

## 인공신경망에 의한 생물공정에서 2차원 형광스펙트럼의 분석 I

### - 자기조직화망에 의한 형광스펙트럼의 분류 -

<sup>1,5</sup>이금일 · <sup>2,4,5</sup>임용식 · <sup>2,4</sup>김춘광 · <sup>2,4,5</sup>이승현 · <sup>1,5</sup>정상욱 · † <sup>3,4,5</sup>이종일  
전남대학교, <sup>1</sup>산업공학과, <sup>2</sup>물질·생물화공과, <sup>3</sup>응용화학공학부, <sup>4</sup>생물공정기술연구소, <sup>5</sup>바이오광 기반기술개발 사업단  
(접수 : 2005. 4. 21., 게재승인 : 2005. 7. 23.)

## Analysis of Two-Dimensional Fluorescence Spectra in Biotechnological Processes by Artificial Neural Networks I

### - Classification of Fluorescence Spectra using Self-Organizing Maps -

Kum-Il Lee<sup>1,5</sup>, Yong-Sik Yim<sup>2,4,5</sup>, Chun-Kwang Kim<sup>2,4</sup>, Seung-Hyun Lee<sup>2,4,5</sup>, Sang-Wook Chung<sup>1,5</sup>, and Jong Il Rhee<sup>3,4,5†</sup>

<sup>1</sup>Department of Industrial Engineering, <sup>2</sup>Department of Material Chemical and Biochemical Engineering,

<sup>3</sup>Faculty of Applied Chemical Engineering, <sup>4</sup>BioProcess Technology Lab., <sup>5</sup>Research Center for Biophotonics,  
Chonnam National University, YongBong-dong 300, GwangJu 500-757, Republic of Korea

(Received : 2005. 4. 21., Accepted : 2005. 7. 23.)

Two-dimensional (2D) spectrofluorometer is often used to monitor various fermentation processes. The change in fluorescence intensities resulting from various combinations of excitation and emission wavelengths is investigated by using a spectra subtraction technique. But it has a limited capacity to classify the entire fluorescence spectra gathered during fermentations and to extract some useful information from the data. This study shows that the self-organizing map (SOM) is a useful and interpretative method for classification of the entire gamut of fluorescence spectral data and selection of some combinations of excitation and emission wavelengths, which have useful fluorometric information. Some results such as normalized weights and variances indicate that the SOM network is capable of interpreting the fermentation processes of *S. cerevisiae* and recombinant *E. coli* monitored by a 2D spectrofluorometer.

**Key Words :** Bioprocess monitoring fermentation, self-organizing map sensors, 2D spectrofluorometer

## 서론

생물공정의 효과적인 제어와 모델링을 위해 공정의 연속적인 모니터링은 매우 중요하다. pH 및 용존산소 센서, 배출가스 분석기 등과 같은 많은 센서들이 미생물 대사상태나 생물반응기내 각종 변수들을 측정하기 위해 사용되고 있다(1, 2). 특히, 광학센서는 비파괴, 비침투적이며 많은 분석물질을 연속적으로 측정할 수 있기 때문에 생물공학분야에 많이 이용되고 있다(3). 광학센서의 하나인 형광센서는 현재 생물공정의 제어와 모니터링에 많이 응용되고 있다(4-6). 최근 여기파장 (Excitation wavelength)과 방출

파장 (Emission wavelength)의 전 영역을 스캔할 수 있는 2차원 형광센서가 개발되었으며, *Enterobacter aerogenes*, *Pseudomonas fluorescences*, *Escherichia coli* 등의 발효공정 모니터링을 위해서 사용되고 있다(7, 8). 2차원 형광센서를 이용한 생물공정 모니터링에 대한 내용은 Schepers 교수의 논문에서 자세히 나타나 있다(9, 10).

한편, 형광센서에 의한 모니터링이 여기파장 250-650 nm와 방출파장 280-650 nm의 영역에서 10 nm 간격으로 이루어질 경우 여기파장과 방출파장의 조합으로 인해 매우 많은 양의 형광 데이터가 얻어진다(10). 예를 들면 스펙트럼 스캔이 5분 간격으로 이루어질 경우 50시간의 배양공정에서 600 (스캔 횟수) x 1558 (방출파장과 여기파장의 조합수)와 같은 많은 양의 데이터를 얻을 수 있다. 그러나 생물공정에서 많은 정보를 가지고 있는 형광 데이터를 얻었다고 할지라도, 다중 회귀분석법과 다중 판별분석법과 같은 고전적인 통계학적 분석방법들을 이용하여 데이터들을 해석하는 것은 쉽지 않다. 따라서 형광데이터의 양을 줄이

† Corresponding Author : Faculty of Applied Chemical Engineering, Chonnam National University, YongBong-dong 300, Gwangju 500-757, Republic of Korea

Tel: +82-62-530-1847, Fax: +82-62-530-0846

E-mail : jirhee@chonnam.ac.kr

고 중요한 정보를 추출하여 공정을 모델링할 수 있는 주성분분석법 (Principal component analysis, PCA), 부분최소자승분석법 (partial least square regression analysis, PLS)과 같은 더 진보된 통계적 알고리즘이 요구된다. 그러나 PCA와 PLS는 선형적 특성을 지니고 있기 때문에, 위의 방법들을 생물공정에서 측정된 형광데이터와 비선형관계에 있는 공정변수들의 해석에 적용하기란 쉽지 않다.

