Reproducibility of Terahertz Peaks in a Frozen Aqueous Solution of 5-Methylcytidine

Yu Heng Tao[1] (ID) · Stuart I. Hodgetts[2,3] · Alan R. Harvey[2,3] · Vincent P. Wallace[1]

Abstract

Abnormally high level of methylation in the epigenome is a biomarker for various forms of cancer. It has been reported that a methylated nucleoside, 5-methylcytidine, when dissolved in water and then frozen, has a resonance absorption peak around 1.6 THz, which could allow for the determination of the degree of methylation using terahertz spectroscopy. This study attempts to replicate the experiment and reproduce this observation independently. Samples of ice were measured at 173 K to establish a baseline and frozen solutions of 5-methylcytidine were measured. Normal (physiological) 0.15 M saline, which is known to have a peak at 1.6 THz, was also measured. The use of a Gaussian function to fit a baseline to the broad (0.2 to 2.5 THz) absorption spectra consistently produced a peak residual around 1.6 THz for all three types of samples, with saline having the highest peak followed by ice. The peak for 5-methylcytidine was found to be smaller than that for ice, which contradicts previously published findings and illustrates the strong dependence of the result on the method of baseline fitting.

Keywords Terahertz · Time-domain spectroscopy · Reproducibility · 5-Methylcytidine

✉ Yu Heng Tao
hank.tao@research.uwa.edu.au

Vincent P. Wallace
vincent.wallace@uwa.edu.au

[1] Department of Physics, The University of Western Australia, 35 Stirling Highway, Perth, WA 6009, Australia

[2] School of Human Sciences, The University of Western Australia, 35 Stirling Highway, Perth, WA 6009, Australia

[3] Perron Institute for Neurological and Translational Science, Verdun Street, Nedlands, WA 6009, Australia

1 Introduction

1.1 DNA Methylation and Terahertz

DNA methylation is one of the key mechanisms in gene regulation; and it is considered to have a largely suppressive effect on gene expression [1]. Two of the four DNA bases, cytosine and adenosine can be methylated, which involves the addition of a methyl group to the nucleobase. Specifically, in cytosine this is at the fifth atom of the pyrimidine ring (see Fig. 1); cytidine found in RNA is the nucleoside formed when a ribose ring is attached to cytosine, and $2'$-deoxycytidine is cytidine with the hydroxyl group on the $2'$ position removed. The methylated nitrogenous base is called 5-methylcytosine and the associated nucleosides are named as in Fig. 1. The attachment of a methyl group can prevent the binding of transcriptional factors to a gene and attracts histone-modifying proteins, such as histone deacetylases, to cause the chromatin to form into a compact structure, making the start site of the gene inaccessible for transcription [2, 3]. Aberrant methylation is considered a biomarker for the diagnosis and prognosis of many cancers, as well as a predictor of patient response to different therapeutic treatments [2, 4].
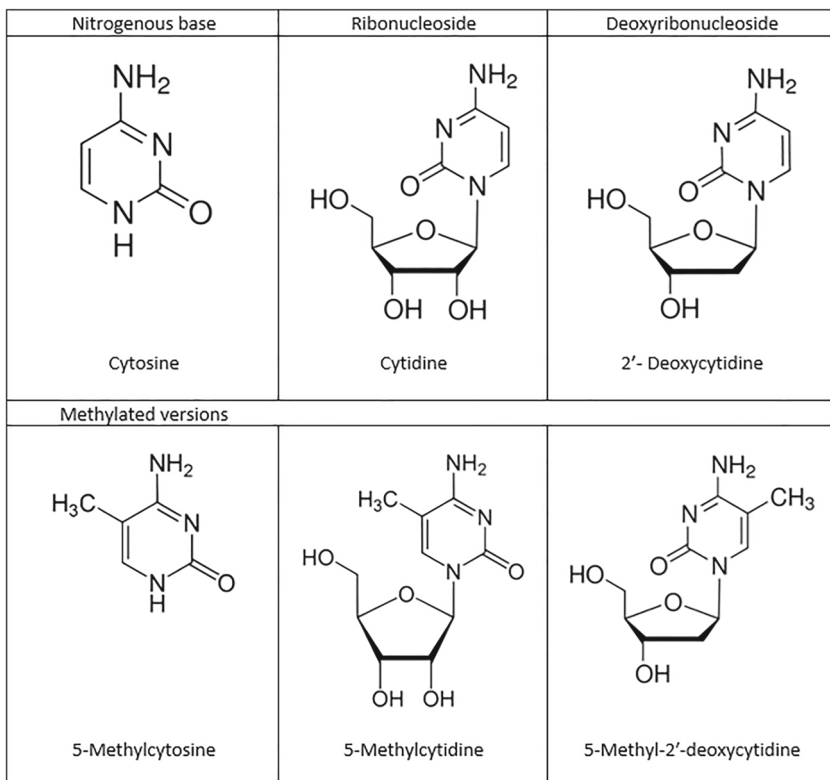

Fig. 1 The nucleosides of cytosine and their methylated forms

Terahertz (THz) radiation lies between the infrared and the microwave regions of the electromagnetic spectrum and has been used in numerous biomedical applications, including cancer imaging [5–7]. In contrast to X-ray crystallography, which provides information on the atomic structure of molecules, terahertz radiation probes higher-order structures and collective vibration modes such as phonon lattice vibrations in crystals [8]. Indeed, dramatic changes in the terahertz spectrum have been reported between polymorphs (different crystal structures) [9, 10] and isomers (different arrangement of atoms) [11, 12] due to the energy of intermolecular interactions being in the terahertz range [13].

Some molecules in the presence of water can form hydrates, e.g., lactose monohydrate where one water is incorporated into the crystal lattice and forms an integral part of it. This type of lactose has characteristic crystal structures that have unique terahertz spectral features different from that of the anhydrous form [14]. Even in longer chain molecules, like fatty acids, distinct spectral terahertz features are apparent as long as there is long-range three-dimensional molecular order [15]. Spectral features in the terahertz range tend to disappear in higher energy materials, for example in amorphous solids, which are disordered by nature and are akin to the liquid state in which all order is lost, and terahertz spectra show no resonant absorption modes [16]. Thus, like all crystalline substances that dissolve in an aqueous environment, the intermolecular interactions no longer occur. The spectrum in the terahertz range is dominated by water absorption [17], which in turn is a featureless spectrum, with a high absorption coefficient (e.g., 220 cm^{-1} at 1 THz [18]) due to dipole reorientations, and hydrogen bond stretching [19]. One way to negate the high absorption of water is to freeze the sample, as in the frozen state, most water molecules arrange themselves into the well-known hexagonal lattice of ice, which absorbs an order of magnitude less terahertz radiation (10 cm^{-1} at 1 THz, see Fig. 2). Freezing has been attempted in tissue samples to reveal underlying structure [20–22], but still, no unique resonance features appear. However, interestingly, unique spectral features appear when solutions of various alkali halides are frozen [23]. Specifically, in the case of saline solution, absorption features are attributed to the formation of sodium chloride dihydrate (NaCl·2H$_2$O) during the freezing process [24, 25]. In its solid form, 5-methylcytidine has characteristic spectral features [26] and it has been suggested that in the presence of water molecules and under the right conditions, i.e., in a frozen solution, it exhibits unique spectral features in the terahertz range [26]. The questions is—does the reduction in THz absorption of water to ice reveal information about the solute and its interaction with the surrounding water molecules?

