Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 229 (2020) 117948

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

Study on glycoprotein terahertz time-domain spectroscopy based on composite multiscale entropy feature extraction method

Pingjie Huang *, Zhangwei Huang, Xiaodong Lu, Yuqi Cao, Jie Yu, Dibo Hou, Guangxin Zhang *

State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou, People's Republic of China

ARTICLE INFO

Article history:
Received 9 September 2019
Received in revised form 2 December 2019
Accepted 10 December 2019
Available online 16 December 2019

Keywords:
Terahertz time-domain attenuated total reflection spectroscopy
Glycoprotein
Tumor marker identification
Composite multiscale entropy (CMSE)
Absorption coefficient
Dielectric loss tangent

ABSTRACT

Tumor genesis is accompanied by glycosylation of related proteins. Glycoprotein is usually regarded as a tumor marker since glycoproteins are consumed remarkably more by the cancer cells than the normal ones. In this paper, the terahertz time-domain attenuated total reflection (ATR) technique is applied to inspect the glycoprotein solution from a concentration gradient of 0.2 mg/ml to 50 mg/ml. A significant nonlinear relationship between the absorption coefficient and the concentrations has been discovered. The influence of the dynamical hydration shell around glycoprotein molecules on the absorption coefficient is discussed and the phenomenon is explained by the concepts of THz excess and THz defect. In order to identify glycoproteins, features are obtained by composite multiscale entropy (CMSE) method and clustered by the K-means algorithm. The results indicate that features extracted by the CMSE method are better than the Principal Component Analysis (PCA) method in both specificity and sensitivity of recognition. Meanwhile, the absorption coefficient and dielectric loss angle tangent are more suitable for qualitative identification. Research shows that the CMSE method has important directive significance for analyzing glycoprotein terahertz spectroscopy. And it has the potential for glycoprotein related tumor markers identification using terahertz technology in medical applications.

1. Introduction

Cancer is a growing threat to human health and has become one of the most common causes of cancer-related death. According to a statistical report released by the World Health Organization (WHO), it is estimated that the number of new cases will reach 19.1 million by 2025, and about 11.5 million people die of cancer each year [1]. At present, clinical detection methods for cancer mainly include imaging examination, endoscopy and tumor marker. The technique of molecular pathology based on tumor markers has been developed rapidly in recent years. Currently, clinical cancer diagnostic tests use mainly soluble protein cancer markers (CEA, PSA, CA125, MUC1, CA15-3) [2]. The best-known biomarker is alpha-fetoprotein (AFP-L3) that is significantly elevated in hepatocellular carcinoma (HCC) [3–5].

Glycomics research has provided the significance of glycoprotein in cancer study [6]. Glycosylation of proteins is a common characteristic of tumors and affects glycoconjugates such as N-glycan and O-glycan. Nigjeh et al. adopted a library-based proteomics approach to analyze galectin-3 binding protein (LGALS3BP) from patients with PDAC and found glycosylation of LGALS3BP [7]. Krishnan et al., detected 703 proteins by using Tandem Mass Tag, providing preliminary evidence of altered glycosylation of several serum proteins prior to pancreatic cancer diagnosis [8].

Terahertz time-domain spectroscopy (THz-TDS) is a powerful non-destructive testing technique that operates in the frequency band from about 0.1 to 10 THz. Terahertz's low energy and non-ionizing properties have demonstrated important academic value and application prospects in the life sciences, especially in the detection of biomacromolecule [9–11]. Zheng et al. presented the identification of isomers by measuring the THz spectra of glucose and fructose from 0.5 to 4.0 THz at room temperature [12]. Bye et al. discussed the absorption of bovine serum albumin (BSA) in water by using terahertz spectroscopy and found it is inconsistent in terahertz absorption with Beer's law [13]. Also, Terahertz (THz) spectroscopic techniques were employed to study the anti-estrogen receptor alpha (AER-α), an important biomarker in breast cancer diagnosis. THz transmission and attenuation could be used to investigate the dielectric properties of antibody–antigen binding reactions [14]. The tumor biomarkers of AFP and CEA are detected by THz spectroscopic techniques [15,16]. In addition, terahertz time-domain attenuated total reflection spectroscopy has become an important inspection method for an aqueous solution of various tumor markers and other cancer-related substances [17–20]. For

* Corresponding authors.
E-mail addresses: huangpingjie@zju.edu.cn (P. Huang), gxzhang@zju.edu.cn (G. Zhang).

https://doi.org/10.1016/j.saa.2019.117948
1386-1425/

instance, the hydration shells of some carbohydrate polymers of commercial and biological importance were studied, employing THz-ATR and differential scanning calorimetry. The aqueous solutions show a non-proportional increase in the absorption coefficient and the hydration number, with a decrease in the carbohydrate concentration. This behavior is consistent with the "chaotropic" or "structure breaking" model of the hydration shell around the carbohydrates [20].

The nonlinear relationship of terahertz absorption with concentration is the main obstacle to detect different proteins [21,22]. Sun et al., used Principal Component Analysis (PCA) method to study the 7 different concentrations dependent terahertz spectra of hemagglutinin proteins [23]. Huang et al. built a molecule classification method based on terahertz absorption spectra by using the factor analysis approach [24].

The above research is significant in the analysis of terahertz biology spectroscopy and protein structure. For example, PCA is a powerful algorithm in terahertz spectral data dimension reduction. However, these methods are affected mainly by the parameters of the terahertz spectrum and the length of the sequence. Aiming at these problems, the CMSE method is proposed for feature extraction of terahertz parameter spectra. CMSE, a nonlinear analysis method, can characterize the complexity of signals and reveal the detailed features of sequences from different spatial and temporal scales. The method has strong robusticity on sequences [25].