자기조직화망 (self-organizing map, SOM)은 비감독신경망 (unsupervised neural network) 알고리즘에 속하며 PCA의 비선형 대리자로서 간주되어진다. SOM은 다양한 분광계로부터 얻어지는 비선형 데이터의 분류와 분석에 있어 매우 강력한 수단이며 응용성이 높은 알고리즘이다. 예를 들면, SOM은 핵자기공명(nuclear magnetic resonance, NMR)스펙트럼 데이터를 분류하기 위해 사용되어졌으며, 효모발효에서 데이터 시각화와 온라인 모니터링을 위해 이온이동분광계와 결합되어 크로마토그래픽값을 예측하고 크로마토그래픽 시스템의 분석을 위해 사용되었다(11-13). 또한 SOM은 침엽수와 활엽수를 분류하는 Fourier- transform 라만 스펙트럼의 분류를 위해서도 성공적으로 사용되었다(14). 그러나 SOM을 이용한 2차원 형광스펙트럼의 분류나 생물공정을 분석하는 체계적인 연구가 아직 이루어지지 않고 있다.

따라서 본 연구에서는 재조합 대장균에 의한 5-amimolevulinic acid (ALA)의 생산공정과 *S. cerevisiae*에 의한 glutathione (GSH)의 생산공정에 SOM을 적용하여 형광 데이터를 분류하고 이를 이용하여 생물공정을 분석하고자 한다.

## 재료 및 방법

### 생물반응기 시스템

미생물 발효를 위해 2.5 L 생물반응기 (KoBiotech Co., Korea)를 사용하였으며, 생물공정에서 발생하는 각종 신호를 온라인 모니터링하기 위해 pH, DO (Mettler Toledo, Germany) 및 온도센서를 이용하였다. 그리고 배출가스 중의 산소와 이산화탄소의 농도분석을 위해 O<sub>2</sub>/CO<sub>2</sub> 가스분석기 (Lokas Co., Korea)를 사용하였다. 또한 공정의 각종 변수제어를 위해서 공기주입기, 교반기, 거품제거 및 pH 제어기 등을 이용하였다. 온라인 데이터의 저장은 LabVIEW 프로그램 (ver. 6.1 National instrument Co., USA)을 사용하였으며 산소소모속도 (Oxygen uptake rate, OUR)와 이산화탄소 생산속도 (Carbon dioxide production rate, CPR)는 반응기 입구와 출구 가스 중의 산소와 이산화탄소의 농도를 이용하여 계산하였다(15). 2차원 형광분광계 (Model F-4500, Hitachi, Japan)는 액체광학전도관 (Lumatec GmbH, Germany)을 생물반응기의 측면에 설치된 직경 19 mm 석영창에 연결하여 온라인 모니터링하였다. 형광분광계의 제어, 데이터저장 및 모니터링된 데이터의 표현을 위해 자체 제작한 프로그램을 이용하였다. 2차원 형광분광계의 측정조건은 스캔속도, 30000 nm·min<sup>-1</sup>; 방출파장과 여기파장 간격, 10 nm; 여기파장 범위, 250-650 nm; 방출파장 범위, 280-650 nm이다. 위의 측정조건에서 1회 전파장 스캔에는 1.5분이 소요되었다.

### 2차원 형광스펙트럼

2차원 형광스펙트럼의 등고선에서 상위 왼쪽 삼각형부분은 무형광의 영역과 빛이 산란되는 영역이다. 수집된 형광스펙트럼 데이터의 형광세기는 색깔변화를 이용하여 나타낼 수 있으며, 스펙트럼의 색깔변화는 특정(생물) 분자의 형광세기 (tryptophan (280/330-360), pyridoxine(310/385-415), NAD(P)H (340/440-460), riboflavin(400/500-530) 등)와 연관지어 표현할 수 있다(4). 형광스펙트럼은 배양액이나 반응기로 인한 배경형광을 포함하므로, 배경형광을 최소화하기 위해 전체 형광스펙트럼으로부터 배양액의 스펙트럼을 제거하였다(9).

### 발효 공정

본 연구에서는 ALA 및 GSH생산을 위해 표 1에서 보는 바와 같이 발효를 수행하였으며, 2차원 형광분광계를 이용하여 온라인 모니터링하였다.

**Table 1.** Fermentation processes for the production of ALA by recombinant *E. coli* and for the production of Glutathione by *Sacchariomyces cerevisiae*

	FM1	FM2	YF1	YF2
Culture medium	MS8	LB	SM	SM
Process operating conditions	pH 6.2 37 °C 1 vvm 450 rpm	pH 6.5 37 °C 1 vvm 450 rpm	pH 5.5 30 °C 1 vvm 350 rpm	pH 5.5 30 °C 1 vvm 350 rpm
Addition of other components	Succinic acid, LA, Glycine, IPTG	Succinic acid, LA, Glycine,	Glut+Gly(at 0 h) Cys(at 11 h)	Glut+Gly(at 11 h) Cys(at 11 h)

ALA생산을 위해 재조합 플라스미드 pFLS45를 *E. coli* BL21 (DE) pLysS (Invitrogen Co., USA)에 도입하였다. 재조합 대장균에 의한 ALA 생산을 위해 최소배지와 복합배지가 사용되었으며, 각각 전구체 (숙신산과 글라이신)와 ALAD의 저해제 (levulinic acid)를 첨가하였다. 대장균 발효는 37°C, 1 vvm, 400 rpm의 조건에서 수행하였으며 균체량, 유기산 및 기질의 분석 등 기타 자세한 사항은 이전논문의 내용과 같다(15).

