

文章编号: 0258-7025(2009)09-2448-07

# 基于 PCA 和 BP 网络的地中海贫血红细胞拉曼光谱判别

陈秀丽<sup>1,2</sup> 王桂文<sup>1</sup> 陶站华<sup>1</sup> 刘军贤<sup>2</sup> 姚辉璐<sup>1</sup> 黄庶识<sup>1</sup> 黎永青<sup>3</sup>

(<sup>1</sup>广西科学院生物物理实验室, 广西南宁 530003; <sup>2</sup>广西师范大学物理与电子工程学院, 广西桂林 541004; <sup>3</sup>美国东卡罗莱纳大学物理系, Greenville, NC27858-4353 USA)

**摘要** 应用激光镊子拉曼光谱(LTRS)技术,俘获单个红细胞并收集其拉曼光谱。将主成分分析(PCA)算法和反向传播 BP 网络预测模型相结合,进行地中海贫血(简称地贫)红细胞类型的判别。PCA 结果显示,正常对照与中间型  $\alpha$ -地贫(HbH-CS)基本可区分,但正常对照与重型  $\beta$ -地贫, HbH-CS 与重型  $\beta$ -地贫间差异不明显。将归一化处理的前 5 个主成分进行 BP 网络训练及预测,结果发现,正常对照与 HbH-CS 间预测正确率高达 97.90%,正常对照与重型  $\beta$ -地贫, HbH-CS 与重型  $\beta$ -地贫间预测正确率分别为 90.72% 和 86.28%。该结果与平均拉曼光谱及主成分分析结果基本吻合。取不同的实验条件下收集的光谱进行同样的分析,3 种组合的预测正确率略有不同,分别为 95.28%, 92.08%, 91.85%, 但呈现基本相同的规律。

**关键词** 拉曼光谱; 判别; 主成分分析; BP 神经网络; 地中海贫血

中图分类号 O657.37 文献标识码 A doi: 10.3788/CJL20093609.2448

## Raman Spectral Discrimination of Thalassemia Erythrocytes Based on PCA Arithmetic and BP Network Model

Chen Xiuli<sup>1,2</sup> Wang Guiwen<sup>1</sup> Tao Zhanhua<sup>1</sup> Liu Junxian<sup>2</sup> Yao Huilu<sup>1</sup>  
Huang Shushi<sup>1</sup> Li Yongqing<sup>3</sup>

<sup>1</sup> Biophysics Laboratory, Guangxi Academy of Sciences, Nanning, Guangxi 530003, China

<sup>2</sup> College of Physics and Electronic Engineering, Guangxi Normal University, Guilin, Guangxi 541004, China

<sup>3</sup> Department of Physics, East Carolina University, Greenville, NC27858-4353, USA

**Abstract** The thalassemias are a group of anemias result from inherited defects in the production of hemoglobin. The current techniques for screening and diagnosis of thalassemia are time consuming and complex. A laser tweezers Raman spectroscopy (LTRS) setup was used to trap single erythrocyte from patients with thalassemias and normal donors, and to collect the Raman scattering of trapped cell. Blood samples obtained from 11 patients with non-deletional HbH disease (HbH-CS), 11 patients with  $\beta$ -thalassemia major, and 11 normal controls, were tested. Principal component analysis (PCA) algorithm combined with back-propagation neural network predictive model was performed to distinguish abnormal erythrocyte. The PCA results reveal that the difference between normal controls and HbH-CSs is significant with the predictive accuracy of BP network as high as 97.90%. The predictive accuracy between normal controls and  $\beta$ -thalassemias major is 90.72%, and 86.28% between HbH-CSs and  $\beta$ -thalassemias major. These results tally closely with the corresponding averaged Raman spectra. Under different experimental condition, the predictive accuracy shows similar results. This pilot study can serve as a useful probe for developing a rapid,

收稿日期: 2008-09-18; 收到修改稿日期: 2008-11-06

基金项目: 国家自然科学基金(30660063), 广西科学院科技创新(桂科院研 0702)资助项目。

作者简介: 陈秀丽(1983—), 女, 硕士研究生, 主要从事分子振动光谱在生物医学方面的应用研究。

E-mail: xiuli\_ch@163.com

导师简介: 王桂文(1969—), 男, 副研究员, 硕士生导师, 主要从事微生物、生物物理等方面的研究。

E-mail: wguiwen@126.com(通信联系人)

simple, reagent-free method for distinguishing of thalassemia erythrocytes.