Research regarding the interaction of terahertz radiation with ice has noted that the absorption generally increases in proportion to the fourth power of frequency due to translational vibration of the ice lattice [27, 28]. A rise in temperature causes an increase in absorption due to increased vibration [29] and polarizability [30]; as seen in Fig. 2, where the THz absorption spectrum of ice is shown at 223 K (red markers) and at 173 K (blue markers). The order of the ice lattice has also been reported to affect the overall absorption strength, with polycrystalline ice having a higher absorption than single-crystalline ice, possibly due to increased scattering [29] and certain vibrational modes not allowed in a perfect crystal [31].
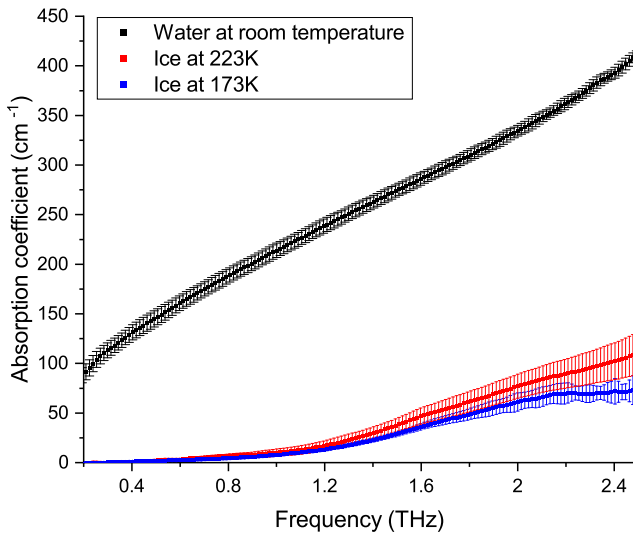
Fig. 2 Absorption coefficient spectra of liquid water at room temperature (black), ice at 223 K (red), and ice at 173 K (blue)

1.2 Measurement of 5-Methylcytidine

Cytidine is the ribonucleoside containing the nitrogenous base cytosine, the methylation of which has been shown to be involved with various cancers [32, 33]. In a study by Cheon et al. [26] 5-methylcytidine (5-mc) and 2′-deoxycytidine (2′-dc) were chosen to represent methylated and unmethylated nucleoside, respectively, and the authors reported distinct resonant absorption of THz radiation by the former. The authors measured the frequency dependent absorption coefficient of 5-methylcytidine dissolved in water at a concentration of 0.27 M, frozen and cooled to 253 K and observed a small peak at approximately 1.6 THz. The absorption by ice in the range of 0.4 to 2 THz was fitted as the left side of a Gaussian function, which, when subtracted from the 5-mc spectrum, revealed a peak that is approximately Gaussian, centered at 1.6 THz, with an amplitude of 5 cm^{-1} and FWHM of 0.4 THz. Such a peak was not observed in a frozen solution of 2′-dc [26]. The magnitude of the 5-mc peak increased to 9 cm^{-1} when the concentration was increased by 2.3 times to 0.62 M. The authors went on to demonstrate a similar observation with artificially methylated DNA and DNA from five different types of cancer cells. The degree of methylation was inferred from the magnitude of the resonant absorption peak at 1.6 THz, which ranged from 3 to 9 cm^{-1}. The relative degree of methylation between the five types of cancer DNA was verified by enzyme-linked immunosorbent assay (ELISA) [26].

Similar findings were reported by the same group more recently [34], where more details were provided regarding the measurements and the data analysis. A variety of fitting functions were explored and compared, all of which showed evidence of excess absorption in the range of 1.2 to 2.5 THz.

To our knowledge, this intriguing and potentially useful resonant feature of 5-mc observed by [26] has not been reproduced anywhere else. Here, we attempted to independently verify the previously published data, measuring not only ice but also frozen saline solution, which also has a peak at 1.6 THz [23, 25, 35]. Successful detection of the 1.6 THz peak in frozen saline would validate our peak detection method, and a comparison between saline and 5-mc would give a sense of how sensitive the peak amplitude is to each solute. Should the peak in physiological saline solution be prominent enough, it could affect the detection of any peak caused by 5-mc, thereby increasing the risk of overestimated degree of methylation in biological or pathological samples.

In the following sections, data from both [26, 34] were extracted for comparison with the result presented here. More specifically, the absorption spectrum of ice was taken from [26], and the measurements related to 5-mc were mostly from [34], as the latter included more details of the measurements on 5-mc, such as error bars and the power spectra, while the spectrum of ice was only available in the former.

2 Methods

2.1 Experimental Design

Terahertz spectra were acquired with a commercial THz-TDS system, Terapulse 4000 (TeraView Ltd, Cambridge, UK) using the fibre coupled emitter/detector devices [15] (see Fig. 3). The devices, the cryostat's liquid cell holder, and the sensing element of a humidity sensor are placed in a nitrogen purged acrylic box. The relative humidity was lowered to < 0.001% for all measurements.

2.2 Sample Preparation

All samples were prepared using distilled water (Milli-Q®, a.k.a. reversed osmosis purified and filtered water with a resistivity of 18.2 MΩ·cm at 25 °C).

At room temperature, a z-cut quartz window (Q-W-32-3, ISP Optics) was wrapped in laboratory film made from polyolefin and wax and is resistant to polar liquids. The window was placed in the liquid sample holder of the cryostat (GS21525, Specac). A Teflon spacer (outer diameter 32 mm, inner diameter 26 mm) 1 mm thick was placed on the window. The purpose of the spacer was to contain the sample while in liquid form. The frozen sample thickness was just over 1 mm due to the volume of ice being greater than that of the liquid water. The thickness was chosen by trial and error as it provides sufficient robustness for sample handling while also maintaining an acceptable spectral bandwidth. It is expected that, within a valid bandwidth, spectral features would be more reliable in thicker samples due to the enhanced likelihood of interaction between the transmitting terahertz beam and the specimen of interest [36]. Another advantage of a thicker sample is that it reduces the number of multiple reflections within the measurement time window, and the increased separation between the main pulse and the first reflection allows more of the main pulse
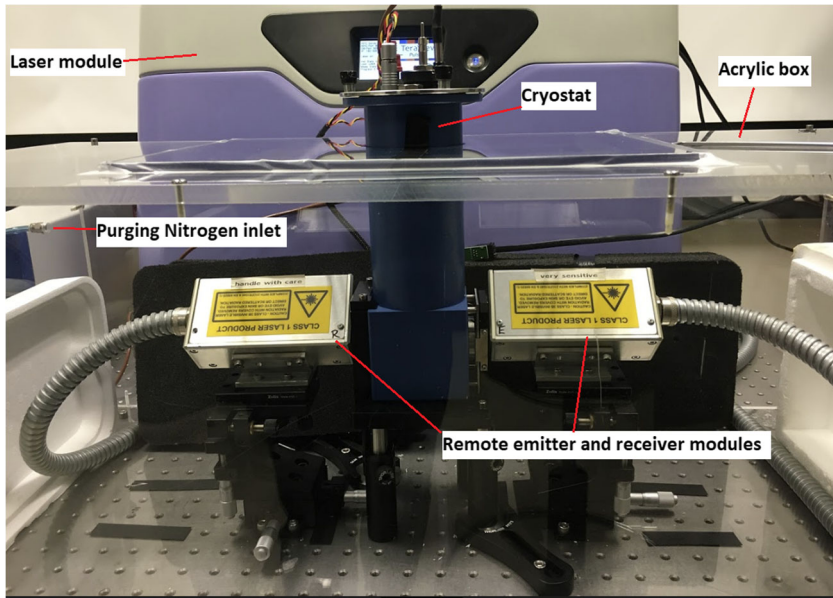
Fig. 3 Terapulse 4000 with remote heads, showing the cryostat cell holder in an acrylic box. The laser and associated optics are contained within the cream and purple unit at the back

available for processing after a filtering window is applied. The appropriate volume of solution per sample was thus determined to be 0.65 ± 0.02 ml.