In this paper, the terahertz attenuation total reflection technology is applied for detecting glycoproteins from different concentrations. The absorption coefficients of glycoprotein solutions at various concentrations are discussed and analyzed. In order to identify different glycoproteins, the CMSE method is introduced for feature extraction and the K-means algorithm is used to cluster the features. Then, the specificity obtained by applying the PCA and CMSE algorithm was compared in both optical and dielectric parameters. Finally, the effects and causes of each optical and electromagnetic parameter in glycoprotein recognition are analyzed and discussed.

2. Samples and experiment

2.1. Sample preparation

The standard glycoprotein ASF and FET used in this study were purchased from Sigma, USA. It was no further treatment. The results [26] show that PBS at PH 7.4 ensures the conformational stability of the protein and has no effect on terahertz detection. ASF and FET were stored in sterilized PBS at PH 7.4. And they were diluted with the following concentrations: 50, 25, 12.5, 6.3, 3.2, 1.6, 0.8, 0.4 and 0.2 mg ml^{-1}.

2.2. Experiment setup

The Tah7500SP terahertz time-domain spectroscopy system produced by ADVANTEST of Japan was used for detection in this study. The system laser source is produced by a fiber femtosecond laser: an output wavelength of 1550 nm, a pulse width of 50 fs, an output power of 20 mW and a repetition rate of 50 MHz. During the detection process, the system continuously pumps in nitrogen gas to reduce the influence of water vapor [27]. The experiments were carried out at room temperature (300 K) and the humidity was kept at about 2%.

In order to prevent the uneven mixing of the solution, in this study, we used a blank pipette tip to blow the cell solution evenly before we dropped the sample. In addition, we obtained 6 valid sample drops for each concentration. The content of each drop was 400 μl, each drop of the sample is dropped at the same height for keeping thickness same. Also, in order to make sure the system was stable, the sample holder was cleaned with alcohol, all results were repeated for three times and the spectrum of distilled water was tested after every 4 drops [28].

3. Method

Costa et al. introduced a multiscale entropy (MSE) method to robustly separate physiologic time series of healthy and pathologic groups [29,30]. Yong et al. tested the simulation data and Bonn epilepsy dataset by using the multivariate multiscale entropy method. The proposed method had a good performance in distinguishing correlation data [31]. In previous medical research, the MSE method is commonly used in electrocardiogram analysis. With the development of THz detection in recent years, the MSE method also has a certain application in terahertz spectrum analysis. Zhang et al. further enhanced the contrast among different biological tissues THz signals by the CMSE method. The result shows the method is very appropriate for distinguishing the THz signal which has no significant absorption peaks [32]. The CMSE algorithm is described as follows:

The $U_i = \{u_1, u_2, \cdots, u_n\}$ represents a one-dimensional series of length N. The τ is scale factor. The coarse-graining sequences

$$x_k^\tau = \left\{x_{k,1}^\tau, x_{k,2}^\tau, \cdots, x_{k,j}^\tau\right\} \tag{1}$$

where $x_{k,j}^\tau = \frac{1}{\tau}\sum_{i=(j-1)\tau+k}^{j\tau+k-1} u_i, 1 \le j \le \frac{N}{\tau} - 1, 1 \le k \le \tau$. Input dimension is set as m. Then the reconstructed vector can be expressed as follow:

$$y_k(i) = \left\{x_{k,i}^\tau, x_{k,i+1}^\tau, \cdots, x_{k,i+m-1}^\tau\right\}, i = 1, 2, \cdots \frac{N}{\tau} - m \tag{2}$$

For every i and j, the number of the distance between the vector y_k (i) and vector y_k(j) is smaller than r is counted as the formulas:

$$B_i^m(r) = \frac{1}{\frac{N}{\tau} - m - 1\{\text{number of } d[y_k(i), y_k(j)] < r\}, i \ne j, 1 \le i, j \le \frac{N}{\tau} - m} \tag{3}$$

where $d[y_k(i), y_k(j)] = Max_{t=0,1,2,\cdots,m-1}[|x_{k,i+t}^\tau - x_{k,j+t}^\tau|]$. Then calculate the average:

$$B^m(r) = \frac{1}{\frac{N}{\tau} - m \sum_{i=1}^{\frac{N}{\tau} - m} B_i^m(r)} \tag{4}$$

The input dimension is set at m + 1. By implementing above Eqs. (1)–(4), $B^{m+1}(r)$ is calculated. The sample entropy can be obtained according to the equation:

$$SampEn_k(m, r, N, \tau) = -\ln\frac{B^{m+1}(r)}{B^m(r)} \tag{5}$$

The value of CMSE is calculated:

$$CMSE(m, r, N, \tau) = \frac{1}{\tau}\sum_{k=1}^{\tau} SampEn_k \tag{6}$$

Due to the value of entropy is greatly affected by τ. So the parameter is found by traversing method based on a series of length N. Finally, the values of entropy at different scales τ are sorted in descending order for all the input U_i:

$$\left\{number\ of\ \gamma * CMSE(m, r, N, \tau \pm 1) \underset{\geq}{\leq} CMSE(m, r, N, \tau)\right\} \tag{7}$$

where $\{\gamma | 0 \le \gamma \le 1\}$, $\{\tau | 1 \le \tau \le N\}$.The CMSE algorithm flow chart is shown in Fig. 1 below:

Common classification methods are SVM, KNN and so on. The SVM method had higher classification accuracy in distinguishing breast

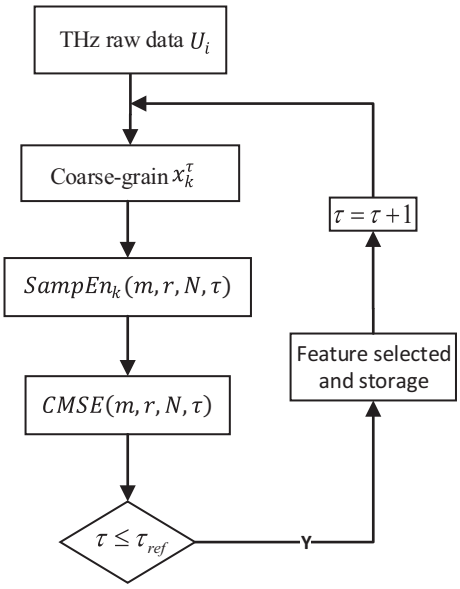

Fig. 1. CMSE algorithm flow chart.

tumors from healthy tissue [26]. However, it is difficult to be divided into training and test set for small sample data. And the supervised learning methods are easy to be over-fitting. Unsupervised learning methods can avoid the above problems. Therefore, K-means clustering is selected for glycoprotein qualitative identification. The glycoprotein recognition framework is shown in Fig. 2 below:

4. Results and discussions

4.1. Glycoproteins expression in the absorption coefficients spectrum

The experimental results are shown in Fig. 3. The absorption coefficients of asialofetuin (ASF) and fetuin (FET) exhibit a nonlinear relationship with concentration, which follows Beer-Lambert's law [33]. The absorption coefficient of ASF with a concentration of about 0.2 and 50 mg/ml is relatively large and is relatively small in 1.6 mg/ml. However, the absorption of FET at a concentration of about 25 mg/ml is relatively large and is relatively small at 50 mg/ml.

In order to explain the relationship between the terahertz absorption coefficient, frequency and concentration, Fig. 4 shows the three-dimensional graph. For two kinds of glycoproteins, the absorption coefficient from 0.2 mg/ml to 25 mg/ml both increase at the early storage time and then decrease later. And the FET decreases again from 25 mg/ml to 50 mg/ml. The trend is more significant especially in the high-frequency band above 1THz, which is consistent with the observation frequency band described by Xu et al. [34].

For further describes independently the relationship between absorption coefficient and concentration, the 1.1THz frequency point was selected to show the corresponding relationship. The results obtained by second-order linear interpolation are shown in Fig. 5. The trend is consistent basically with the measurement of protein by Li et al. [35]. The maximum absorption coefficient and its corresponding concentration are different for both glycoproteins. FET reaches its maximum absorption (about 245 cm^{-1}) at a concentration of about 25 mg/ml. But ASF is >50 mg/ml.

The concepts of THz excess and THz defect proposed by Benjamin Born et al. [36] are widely used for explaining this phenomenon.

Taking FET as an example, the solution can be divided into solute and free water at a low concentration of 0.2 mg/ml to 0.4 mg/ml. In general, the absorption of terahertz waves by hydrogen bond network vibrations is significantly more than that of biomolecules in aqueous solution. So, the terahertz absorption coefficient decreases as the concentration of solute increases.

At high concentrations, the solution can divide into three parts: solute, free water and hydration water. The terahertz absorption of the hydration water above that of free water and both are much higher than the solute. As the concentration increases, free water continuously adsorbs on the surface of the glycoprotein to become hydration water, thereby causing the terahertz absorption coefficient to gradually increasing (Fig. 5 FET at 0.4–25 mg/ml). Until the solute reaching a certain concentration, the glycoprotein hydration layer overlaps with each other. At this time, the hydration layer does not play a leading role, so the absorption coefficient decreases with an increase in the solute concentration (Fig. 5 FET at 25 mg/ml to 50 mg/ml).

Viewed from the trend of the absorption coefficient and solution concentration, ASF and FET are basically consistent. But, there are significant differences in the concentration corresponding to the maximum absorption. Since the sample of ASF and FET is only differed

