

Identification of hemagglutinating antigen-antibody interaction using principle component analysis method in terahertz frequency

Yiwen SUN, Junlan ZHONG and Shengxin YANG

National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong, Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Department of Biomedical Engineering, School of Medicine, Shenzhen University, Shenzhen 518060, China

Abstract—In this paper, we concerned with the spectrum analysis of binding interaction of recombinant hemagglutinating (HA) antigen against its own antibody in terahertz frequency using principle component analysis (PCA). 21 liquid samples with different component were divided into positive and negative control groups and measured using THz-TDS in transmission geometry. The spectral pretreatment methods were adopted to make our results more explicit in this study. As a result, HA complexes are distinguished by different scores which indicate the PCA method can be used to identify the antigen-antibody binding interaction in THz frequency.

I. INTRODUCTION

NOWADAYS terahertz spectroscopy is the state-of-the-art techniques and demonstrated its potential applications on varieties fields such as non-destructed detection, quality control and sample identification. Several authors have succeeded to prove the frequency-dependent collective structural vibration of proteins lay in the terahertz frequency range, and the vibrational modes are thought to be essential for conformational dynamics of protein function, which allows the probing biological materials for identification and for medical diagnostics.

Influenza (flu) is a respiratory infection in mammals and birds, which remains a major medical problem and is a constant threat to human health. Influenza viruses are characterized by two types of surface antigen hemagglutinin (HA) and neuraminidase (NA) that they carry [1]. Particularly, H9N2 subtype influenza A virus is recognized as the most possible pandemic strain, as it has crossed the species barrier, infecting swine and humans asymptotically [2]. HA protein comprises over 80% of the envelope proteins present in the virus particle and is involved in two major functions: recognition of target cells by binding to their sialic acid-containing receptors, and fusion of the viral and the endosomal membranes succeeding endocytosis. In this paper we use principle component analysis (PCA) to study the concentration dependent terahertz spectroscopy of HA protein, as well as the detection of binding interaction of HA with the broadly neutralizing monoclonal antibody F10 in liquid phase. The irrelevant antibody (irmAb) was used as negative control in this study

II. MATERIALS AND METHODS

Terahertz spectra were acquired using a commercial Z-3™ Time Domain THz spectrometer (Zomega THz Corp., Troy, USA). In this system optical excitation is achieved by a mode-locked Mai-Tai laser which emits less than 120 fs pulses centered at a wavelength of 800 nm, with an 80 MHz repetition rate and an average power of ~1W. The measurements were conducted in the typical transmission geometry at room temperature (~298 K), the optics were purged using nitrogen

gas to remove the water vapour from the air to decrease the humidity down to less than 1%. The usable frequency range of the system is from about 0.1 THz to 2.5 THz but the valid range tends to decrease when the sample is in an aqueous phase due to signal attenuation by the sample.

The sample holder was fabricated from TOPAS™ 5013L-10, this material (which is a Cyclic Olefin Co-polymer) was chosen as it has very low attenuation at THz frequencies [3]. The hydrated protein formulation was pipetted into the sample holder with a liquid layer which is of the order of a few micrometers thick for the THz measurement. The clean homogeneous empty sample cell was also measured as a reference (refractive index ~1.53) enabling the spectroscopic properties of the sample to be accurately determined. Each sample was measured 4 times. For each measurement, the sample was extracted and re-pipetted in the sample holder. By averaging the spectra, systematic errors produced by wrong positioning, as well as present heterogeneities in the sample, were minimized.

The recombinant hemagglutinin (HA) protein was purchased from Sino Biological (Cat no. 11229-V08H). Briefly, the DNA sequence encoding the extracellular domain of Influenza A virus (A/HongKong/1073/1999; H9N2) hemagglutinin was synthesized and expressed in Human Embryonic Kidney 293 cells. The recombinant HA protein was not infectious and had no biohazard risk for any in-vitro study. Fully human monoclonal antibody F10 was expressed in 293T cells using TCAE5.3 vector [4] containing corresponded immunoglobulin heavy chain (VH) and light chain (VL) gene and purified using protein A beads (GE Healthcare). The concentration of the proteins was determined using Protein A280 application module by NanoDrop spectrophotometer (Thermo Scientific). Fully human anti-CXCR4 monoclonal antibody (purchased from NBGen) was used as irrelevant antibody control (irmAb) in this study. All of the H9 HA, F10 antibodies and irmAb were stored in sterilized phosphate buffered saline (PBS) at pH7.4. H9 HA was prepared at the following concentration using PBS as diluents: 7.5, 15, 29, 56, 113, 225 and 430 ug/ml.

III. RESULTS AND ANALYSIS

21 liquid samples with different mass ration or component as mentioned above were divided into positive and negative control groups and measured using THz time-domain spectrometers. In order to investigate the feasibility of automatic spectral recognition of the effect on antibody-antigen interaction by terahertz spectroscopy, PCA was employed to cluster spectra.

PCA was proposed by Karl Pearson in 1901[5], and then well established in many fields of spectroscopy. PCA is a useful statistical technique that has found application in fields such as

face recognition and image compression, and is a common technique for finding patterns in data of high dimension [6]. Its method is mainly through the decomposition of the covariance matrix characteristics to derive data for the main components (feature vectors) and their weights. PCA can use fewer new variables of original ones to figure out the difficulty of analysis incapable as superposition of spectral band.

Spectrum pretreatment and band selection plays a vital role in the THz spectroscopy analysis before using PCA, due to the original spectrum contains a large amount of interference information. In order to compress variables and extract useful information, we use a variety of pretreatment methods, such as second derivative, multiplicative scatter correction (MSC), least squares polynomial fitting derivation, standard normalization, smoothing, moving window median filtering before PCA analysis. We even considered MSC + smoothing + SG second derivative + median filtering as the optimized pretreatment method finally.