**Key words** Raman spectroscopy; identification; principal component analysis; back-propagation network; thalassemias

## 1 引言

地中海贫血<sup>[1]</sup>(简称地贫)是一种遗传性溶血性贫血病,主要分为 $\alpha$ -地贫和 $\beta$ -地贫两种类型。筛查地贫患者,避免重型地贫患儿的出生是控制本病的关键。但常规的基于细胞群体或分离血红蛋白的检测方法(如血常规检测、血红蛋白电泳分析、脱氧核糖核酸(DNA)分析等),操作繁琐,工作量大,费用相对昂贵,不适合大面积人群的筛查。因此,发展一种快速、准确、低成本的检测方法具有重要的临床应用意义。

与其他物质的拉曼散射相比,水的拉曼散射非常微弱,因此拉曼光谱拥有在水溶液环境中开展研究的独特优势<sup>[2]</sup>,适合生物大分子、细胞等的研究。而将光镊和显微拉曼光谱相结合的激光镊子拉曼光谱(Laser tweezers Raman spectroscopy, LTRS)技术,单个细胞被囚禁在激光束的焦点处,无需机械固定或引入化学标记等,消除了普通显微拉曼光谱技术将细胞固定在玻片上所引起的不良影响。同时,光镊可提升被囚禁的细胞,减少了来自玻片的信号干扰,从而获得信噪比更高的光谱。该技术已成功应用于识别非生物微粒<sup>[3]</sup>、红细胞<sup>[4~7]</sup>、脂质体膜<sup>[8]</sup>、酵母细胞和细菌细胞<sup>[9,10]</sup>及芽孢<sup>[11]</sup>等。

收集拉曼光谱后,需对光谱数据进行分析。传统的方法是对光谱做平滑预处理,对特征谱线做归属分析,比较不同光谱在某些波数范围内的相对峰强和拉曼位移等,得到一些关于分子构像变化的信息。但由于 CCD 探测器存在量子起伏或样品存在荧光物质和光吸收,会造成拉曼光谱强度起伏和提高荧光背景。主成分分析(Principal component analysis, PCA)和层序聚类分析(Hierarchical clustering analysis, HCA)已成功用于鉴别不同培养环境的细菌<sup>[12]</sup>。小波变换及主成分分析压缩拉曼光谱数据,再结合神经网络可进行动脉硬化症的诊断<sup>[13]</sup>。

应用 LTRS 通过分析细胞的特征拉曼谱峰差异性,有效地判别不同物种的血小板<sup>[14]</sup>、红细胞<sup>[7]</sup>。对于不同类型地贫红细胞,由于细胞的主要成分是血红蛋白,常规的拉曼光谱数据处理技术难以直观、快速准确地判别红细胞类型。因此,本研究利用 LTRS 系统收集 HbH-CS, 重型 $\beta$ -地贫及正常对照的样本各 11 个,每个样本 30 个红细胞的拉曼光谱。

在平滑去噪的基础上,对光谱背景进行拟合并扣除,同时引入了 PCA 算法用于高维光谱数据的压缩,提取主要信息。然后将归一化后的前 5 个主成分输入反向传播(BP)网络进行训练和预测。

## 2 材料与方法

### 2.1 仪器与样品

血液样本经 Hb 电泳分析和 DNA 分析确诊,分为 3 组,每组 11 个病人(正常人)的血液标本。第 1 组:重型 $\beta$ -地贫(Major),样品编号为 Major1~Major11;第 2 组:HbH-CS,样品编号为 HbH-CS1~HbH-CS11;第 3 组:正常对照,样品编号为 Norm1~Norm11。