효모 *S. cerevisiae* ATCC7754 (American Type Cell Collections, USA)는 glutathione (GSH)의 생산을 위해 사용되었다. 발효는 pH 5.5, 30°C, 350 rpm, 1 vvm의 조건에서 수행하였다. 배양배지는 최소배지를 이용하였으며, 포도당, 염 (NaCl, KH<sub>2</sub>PO<sub>4</sub>, KCl, MgSO<sub>4</sub>), 미량원소 수용액, 비타민 수용액 및 GSH의 생산을 위한 전구체 (글루탐산, 시스테인, 글라이신)가 첨가되었다(16). 세포내 생산물인 GSH의 분석은 세포를 3차 증류수로 세척, 원심분리하고 인산 완충용액에서 재현탁시킨 후, ultrasonic processor (20 KHz: output power 30 w, sonics & materials Inc., USA)를 이용하여 15초간, 3회 초음파 파쇄하였다. 초음파 파쇄 후 상등액과 glutathione reductase, 5,5-dithiobis (2-nitrobenzoic acid)와의 효소반응을 이용하여 GSH의 농도를 측정하였다(17). 시스테인의 농도는 copper ion (II), iron ion (III)과 1.10-phenmonohydrate 사이의 반응에 기초한 발색법을 이용하였다(18).

### 자기조직화망 (SOM)

자기조직화망은 고차원의 입력 데이터들을 저차원의 공간으로 투영시킨다. 즉, SOM은 입력벡터들 간에 기하학적인 관

제를 유지시키면서 ‘뉴런’이라 불리는 처리요소를 이용하여 고차원의 입력벡터들을 분류한다(19). SOM은 두개의 층(입력층, 출력층)으로 구성되며 입력벡터에서 각 요소들은 출력층의 뉴런에 모두 연결되어있다. SOM에서 노드들은 가중벡터( $w_i$ )라고 불리우며 입력벡터( $x_j$ )처럼 요소간에 같은 차원을 갖는다. 그리고  $x_j$ 와  $w_i$  간의 유사성 정도를 측정하기 위해 훈련 중  $k$ 번째 실행에서 거리  $d(x_j, w_i)$ 를 정의하게 된다. 거리  $d(x_j, w_i)$ 는 유클리디안 거리행렬에 의하여 계산되며, 입력벡터에 가장 근접한 뉴런이 승자뉴런으로 채택되고  $w_p$ 로 나타낸다. 일단 승자뉴런이 선택되면  $k+1$ 번째 실행에서는 다음 식(1)에 의해서 가중치( $w_i(k+1)$ )가 계산되어진다.

$$w_i(k+1) = w_i(k) + \epsilon(k) * h_p(i, k) * \{x_j(k) - w_i(k)\} \quad (1)$$

식 (1)에서  $\epsilon(k)$ 는 학습율로써  $k$ 의 함수이며 훈련횟수가 증가할수록 감소한다. 또한,  $h_p(i, k)$ 는 승자뉴런 이웃들을 나타내는 파라미터로  $k$ 번째 실행에서 승자뉴런  $p$ 이웃들의 크기를 나타내며, 훈련횟수가 증가할수록 감소한다(20). 전체과정은 최대반복횟수에 도달할 때까지 반복되며, 이러한 방법으로 많은 데이터 쌍들에 숨겨진 기하학적인 관계가 분류되게 된다(21). 생물공정에서 수집된 스펙트럼 데이터들은 배양시간과 같은 일차원적인 네트워크 상으로 투영되며, 출력층의 숫자는 스펙트럼 데이터에서 분류되어진 숫자와 같다. 스펙트럼 데이터의 분류에 관한 전체적인 작업과정은 Fig. 1에 나타내었으며, 다음과 같은 단계로 이루어진다.

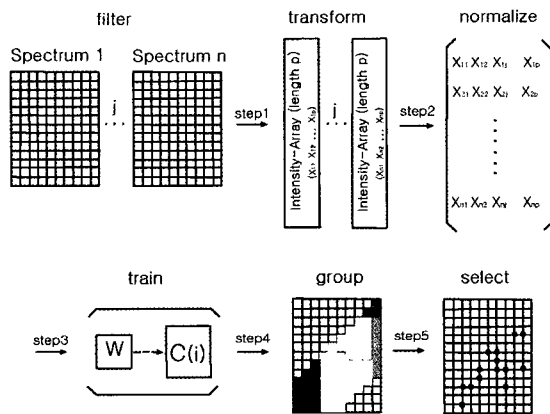


Figure 1. The overall working process for the classification of the diverse 2D fluorescence spectral data within the SOM network.

- 제 1 단계: 빛이 산란된 스펙트럼 데이터들을 필터링한 후에 데이터를 여기/방출파장의 조합에 따라 1차원의 입력 데이터쌍으로 변환시킨다(493개 입력자료).
- 제 2 단계: 변환된 스펙트럼 데이터는 벡터형태로 입력층의 노드로 입력되며 (예, 493 x 600), 입력층에서 스펙트럼과 발효시간은 각각 하나의 행 (column)과 열 (row)을 구성하게 된다. 하나의 행과 열로 구성된 스펙트럼 데이터는 행렬에서 일정한 크기를 가지도록 정규화되며 자기조직화망의 입력자료로 쓰이게 된다.
- 제 3 단계: 자기조직화망은 임의의 초기화에 의해 정렬되고, 가중치 벡터가 결정된다.

