Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC-PapersOnLine 49-7 (2016) 195–199

Chemical Recognition using the Time-dependent Cellular Response Profiles ⋆

Jiao Chen *,** **Tianhong Pan** * **Haoran Li** * **Kaili Xu** * **Zhengming Li** *

* *School of Electrical and Information Engineering, Jiangsu University, Zhenjiang, Jiangsu, China (e-mail: thpan@ujs.edu.cn, lihaoran32@qq.com, kaili_xu@foxmail.com, lzming@ujs.edu.cn).*
** *School of Electronics and Electrical Engineering, Changzhou College of Information and technology, Changzhou, Jiangsu, China (e-mail: luckycj79@163.com).*

Abstract: To distinguish the chemicals on the cellular level, a pattern recognition approach, which uses the time-dependent cellular response profiles (TCRPs), is proposed in this paper. Firstly, the TCRPs is collected from the xCELLigence real time cellular analyzer high throughput (RTCA HT) system. Secondly, based on the traditional cellular toxic-effect evaluation, the dose-response curves is generated from the multi-concentration TCRPs. And then features are extracted from the produced dose-response curves. Thirdly, an improved k-means cluster is used to classify the extracted features. The proposed method can provide a useful solution and a high throughput screening for chemical recognition at the cellular level.

Keywords: TCRPs, Cytotoxicity, Real time cellular analyzer high throughput (TRCA HT) system, Singular Value Decomposition(SVD), k-means cluster.

1. INTRODUCTION

The success of high throughput screening for chemicals was based on the assumption that chemicals with similar biological activity would produce similar *TCRPs*. Chemicals with similar Mechanism of Actions (MoAs) show similar patterns as indicated by *TCRPs* **?**. Several methods, such as SVM algorithm based on cellular state variable identification (CSVID), a majority vote of the class labels **?**, and a model-based hierarchical classification approach incorporating PCA and functional data analysis **?**, are applied to classify the chemical compounds on cellular level. But the method would lose the effectiveness when the cellular response curves show different shapes not shown in the training data.

In order to improve the accuracy of pattern recognition using *TCRPs*, a novel method integrating the dose-response curves, feature extraction, orthogonal projection and k-means clustering algorithm is proposed in this paper.

2. PROBLEM STATEMENT

The *TCRPs* obtained from the RTCA HT system reflect the chemicals' MoA and can be used to identify the tested chemicals (shown in Fig.1) **?**.

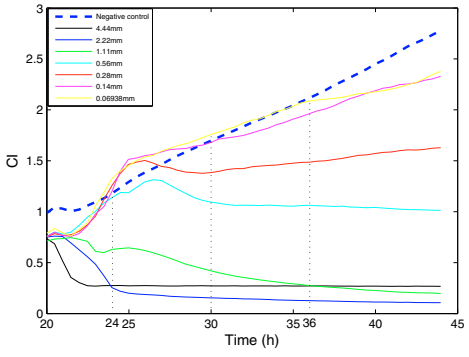


Fig. 1. The TCRPs obtained from the RTCA HT system .

Inspired by traditional cellular toxic-effect evaluation, dose-response curves at typical time-points can be generated from the multi-concentration *TCRPs*. The toxic-effect is calculated as equation (1). See **?**Pan et al. (2013).

$$\begin{cases} GI(x,t) = \frac{CI(x,t)-CI(x,0)}{CI(NC,t)-CI(NC,0)} \times 100, if\ CI(x,t) \geq CI(NC,0) \\ LC(x,t) = \frac{CI(x,t)-CI(x,0)}{CI(NC,0)} \times 100, if\ CI(x,t) < CI(NC,0) \end{cases} \quad (1)$$

where x is the concentration of tested chemical, $CI(x,t)$ is the cell index at the t^{th} sampling instant after adding the test substance with concentration x, $CI(x,0)$ and $CI(NC,0)$ are the cell indices of the chemical and Negative Control (NC) at the 1^{st} sampling instant after adding the substance, respectively, and $CI(NC,t)$ is the cell index of

⋆ The work was supported by the National Nature Science Foundation of China [grant number 61273142], the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) and Foundation for Six Talents by Jiangsu Province [grant number 2012-DZXX-045].

10.1016/j.ifacol.2016.07.247

the negative control at the t^{th} sampling instant. $GI(x,t)$ and $LC(x,t)$ are the cell growth inhibition index and cell lethal growth inhibition index of the test substance with concentration x at the t^{th} sampling time, respectively, and t is sampling time in hours.

Using the calculated $LC(x,t)$ and $GI(x,t)$, the dose-response curve at the special exposure time point described by the Hill model is fitted as equation (2). See Dinse (2011).

$$LC(x,t) = p_1(t) + \frac{p_2(t) - p_1(t)}{1 + \exp(-(x - p_3(t))/p_4(t))} \tag{2}$$

where $p_1(t)$ is the baseline response at the t^{th} sampling instant (at concentration 0), $p_2(t)$ is the maximum response at the t^{th} sampling instant (at an infinite concentration), $p_3(t)$ is the concentration producing a response halfway between $p_1(t)$ and $p_2(t)$ at the t^{th} sampling instant, and $p_4(t)$ is the shape parameter at the t^{th} sampling instant.

