Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Infrared spectra of the SARS-CoV-2 spike receptor-binding domain: Molecular dynamics simulations

Jianbin Du, Ning Yao, Xiangyun Ma, Huijie Wang, Qifeng Li, Zhifang Feng

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has spread around the world rapidly, which seriously threatens to human health and safety. The rapid detection of the virus in the early stage is very important to prevent the cross infection and transmission. It is also a key link in the post-treatment examination. This paper has explored the infrared (IR) spectra of spike protein receptor-binding domain (RBD) for SARS-CoV-2 using molecular dynamics simulations, and the absorption bands are assigned. The calculated IR spectra of water and insulin are compared with that measured in the related literatures. The results showed that O–H stretching vibration generated a strong absorption band located around 3591 cm\(^{-1}\), the oscillator strength of 310 K is slightly higher than that at 298 K. The absorption peaks have a small red shift or blue shift with the change of temperature. As a theoretical basis for the optical detection of SARS-CoV-2 virus, this work will play a positive role in promoting the development of new virus detection technology.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is spreading rapidly all over the world [1,2]. It has caused a pandemic of coronavirus disease 2019 (COVID-19) in a short period of time by its highly contagious in humans [3]. Until 5 March 2022, the World Health Organization (WHO) has reported that more than 400 million people had been infected with SARS-CoV-2, resulting in more than 5 million deaths [4].

The diameter of SARS-CoV-2 is about 50–200 nm, which contains four structural proteins, namely S (spike) protein, E (envelope) protein, M (membrane) protein and N (nucleocapsid) protein [2,3]. The high affinity combination of receptor binding domain (RBD) of S protein with the angiotensin converting enzyme 2 (ACE2) in human body [3,5] is the key role to the strong infection of the virus. RBD is the crux of the SARS-CoV-2 infection, which is the important research problem of SARS-CoV-2. Du et al. studied the reaction between six antibodies of SARS-CoV and RBD of SARS-CoV-2 [6]. Sun and his colleagues discussed the RBD key residues in the interaction between SARS-CoV-2 and ACE2 [7]. Su et al. compared the binding of RBD with ACE2 from human, pets, farm animals, and putative intermediate host [8]. Several research groups analyzed the crystal structure of RBD [5,9,10], which provides a basis for further research on SARS-CoV-2. Kräi and Han designed SARS-CoV-2 peptide inhibitors based on human ACE2 by molecular dynamics simulations [3].

As SARS-CoV-2 is highly infectious among humans, early detection, early quarantine and early medical treatment are main measures to prevent the spread of the epidemic. Rapid and accurate virus detection has become a significant demand of the battle against SARS-CoV-2. For emerging pathogens, real-time reverse transcription-polymerase chain reaction (RT-PCR) is the main diagnostic method. Currently, real-time RT-PCR is used to detect SARS-CoV-2 [11], but including the preparation of virus RNA, real-time RT-PCR molecular diagnosis needs at least 3 h [11]. In addition, the preparation of RNA can affect the accuracy of diagnosis. Therefore, it is vital to develop a rapid, convenient and sensitive detection technology for SARS-CoV-2. IR spectroscopic detection is a powerful method for qualitative and quantitative analysis of substances, based on the characteristic molecular absorption fingerprints [12–14]. It has the rather attractive advantage of fast, multi-parameter, high-specificity and non-invasive measurement, having been widely used in numerous fields, including biomedical diagnosis, pharmacy, biotechnology, chemical detection, material analysis and environmental monitoring [15–19]. The simulations of IR spectra for large systems have been studied [20]. In this paper, we used MD simulations to explore...
the IR spectra of RBD, which has a positive role in promoting the application of spectral technology in the detection of SARS-CoV-2.