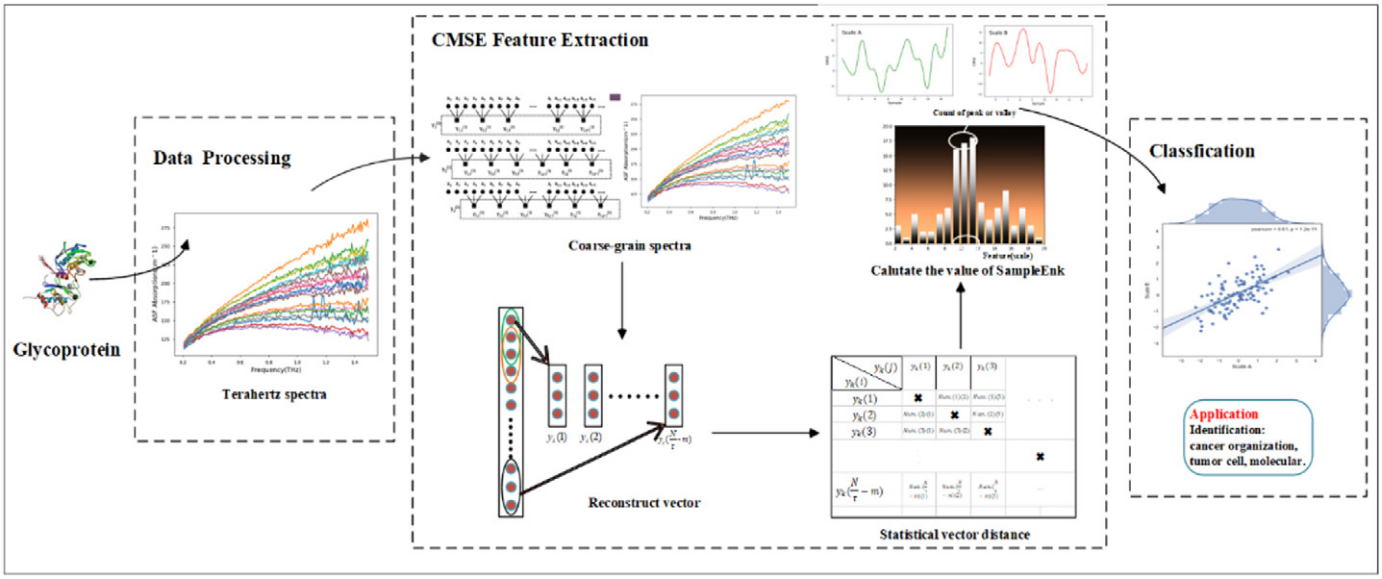

Fig. 2. Glycoprotein recognition framework.

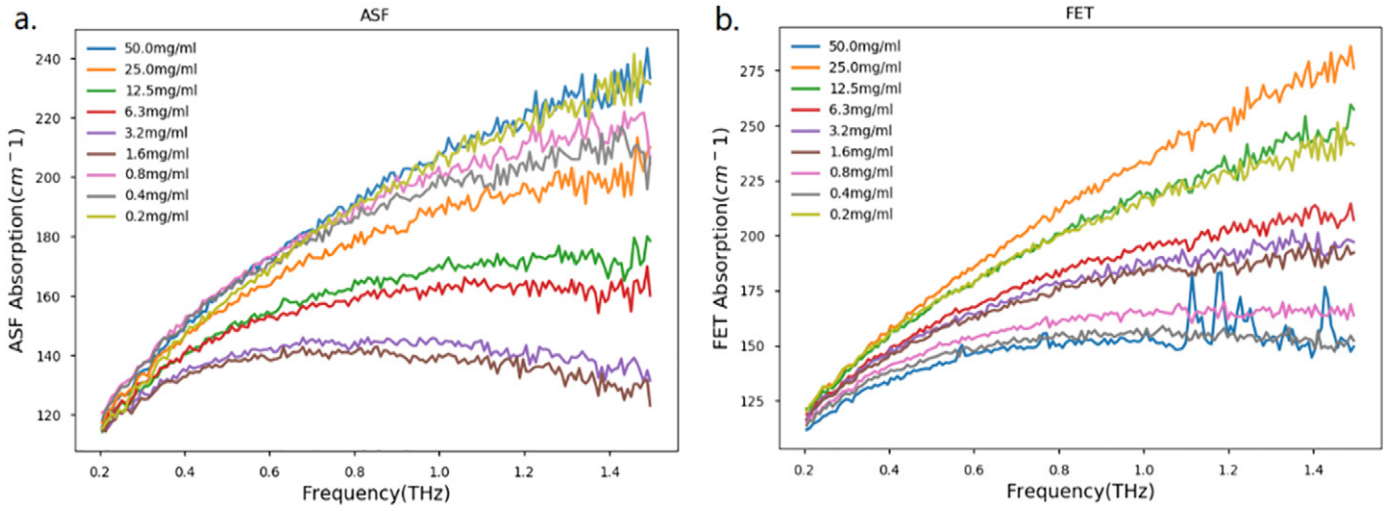

Fig. 3. Average absorption spectra of (a) ASF and (b) FET.

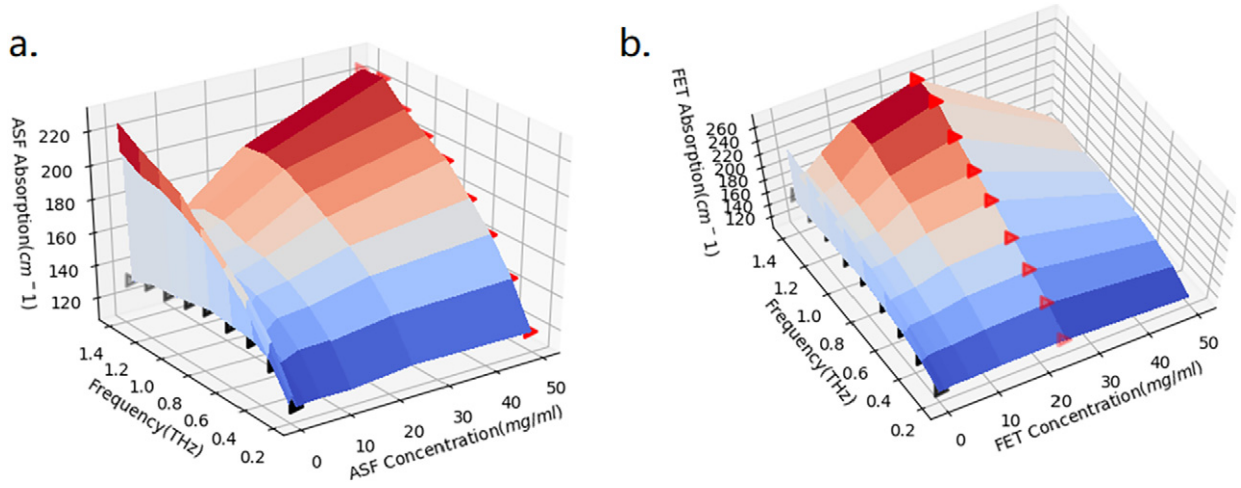

Fig. 4. A three-dimensional graph of the terahertz absorption coefficient, frequency and glycoprotein concentration between the sample of (a) ASF. (b) FET.

from the sialic acid group, and all of the experimental conditions and treatments are the same. So this difference is most likely caused by terminal sialic acid.