The sample solution was placed on the quartz window and another quartz window wrapped in laboratory film was gently placed on top of the spacer. The assembly was then placed in a standard freezer at $- 15$ °C for 30 to 40 min. Once fully frozen, both quartz windows were separated from the frozen sample and the sample was placed back into the sample holder. The cap of the sample holder was screwed on to secure the frozen sample in position. The sample holder was then placed in the cryostat primed with liquid nitrogen. Once the temperature set point was reached the sample terahertz data was collected. Transmitted pulses were recorded for 33.37 ps at 8.15 fs resolution. For each sample five hundred transmitted pulses were averaged, which at the spectral resolution of 0.94 cm^{-1} takes approximately 30–33 s to complete.

Preliminary measurements of ice show, as expected, that the overall absorption is lower at lower temperatures, and that the higher frequencies are more sensitive to this thermal effect (see Fig. 4). Given the design of the cryostat and the experimental setup, the time required to cool the sample to 173 K is typically 10 to 15 min as shown in the inset of Fig. 4. To increase the visibility of the 1.6 THz peak, 173 K was chosen to be the temperature setting for all measurements for a reasonable compromise between sufficiently low background absorption of ice and an adequate cooling time for sample throughput.

Preliminary measurement of 5-mc at 0.5 M (88.2 mg/0.7 ml), a concentration within the range of those reported in [26], did not yield any peak feature. Therefore,
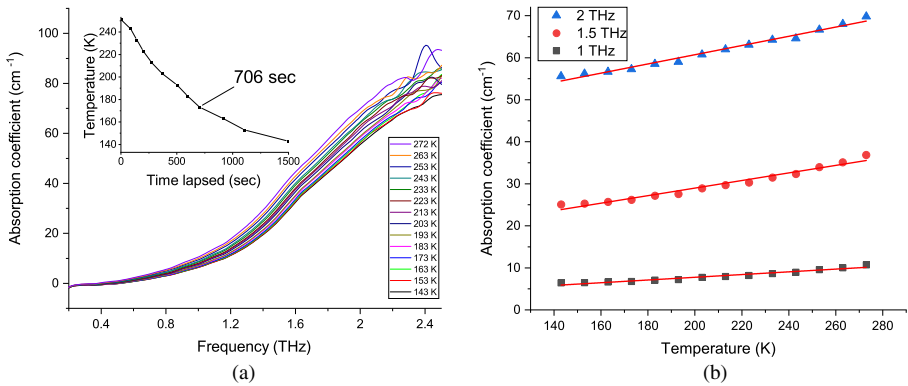
Fig. 4 Thermal effect on the absorption spectrum of ice. **a** Absorption spectrum of ice from 143 to 272 K. Inset: cooling time as a function of temperature; time required to reach 173 K is at least 706 s. **b** Absorption of ice at various frequencies as a function of temperature. The slope increases with frequency, which is 0.03, 0.09, 0.11 for 1, 1.5, and 2 THz, respectively

two batches of 1 M solution 5-mc were prepared. The first batch was prepared by dissolving 480.6 mg of 5-mc powder (Sapphire Bioscience) in 1.868 ml of distilled water. The second batch was prepared by dissolving 359 mg of 5-mc power (Sapphire Bioscience) in 1.395 ml of distilled water.

Normal (physiological) saline (0.154 M) was prepared by dissolving 0.9 g of NaCl in 100 ml of distilled water.

Ten samples of ice, five samples of 5-mc, and ten samples of normal saline were measured. Other concentrations of saline ranging from 0.1 to 1 M were also prepared and measured.

2.3 Data Processing

Sample thicknesses were determined by the time difference between the main pulse and the first Fabry-Pérot reflection. The refractive index of ice of 1.8 [28, 37], is assumed for all measurements; this value was confirmed by iterative calculations of sample properties in the commercial software, Teralyzer (Menlo Systems, Germany).

The time-domain transmission data was used as input to a simplified version of the Duvillaret algorithm [38] implemented in MATLAB (The MathWorks Inc., USA), the outputs of which are the optical constants including the absorption coefficients.

All curve fitting and inference statistics were performed in OriginPro 2020b (OriginLab Corp, USA). Spectral fitting was conducted using a Gaussian function as given below:

$$y = A \times exp\left[-\left(\frac{(x - f_c)^2}{2w^2}\right)\right] \tag{1}$$

where A is the amplitude of the Gaussian peak, f_c is the centre frequency of the Gaussian peak, and w relates to the full width at half maximum (FWHM) of the Gaussian peak (FWHM $\approx 2.355w$).

To reduce the number of iterations required boundary conditions were set as follows: $A \leq 200$, $f_c \geq 1.8$, $w > 0.5$.

The fit of a Gaussian function is susceptible to distortion by the presence of peaks in the measured spectra. Such distortion is illustrated in Fig. 5a, where the absorption spectrum of a 1 M saline solution acquired at 173 K was fitted with a Gaussian function. The fit appears distorted and it is due to the salient peak at around 1.6 THz, which reduces the amplitude (A) of the Gaussian from 85 to 64 and shifts the centre peak f_c to the left (from 2.7 to 2.1 THz). One way to minimize such distortion is to exclude the frequency range containing the peak from the fitting process. The exclusion was achieved by setting the difference between the fit line and the measured absorption values to zero over a frequency range around the peak. The improvement in the fit is clearly visible in Fig. 5b where the frequencies 1.45 to 1.83 THz have been excluded.

In comparing the goodness of fit, we report residuals sums of squares (RSS) instead of the coefficients of determination (R^2). Although the advantage of R^2 over RSS is that it allows comparison between fits on different scales, the advantage does not apply here as the fits being compared are on the same scale. What is of interest is the size of any peaks over the fitted lines, the magnitudes of which are in the order of less than 10 cm^{-1} (i.e., peak amplitudes reported in [34]). Such magnitude is much smaller compared to how much the absorption coefficient varies in the fitting range, which changes from less than 1 cm^{-1} at 0.2 THz to more than 120 cm^{-1} at 2.5 THz [34]. The total sums of squares (TSS) would thus always be very large compared to RSS, leading to a R^2 value always very close to one (see Eq. 2).
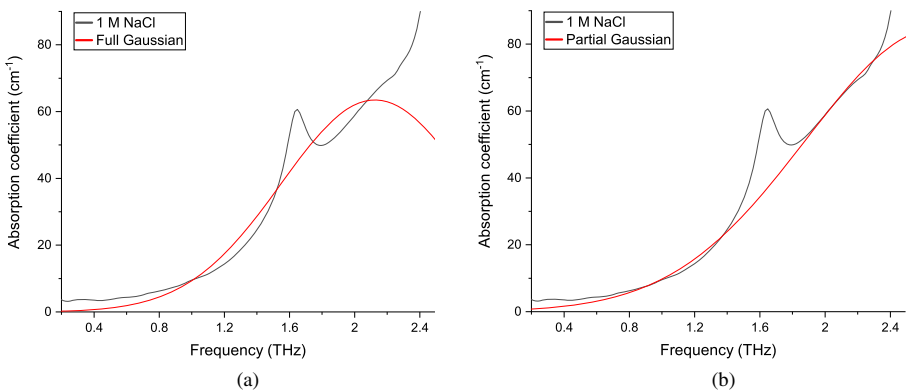