- 제 4 단계: 훈련이 끝난 후 입력 자료에 대한 유용한 정보를 제공하는 그룹분포표가 얻어진다.
- 제 5 단계: 그룹분포표에서 각 그룹의 중심영역은 유클리디안 거리를 계산한 후에 얻어진다. 즉, 그룹 내에서 가장 작은 거리값을 갖는 스펙트럼요소가 그룹의 중심영역에 포함된다. 각 그룹의 중심영역은 각 그룹의 시간에 따른 변화특성을 잘 나타내어주며, 발효공정에 대한 유용한 정보를 제공한다.

가중치 벡터의 초기화나 초기학습속도와 같은 변수들은 참고문헌의 자료를 이용하였으며, SOM 알고리즘의 학습단계는 500회 반복 후에 중단되었다(21, 22). SOM 훈련은 Matlab Neural Network Toolbox (vers. 6.1, The MathWorks, Inc., USA)를 이용하였으며 512 MB의 Pentium V 프로세서 컴퓨터가 사용되었다.

그룹분포표에서 최적그룹수는 그룹내 모든 스펙트럼요소의 형광세기 산란정도를 이용하여 계산하였다. 특정시간에 어느 한 그룹에서 스펙트럼요소의 형광세기 ( $I(s, r)$ )는 임의의 변수로 설정된다. 그리고 스펙트럼요소의 가중치 ( $W(r)$ )는 각 그룹의 평균가중치와 유사하다고 생각할 수 있으며, 특정그룹에서 형광세기의 시간의존성 변수 ( $VAR_i(r)$ )는 다음과 같이 정의된다.

$$VAR_i(r) = \frac{1}{n} \sum_{s=1}^n \{I(s, r) - w(r)\}^2 \quad (2)$$

위식에서  $s$ 는 그룹 ( $i$ )에서 스펙트럼 요소와 스펙트럼 요소의 수를 나타내며  $r$ 은 형광스펙트럼의 현재수를 나타낸다. 또한, 그룹분포표에서 스펙트럼 요소에 대한 형광세기의 평균분산 ( $VAR_m$ )은 다음 식과 같다.

$$VAR_m = \frac{1}{T} \times \frac{1}{M} \sum_{i=1}^M \sum_{r=1}^T VAR_i(r) \quad (3)$$

위식에서  $T$ 는 특정그룹에서 스펙트럼 요소의 수를 나타내고  $M$ 은 그룹의 수를 나타낸다. 평균분산은 최적그룹수를 결정하는 기준으로 사용되며 일반적으로 그룹수가 증가할수록 감소한다. 연속적인 두 그룹의 평균분산의 차이가 5% 미만일 경우 하위그룹수가 최적그룹수로 채택되어지며, 채택된 그룹에 의해서 전범위의 스펙트럼 데이터가 분류된다.

### 결과 및 고찰

발효 중 형광세기는 배양배지나 공정변수 등에 의존하며 형광세기의 변화는 전체 형광스펙트럼과 배경 형광스펙트럼의 차이로써 나타낼 수 있다. 특정성분의 첨가나 조작조건의 변화 전후에 나타나는 스펙트럼의 차이는 어떤 스펙트럼영역이 특정성분 혹은 조작변수에 민감한가를 나타낸다. 즉 배양공정에서 특정 화학종의 첨가는 평균 형광세기를 반영하는 스펙트럼 추출기법을 이용하여도 잘 해석될 수 있다는 것이다. 그러나 스펙트럼 추출기법이 배경

형광효과를 최소화하고 스펙트럼영역의 형광세기들을 공 정변수들과 상관시키는데 효과적이라 할지라도, 유용한 스펙트럼 데이터를 많은 양의 스펙트럼으로부터 찾아내는 것은 쉽지 않다. 따라서 이러한 단점들을 극복하고 전체 형광스펙트럼 해석을 위해서는 SOM과 같은 전체적인 접근법이 필요하다.

**SOM 알고리즘을 이용한 2차원 형광스펙트럼의 분류**

SOM 알고리즘은 형광데이터 분류를 위한 매우 강력한 도구로써 발효공정에 대한 사전지식이 없이도 전체 스펙트럼 데이터의 분류가 가능하다. 창이동기법 (moving-window technique)을 이용하여 스펙트럼데이터를 필터링한 후에 형광스펙트럼을 최대/최소 세기값의 범위에서 0과 1 사이로 정규화한다.

최소배지에 IPTG를 첨가한 재조합 대장균의 발효에서 수집된 형광스펙트럼 데이터를 2, 5 및 8과 같은 몇몇 그룹으로 분류하였다. 전범위 스펙트럼 데이터의 분류를 위해 필요한 최적그룹수를 결정하기 위해서 식 (3)을 이용하여 평균분산을 계산하였으며, 그룹수에 따른 평균분산의 변화를 Fig. 2에 나타내었다. 5개 이상의 그룹으로 분류되어진 그룹분포에서 그룹숫자가 증가하여도 평균분산은 많이 증가하지 않는다. 즉, 형광데이터를 5개 그룹과 6개 그룹으로 분류하였을 경우 두 그룹 간 평균분산의 차이는 5% 미만이다. 따라서 발효공정에서 수집된 스펙트럼 데이터를 5개 그룹으로 분류하였다.