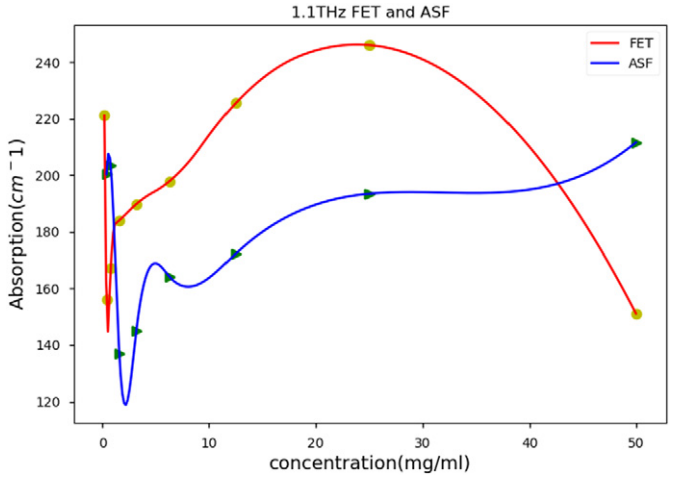

Fig. 5. Relationship between absorption coefficient and solubility of ASF and FET at 1.1THz point.

4.2. Glycoprotein identification

The absorption peak is one of the important features in traditional spectral analysis. However, the complex molecular structure, the broad total response band and the superposed absorption bands make glycoproteins have no significant absorption peak and result in the difficulty of identifying. CMSE method is an effective way to solve this problem. Taking the absorption spectrum as an example, the parameters of the CMSE method are m = 2, r = 0.2*std. (std is the standard deviation of the original sequence). The calculation results are shown in Fig. 6. The FET and ASF both have a rising trend in scale from 1 to 8. However, there is a significant difference in scale 9 and 11. The entropy of FET has peaked but ASF has a valley at scale = 9 in Fig. 6(a). As shown in Fig. 6(b), the entropy of FET and ASF has a different result in different concentrations. At the same time, there are significant differences in the value of entropy for the same substance at different concentrations between Fig. 6(a) and (b). In summary, the significant entropy peaks or valleys were constructed through the conversion of the CMSE method in the different samples.

CMSE is a nonlinear analysis method that characterizes the complexity of signals from multiple scales. PCA is a dimension reduction method that converting the original variable into a small number of new variables. It can characterize the data differences of original variables in the maximal degrees. Fig. 7 shows the 2D and 3D features, which

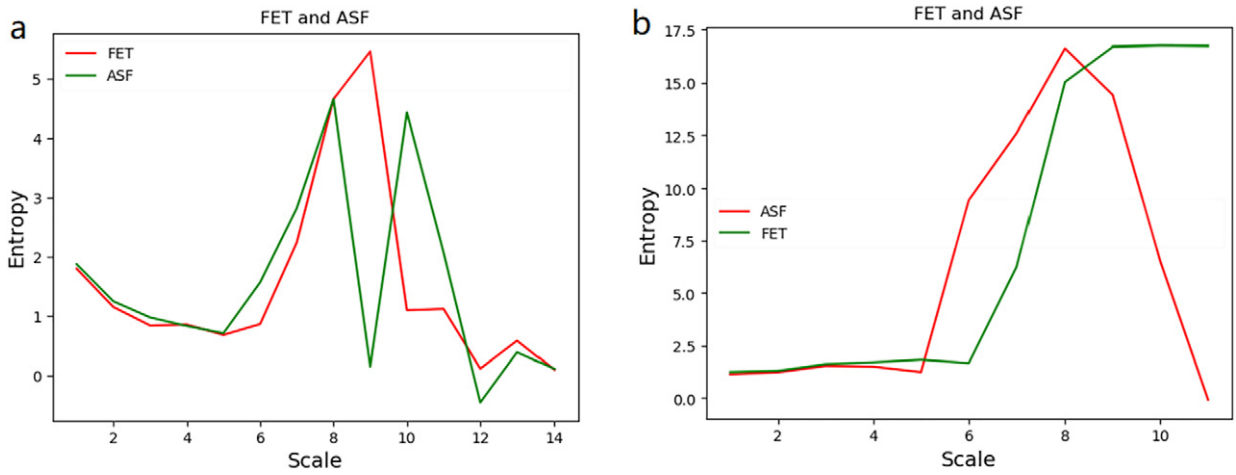

Fig. 6. (a) ASF and FET with concentration of 12.5 mg/ml (b) FET is 0.8 mg/ml and ASF is 50 mg/ml.

extracted respectively by CMSE and PCA algorithms. The clustering result of the 2D feature extracted based on PCA is significantly weaker than the CMSE. A potential explanation to the phenomenon is that PCA maps the input x to an approximately reconstructed output r, which implemented by the formula $r = \omega^{T}\omega * x$. Discarding multiple eigenvectors corresponding to small eigenvalues can reduce the dimensionality. But it also loses partial information. Therefore, the classification accuracy of clustering is significantly improved by increasing the number of principal components. Nevertheless, the CMSE method is paid attention to extracting specific features from a sequence. It has advantages in terahertz nonlinear spectra analysis. Both algorithms have a good clustering accuracy based on the 3D features in Fig. 7. But the CMSE algorithm is more closely in terms of signal density.