Fig. 5 Comparison of full versus partial Gaussian fit applied to the absorption spectrum of a 1 M frozen saline sample, which has a prominent peak at 1.6 THz. **a** Full Gaussian fit. **b** Partial Gaussian fit, with frequency range 1.45–1.83 THz excluded from the fitting process

All error bars displayed in the present study are standard errors unless otherwise specified.

$$R^2 = 1 - \frac{RSS}{TSS}$$
$$= 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2} \qquad (2)$$

3 Results

3.1 Spectra

The frequency power spectra and standard errors of the ice and the 5-mc samples (all around 1.3 mm thick) are shown in Fig. 6a and b, respectively. The power spectra extracted from [34] is overlaid on Fig. 6b for comparison. The reference power spectra were measured with the sample cell removed (i.e., nitrogen gas in the beam path). The relative humidity for all measurements was < 0.001%. The sample sizes for ice and 5-mc are ten and five, respectively. In [34], the reference was a single piece of 2 mm z-cut quartz window, and the sample thickness was 0.3 mm. The temperature in [34] was 253 K, compared to 173 K in the present study.

The signal intensity is lower at all frequencies in the present study due to the thicker samples. However, with a lower noise floor, a signal-to-noise ratio (SNR) of 10 is achieved at 2.5 THz, which is close to the SNR of 15 achieved in [34]. Interestingly, the standard errors reported in [34] is significantly higher even for reference measurements. At 2 THz, the error in the reference [34] is more than six times those in the present study (i.e., 4.97 dB vs. 0.73 dB). Based on the SNRs and standard errors, we deemed it appropriate to assume the same valid bandwidth as in [34] of 0.2 to 2.5 THz.

Figure 7 shows the mean absorption spectrum of ice at 173 K, the preliminary ice sample at various temperatures, and the absorption spectrum of ice extracted from [26]. The thickness of the ice sample in [26] is presumably 0.3 mm, whereas the thickness in the present study is 1.3 ± 0.1 mm. Note that the overall absorption of ice is lower in this study compared to [26] at all temperatures. In our experience, the overall absorption was often higher when the sample surface had visible imperfections, typically introduced during the separation of the film-wrapped windows and the sample. Such surface imperfections may cause diffuse reflection and increased scattering through the sample, leading to an overestimation of absorption. Thinner samples would also be less tolerant to surface imperfections and thickness imprecisions. The apparently higher overall absorption reported in [26] could be a result of higher scattering and should be apparent in a larger sample size (e.g., $n = 10$, as per the present study). We therefore believe the spectra measured in the present study to be closer to the true absorption spectrum of ice in comparison to the spectrum published in [26].
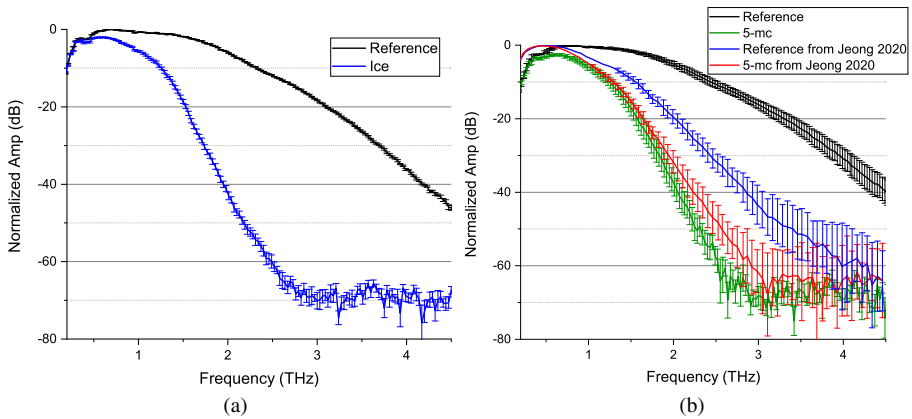
Fig. 6 Power spectra of ice versus frozen solution of 5-mc. **a** Power spectrum of ice (blue, $n = 10$) and reference (black). The sample size is 10. **b** Power spectrum of 5-mc in the present study (green, $n = 5$) and that extracted from [34] (red, $n = 5$). The reference spectrum of the present study (black) and that extracted from [34] (blue)

Also note that in the extracted spectrum (Fig. 7, black squares), the absorption starts dropping off at 2 THz. The authors in [26] attributed the drop off in the spectrum to noise and defined their valid bandwidth for ice to be 0.4–2.0 THz. The spectrum of ice we measured at 173 K has relatively small standard error up to 2.5 THz (2–3% over the range of 1.6–1.7 THz).
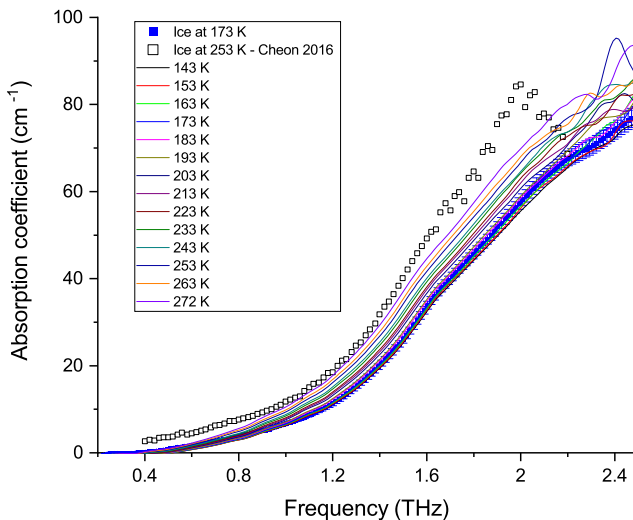

Fig. 7 Mean absorption of ice at 173 K with standard error, and the absorption spectrum of ice at 253 K extracted from [4]. The samples size is ten ($n = 10$) in the present study but unspecified in [4]. Thermal effect on the spectrum is also shown with a sample measurement of ice from this study and [26]

Figure 8a shows the mean absorption spectrum of 5-mc compared to ice in the present study, and Fig. 8b compares those obtained from [26, 34]. In both studies the overall absorption of 5-mc is higher than that of ice by about 15 to 20%. Note that the standard error in the spectrum from [34] is much higher and increases with frequency, whereas the standard error in the present study is much lower and remains relatively stable.

In Fig. 9a the spectrum of 5-mc is compared to the spectrum of normal saline; number of samples of normal saline measured was 10. The overall absorption of normal saline is similar to that of ice but lower than that of 5-mc. When the absorption spectrum of ice is subtracted from the absorption spectrum of saline, the peak at around 1.6 THz in saline is clearly visible (see Fig. 9b), whereas there are no distinct features when the same method is applied to 5-mc. The source of the peak in saline has been suggested as sodium chloride dihydrate ($NaCl \cdot 2H_2O$) based on molecular dynamics simulations [24, 25].