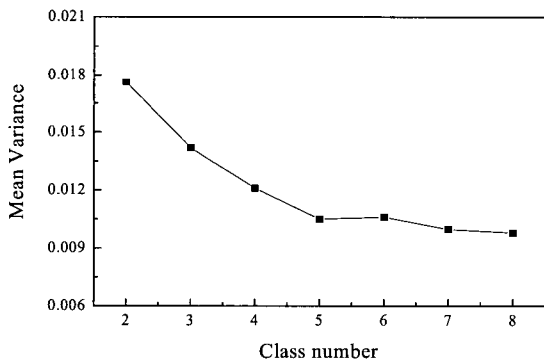
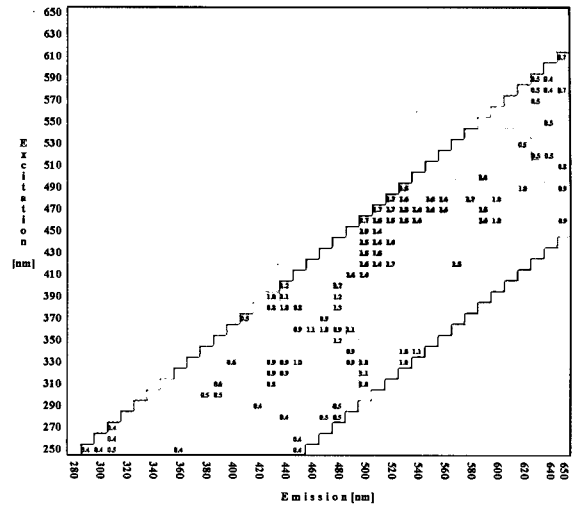


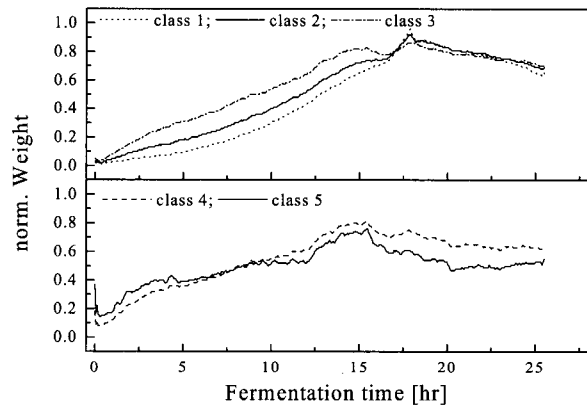
Figure 2. Mean variances with respect to the number of classes in a class distribution card.

Fig. 3은 중심영역이 포함된 5개 그룹의 그룹분포와 각 그룹의 분산과 정규화된 가중치 (weight)의 변화를 나타내고 있다. 5개 그룹으로 분류된 분포표와 2개 그룹으로 분류된 분포표에서, 2개 그룹으로 분류된 분포표의 그룹1은 5개 그룹으로 분류된 분포표에서 그룹1과 그룹2로 나누어진다. 또한 2개 그룹으로 분류된 분포표에서 그룹2는 5개 그룹으로 분류된 분포표에서 그룹 3, 4, 그리고 5로 분류되어진다. 하나의 분포표에서 몇몇 그룹의 분포는 종종 전형적인 단백질, 비타민 및 보조효소의 형광영역에 해당된다. 예를들면 8개 그룹으로 분류된 분포표에서 그룹 4는 트립토판, 티로신, 페닐알라닌등과 같은 아미노산의 형광영역에 해당하며 이 영역의 파장범위는 260-300 nm(ex)와

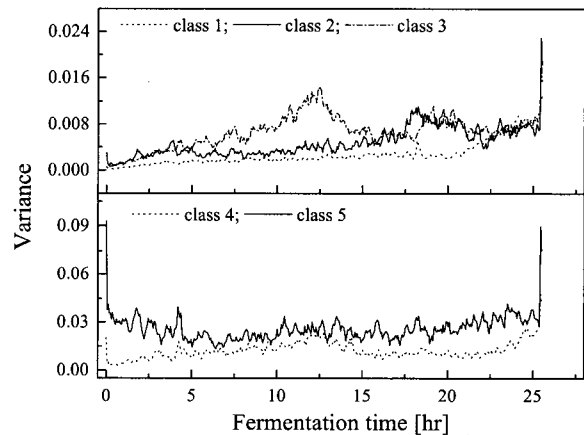
280-370 nm(em)이다.



(a)



(b)



(c)

Figure 3. (a) Class distribution card of 5 classes including central regions, and time courses of (b) the normalized weights and (c) the variances of each class (□ class 1; □ class 2; □ class 3; □ class 4; ■ class 5).

상호 유클리드공간에서 형성된 요소들로 구성된 행렬은 상관표현 (relationship presentation) 행렬으로써 정의 될 수 있으며 각 그룹에서 가장 큰 요소들 중 20% 내의 여기파장과 방출파장의 조합이 각 그룹의 중심영역으로 나타난다.

각 그룹의 중심영역은 SOM에서 “피크”로써 시각화할 수 있으며 시간의존성인 스펙트럼데이터에서 큰 변화가 일어날 것이라고 예측되는 영역을 나타낸다. Fig. 3의 5개 그룹으로 분류되어진 분포표에서 그룹 1은 34개, 그룹 2는 14개, 그룹 3은 21개, 그룹 4는 13개 그리고 그룹 5는 16개의 중심영역이 있다. 그러나 2개 그룹으로 분류되어진 분포표에서 그룹 1은 51개, 그룹 2는 47개의 중심영역을 가지고 있다(데이터 없음). 위의 결과들은 PCA 등과 같은 다른 통계적 방법들의 결과와도 비교할 수 있으며, 데이터의 축소 면에서 좋은 연관성을 나타낸다.