In order to explore the most suitable terahertz spectral parameters for classification, the algorithm showing in Fig. 2 was used to identifying. The results are shown in Table 1. The refractive index is slightly better than the complex permittivity. But both them exhibits a relatively low recognition accuracy. The absorption coefficient and dielectric loss tangent are significantly better than other parameters in the identification precision. The cause may be the diversity in the formation of

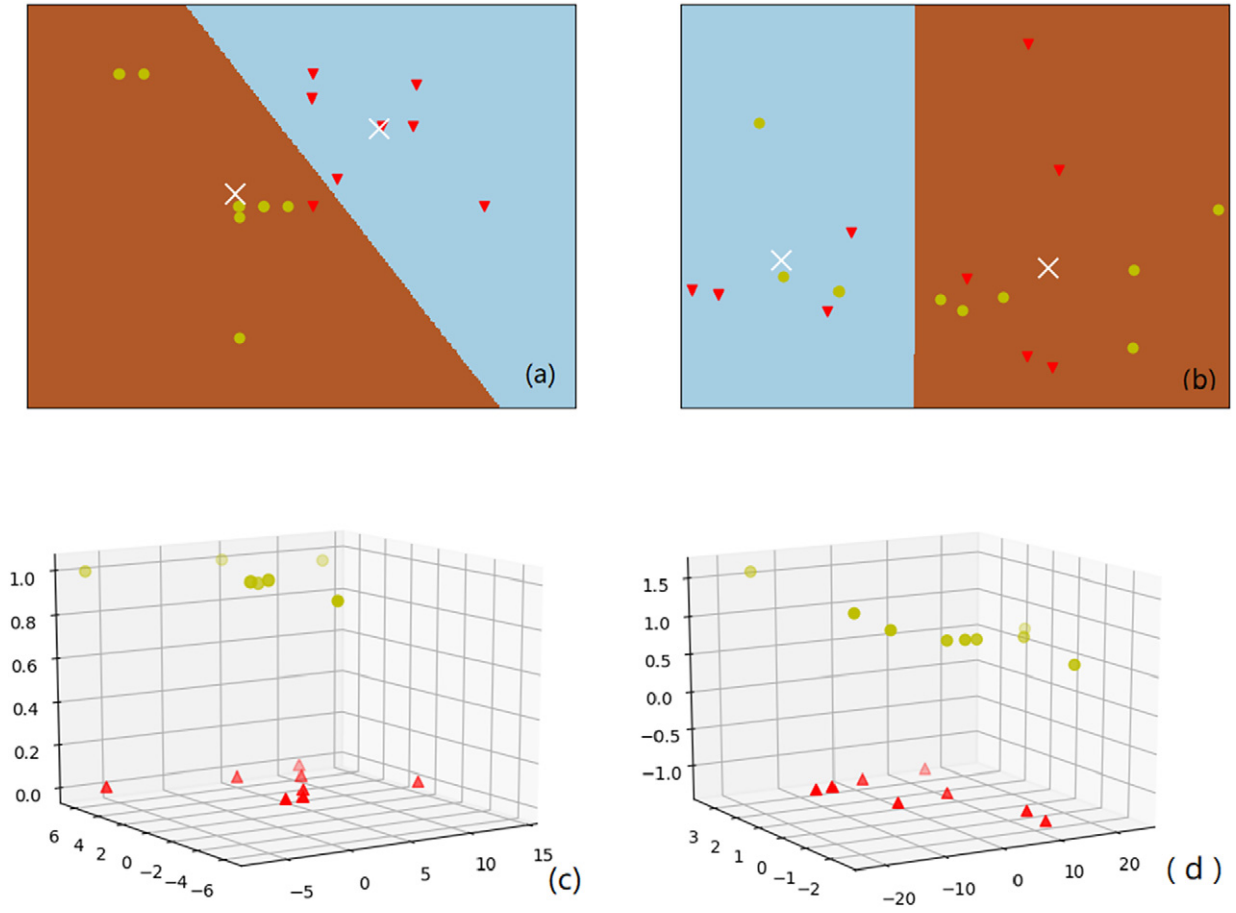

Fig. 7. Yellow point is FET and red point is ASF (a) 2D Feature by CMSE (b) 2D Feature by PCA (c) 3D Feature by CMSE (d) 3D Feature by PCA.

Table 1
Prediction accuracy based on different terahertz spectral parameters.

Parameters	CMSE			PCA		
	Specificity	Sensitivity	Accuracy	Specificity	Sensitivity	Accuracy
Absorption coefficient	90%	100%	94.44%	55.6%	66.67%	55.6%
Refractive index	77.8%	90%	83.33%	66.7%	77.8%	72.3%
Complex permittivity(real)	66.7%	77.8%	72.3%	55.6%	77.8%	66.7%
Complex permittivity(imaginary)	77.8%	77.8%	77.8%	66.7%	77.8%	76.5%
Dielectric loss tangent	90%	100%	94.44%	90%	90%	90%

hydrated shells of different glycoproteins. The dielectric loss tangent represents the internal loss of electromagnetic energy when it is converted into other energy and is very sensitive to the dynamic changes of the glycoprotein.