3.2 Full Gaussian Fit

Figure 10a shows the full Gaussian fit to the spectrum of ice, 5-mc, and normal saline at 173 K. The fit clearly deviates around 1.6 THz for saline, indicating a peak, which is the same peak as seen at higher concentrations of saline (see Fig. 5).The deviation at 1.6 THz is less clear for ice, and almost imperceptible for 5-mc. The fitted parameters of the Gaussian function are presented in Table 1. Note that saline has the largest RSS due to the pronounced peak at 1.6 THz, whereas 5-mc has the lowest RSS (lower than that of ice).

Figure 10b shows the residuals of the fits, i.e., measured data minus predicted absorption coefficient from the Gaussian fit. Three standard deviations are shown as the error bars of the ice residuals in order to show the limit of detection, as commonly defined in the pharmaceutical industry [39]. All three sets of residuals show a similar, non-random pattern where a prominent peak occurs at 1.67 THz. The 1.67
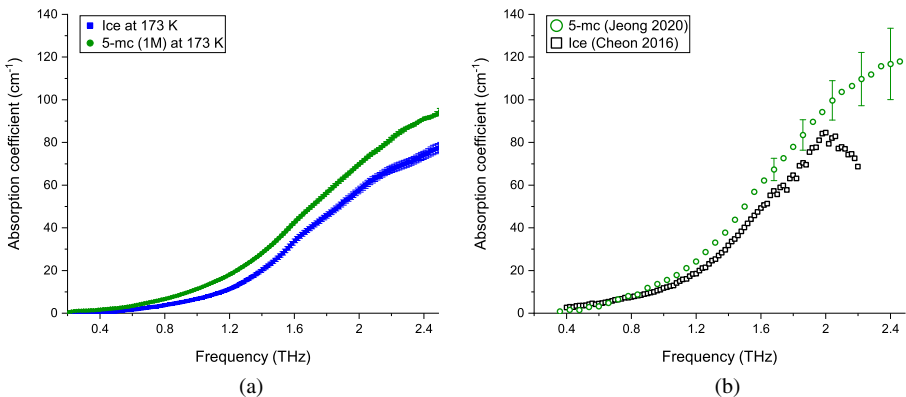

Fig. 8 Comparison of absorption spectrum of 5-mc versus ice. **a** Absorption spectra of frozen 1 M 5-mc solution (green, $n = 5$) and ice (blue, $n = 10$) at 173 K. **b** Absorption spectra of frozen 0.78 M 5-mc solution (green, $n = 5$) extracted from [34] and ice (black) extracted from [26] at 253 K
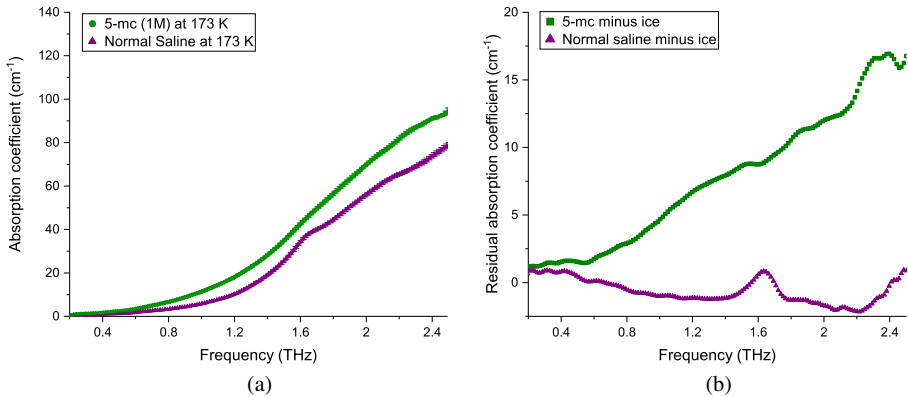
Fig. 9 Comparison of absorption spectra of frozen solutions of 5-mc versus saline. **a** Absorption spectrum of normal saline (purple, $n = 10$) and 5-mc (green, $n = 5$) at 173 K. **b** Absorption of ice subtracted from that of 5-mc (green), saline (purple). Absorption of ice reported in [26] subtracted from that of 5-mc reported in [34] also shown in red squares connected by black lines

THz peak is highest for normal saline, followed by ice, with 5-mc having the smallest peak. Note that the peak for 5-mc lies within three standard deviations of the ice peak, whereas the peak for normal saline lies outside it. The peak for normal saline is clearly detectable whereas the residuals of ice and 5-mc are statistically indistinguishable.
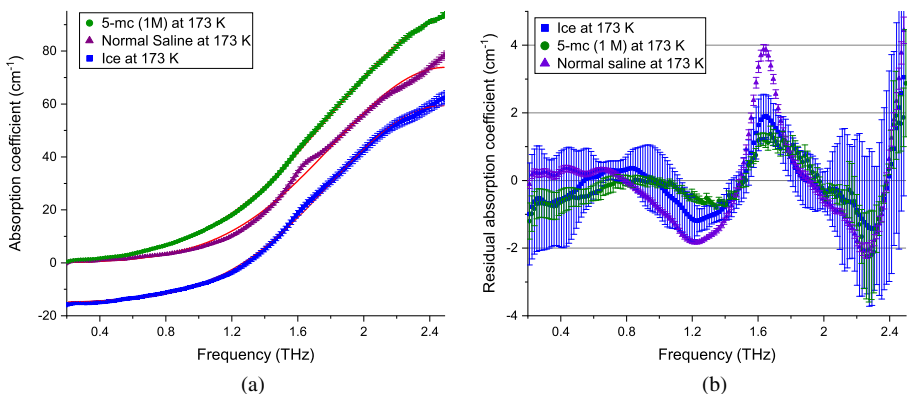

Fig. 10 Comparison of absorption spectra and full Gaussian fits of ice, 5-mc, and saline. **a** Absorption spectrum of frozen solution of 5-mc (green), saline (purple), and ice (blue). The spectrum of ice is lowered by 15 cm^{-1} for display purpose. Full Gaussian fit lines in red. **b** Residuals of the full Gaussian fit of 5-mc (green), saline (purple), and ice (blue). The error bars of the residuals of ice are shown as three standard deviations (a.k.a. detection limit)

Table 1 Parameters with standard errors of the full Gaussian fit for ice, 5-mc (1M), and normal saline at 173 K

Sample Type	A	f_c	w	RSS
Ice	74.34 ± 0.29	2.48 ± 0.01	0.68 ± 0.00	112.43
5-mc (1 M)	91.78 ± 0.36	2.60 ± 0.01	0.77 ± 0.00	87.10
Normal saline	73.86 ± 0.52	2.50 ± 0.01	0.68 ± 0.01	303.05

3.3 Partial Gaussian Fit

As described in Section 2.3, to more accurately fit the Gaussian to the data either side of the 1.6 THz peak we eliminate some frequencies from the fitting procedure. Figure 11a shows the partial Gaussian fit to the spectrum of ice, 5-mc, and normal saline at 173 K. Here, the fitting algorithm is set to ignore the data points from 1.2 to 2.3 THz, which contains most of the 5-mc peak features reported in [34]. The deviation of all three fits is now visible but that in 5-mc is still the smallest. There is no longer a deviation near 2.3–2.5 THz for any of the three fits.