전체범위의 스펙트럼 데이터가 5개 그룹으로 분류될 경우 5개의 정규화된 가중치의 변화를 얻을 수 있다. 5개의 그룹으로 분류되어진 분포표에서 그룹 1과 2의 일반화된 가중치는 18시간에 최대값이 되지만 그룹 4와 5는 15시간에 최대값이 된다. 정규화된 가중치의 경향은 세포대사, 기질소모, 대사물질이나 효소의 생산 및 생물반응기에 특정성분의 첨가와 같은 정량적인 정보를 제공한다.

각 그룹에서 형광세기의 분산은 식 (2)를 이용하여 계산하였으며 시간에 따른 분산의 변화와 형광세기의 산란정도는 배양공정에 대한 정보를 제공한다. 예를 들면 5개 그룹으로 분류되어진 분포표에서 그룹 3의 분산변화는 배출가스중 이산화탄소의 농도변화를 나타낸다.

**SOM을 이용한 발효공정의 해석**

발효공정에서 수집된 스펙트럼 데이터의 분석을 위해 SOM을 이용하였으며, 최소배지와 LB 복합배지에서 각각 전체 형광스펙트럼의 변화를 조사하였다. 최소배지에서 재조합 대장균의 발효는 4.1과 다르게 IPTG를 첨가하지 않았다. 그리고 *S. cerevisiae*의 발효는 각기 다른 전구체 (시스테인, 글루탐산, 글리신)가 첨가된 최소배지에서만 수행하였다.

**Table 2.** Numbers of central regions for the fermentation processes of FM1 and FM2

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8
FM1	13	5	12	13	14	8	20	13
FM2	13	14	5	12	5	18	11	20

재조합 대장균 발효에서 얻어진 스펙트럼 데이터들을 분류하고 그룹숫자에 따른 평균분산을 계산하였다. 재조합 대장균의 발효에서 평균분산은 8개 이상으로 그룹수가 증가함에 따라서 감소하며 8개 그룹과 9개 그룹사이에서 평균분산의 차이는 5% 미만이다. 그러므로 발효에서 수집된 형광스펙트럼 데이터를 8개 그룹으로 분류하였다. 최적그룹수를 8개 그룹으로 분류한 후에 각 그룹의 중심영역을 결정하였으며 두 그룹분포표에서 각 그룹의 중심영역의

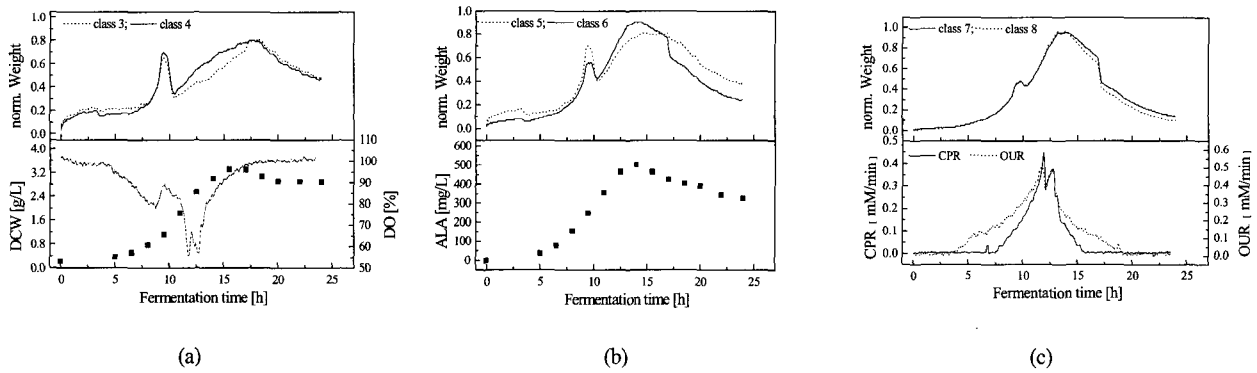


Figure 4. Comparison of the normalized weights of some classes with on- and off-line measurement data during the fermentation process FM1 ((a); class 3-4 vs. DCW & DO; (b); class 5-6 vs. ALA; (c); class 7-8 vs. CPR & OUR).

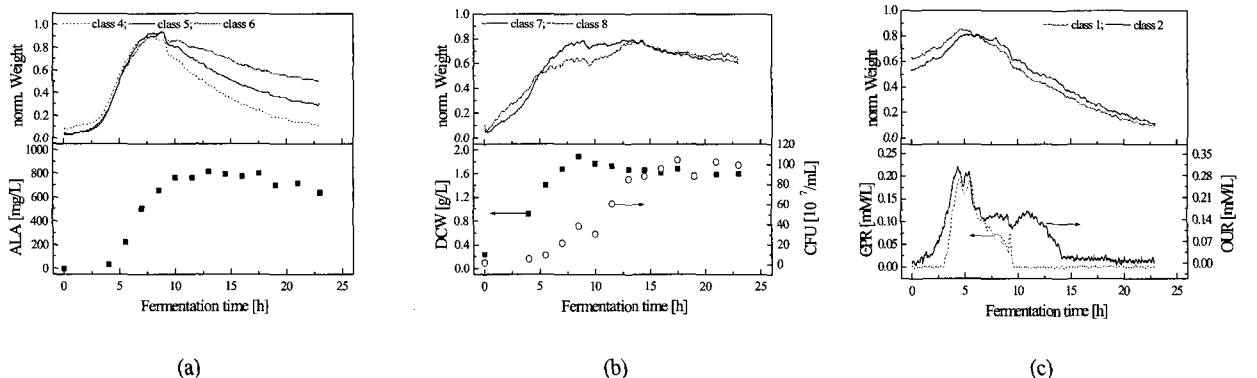


Figure 5. Comparison of the normalized weights of some classes with on- and off-line measurement data during the fermentation process FM2 ((a); class 4-6 vs. ALA; (b); class 7-8 vs. DCW & CFU; (c); class 1-2 vs. CPR & OUR).