In this paper, three typical exposure time points, i.e. t=24hr, 30hr, 36hr, were selected to assess the cytotoxicity. One example is shown in Fig.2.

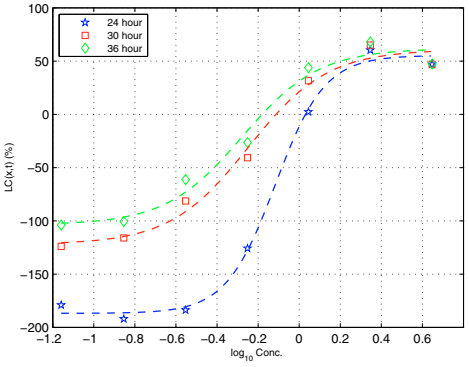


Fig. 2. Dose-response curves of 1-Naphthol at the exposure time 24hr, 30hr, 36hr.

3. METHODS

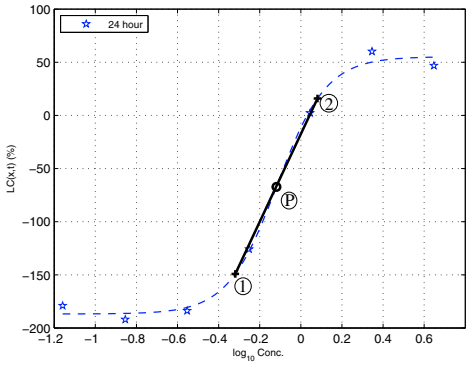
3.1 Feature extraction

Feature extraction from complete dose-response curves As shown in Fig. 1 and Fig. 2, the dose-response curves of 1-Naphthol at three typical exposure time points have a sigmoidal shape. Consider the dose-response curve at 24hr (Fig. 3 (a)). The tangential points of the dose-response curve reflect an increased rate of cytotoxicity (shown in Fig. 3 (b)), which is formulated as follows:

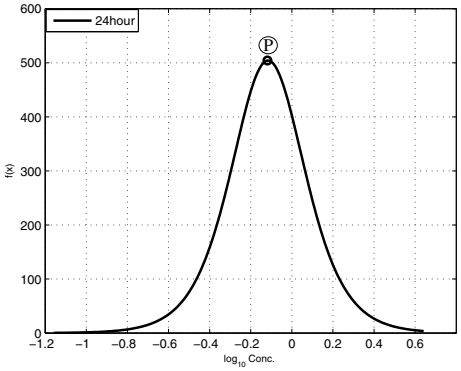
$$f(x,t) = \frac{\partial LC(x,t)}{\partial x}, t = 24, 30, 36 \tag{3}$$

where $f(x,t)$ denotes the response rate of cytotoxic intensity.

As shown in Fig. 3 (b), the curve of the derivative has a peak which indicates that the maximum rate of toxicity increases at 24hr. Different chemicals reach peaks at different concentration levels. The logarithm of concentration



(a) The two corners and the point Ⓟ in the dose-response curve of 1-Naphthol at the exposure time 24hr.



(b) The peak Ⓟ in the first derivative of dose-response curve of 1-Naphthol.

Fig. 3. Feature extraction from complete dose-response curves

$(P(t), t = 24, 30, 36)$ which achieves the peak is selected as a feature (point Ⓟ shown in Fig. 3 (a) and Fig. 3(b)). $P(t)$, which is the logarithm of the concentration for the curve, denotes the cytotoxic intensity of the tested chemical. $P(t)$ is defined by

$$P(t) = \arg\max\{f(x,t)\}, t = 24, 30, 36 \tag{4}$$

Moreover, the middle part of the dose-response curve reflects the cytotoxicity sensitivity of the tested chemical, which is contained in the slope $S(t)$.

$$S(t) = \frac{LC(x_2(t),t) - LC(x_1(t),t)}{x_2(t) - x_1(t)}; t = 24, 30, 36 \tag{5}$$

where $x_1(t)$ and $x_2(t)$ are the logarithm of concentration at points ① and ②, respectively, in the x-axis in Fig. 3, and calculated by the following equation.

$$\begin{cases} x_1(t) = P(t) - \Delta conc \\ x_2(t) = P(t) + \Delta conc \end{cases}; \ t = 24, 30, 36 \tag{6}$$

where $\Delta conc$ is a predetermined range of concentrations. Therefore, the feature vector of 1-Naphthol can be formulated as: $Y = [S(24), S(30), S(36), P(24), P(30), P(36)]$.

Feature extraction from incomplete dose-response curves Owing to the problem of concentration selection or dilution