2. Methodology and model

2.1. Methodology

The total dipole moment $M(t)$ of system at the time $t$ can be calculated by the equation (1).

$$M(t) = \sum_{i=1}^{n} q_i \vec{r}_i$$  \hspace{1cm} (1)

where $n$ is the total number of atoms, $q_i$ and $\vec{r}_i$ are the charge and position vector of the $i$th atom. The IR spectra for a system in equilibrium can be calculated using linear response theory by Fourier transforming the total dipole moment auto-correlation function [21].

$$I_c(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt e^{-i\omega t} < \overline{M(0)M(t)} >$$  \hspace{1cm} (2)

where $I_c(\omega)$ is the classical spectral intensity, $\omega$ is the angular frequency of IR radiation. Using Parseval theorem, Wiener-Khintchine theorem and Fourier transforms, the spectral intensity can be transformed into the equation (3) [22,23].

$$I_c(\omega) = \frac{1}{2\pi\alpha^2} \int_{-\infty}^{\infty} dt e^{-i\omega t} \left\langle \frac{d\overline{M(0)} d\overline{M(t)}}{dt} \right\rangle$$

$$= \frac{1}{2\pi\alpha^2} \int_{-\infty}^{\infty} dt e^{-i\omega t} \left( \sum_{j=1}^{n} q_j \vec{v}_j(0) \cdot \sum_{j=1}^{n} q_j \vec{v}_j(t) \right)$$  \hspace{1cm} (3)

where $\vec{v}_j(t)$ refers to the speed of the $j$th atom at time $t$.

For the classical spectral intensity $I_c(\omega)$, several quantum correction methods have been suggested in the literature [24,25]. In this paper, the quantum correction factor $Q_{th}$ based on the “harmonic approximation” [24,25] is applied to the calculation of spectral intensity $I_c(\omega)$, which is effective for H-bond system [24]. Based on this method, the spectral intensity can be calculated by the equation (4).

$$I(\omega) = Q_{th} I_c(\omega)$$

$$= (\frac{\beta \hbar \omega}{1 - \exp(-\beta \hbar \omega)}) \frac{1}{2\pi\alpha^2} \int_{-\infty}^{\infty} dt e^{-i\omega t} \left( \sum_{j=1}^{n} q_j \vec{v}_j(0) \cdot \sum_{j=1}^{n} q_j \vec{v}_j(t) \right)$$  \hspace{1cm} (4)

where $\beta = \frac{1}{kT}$, $k$ is the Boltzmann constant, and $T$ is the thermodynamic temperature. Finally, the IR spectra of the protein can be obtained by the charges $q_i$ and velocities $\vec{v}_i$ of the atoms.

2.2. Model

Fig. 1 showed the simulation system. The crystal structure of RBD and ACE2 complex was obtained from the protein data bank (PDB: 6M0J). RBD was separated from the crystal structure 6M0J and its residues were completed by homologous modeling. Then, the protein was added to the simulation box, which was at least 10 Å away from the edge of the box. Finally, 12,494 water molecules and 3Cl atoms were filled into the box, which were randomly distributed throughout the system. The purpose of adding Cl atoms was to maintain the electrical neutrality of the system.

GROMACS is a versatile package to perform molecular dynamics. It is primarily designed for biochemical molecules like proteins, lipids and nucleic acids. It is not only widely used but also powerful, supporting many commonly used force fields and algorithm. All simulations were performed in the GROMACS 5.1.4 software package [26] in combination with the CHARMM36 force field [28,29], and SPC/E water model [28,29]. The periodic boundary conditions (PBC) were employed on all three dimensions. The cutoff values of Lennard-Jones (LJ) and Coulomb (Coul) interaction were 10 Å. The long-range electrostatic interaction was calculated by particle mesh Ewald (PME) method [30]. Short range interactions used a cutoff of 12 Å with a dispersion correction of both energy and pressure. All the chemical bonds in the system used linear unconstrained algorithm, and the leapfrog algorithm was used to integrate Newton’s motion equation [31].