숫자는 표 2에 나타내었다. FM1에서 여기과장 310-530 nm, 480-600 nm 그리고 방출과장 370-530 nm와 560-650 nm에 분포된 그룹 7은 그룹분포표에서 가장 많은 중심영역을 가지고 있으며, 세포대사나 발효공정에 대해서도 다른 그룹보다 더 많은 정보를 제공한다. 그리고 FM2에서 그룹 6 (Ex; 430-530 nm, Em; 470-600 nm)과 그룹 8 (Ex; 250-290 nm & 500-610 nm, Em; 270-490 nm & 560-650 nm)은 공정에 대해서 약 40%의 정보를 가지고 있다.

FerM1과 FerM2의 그룹분포표에서 시간에 따른 분산과 정규화된 가중치의 변화를 얻을 수 있었다. 발효공정에 대한 정보는 온라인 데이터와 오프라인 데이터를 시간에 따른 분산과 정규화된 가중치의 변화와 비교함으로써 얻을 수 있다.

최소배지를 이용한 재조합 대장균의 발효공정 (FM1)에서 Fig. 4(a)의 그룹 3과 4의 시간에 따른 정규화된 가중치의 변화는 건조균체량과 연관된다(이 스펙트럼 요소들은 NAD(P)H의 스펙트럼영역과 잘 일치됨). 그러나 그룹 5와 6의 시간에 따른 정규화된 가중치의 변화는 Fig. 4(b)에서와 같이 ALA의 농도와 일치한다. CPR과 OUR은 Fig. 4(c)에서와 같이 그룹 7과 8의 시간에 따른 정규화된 가중치의 변화와 연관시킬 수 있다. 그리고 9.5시간에 일반화된 가중치가 갑자기 증가하는 것은 발효공정에서 많은 거품이 생성되고 이에 따라 생물반응기로 거품제거 시약이 첨가되었기 때문일 것이다.

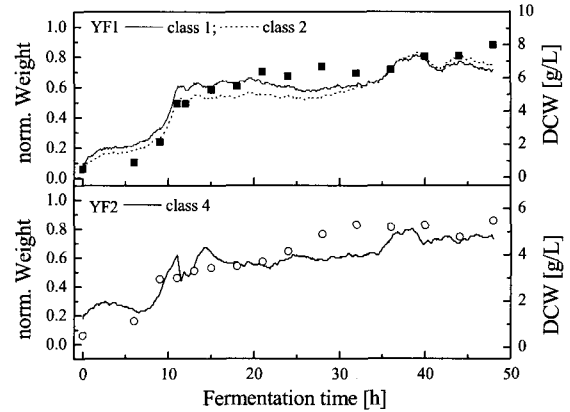
LB 복합배지를 이용한 재조합 대장균의 발효 (FM2)에서 ALA의 농도는 Fig. 5(a)에서 그룹 4와 6의 시간에 따른 정규화된 가중치의 변화와 연관시킬 수 있다. 그룹 4와 6에서 몇몇 스펙트럼 요소들은 EGFP의 형광영역에 놓여 있다. 그룹 7과 8의 시간에 따른 정규화된 가중치들의 변화는 건조균체량과 CFU (colony forming units)의 변화와 각각 일치한다. 그러나 그룹분포표에서 그룹 7과 8은 NAD(P)H의 형광영역에 속하지 않는데, 그 이유는 LB 복합배지에서 아미노산들의 분해 또는 생합성에 의한 것으로 생각된다. 그리고 Fig. 5(c)에서 CPR과 OUR은 NAD(P)H 영역에 있는 그룹 1과 2의 시간에 따른 정규화된 가중치의 변화와 연관시킬 수 있다.

**Table 3.** Numbers of central regions for the fermentations of *S. cerevisiae* with the addition of different precursors (YF1 and YF2)

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
YF1	27	19	25	10	17	-	-
YF2	8	11	9	10	23	19	18

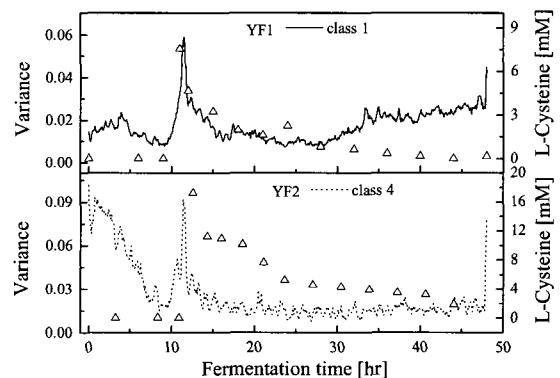
*S. cerevisiae* 발효공정 (YF1, YF2)에서 그룹수에 따른 평균분산을 계산하였다. YF1에서 그룹 5와 6의 평균분산 차이는 5% 미만이다. 또한 YF2에서 그룹 7과 8의 평균분산 차이는 5% 미만이다. 따라서 YF1과 YF2에서 얻어진 형광 스펙트럼 데이터는 각각 5개와 7개 그룹으로 분류하였으며, 각 그룹의 중심영역은 그룹분포표에 나타내었다. YF1에서 그룹 1 (Ex; 290-340 nm; Em; 330-490 nm)과 그룹 3 (Ex; 340-410 nm & 370-590 nm; Em; 380-460 nm & 560-650 nm)은 52개의 중심영역이 있으며 공정에 대해

50% 이상의 유용한 정보를 가지고 있다. 그러나 YF2에서 그룹 5와 6은 340-560 nm(ex)와 400-640 nm(em)의 과장범위에서 공정에 대해 43% 정도의 환경적, 생물학적 정보를 제공한다.