激光镊子拉曼光谱系统如文献[14]所述。

### 2.2 光谱收集和预处理

实验 1,取重型 $\beta$ -地贫样本 Major1~Major5, HbH-CS 样本 HbH-CS1~HbH-CS5,正常对照样本 Norm1~Norm5。每个样本取 10  $\mu$ L 用生理盐水稀释至 3000 倍,取 200  $\mu$ L 置于样品槽。激光器波长为 785 nm,功率为 30 mW,曝光时间为 10 s,光镊将红细胞囚禁于载玻片上方约 8  $\mu$ m 位置,收集细胞的拉曼光谱  $S_0(\nu)$ ;移开红细胞,保持参数和光镊高度不变,收集生理盐水溶液光谱为  $S_1(\nu)$ 。将细胞光谱减去溶液光谱,并经响应函数  $R(\nu)$  校正,得到实际光谱  $S(\nu) = [S_0(\nu) - S_1(\nu)]R(\nu)$ 。每个样本收集 30 个细胞光谱和 4 个溶液光谱。将光谱数据转换为 ASCII 数据,利用软件 Matlab7.0 进行 5 点平滑去噪,截取信号峰集中的 600~1800  $\text{cm}^{-1}$  指纹区。

实验 2,取重型 $\beta$ -地贫样本 Major6~Major11, HbH-CS 样本 HbH-CS6~HbH-CS11,正常对照样本 Norm6~Norm11。激光波长为 780 nm,功率为 7 mW,曝光时间为 20 s。其他操作和处理方法如实验 1。

### 2.3 背景拟合

为消除各细胞光谱由于背景不同而导致的光谱差异,利用最小二乘多项式拟合并扣除光谱背景。拟合数据点数为 6(见图 1),分别对应卟啉呼吸振动模 752  $\text{cm}^{-1}$  ( $\nu_{15}$ ), 苯丙氨酸对称呼吸振动模 1001  $\text{cm}^{-1}$ ,  $C_mH$  变形模 1212  $\text{cm}^{-1}$  ( $\nu_5 + \nu_{18}$ ), 来自氨基酸侧链的  $\text{CH}_2/\text{CH}_3$  的变形模 1456  $\text{cm}^{-1}$ ,  $C_\beta C_\beta$

的振动模  $1547\text{ cm}^{-1}$  及  $C_aC_b$  的振动模  $1622\text{ cm}^{-1}$ [15] 等峰的峰谷。由于 CCD 各个探测器是独立的,即使是恒定光源,光场打在探测器上的强度也会有相对起伏。此外,探测器的电子信号转化为拉曼光谱信号的过程中还存在读出噪声(readout noise),造成任意一个像数点的强度有波动。为了消弱这些噪声影响,采用了平均值取点,即在拟合点左右各取 3 个点(共 7 个点)的平均值作为这个点的值。通过比较不同阶数的拟合结果表明,三阶的拟合曲线最平滑,因此选取拟合阶数为 3。

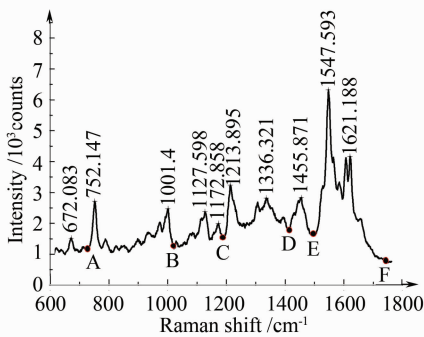


图 1 单个红细胞的原始拉曼光谱及背景拟合数据点的选取  
Fig. 1 Raw Raman spectrum of single red blood cell and the points chosen for curve-fitting

## 2.4 主成分分析

经上述处理的光谱维数较高,即截取后的光谱仍然有 850 维,如此高的维数进行网络训练非常困难。因此,采用主成分分析方法实现降维,通过对数据去相关,提取出数据的最主要特征,经变换后得到的新变量作为主成分,按信息量的大小排序后,前几个主成分包含了绝大部分信息量,后面的主成分主要是噪声。

## 2.5 BP 神经网络模型的设计

BP 网络是一种多层前馈神经网络,采用调整网络权值的向后传播学习算法。S 型激活函数的多层 BP 网络所划分的区域比较柔和、光滑,具有较强的非线性预测能力和较好的网络容错性。

这种网络结构模型具有两条指导原则:

1) 对于一般的模式识别问题,3 层网络可以很好地解决。

2) 3 层网络中,隐含层单元个数  $N_2$  和输入层神经元个数  $N_1$  之间有以下近似关系:  $N_2 = 2N_1 + 1$ [16]。

网络训练及预测发现,输入层节点不宜过多,以避免隐含层节点过多导致训练过慢,但同时也需要保证输入的信息尽可能包含主要的信息。为此,比

较了不同输入节点(节点数取 3~7,即取前 3~7 个主成分)的训练情况及其预测结果发现,在保证所取主成分比重高于 90% 以上,一般取前 5 个主成分进行网络训练的效果较好。但在不同的实验参数下收集的光谱数据,可根据主成分的比重和网络的收敛情况适当调整输入层及隐含层神经元数目。

采用 3 层网络,输入层单元个数是 5;隐含层单元个数为 11,传递函数为对数函数 logsig;训练函数为 trainlm;根据输出目标的数量,确定输出层单元个数为 3,传递函数仍采用对数函数 logsig,目标输出形式如下:

正常对照:(1, 1, 1); 重型  $\beta$ -地贫:(1, 0, 1); HbH-CS:(0, 1, 0)。

## 3 结果与分析

### 3.1 谱图分析

将所有预处理的光谱进行背景扣除后,光谱基线基本一致,提高了可比性。图 2 是实验 1 中 3 种类型样本的平均拉曼光谱,即每种类型 5 个血液样本,每个样本 30 个细胞,共 150 个细胞光谱的平均值。图中显示,正常样本的平均光谱强度最高,重型  $\beta$ -地贫的次之,而 HbH-CS 强度最弱。3 种类型的光谱除了强度存在一些区别外,峰形和峰位几乎一致。由于测量误差、样本个体差异和细胞差异等,可能出现不同病种间光谱强度差异较小的情况,导致直接观察某一细胞光谱强度难以辨识其所属类型。因此,需要借助其他方法,如多元统计方法等,做进一步的处理,以寻求一种高效的、准确的模式识别方法。

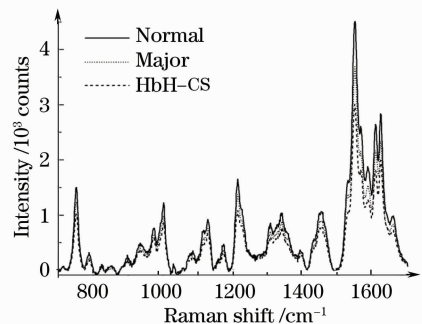


图 2  $\beta$ -地贫重型, $\alpha$ -地贫中间型和正常对照红细胞的平均拉曼光谱

Fig. 2 Averaged Raman spectra of single red blood cell from normal controls and patients with  $\beta$ -thalassemia major, HbH-CS

### 3.2 PCA 的结果及归一化

为了更清楚地显示各种类型红细胞间的差别,探讨了 3 种类型两两组合间的情况。第一组合:正

常对照与 HbH-CS; 第二组合: 正常对照与重型  $\beta$ -地贫; 第三组合: HbH-CS 与重型  $\beta$ -地贫。调用 Matlab7.0 的函数 princomp(), 求出主成分空间中的数据, 即得分(score), 以及各主成分(PC)的比例。

在实验 1 中, 第一组合前五维主成分的质量分数

分别为 94.678%, 1.1216%, 0.82521%, 0.36652%, 0.22532%, 占总方差的 97.21%; 第二组合前五维主成分的总质量分数为 95.21%; 第三组合的为 95.05%, 均保证了光谱中的主要信息被抽取。其前三维得分的空间分布如图 3 所示。