In general, the performance of the CMSE algorithm is better than that of the PCA in the above terahertz spectral parameters. So, the CMSE algorithm has specific advantages for feature extraction of nonlinear spectra.

5. Conclusions

In this paper, the THz-ATR technique is applied to detect glycoproteins at different concentrations. The experiment results show that the absorption coefficient of the sample has a nonlinear relationship with concentration. The main reason comes from the state of the bound water, which plays an important role in the absorption of terahertz. The bound water also leads to the difficulty in the identification of glycoprotein. The CMSE algorithm is used to extracted features in terahertz parameter spectra. A small number of features distinguish glycoproteins effectively. By comparing with the PCA method, it is found that the CMSE method has a good advantage in the nonlinear spectra analysis. Especially for some molecular and tissue detection, due to individual differences, terahertz optical parameters are easily interfered with by the concentration of substances and so on. Through multi-scale information entropy, we can pay more attention to the trend of sequence changes and mitigate these disturbances. Also, the absorption coefficient and dielectric loss tangent are more accurate for qualitative analyzed. On the other hand, due to the complexity of glycoprotein, the mechanism of glycoprotein and water still need further investigation. In addition, the algorithm not be compared through a large amount of data. But we found that CMSE pays more attention to the changes in the sequence itself, can provide the characteristics of individual sequence changes, and may help alleviate the problem of individual differences in tumor marker detection in the future. It provides a method to identify cancer for future medical applications.

Contributions

Zhangwei Huang, Xiaodong Lu and Yuqi Cao performed the measurements provided in this manuscript. Pingjie Huang, Jie Yu and Dibo Hou presented the main technology ideas and mapping methods. Guangxin Zhang provided overall supervision and guidance on the experimental and theoretical aspects. All authors contributed to the writing and polishing of the manuscript.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grant No. 61473255, 51504228, 61873234).

Declaration of competing interest

The authors declare no conflicts of interest.

References

[1] S. McGuire, World Cancer Report 2014, World Health Organization, International Agency for Research on Cancer, WHO Press, Geneva, Switzerland, 2015https://doi.org/10.3945/an.116.012211(2016).
[2] N. Taniguchi, Toward cancer biomarker discovery using the glycomics approach, Proteomics 8 (16) (2008) 3205–3208, https://doi.org/10.1002/pmic.200890056.
[3] G.Y. Locker, S. Hamilton, J. Harris, et al., ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer, J. Clin. Oncol. 24 (33) (2006) 5313–5327, https://doi.org/10.1200/JCO.2006.08.2644.
[4] L. Harris, H. Fritsche, R. Mennel, et al., American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer, J. Clin. Oncol. 25 (33) (2007) 5287–5312, https://doi.org/10.1200/JCO.2007.14.2364.
[5] N.L. Henry, D.F. Hayes, Cancer biomarkers, Mol. Oncol. 6 (2) (2012) 140–146, https://doi.org/10.1016/j.molonc.2012.01.010.
[6] S.S. Pinho, C.A. Reis, Glycosylation in cancer: mechanisms and clinical implications, Nat. Rev. Cancer 15 (9) (2015) 540, https://doi.org/10.1038/nrc3982.
[7] E.N. Nigjeh, R. Chen, Y. Allen-Tamura, et al., Spectral library-based glycopeptide analysis–detection of circulating galectin-3 binding protein in pancreatic cancer, PROTEOM. Clin. Appl. 11 (9–10) (2017), 1700064. https://doi.org/10.1002/prca.201700064.
[8] S. Krishnan, H.J. Whitwell, J. Cuenco, et al., Evidence of altered glycosylation of serum proteins prior to pancreatic cancer diagnosis, Int. J. Mol. Sci. 18 (12) (2017) 2670, https://doi.org/10.3390/ijms18122670.
[9] D.B. Hou, X. Li, J.H. Cai, et al., Terahertz spectroscopic investigation of human gastric normal and tumor tissues, Phys. Med. Biol. 59 (18) (2014) 5423, https://doi.org/10.1088/0031-9155/59/18/5423.
[10] L.J. Xie, W. Gao, J. Shu, et al., Extraordinary sensitivity enhancement by metasurfaces in terahertz detection of antibiotics, Sci. Rep. 5 (2015) 8671, https://doi.org/10.1038/srep08671.
[11] Y. Zhu, S. Zhuang, Terahertz electromagnetic waves emitted from semiconductor investigated using terahertz time domain spectroscopy, Chin. Opt. Lett. 9 (11) (2011) 110007, https://doi.org/10.3788/COL201109.110007.
[12] Z.P. Zheng, W.H. Fan, Y.Q. Liang, et al., Application of terahertz spectroscopy and molecular modeling in isomers investigation: glucose and fructose, Opt. Commun. 285 (7) (2012) 1868–1871, https://doi.org/10.1016/j.optcom.2011.12.016.
[13] J.W. Bye, S. Meliga, D. Ferachou, et al., Analysis of the hydration water around bovine serum albumin using terahertz coherent synchrotron radiation, J. Phys. Chem. A 118 (1) (2013) 83–88, https://doi.org/10.1021/jp407410g.
[14] M. Li, T. Chang, D. Wei, et al., Label-free detection of anti-estrogen receptor alpha and its binding with estrogen receptor peptide alpha by terahertz spectroscopy, RSC Adv. 7 (39) (2017) 24338–24344.
[15] Z. Geng, X. Zhang, Z. Fan, et al., A route to terahertz metamaterial biosensor integrated with microfluidics for liver cancer biomarker testing in early stage, Sci. Rep. 7 (1) (2017) 16378.
[16] X. Lu, P. Xie, Y. Sun, Label-free Detection of the Carcinoembryonic Antibody Protein Based MoS2 Nanosheets Using Terahertz Spectroscopy[C]//Infrared, Millimeter-Wave, and Terahertz Technologies V, 10826, International Society for Optics and Photonics, 2018 1082607.
[17] K. Shiraga, Y. Ogawa, N. Kondo, et al., Evaluation of the hydration state of saccharides using terahertz time-domain attenuated total reflection spectroscopy, Food Chem. 140 (1–2) (2013) 315–320.
[18] F. Wang, D. Zhao, H. Dong, et al., Terahertz spectra of DNA nucleobase crystals: a. joint experimental and computational study, Spectrochim. Acta A Mol. Biomol. Spectrosc. 179 (2017) 255–260.
[19] J. Shikata, H. Handa, A. Nawahara, et al., Terahertz ATR spectroscopy of liquids using THz-wave parametric sources[C]//lasers and electro-optics - Pacific Rim, CLEO/Pacific Rim 2007, Conference on. IEEE 2007 2007, pp. 1–2.
[20] J.A. Morales-Hernández, A.K. Singh, S.J. Villanueva-Rodriguez, et al., Hydration shells of carbohydrate polymers studied by calorimetry and terahertz spectroscopy, Food Chem. 291 (2019) 94–100.
[21] Q. Sun, Y. He, K. Liu, et al., Recent advances in terahertz technology for biomedical applications, Quant. Imaging Med. Surg. 7 (3) (2017) 345, https://doi.org/10.21037/qims.2017.06.02.
[22] X. Yang, X. Zhao, K. Yang, et al., Biomedical applications of terahertz spectroscopy and imaging, Trends Biotechnol. 34 (10) (2016) 810–824, https://doi.org/10.1016/j.tibtech.2016.04.008.
[23] Y. Sun, J. Zhong, J. Zuo, et al., Principal component analysis of terahertz spectrum on hemagglutinin protein and its antibody, Acta Phys. Sin. 64 (16) (2015), 168701. https://doi.org/10.7498/aps.64.168701.

