The Terahertz Identification of the Mixtures of Amino Acids by Principle Components Analysis

Jian Zuo[a], Qingli Zhou[a], Fei Yu[a], Kaijun Mu[b], Liangliang Zhang[a], Zhenwei Zhang[a], and Cunlin Zhang[a]
[a]Physics of Department, Capital Normal University, Key Laboratory of Terahertz Optoelectronics, Ministry of Education, Beijing Key Laboratory for Terahertz Spectroscopy and Imaging, Beijing, 100048, P.R.China
[b]School of Physical Science and Engineering, Zhengzhou University, Zhengzhou 450052, P.R.China

***Abstract*—Terahertz time-domain spectroscopy has been used to analyze the mixed samples of different mass ratio of L-Tyrosine and L-Phenylalanine. The total dipoles of theses mixtures are given by absorption line shape function. Moreover, the statistical method of principle components analysis is employed to spectral recognition of these samples.**

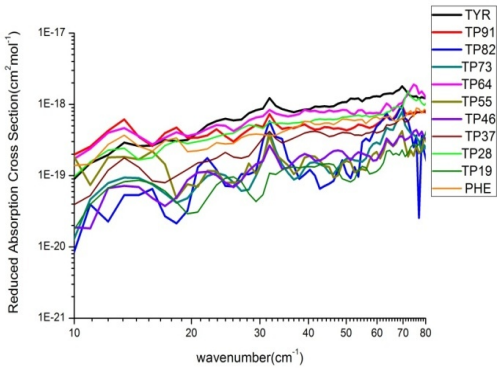
I. Introduction And Background

TERAHERTZ spectroscopy technology is increasingly being used in a host of pharmacology, medical and biological applications. The extracted spectral parameters in terahertz time-domain spectroscopy (THz-TDS) include the absorption coefficient and the refractive index. However, due to scattering from the tablets surface and inhomogeneous within the sample, there are different exhibitions of the absorption features even for the same sample, so that the absorption dependent identifications of the biomolecules are not always valid. Here, absorption line-shape function (ALF) is introduced. ALF was defined as the physical parameter which is proportional to the product of the absorption coefficient and refractive index[1]. The statistical combination of the two physical quantities, n and α, enables us to discuss rigorously dynamics of the total dipole moments of the compounds in terahertz region. It can contribute to distinguish one chemical compound with another with similar function-groups or the mixtures with different mass ratios. Experiments are carried out on L-Tyrosine and L-Phenylalanine and the different mass ratios of their mixtures as an example of the application of ALF.

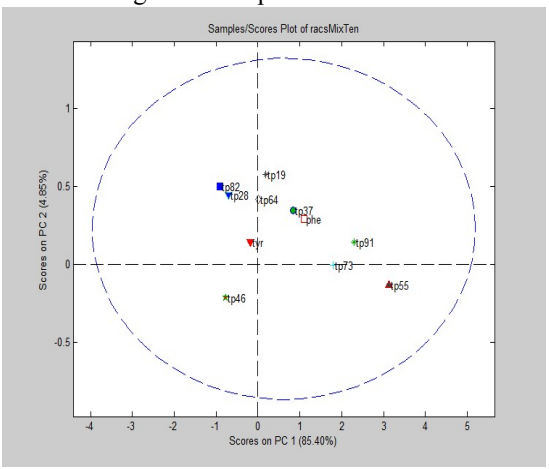
Principle component analysis (PCA) is a statistical technique which reduces the number of dimensions within the data while retaining as much of the overall variation as possible. It is usually used to identify the groups within the data. In this work, the mixtures of different mass ratios are identified by PCA. To visualize spectral differences, the scores of different PCs can be plotted against one another in a scatter plot. The samples that are similar and hence have closely matching spectra will be located in one cluster while samples that contain differences form separate clusters. The first two or three principal components extracted from the data can often be used to interpret the main phenomena in the data since these components usually contain most of the systematic variation present in the data.

II. RESULTS

The absorption line-shape function curves of different mass ratio mixtures are given in figure 1. The slopes of these curves for eleven mixtures are different.



The PCA analysis of ALF values from these mixtures was shown in figure 2. The PCs- PC1 and PC2 of the data set were found to describe 90.25% of the variance within the data. The differences among these samples are obvious.



This work was funded by the Science and Technology Projects of Beijing Municipal Commission of Education (No. 11224010011), Excellent Persons Projects of the Organization Department of the CPC Beijing Municipal Committee (No. 2010D005016000010), Project 11204190 supported by National Natural Science Foundation of China.

References

[1] Donald A. Mcquarrie, "Statistical Mechanics", 2nd, University Science Books, 2000, pp. 467–476.

978-1-4673-4717-4/13/$31.00 ©2013 IEEE