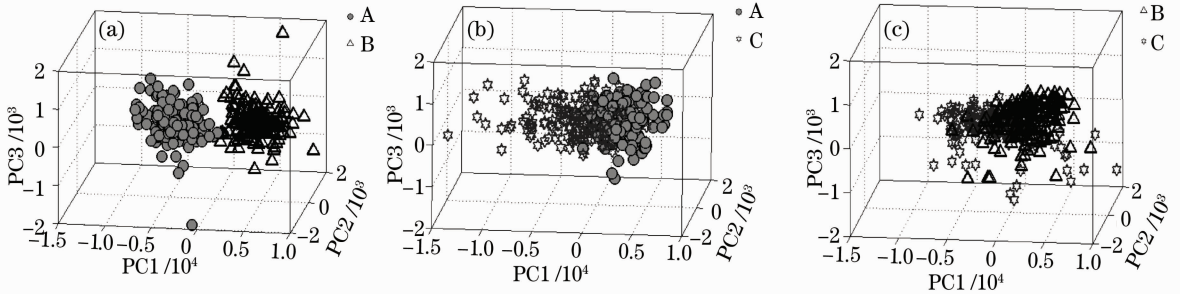


图 3 实验 1 中 3 种类型红细胞的 PCA 结果示意图。(a) 正常对照(A)与 HbH-CS(B); (b) 正常对照(A)与重型  $\beta$ -地贫(C); (c) HbH-CS(B)与重型  $\beta$ -地贫(C)。实验条件: 激光波长 785 nm, 功率 30 mW, 收集时间 10 s

Fig. 3 PCA plots generated from the Raman spectra of single erythrocytes from different individuals in experiment 1. (a) Normal controls (A) and HbH-CSs (B); (b) normal controls (A) and  $\beta$ -thalassemias major (C); (c) HbH-CSs (B) and  $\beta$ -thalassemias Major(C). Laser wavelength 785 nm, 30 mW, 10 s acquisition time

网络训练过程表明, 归一化的数据能使网络更快地收敛到目标误差值。因此, 取得分的前 30 维, 即得到  $n \times 30$  的矩阵( $n$  为细胞光谱个数), 进行归一化处理, 有

$$a_{i,j} = (x_{i,j} - x_{\min,j} + 0.01) / (x_{\max,j} - x_{\min,j} + 0.01)$$

$$(i = 1, 2, 3 \dots 30, j = 1, 2, 3 \dots n)$$

式中  $x$  为主成分得分值。然后取归一化结果的前五维作为输入层, 进行 BP 网络训练。

### 3.3 基于 PCA 的 BP 神经网络预测结果

为了更好地说明网络预测结果的准确性, 采用循环预测的方法。实验 1, 在每种类型样本(例如: Norm1~Norm5)中任意取一个样本做预测集(共 30 个细胞光谱), 其余 4 个样本做训练集(共 120 个细胞光谱)。而且, 每一个预测集均统计了 6 次预测结果的平均值, 避免一次可能因训练初始随机权值不当造成预测结果不佳的情况。

图 4 是第一组合中某一次网络训练误差曲线。经过 38 次网络权值修正即达到了设定的目标误差值。本次预测集中, 1~30 号属于 Norm5, 31~60 号属于 HbH~CS5。通过对比预测值和目标值(具体数据没给出)发现, 在目标误差范围内, 只有一个预测值出错, 正确率为 98.33%。

3 种组合的预测正确率统计结果见表 1。表中显示, 正常对照与 HbH-CS 的预测结果最理想, 正

确率高达 97.90%。而正常对照与重型  $\beta$ -地贫, HbH-CS 与重型  $\beta$ -地贫间的预测结果略低, 分别为 90.72% 和 86.28%。这一结果与 3 种类型细胞的平均拉曼光谱(图 2)基本相符, 即正常对照与 HbH-CS 的光谱强度差别较大, 重型  $\beta$ -地贫的光谱强度处于两者之间。同时, 前三维得分空间图[图 3 (a)]也显示正常对照和 HbH-CS 明显落入不同的空间, 重叠数目非常少, 而其他两组虽然也大致落入两个区域, 但重叠程度很高, 界限不明显[图 3 (b), (c)]。组合中各个样本的预测正确率也有差异, 如正常对照与 HbH-CS 组合中不同样本的预测正确率不同, 最高达到了 100%, 最低是 94.44%。此外, 个别样本的网络预测结果不佳, 如表 1 中 Major4 和 HbH-CS4 的预测准确率只有 73.61%。

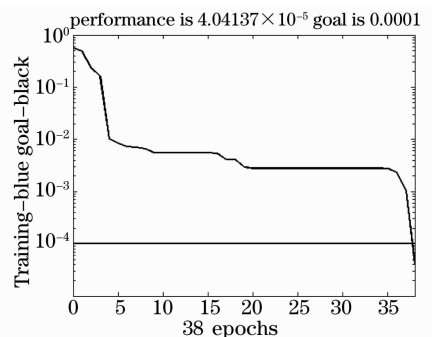


图 4 网络训练误差曲线

Fig. 4 The error curve of network training

表1 实验1中3个组合的预测正确率(%)