Figure 11b shows the residuals of the fits. The detection limit is again shown as three standard deviations of the ice residuals. All three sets of residuals share a similar, non-random pattern, but the pattern has changed slightly due to the weighting—the peak at 1.7 THz has increased in amplitude, and there appears a second peak at 1.9 THz that is smaller in amplitude but broader in bandwidth. The standard deviation of the residuals particularly that of ice, has increased rather dramatically, which supports the notion that peak identification via baseline fitting is sensitive to the baseline function. However, even with the increase standard deviations, the peak in normal saline still exceeds the limit of detection at one data point (1.633 THz), as shown in the inset.

3.4 Fitting of Published Data

Having used the Gaussian fitting method on the data collected in this present study, we now refit previously published data extracted from [26] and [34]. In those studies, the spectrum of ice is taken to be a Gaussian curve with no peaks in the range 0.4 to 2 THz, whereas it has been suggested that frozen 5-mc solution has at least one, and possibly two peaks from 1.6 to 2 THz.

Figure 12a shows the extracted spectra and the full Gaussian fit. Figure 12c shows the residuals of the fits. The residuals of the ice fit deviate due to data fluctuations from 1.7 THz and above; however, both fits have a peak residual of around 2 cm^{-1} at 1.6 THz.

Figure 12b shows the partial Gaussian fit to the same spectra from [26] and [34]. The frequencies excluded from the fitting for 5-mc are between 1.2 and 2.3 THz (i.e., the same range as used in Fig. 11), which covers the reported peaks. It should be noted that the 5-mc data from [34] covers 0.36–2.7 THz and the ice data from [26] is 0.4–2.2 THz but the authors considered the data above 2 THz as noise. As a result,
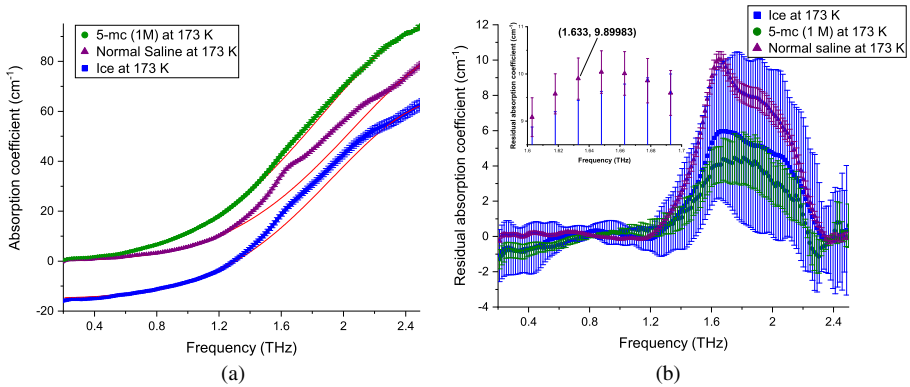
Fig. 11 Comparison of absorption spectra and partial Gaussian fits of ice, 5-mc, and saline. Frequencies 1.2–2.3 THz were excluded from the fitting algorithm. **a** Absorption spectra with partial Gaussian fit lines in red. **b** Residuals of the partial Gaussian fits. The error bars of the residuals of ice are shown as three standard deviations (a.k.a. detection limit). Inset: Expanded view of 1.6–1.7 THz. The data point marked is the residual of saline that lies completely outside the detection limit (i.e., lower error bar also outside limit)

the same frequencies could not be excluded from the ice data [26] due to the smaller spectral content and so the range of 1.4–1.85 THz was chosen to provide sufficient data points outside the range for a good fit.

Figure 12d shows the residuals of partial Gaussian fits. The pattern is similar to the full fits except that the peak height at 1.6 THz has increased due to the frequency exclusion. For the partial Gaussian fits, both ice and 5-mc show a peak residual of around 4 cm^{-1} at 1.6 THz.

For 5-mc, the partial Gaussian fit produced a residual peak much smaller than that reported in [34] (Fig. 12d). The likely reason is the difference in the baseline used although the method of generating the baseline is not specified in [34].

Data points on the baseline reported [34] were extracted and the Gaussian parameters thereof calculated. Figure 13 shows the extracted baseline and parameters, and compares them to those of the partial Gaussian baseline used in the present study. The frequency range of 1.2–2.3 THz is excluded from both fits. The relevant parameters are compared in Table 2. The extracted baseline has a much higher RSS. Note that any deviation in the range of 1.2–2.3 THz does not contribute to the calculated RSS. The absorption spectrum of 5-mc outside the frequency range of the purported peak fits the Gaussian baseline more closely, and the partial Gaussian baseline should be used instead of the baseline in [34] in the detection of peaks located between 1.2 and 2.3 THz.

4 Discussion

The findings in the present study are as follows. The spectrum of ice deviates from that of a Gaussian fit and the residual plots consistently exhibits a peak around
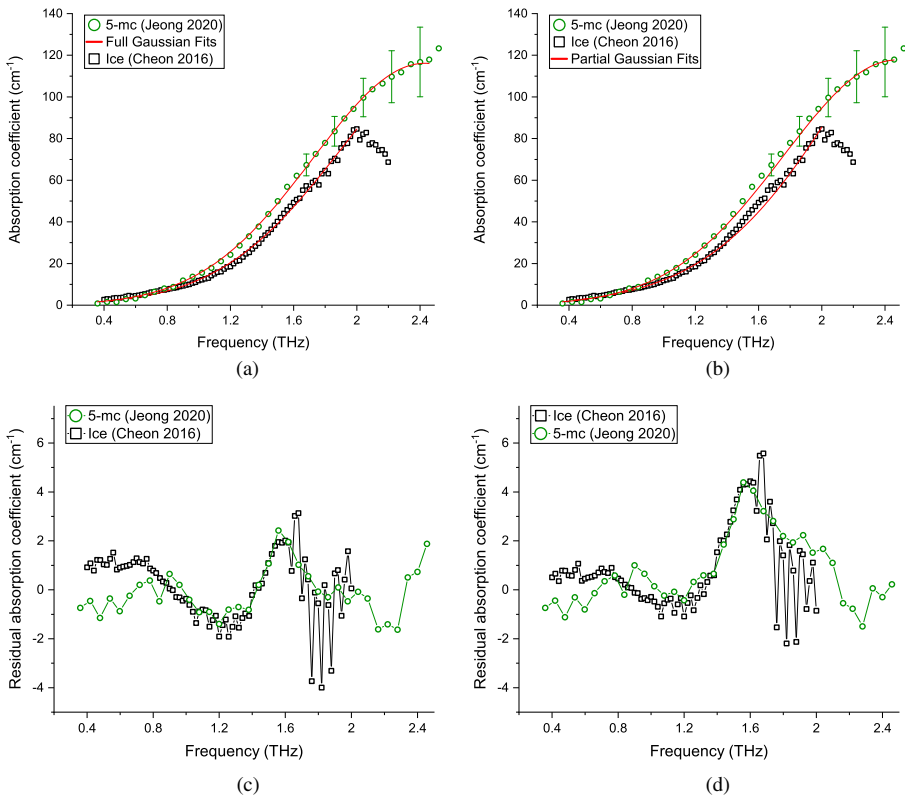
Fig. 12 Full and partial Gaussian fits on the absorption spectra of ice and 5-mc extracted from [26] and [34], respectively. **a** Full Gaussian fit (red) on the absorption spectrum of ice (black squares) and 5-m (green circles). **b** Partial Gaussian fits. **c** Residuals of the full Gaussian fits. **d** Residuals of the partial Gaussian fits. The fitting range for ice is 0.4–2 THz (same as [26]). The fitting range of 5-mc is 0.2–2.5 THz (same as [34])