Figure 1 indicates the absorption spectrum of HA proteins with 7 different concentrations in terahertz frequency from 0.1 to 1.4 THz and the principal component scores plots before and after the optimized pretreatment method. The correlations between the concentrations of the 7 HA samples and the scores of principal component 1(PC1) were presented as the inset of the Figure 1(d).

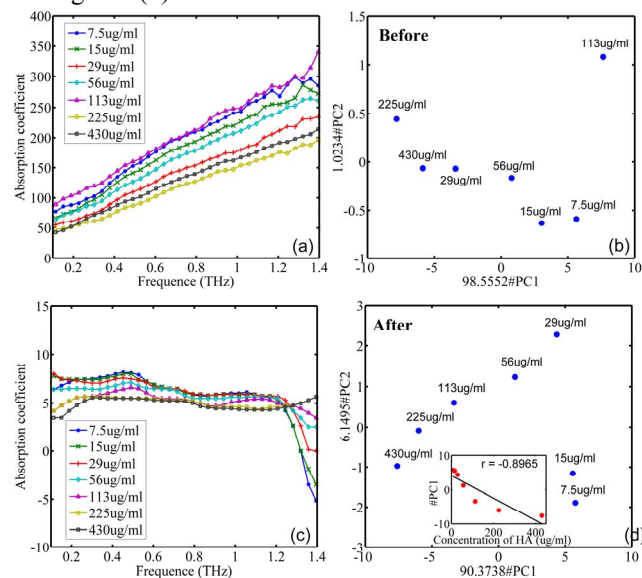


Fig. 1. THz absorption spectrum of HA proteins and the principal component scores plots (a) original spectrum; (b) principal component scores plot of original spectrum; (c) preprocessed data; (d) principal component scores plot of preprocessed data.

The correlation factors r of refractive index (n), absorption coefficient (a), reduced absorption cross-section (RACS) and dielectric loss angle tangent ($\tan\delta$) were calculated respectively for correlation investigation in order to find the most sensitivity parameters for THz spectroscopy analysis as shown in Table 1. Within the significant correlation range, we further divide the correlation into three category, moderate to strong ($|r|>0.5$), weak to moderate ($0.3 < |r| < 0.5$), and weak ($|r| < 0.3$).

Table 1 Correlation coefficient of HA's concentration in different parameter with PC1 and PC2

	n	a	RACS	$\tan\delta$
PC1	0.7014	-0.8965	-0.9265	0.8328
PC2	-0.6561	-0.1957	0.2738	0.3093

As for the antigen-antibody interaction, principal component scores plots of refractive index, absorption coefficient, reduced absorption cross-section and dielectric loss angle tangent of HA, H9-F10 and HA/irmAb were further compared before and after spectrum pretreatment. Taking the principal component scores plots of dielectric loss angle tangent as the best example was presented in Figure 2.

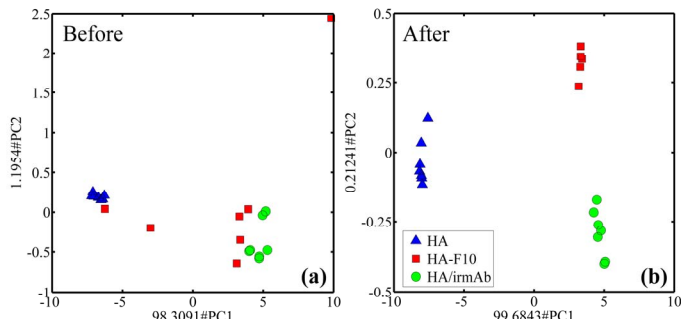


Fig. 2. Principal component scores plots for dielectric loss angle tangent of HA, H9-F10 and HA/irmAb in terahertz frequency (a) before and (b) after spectrum pretreatment.

The results indicate that the reduced absorption cross-section presents the highest correlation response with the concentration variations of HA protein, and the dielectric loss angle tangent appears to be more appropriate on qualitative analysis of HA-antibody binding interaction. PCA method provides a feasible and effective way to find the sensitive parameters for the further analysis on the function of protein and the antigen-antibody interaction using terahertz spectroscopy, whereas an appropriate pretreatment method is required.

IV. SUMMARY

This study may provide a feasible and effective way to study the element of protein and the processes of antigens with antibodies.

REFERENCES

- [1] S. Jianhua, C.H. William, P. Sandra, G. Wei, A. Daniel, et. al., "Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses," *Nature Structural & Molecular Biology*, 16(3), 265-73, 2009.
- [2] T. Horimoto, and Y. Kawaoka, "Pandemic Threat Posed by Avian Influenza A Viruses," *Clinical Microbiology Reviews*, 14(1), 129-149, 2001.
- [3] J. Balakrishnan, B. M. Fischer and D. Abbott, "Sensing the hygroscopicity of polymer and copolymer materials using terahertz time-domain spectroscopy," *Applied Optics*, 48, 2262-2266, 2009.
- [4] M. E. Reff, K. Carner, K. S. Chambers, P. C. Chinn, J. E. Leonard, R. Raab, R. A. Newman, N. Hanna and D. R. Anderson, "Depletion of B cells in vivo by a chimeric mouse human monoclonal antibody to CD20." *Blood*, 83, 435-445, 1994.
- [5] K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine*, 2, 559-572, 1901.
- [6] L.I. Smith, *A tutorial on Principal Components Analysis*, 2002.