Table 1 Prediction accuracy of different erythrocytes from normal controls, HbH-CSs and  $\beta$ -thalassemias major in experiment 1(%)

Normal controls and HbH-CSs		Normal controls and $\beta$ -thalassemias major		HbH-CSs and $\beta$ -thalassemias major	
Norm1-HbH-CS1	94.44	Norm1-Major1	91.94	Major1-HbH-CS1	85.28
Norm2-HbH-CS2	98.33	Norm2-Major2	95.56	Major2-HbH-CS2	87.78
Norm3-HbH-CS3	100	Norm3-Major3	84.81	Major3-HbH-CS3	90.83
Norm4-HbH-CS4	98.61	Norm4-Major4	96.67	Major4-HbH-CS4	73.61
Norm5-HbH-CS5	98.13	Norm5-Major5	84.63	Major5-HbH-CS5	93.89
Mean accuracy	97.90	Mean accuracy	90.72	Mean accuracy	86.28

Notations: In Table 1, Table 2, the accuracy for each couple is the averaged accuracy of six trainings, and the mean accuracy is the averaged data of all trainings in each groups.

采用同样的处理方法分析实验2的光谱数据, 样本为 Norm6~Norm11, Major6~Major11, HbH-CS6~HbH-CS11。本次实验的激光波长和功率以及采集时间均与实验1不同。在本次训练网络中,

每一类型样本中取5个做训练集(150个细胞), 剩下一个做预测集(30个细胞)。通过网络训练及预测, 得到的预测正确率见表2所示。

表2 实验2中3个组合的预测正确率(%)

Table 2 Prediction accuracy of different erythrocytes from normal controls, HbH-CSs and  $\beta$ -thalassemias major in experiment 2(%)

Normal controls and HbH-CSs		Normal controls and $\beta$ -thalassemias major		HbH-CSs and $\beta$ -thalassemias major	
Norm6-HbH-CS6	96.67	Norm6-Major6	76.94	Major6-HbH-CS6	93.89
Norm7-HbH-CS7	88.89	Norm7-Major7	95.83	Major7-HbH-CS7	86.39
Norm8-HbH-CS8	97.78	Norm8-Major8	99.17	Major8-HbH-CS8	99.17
Norm9-HbH-CS9	98.33	Norm9-Major9	97.78	Major9-HbH-CS9	97.78
Norm10-HbH-CS10	97.5	Norm10-Major10	88.06	Major10-HbH-CS10	90.00
Norm11-HbH-CS11	92.5	Norm11-Major11	94.17	Major11-HbH-CS11	83.89
Mean accuracy	95.28	Mean accuracy	92.08	Mean accuracy	91.85

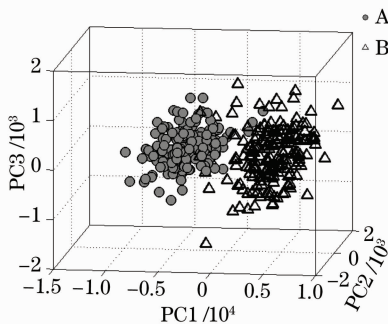


图5 实验2中,正常对照(A)与HbH-CS(B)的PCA结果示意图。实验条件:激光波长780 nm,功率7 mW,收集时间20 s

Fig. 5 PCA plots generated from the Raman spectra of single erythrocytes from normal controls. (A) and patients with HbH-CS (B) in experiment 2. Laser wavelength 780 nm, 7 mW, 20 s acquisition time

表2的数据表明,仍然是第一组合的正确率最高,平均结果为95.28%,第二和第三组合的结果次

之,分别为92.08%和91.85%,呈现出与实验1基本相同的规律,即正常对照与 $\alpha$ -地贫HbH-CS间的区别比其他两组合的大。但在本实验中,同组合中不同样本的预测正确率差异较大,如表2中第二组合的最低预测正确率为76.94%,最高预测正确率为99.17%,相差22.23%。其他组合的预测正确率也参差不齐。通过比较图3(a)及图5也发现,实验2中同一类型各样本的前三维得分空间分布较疏散,不同类型的重叠数目较多。