1.6 THz, the shape and size within error of which is like that described in [26] and [34] as related to 5-mc. At 1 M concentration of 5-mc, the maximum peak height at 1.6 THz is 3.4±0.9 cm^{-1}, which is lower than that of ice (6.0±0.4 cm^{-1}), and much lower than the 9 cm^{-1} at 0.78 M reported in [34]. When the 5-mc data extracted from [34] was refitted, the peak amplitude was found to be much lower (4.4 cm^{-1}) than reported (9.4 cm^{-1}). Comparing the current fit to the reported fit, the reported fit has a four times larger residual sums of squares. The 1.6 THz peak in frozen saline is prominent even at 0.15 M (physiological concentration) and increases linearly with NaCl concentration. Under the partial Gaussian fit, the peak in 0.15 M saline is as high as 10 cm^{-1}, which is significantly higher than that of ice and 5-mc. The peak in normal saline is close to the limit of detection of our system, which we calculated to be 0.13 M at three standard deviations relative to the peak in ice.
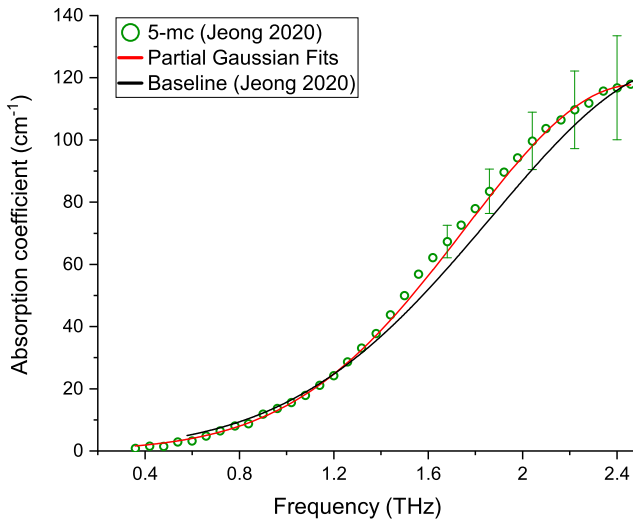
Fig. 13 Partial Gaussian fit of extracted 5-mc (0.78 M) versus the (unspecified) Gaussian fit as reported in [34]. The frequency range of 1.2–2.3 THz was excluded in both fits

In addition, we found that the overall absorption over the entire spectral range increased by approximately 20% in the presence of 5-mc (1 M) but not in the presence of NaCl (0.154 M). The increase in overall absorption possibly contributes to the smaller residual peak. The reason for the increased overall absorption is not clear but could potentially be due to the absorption of all frequencies by the 5-mc molecules and water solute and/or an increase in scattering. Scattering may have increased due to an increase in scattering particles, which can include ice crystallites, and 5-mc precipitate due to the change in solubility at lower temperatures.

The residuals of Gaussian fitting for the spectrum of ice shows a non-random pattern, which consists of a broad peak at around 0.7 THz, and a sharper peak at 1.6 THz. Such a pattern is also found in the ice spectrum extracted from [26]. The existence of a pattern suggests that a better fitting function may exist for the spectrum of ice.

Since the Gaussian fit of the ice spectrum always contains a 1.6 THz peak, the question is no longer whether there exists a peak in 5-mc, but whether the peak is larger than the one in ice. The answer is negative based on the findings presented

Table 2 Parameters of the fitted Gaussian for the 5-mc spectrum reported in [34]

Baseline function	A	f_c	w	RSS
As reported in [34]	79.61 ± 0.48	2.68 ± 0.01	0.75 ± 0.01	12.287
Partial Gaussian	95.77 ± 0.17	2.66 ± 0.01	0.80 ± 0.00	2.981

Baseline reported in [34] was extracted for comparison with the baseline generated by the partial Gaussian fit used in the present study. Frequency range 1.2–2.3 THz was excluded from the fitting process and RSS calculations

here—the peak in 5-mc is consistently lower than that found in ice, which suggests the addition of 5-mc dampens or obscures the 1.6 THz resonance.

A 1.6 THz peak is also found in the spectrum of ice extracted from [24], which could have been easily overlooked given the amount of noise in their data. It is possible that the data above 1.8 THz were considered too "noisy" and thus given less weight in the fitting process, as the fit shown in [24] consistently overestimates the data above 1.8 THz. In the present study, the absorption data up to 2.5 THz is reliable given the small standard error, and a 1.6 THz peak in the residuals is consistently found.

Determining the frequencies to be excluded is only possible knowing of the location of the peaks. The present study has the benefit of knowing where the purported peaks are from previous studies, however, in a blind test of unknown samples it is more appropriate to use full fitting rather than excluding any frequencies. Regardless, the same frequencies were excluded in both the sample (i.e., 5-mc) and the baseline (i.e., ice) spectrum, and when this is done the peak residuals in 5-mc never exceeds that of ice.

The amplitudes of residuals clearly depend on the baseline used, and the methods of generating the baseline described in [26, 34] might lack sufficient details to re-generating the baselines. Neither partial nor full Gaussian produced the baseline reported, though judging by the RSS, i.e., how well the model fits the data, the partial Gaussian baseline is just as good if not better than the reported baseline in [34] at representing the purported featureless part of the 5-mc spectrum. It is possible that a weighting related to the standard error was used in previous reports. If so, such weighting would not change the contradictions found, and the standard error of the present study is not only lower than [34] but remains more stable over the fitting range. However, using a Gaussian to fit the data, it is somewhat a moot point, given that we have shown that there is a feature in the spectrum of ice around 1.6 THz. Thus, fitting the whole ice spectrum from 0.4 to 2.4 THz to a smooth curve like a broad Gaussian is not ideal. If one assumes that features of interest lay over a background, then a more sensible approach is to use an empirical method of determining that background spectrum by measuring it, and then subtracting that from the spectrum of interest as we have done in Fig. 9b. Subtracting the spectrum of ice from that of a frozen solution of 5-mc revealed no special spectral features around 1.6 THz over and above variation of the measurement.

5 Conclusion

We have measured ice and 5-mc in ice over a broader terahertz frequency range than previously reported. The 1.6 THz peak is found in both the frozen solution containing 5-mc and pure ice. The spectra of frozen 5-mc solution and ice are therefore indistinguishable based on the 1.6 THz peak alone. In fact, the addition of 5-mc appears to have a slight dampening effect on the 1.6 THz peak. On the other hand, the 1.6 THz peak in normal saline is more than 60% higher than that in ice, which is clearly distinguishable even at physiological concentration. If the amplitude of the 1.6 THz peak is used to determine the concentration of 5-mc samples, even minute amount of

NaCl could significantly bias the result. With our system we determined the detection limit of frozen NaCl solution is 0.13 M.

The Gaussian function was found to fit the absorption spectrum of ice rather poorly judging by the consistent characteristics of the residuals and is primarily due to a slight flattening of the spectra starting around 1.6 THz. The reason for such flattening is unknown. The flattening was observed in both ice and 5-mc samples. We used a Gaussian function to fit the data to attempt to replicate the baselines in previous work [26, 34], however, there is no physical reason to assume the sloping background as the left side of a Gaussian function. Other types of fits may be equally valid such as one based on scattering [40].

In lieu of a better function(s) it is necessary to include both the residuals and the Gaussian fit for an accurate representation of the spectrum of ice. Peaks found using the Gaussian fit must be calibrated against existing peaks in ice.