IR spectra were computed by performing constant-pressure (NPT) MD simulations with target temperatures of 298 K and 310 K at 1 bar pressure. In order to obtain the initial equilibrium structure, the steepest descent method was used to minimize the energy of the system during 50,000 steps. Then, to relax the solvent molecules with position restraints on the backbone atoms of RBD, constant-volume (NVT) MD simulations were performed for 1 ns. Next, the system was pre-equilibrated with 30 ns MD runs in the NPT ensemble. In the simulations, the temperatures of the system were kept constant using the V-rescale thermostat and the Parrinello-Rahman method was used to control the system pressure. Linear constraint algorithm was used for all chemical bonds in all pre-equilibrium simulations. The root mean square deviation (RMSD), root mean square fluctuation (RMSF) and energy changes were collected during the NPT pre-equilibrium simulations, as shown in the Figs. 4, 5 and 6. Finally, the chemical bond linear constraint algorithm was stopped. The MD simulations were performed in the NPT ensemble for 140 ps with data collected at every 4 fs [31]. From the trajectory of the system, it can be seen that the structure of RBD remained relatively stable during the 140 ps spectral simulations. The temperatures of the system were kept constant using the Nosé-Hoover thermostat.

The spectral resolution $\Delta V$ is determined by the time. According to the sampling theorem, $\Delta V > 2/T$, $T$ are the total time. If we want to get a higher resolution spectrum, we need to increase the simulation time. If the required spectral resolution is 1 cm$^{-1}$, the total time must be greater than 66 ps. Because most of the correlation functions can only get half of the total valid data, so the total time is about 140 ps. The maximum frequency of the spectrum is determined by the time interval $f_{max} < 1/2\Delta t$, $\Delta t$ is the time interval. We calculated the IR spectra in the range of 400–4000 cm$^{-1}$, so as long as the time interval was 4 fs.

In addition, the force constants of chemical bonds are given in the
CHARMM36 force field. The stretching vibration frequency of chemical bond can be calculated according to.

\[
\tilde{\nu} = \frac{1}{2\pi c} \sqrt{\frac{\kappa}{\mu}}
\]

where \(\tilde{\nu}\) is the stretching vibration frequency, \(\kappa\) is the force constant, \(c\) is the speed of light, and \(\mu\) is the reduced mass.

In order to confirm the accuracy of the simulation results, we used the same method to calculate the IR spectra of water and insulin, as shown in the Fig. 2 and Fig. 3. 852 water molecules were filled into the simulation box, which was used for simulating the infrared spectrum of water. In the simulation model of insulin, the insulin adopted 3INC crystal structure, 5729 water molecules and 4Na\(^+\) were filled into the box. Both simulations used the combination with the CHARMM36 force field and SPC/E water model. In Fig. 2, the absorption bands around 700 cm\(^{-1}\) are assigned to O–H rocking vibration mode. It is greatly affected by the environment, and the measured peak position is slightly red shifted compared with the calculated peak position. The absorption bands around 1600 cm\(^{-1}\) are generated by O–H bending vibration, the simulated strength is higher than the measured value because the experimental sample is gaseous water and the simulation model is liquid water. The absorption peaks near 3400 cm\(^{-1}\) are assigned to O–H stretching vibration. The simulation results show that there are two sharp peaks, while the spectra in the literature have only one absorption peak, because the measured spectra are fitted. In Fig. 3, comparing the two results, it is found that the peaks of the measured spectrum are wider. This is because the experimental sample is solid insulin, while the MD simulation model is an insulin molecule. The transmittance of the absorption bands around 700 cm\(^{-1}\) are obviously different, which is due to the sensitivity of the bands to the external environment such as solvent polarity. Multiple peaks appeared near 1600 cm\(^{-1}\) because Fourier transform infrared (FTIR) spectrum was measured by insulin and KBr mixed pellets in air. Combining Fig. 2 and Fig. 3, the measured spectra and MD results show that the peak position, peak shape, peak distribution and relative strength of the two are basically consistent [33–35], which indicate that our method is reliable.