## 4 讨论

利用拉曼光谱进行微生物及细胞鉴别的研究<sup>[7,9,14,17~19]</sup>,研究对象的拉曼光谱特征峰由于胞内物质的差异,已存在较明显的差别,只需对光谱做谱峰归属、比较谱峰的强度、峰强的比值及拉曼位移等,就能较好地分辨出不同的物质。然而,对于本实验系统所收集的不同类型地贫红细胞的拉曼光谱,



除了在强度上略有区别外,拉曼位移和峰形几乎一致,没有明显的特征性,应用这些常规方法难以识别其类型。此外,实验过程产生的测量误差,使得鉴别更加困难。

面对收集到的大量光谱数据,利用多元统计方法进行信息的提取,并结合一定的模式识别算法,有望从繁琐的光谱数据中抽取主要信息,实现对不同类型红细胞的光谱分类。主要方法有主成分分析、聚类分析、判别分析及神经网络模型等。主成分分析的结果一般是通过得分空间分布图来表现。但对于差异不明显的光谱,利用三维空间难以达到显著分类效果,如图 3(b), (c)。而且,空间图形表现出来的结果只能是一种视觉上的区分度,无法进行量化。而采用 BP 网络预测模型,能根据样品的数据特征将其训练到一个目标上。对于检验集,可以通过比较预测结果和目标值来统计预测正确率,得到一个量化的数值。此外,采用 S 型激活函数的三层网络具有较好的容错性,能对偏离的个别细胞光谱进行准确预测。

应用主成分分析提取光谱的主要信息,将得分归一化到(0,1),输入已训练好的 BP 神经网络预测模型,能较好地进行不同类型地贫红细胞的预测。两次实验结果均得出,不同类型地贫间的预测正确率不同,其中正常对照与 HbH-CS 组的结果最理想。多次实验发现, HbH-CS 红细胞的平均拉曼光谱强度明显低于正常对照的。而本预测模型正是基于光谱强度,只是经过主成分分析及归一化等变换,没有改变其本质差别。因此,预测正确率与原光谱的强度差异程度及主成分得分空间分布图基本吻合。

然而,组合中取不同样本做预测集,预测结果参差不齐。个别样本的预测正确率较低。造成这些结果的主要因素有:1)拉曼光谱系统本身的局限性,如量子起伏噪声、探测器信号的读出噪声等;2)样本个体差异及光谱采集存在测量误差,如细胞所处的高度不同导致光谱强度的差异;3)可能存在采血时间不同造成的差异性;4)预测模型没有达到最优化,如易陷入局部误差及随机权值不当等。此外,通过对比发现,实验 2 中各个样本预测正确率差异程度较大,最高可达 22.23%。这主要因为实验 2 的数据是取自 3 次不同实验时间的测量结果;而实验 1 中所有数据均在相同的实验状态下相继收集的。这说明不同测量时间,激光波长发生漂移,可能会导致相同类型样本间的差异较大,使得个别样本偏离训练

集,预测正确率低。

## 5 结 论

应用 LTRS 技术收集单个红细胞的拉曼光谱,将 PCA 算法和 BP 网络预测模型相结合,对重型  $\beta$ -地贫、HbH-CS 与正常对照红细胞进行判别,但分析结果距临床应用还存在较大差距。而要进一步提高地贫识别的准确率,除了发展更好的预测模型(如改进网络训练函数,避免网络陷入局部误差等)外,关键在于控制好实验条件,优化实验系统,消除非本质的干扰因素,提高光谱的信噪比。本文采用两两组合的方式,意在比较不同类型间的差异程度。若能进一步提高光谱的信噪比,还可以将三种或更多类型的样本组合在一起,通过设置不同的训练目标,建立多类型样本的预测模型,有可能实现地贫的快速诊断。