Although the fitting was performed for spectra acquired at 173 K for a balance of low background absorption and high sample throughput, we do not believe it is critical to the conclusions reached here in this study. As shown in Fig. 4, the shape of the spectrum remains the same and only the overall absorption strength is affected by temperature. We therefore expect the same ranking of residual peak amplitudes at higher temperatures, and with 5-mc within the detection limit of ice.

Attempts in reproducing the results reported in [26], or in searching for fine terahertz features in biomolecules and biological samples should proceed with close attention to the accuracy and standard deviation of the baseline data, especially if baseline subtraction is involved. We therefore concluded that the use of THz radiation to analyse specific molecular characteristics of biological samples, while potentially of diagnostic value, requires considerably further research before its more general application is warranted.

Declarations

Conflict of Interest The authors declare no competing interests.

References

1. G. Auclair and M. Weber, Biochimie **94**, 2202 (2012).
2. B. Jin and K. D. Robertson, "DNA methyltransferases, DNA damage repair, and cancer", in *Epigenetic alterations in oncogenesis*, edited by A. R. Karpf (Springer New York, New York, NY, 2013), pp. 3–29.
3. A. M. Deaton and A. Bird, Genes & development **25**, 1010 (2011).
4. W. H. Lee, R. A. Morton, J. I. Epstein, J. D. Brooks, P. A. Campbell, G. S. Bova, W. S. Hsieh, W. B. Isaacs, and W. G. Nelson, Proceedings of the National Academy of Sciences **91**, 11733 (1994).
5. E. Pickwell and V. P. Wallace, Journal of Physics D: Applied Physics **39**, R301 (2006).

6. A. D'Arco, M. Di Fabrizio, V. Dolci, M. Petrarca, and S. Lupi, Condensed Matter **5**, 25 (2020).
7. O. A. Smolyanskaya, N. V. Chernomyrdin, A. A. Konovko, K. I. Zaytsev, I. A. Ozheredov, O. P. Cherkasova, M. M. Nazarov, J. P. Guillet, S. A. Kozlov, Y. V. Kistenev, J. L. Coutaz, P. Mounaix, V. L. Vaks, J. H. Son, H. Cheon, V. P. Wallace, Y. Feldman, I. Popov, A. N. Yaroslavsky, A. P. Shkurinov, and V. V. Tuchin, Progress in Quantum Electronics **62**, 1 (2018).
8. E. P. J. Parrott, Y. Sun, and E. Pickwell-MacPherson, Journal of Molecular Structure **1006**, 66 (2011).
9. C. J. Strachan, P. F. Taday, D. A. Newnham, K. C. Gordon, J. A. Zeitler, M. Pepper, and T. Rades, Journal of Pharmaceutical Sciences **94**, 837 (2005).
10. P. F. Taday, I. V. Bradley, D. D. Arnone, and M. Pepper, Journal of Pharmaceutical Sciences **92**, 831 (2003).
11. M. Walther, B. Fischer, M. Schall, H. Helm, and P. U. Jepsen, Chemical Physics Letters **332**, 389 (2000).
12. P. C. Upadhya, Y. C. Shen, A. G. Davies, and E. H. Linfield, Journal of Biological Physics **29**, 117 (2003).
13. M. T. Ruggiero, Journal of Infrared, Millimeter, and Terahertz Waves, https://doi.org/10.1007/s10762-019-00648-3 (2020).
14. E. R. Brown, J. E. Bjarnason, A. M. Fedor, and T. M. Korter, Applied Physics Letters **90**, 061908 (2007).
15. S. Fan, M. T. Ruggiero, Z. Song, Z. Qian, and V. P. Wallace, Chemical Communications **55**, 3670 (2019).
16. J. Sibik, K. Löbmann, T. Rades, and J. A. Zeitler, Molecular Pharmaceutics **12**, 3062 (2015).
17. O. P. Cherkasova, M. M. Nazarov, M. Konnikova, and A. P. Shkurinov, Journal of Infrared, Millimeter, and Terahertz Waves **41**, 1057 (2020).
18. X. Jing, W. P. Kevin, and S. J. Allen, The Journal of Chemical Physics **124**, 036101 (2006).
19. J. T. Kindt and C. A. Schmuttenmaer, The Journal of Physical Chemistry **100**, 10373 (1996).
20. G. Png, R. Flook, B.-H. Ng, and D. Abbott, Electronics Letters **45**, 343 (2009).
21. H. Hoshina, A. Hayashi, N. Miyoshi, F. Miyamaru, and C. Otani, Applied Physics Letters **94**, 123901 (2009).
22. Y. He, B. S. Y. Ung, E. P. J. Parrott, A. T. Ahuja, and E. Pickwell-MacPherson, Biomedical optics express **7**, 4711 (2016).
23. R. Rungsawang, Y. Ueno, and K. Ajito, *Detecting a sodium chloride ion pair in ice using terahertz time-domain spectroscopy*, Vol. 23 (2007), pp. 917–20.
24. L. Chen, G. Ren, L. Liu, P. Guo, E. Wang, L. Zhou, Z. Zhu, J. Zhang, B. Yang, W. Zhang, Y. Li, W. Zhang, Y. Gao, H. Zhao, and J. Han, The Journal of Physical Chemistry Letters **11**, 7146 (2020).
25. P. Banks, L. burgess, and M. Ruggiero, https://doi.org/10.26434/chemrxiv.13473063.v1 (2020).
26. H. Cheon, H.-J. Yang, S.-H. Lee, Y. Kim, and J.-H. Son, Scientific Reports 6, https://doi.org/10.1038/srep37103 (2016).
27. J. E. Bertie, H. J. Labbé, and E. Whalley, The Journal of Chemical Physics **50**, 4501 (1969).
28. K. Takeya, T. Fukui, R. Takahashi, and K. Kawase, Journal of Optics **16**, 094005 (2014).
29. T. Matsuoka, S. Fujita, and S. Mae, Journal of Applied Physics **80**, 5884 (1996).
30. G. P. Johari and S. J. Jones, Nature **263**, 672 (1976).
31. J. E. Bertie and E. Whalley, The Journal of Chemical Physics **46**, 1271 (1967).
32. Y. Zhou, Y. Kong, W. Fan, T. Tao, Q. Xiao, N. Li, and X. Zhu, Biomedicine & Pharmacotherapy **131**, 110731 (2020).
33. C.-H. Shen, "Chapter 12 - genome and transcriptome analysis", in *Diagnostic molecular biology*, edited by C.-H. Shen (Academic Press, 2019), pp. 303–329.
34. S.-Y. Jeong, H. Cheon, D. Lee, and J.-H. Son, Optics Express **28**, 3854 (2020).
35. K. Ajito, Y. Ueno, J.-Y. Kim, and T. Sumikama, Journal of the American Chemical Society **140**, 13793 (2018).
36. W. Withayachumnankul, B. M. Fischer, and D. Abbott, Optics Express **16**, 7382 (2008).
37. M. Nagai, H. Yada, T. Arikawa, and K. Tanaka, International Journal of Infrared and Millimeter Waves **27**, 505 (2006).
38. L. Duvillaret, F. Garet, and J. L. Coutaz, IEEE Journal of Selected Topics in Quantum Electronics **2**, 739 (1996).
39. V. Thomsen, D. Schatzlein, and D. Mercuro, Spectroscopy **18**, 112 (2003).
40. Y. C. Shen, P. F. Taday, and M. Pepper, Applied Physics Letters **92**, 051103 (2008).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.