N.-T. Van-Oanh, J. Wang, T. Kawasaki, P. Derreumaux, P. H. Nguyen, Breaking down cellulose fibrils with a mid-infrared laser, *Cellulose* 25 (2018) 5553–5568.
- [45] E. Goormaghtigh, V. Cabiaux, Lean marie ruysschaert, determination of soluble and membrane protein structure by fourier transform infrared spectroscopy III. Secondary structures, *Subcell. Biochem.* 23 (1994) 405–450.
- [46] S. Sukumaran, Protein Secondary Structure Elucidation Using FTIR Spectroscopy, Thermo Fisher Scientific, USA, 2017.
- [47] G. Zandomenighi, M.R. Krebs, M.G. McCammon, M. Fändrich, FTIR reveals structural differences between native β -sheet proteins and amyloid fibrils, *Protein Sci.* 13 (2004) 3314–3321.
- [48] L. Fu, G. Ma, E.C.Y. Yan, In situ misfolding of human islet amyloid polypeptide at interfaces probed by vibrational sum frequency generation, *J. Am. Chem. Soc.* 132 (2010) 5405–5412.
- [49] J. Kubelka, T.A. Keiderling, Differentiation of β -sheet-forming structures: ab initio-based simulations of IR absorption and vibrational CD for model peptide and protein β -sheets, *J. Am. Chem. Soc.* 123 (2011) 12048–12058.
- [50] Y. Zou, Y. Li, W. Hao, X. Hu, G. Ma, Parallel β -sheet fibril and antiparallel β -sheet oligomer: new insights into amyloid formation of hen egg white lysozyme under heat and acidic condition from FTIR spectroscopy, *J. Phys. Chem. B* 117 (2013) 4003–4013.
- [51] S.J. Roeters, A. Iyer, G. Pletikapic, V. Kogan, V. Subramaniam, S. Woutersen, Evidence for intramolecular antiparallel beta-sheet structure in alpha-synuclein fibrils from a combination of two-dimensional infrared spectroscopy and atomic force microscopy, *Sci. Rep.* 7 (2017) 1.
- [52] J. Hermans, H.J.C. Berendsen, W.F.V. Gunsteren, J.P.M. Postma, A consistent empirical potential for water–protein interactions, *Biopolymers* 23 (1984) 1513–1518.
- [53] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, W.L. Jorgensen, Evaluation and reparameterization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides, *J. Phys. Chem. B* 105 (2001) 6474–6487.
- [54] J. Torres, A. Kukol, J.M. Goodman, I.T. Arkin, Site-specific examination of secondary structure and orientation determination in membrane proteins: the peptidic $^{13}\text{C},^{18}\text{O}$ group as a novel infrared probe, *Biopolymers* 59 (2001) 396–401.
- [55] J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Zhang, X. Wang, Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor, *Nature* 581 (2020) 215–220.
- [56] Y. Zhang, X. He, J. Zhai, B. Ji, V.H. Man, J. Wang, In silico binding profile characterization of SARS-CoV-2 spike protein and its mutants bound to human ACE2 receptor, *Briefings Bioinf.* 22 (6) (2021) bbab188.
- [57] Y. Jiang, L. Lu Wang, Development of vibrational frequency maps for nucleobases, *J. Phys. Chem. B* 123 (27) (2019) 5791–5804.
- [58] E. Lee, J. Koo, T. Berger, UVB phototherapy and skin cancer risk: a review of the literature, *Int. J. Dermatol.* 44 (5) (2014) 355–360.
- [59] G.L. Willis, C. Moore, S.M. Armstrong, A historical justification for and retrospective analysis of the systematic application of light therapy in Parkinson's disease, *Rev. Neurosci.* 23 (2) (2012) 199–226, 2012.
- [60] Nicholas Hanford, Mariana Figueiro, Light therapy and alzheimer's disease and related dementia: past, present, and future, *J. Alzheim. Dis.* 33 (4) (2013) 913–922.
- [61] T. Kawasaki, V.H. Viet Hoang Man, Y. Sugimoto, N. Sugiyama, H. Yamamoto, K. Tsukiyama, J. Wang, P. Derreumaux, P.H. Nguyen, Infrared laser-induced amyloid fibril dissociation: a joint experimental/theoretical study on the GNNQQNY peptide, *J. Phys. Chem. B* 124 (29) (2020) 6266–6277.
- [62] S.H. Shim, D.B. Strasfeld, Y.L. Ling, M.T. Zanni, Automated 2D IR spectroscopy using a mid-IR pulse shaper and application of this technology to the human islet amyloid polypeptide, *Proc. Natl. Acad. Sci. USA* 104 (36) (2007) 14197–14202.