致谢 感谢广西医科大学陈萍博士提供血液样本。

## 参 考 文 献

- 1 Zhang Junwu, Long Guifang. Hemoglobin & Hemoglobinopathies [M]. Nanning: Guangxi Science & Technology Publishing House, 2003. 151~235  
张俊武, 龙桂芳. 血红蛋白与血红蛋白病[M]. 南宁: 广西科学技术出版社, 2003. 151~235
- 2 S. Keller, B. Schrader, A. Hoffmann *et al.*. Application of near-infrared-Fourier transform Raman spectroscopy in medical research[J]. *J. Raman Spectrosc.*, 1994, **25**: 663~667
- 3 C. A. Xie, Y. Q. Li. Raman spectra and optical trapping of highly refractive and nontransparent particles[J]. *Appl. Phys. Lett.*, 2002, **81**(6): 951~953
- 4 C. A. Xie, M. A. Dinno, Y. Q. Li. Near-infrared Raman spectroscopy of single optically trapped biological cells[J]. *Opt. Lett.*, 2002, **27**(4): 249~251
- 5 K. Ramser, K. Logg, M. Goksör *et al.*. Resonance Raman spectroscopy of optically trapped functional erythrocytes[J]. *J. Biomedical Optics.*, 2004, **9**(3): 593~600
- 6 K. Ramser, K. Logg, J. Enger *et al.*. Resonance Raman study of the oxygenation cycle of optically trapped single red blood cells in a microfluidic system[C]. *SPIE*, 2004, **5514**: 560~568
- 7 Yao Huilu, Wang Guiwen, He Bijuan *et al.*. Raman spectra of single blood cells[J]. *Journal of Jinan University (Sci. & Tech.)*, 2005, **19**(4): 328~330  
姚辉璐, 王桂文, 何碧娟 等. 单个红细胞的拉曼光谱研究[J]. 济南大学学报(自然科学版), 2005, **19**(4): 328~330
- 8 J. M. Sanderson, A. D. Ward. Analysis of liposomal membrane composition using Raman tweezers[J]. *Chem. Commun.*, 2004, **7**: 1120~1121
- 9 C. A. Xie, Y. Q. Li, W. Tang *et al.*. Study of dynamical process of heat denaturation in optically trapped single microorganisms by near-infrared Raman spectroscopy[J]. *J. Appl. Phys.*, 2003, **94**(9): 6138~6142
- 10 C. A. Xie, C. Goodman, M. A. Dinno *et al.*. Real-time Raman spectroscopy of optically trapped living cells and organelles[J]. *Opt. Express*, 2004, **12**(25): 6208~6214

- 11 J. W. Chan, A. P. Esposito, C. E. Talley *et al.*. Reagentless identification of single bacterial spores in aqueous solution by confocal laser tweezers Raman spectroscopy[J]. *Anal. Chem.*, 2004, **76**: 599~603
- 12 Q. G. Zhu. Development of oral bacteria concentration measurement system using Raman spectroscopy[D]. New York: University of Rochester, 2005, 30~37
- 13 A. R. De Paula Jr, S. Sathaiah. Raman spectroscopy for diagnosis of atherosclerosis: a rapid analysis using neural networks [J]. *Medical Engineering & Physics*, 2005, **27**: 237~244
- 14 Wang Guiwen, Yao Huilu, He Bijuan *et al.*. Raman micro-spectroscopy of single blood platelets [J]. *Spectroscopy and Spectral Analysis*, 2007, **27**(7): 1347~1350  
王桂文,姚辉璐,何碧娟等. 单个血小板的拉曼光谱分析[J]. *光谱学与光谱分析*, 2007, **27**(7): 1347~1350
- 15 B. R. Wood, P. Caspers, G. J. Puppels *et al.*. Resonance Raman spectroscopy of red blood cells using near-infrared laser excitation[J]. *Anal. Bioanal. Chem.*, 2007, **387**: 1691~1703
- 16 Sun Zhiqiang, Ge Zhexue. Neural Network Theory and Achieving in Matlab7. 0 [M]. Beijing: Publishing House of Electronics Industry, 2005  
孙志强,葛哲学. *神经网络理论与 Matlab7 实现*[M]. 北京:电子工业出版社, 2005
- 17 P. Rösch, M. Schmitt, W. Kiefer *et al.*. The identification of microorganisms by micro-Raman spectroscopy[J]. *J. Molecular Structure*, 2003, **661-662**: 363~369
- 18 C. A. Xie, D. Chen, Y. Q. Li. Raman sorting and identification of single living micro-organisms with optical tweezers[J]. *Opt. Lett.*, 2005, **30**: 1800~1802
- 19 H. Y. Tang, H. L. Yao, G. W. Wang *et al.*. NIR Raman spectroscopic investigation of single mitochondria trapped by optical tweezers[J]. *Opt. Express*, 2007, **15**: 12708~12716