3. Results and discussion

3.1. RBD pre-equilibrate

The system was pre-equilibrated with 30 ns MD runs in the NPT ensemble, as shown in Figs. 4, 5 and 6. It can be seen from Fig. 4 that after 6 ns, RMSD tend to converge at both temperatures, which indicates that RBD has equilibrated. Fig. 5 shows that RMSF at 310 K is greater than 298 K, especially the residues 475–490, which is the region where RBD binds to ACE2. This phenomenon shows that temperature has an obvious effect on human infection with SARS-CoV-2. Fig. 6 shows that LJ energy, Coul energy and potential energy converge, which also shows that the structure of RBD is stable.
3.2. IR spectra of RBD

The simulation data are convoluted by a Lorentzian whose full width at half maximum (FWHM) is 20 cm$^{-1}$ [36]. The IR spectra of RBD at 298 K and 310 K are shown in Fig. 7. At both temperatures, there is a strong absorption band around 3591 cm$^{-1}$, which is generated by the stretching vibration of O–H in RBD. The peak of 310 K is slightly higher than that at 298 K. It can be explained that the temperature rising results the increase of molecular energy, thereby increasing vibrational amplitude, and oscillator strength. Comparing the positions of the absorption peaks of O–H at the two temperatures, there is a slight red shift at 310 K. In general, because the aqueous solution is polar, the stretching vibration of O–H shows a small red shift, and its intensity increases slightly. Except the absorption peaks of O–H stretching mode, there are absorption bands around 3154 cm$^{-1}$, which are assigned to C–H stretching mode in the protein. The absorption bands around 1594 cm$^{-1}$ are generated by C–O stretching vibration. A series of absorption peaks near 1431 cm$^{-1}$ are assigned to C–O stretching mode and C–H in-plane bending mode. The bands located around 972 cm$^{-1}$ are attributed to a large number of C–H out-of-plane bending vibration. Comparing the absorption peaks, they have a small red shift or blue shift with the change of temperature. It shows that the temperature change has a slight impact on their vibration frequency.

Compared with Fig. 2 and Fig. 7, it can be seen that the main IR absorption peaks of RBD and water do not overlap. The spectra of RBD are also compared with those of severe acute respiratory syndrome (SARS) coronavirus envelope protein from 1520 to 1800 cm$^{-1}$ [37] and HIV-infected cervical cells from 950 to 1300 cm$^{-1}$ [38], which indicates the difference in spectral distribution. Even if RBD has spectral overlap with water or other proteins, we can identify it by convolutional neural network (CNN) method [39–41] in chemometrics through the information of peak height, peak width, peak shape and peak distribution.

In addition, the stretching vibration frequencies of O–H, C–H and C = O are 3693 cm$^{-1}$, 2946 cm$^{-1}$ and 1605 cm$^{-1}$, respectively, which are calculated according to Eq.5. Compared with the molecular dynamics simulations, the errors are 2.9 %, 6.6 % and 0.7 %, respectively, which indicates the reliability of our simulations. In this sense, the IR absorption characteristics can be the guide to RBD detection.

4. Conclusion

In this paper, we discussed the IR spectra of SARS-CoV-2 RBD by the MD simulations. The results showed that the strong absorption bands located around 3591 cm$^{-1}$ is assigned to O–H stretching mode, the oscillator strength of 310 K is slightly higher than that at 298 K. The absorption peaks have a small red shift or blue shift with the change of temperature. The IR absorption characteristics can be the guide to RBD detection. This work provides a theoretical basis for the optical detection of SARS-CoV-2, and can also guide the physical detection of other viruses.

CRediT authorship contribution statement

Jianbin Du: Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. Ning Yao: Visualization, Investigation. Huijie Wang: Software. Qifeng Li: Software, Visualization. Zhifang Feng: .

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by National Key Research and Development Program of China (2017YFC0803600), National Natural Science Foundation of China (NSFC22174098) and Key Projects in the Science & Technology Pillar Program of Tianjin (20YFZCSN00530).

References

[1] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, Clinical features