[24] J. Huang, J. Liu, K. Wang, et al., Classification and identification of molecules through factor analysis method based on terahertz spectroscopy, Spectrochim. Acta A Mol. Biomol. Spectrosc. 198 (5) (2018) 198–203, https://doi.org/10.1016/j.saa.2018.03.017.
[25] W. Jun, L. Qian, Multiscale entropy based study of the pathological time series, Chin. Phys. B 17 (12) (2008) 4424DOI: 1674-1056/2008/17(12)/4424-04.
[26] B.C.Q. Truong, H.D. Tuan, A.J. Fitzgerald, et al., A dielectric model of human breast tissue in terahertz regime, IEEE Trans. Biomed. Eng. 62 (2) (2015) 699–707, https://doi.org/10.1109/TBME.2014.2364025.
[27] R. Albert, C. Xi, Zhang. Self-referenced method for terahertz wave time-domain spectroscopy, Opt. Lett. 36 (17) (2011) 3308–3310, https://doi.org/10.1364/OL.36.003308.
[28] qi C. Yu, et al., Terahertz spectral unmixing based method for identifying gastric cancer, Phys. Med. Biol. 63 (3) (2018) 35016, https://doi.org/10.1088/1361-6560/aa9e1a.
[29] M. Costa, A.L. Goldberger, C.K. Peng, Multiscale entropy analysis of biological signals, Phys. Rev. E 71 (2) (2005), 021906. https://doi.org/10.1103/PhysRevE.71.021906.
[30] M. Costa, A.L. Goldberger, C.K. Peng, Multiscale entropy analysis of complex physiologic time series, Phys. Rev. Lett. 89 (6) (2002), 068102. https://doi.org/10.1103/PhysRevLett.89.068102.
[31] H.X. Yong, J. Cui, W. Hong, H.J. Liang, Automatic classification of epileptic electroencephalogram signal based on improved multivariate multiscale entropy, J. Biomed. Eng. 32 (2) (2015) 256–262, https://doi.org/10.7507/1001-5515.20150047.
[32] R. Zhang, Y. He, K. Liu, et al., Composite multiscale entropy analysis of reflective terahertz signals for biological tissues, Opt. Express 25 (20) (2017) 23669–23676, https://doi.org/10.1364/OE.25.023669.
[33] H.X. Su, Z.H. Zhang, X.Y. Zhao, et al., The Lambert-Beer's law characterization of formal analysis in Terahertz spectrum quantitative testing, Spectrosc. Spectr. Anal. 33 (12) (2013) 3180, https://doi.org/10.3964/j.issn.1000-0593(2013)12-3180-07.
[34] Y. Xu, M. Havenith, Perspective: watching low-frequency vibrations of water in biomolecular recognition by THz spectroscopy, J. Chem. Phys. 143 (17) (2015), 170901. https://doi.org/10.1063/1.4934504.
[35] M. Li, T. Chang, D. Wei, et al., A label-free detection of anti-estrogen receptor alpha and its binding with estrogen receptor peptide alpha by terahertz spectroscopy, RSC Adv. 7 (39) (2017) 24338–24344, https://doi.org/10.1039/C6RA28754A.
[36] B. Born, M. Havenith, Perspective: terahertz dance of proteins and sugars with water, J. Infrared Millimeter Terahertz Waves 30 (12) (2009) 1245–1254, https://doi.org/10.1007/s10762-009-9514-6.