**Figure 6.** Time courses of the normalized weights of class 1 and class 2 in YF1 and class 4 in YF2, and cell mass concentrations during the fermentations of *S. cerevisiae* with YF1 and YF2.

YF1과 YF2에서 시스테인은 모두 11시간에 첨가되었다. 그러나 글루탐산과 글리신은 YF1에서 발효초기에 첨가되었으며 YF2에서는 11시간에 첨가되었다. 두 아미노산의 첨가시간 차이는 YF1과 YF2의 그룹 분포표에서 몇몇 그룹들의 정규화된 가중치들을 비교함으로써 알 수 있다. Fig. 6에서 YF1의 그룹 1과 2 그리고 YF2의 그룹 4에서 정규화된 가중치들은 세포농도와 잘 일치한다. NAD(P)H의 형광영역은 YF1의 그룹 1과 2 그리고 YF2에서 그룹 4의 영역과 일치한다. YF2의 그룹 4에서 11시간에 정규화된 가중치가 급격히 변화하는 것은 글루탐산과 글리신이 첨가되었기 때문이다. 그리고 YF2에서 가중치가 급격히 감소하는 것은 시스테인, 글루탐산 및 글리신이 첨가되어 세포내부의 대사과정이 변화하였기 때문일 것이다. YF1과 YF2에서 11시간에 첨가된 시스테인의 농도변화는 YF1의 그룹 1과 YF2에서 그룹 4의 시간에 따른 분산 변화와 잘 일치하며 그 결과를 Fig. 7에 나타내었다.



**Figure 7.** Comparison of the variances of class 1 in YF1 and class 4 in YF2 with concentrations of cysteine.

발효공정에서 수집된 많은 형광스펙트럼 데이터의 분석에는 일반적으로 스펙트럼 추출기법이 사용된다. 스펙트럼 추출기법은 스펙트럼 데이터로부터 유용한 데이터를 가려내고, 형광스펙트럼 데이터와 이산화탄소 농도와 같은 공정변수들을 연관시키는 방법이다(7-10). 그러나 이 방법은 전범위 스펙트럼 데이터 분석에 한계점을 가지고 있다. 이와같은 한계점을 극복하기 위해서 Tartakovsky 등은 생물공정 모니터링에서 유용한 정보를 추출하기 위해 단계적인 선형 다중회귀법을 이용하였다(6). 이들은 대장균이나 *S. cerevisiae*의 발효공정에서 공정변수들을 3개 혹은 4개의 스펙트럼 요소와 잘 연관시킬 수 있다는 것을 발견하였다. 또한 Skibsted 등은 *Pseudomonas fluorescences*의 발효공정 모니터링을 위해서 다중과장 형광센서를 사용하였으며, 배양액 중에 숙신산이나 단백질 그리고 배출가스내의 산소나 이산화탄소에 대한 모델을 얻기 위해서 부분최소자승(PLS) 회귀법을 정립하였다(8).

그리고 본 연구에서 형광데이터의 분류와 분석을 위해 사용된 SOM은 실시간 학습변화에 대해 연속적인 적응력을 가지고 있으며 “look-up table”에 빠르고 직접적인 접근을 할 수 있는 장점이 있다. 또한 중요한 스펙트럼 데이터는 전체영역의 스펙트럼 데이터로부터 추출할 수 있으며 (493 스펙트럼 요소 중 98개), 이렇게 추출된 데이터들은 발효공정 모델링을 위한 ANN의 입력데이터로 사용된다. 비록 SOM이 세포의 상태나 배양공정에 대한 정량적인 정보를 제공하지 않는다고 할지라도, SOM 알고리즘으로부터 얻은 분산이나 정규화된 가중치는 온라인 및 오프라인 데이터와 잘 연관시킬 수 있다.

## 요 약

본 연구는 재조합 대장균과 *S. cerevisiae*의 발효공정에서 형광스펙트럼 데이터를 수집하였으며, SOM을 이용하여 형광스펙트럼 데이터를 특정 그룹으로 분류하고 발효공정을 분석하고자 하였다. 배출가스 내 이산화탄소농도와 세포농도 같은 공정변수들은 SOM 알고리즘으로부터 얻은 분산 및 정규화된 가중치들과 좋은 연관성을 나타내었다. 전체 스펙트럼 데이터의 분류는 생물공정 모델링을 위한 매우 중요한 단계인데 그 이유는 몇몇 여기과장과 방출과장의 유의한 조합들이 전체영역의 스펙트럼 데이터로부터 추출되기 때문이다. 예를 들면, 본 연구에서 SOM을 이용하여 추출한 98개의 스펙트럼 데이터의 예제들은 부분최소자승법이나 감독신경망 (supervised neural network)을 이용한 공정의 모델링에 사용될 수 있다.

## 감 사

본 연구는 한국과학재단 (KOSEF)의 2002년도 특정기초연구과제 (과제번호: R01-2002 -000-00027-0)에 의해 이루어졌으며 이에 감사드립니다.

## NOTATION

ALA	5- Aminolevulinic acid (mg/L)
ALAS	Aminolevulinic acid synthetase
BPNN	Backpropagation neural network
C	Class
CFU	Colony forming units
CPR	Carbon dioxide production rate (mM/min)
d	Distance between the winning neuron and the input neuron I
DCW	Dry cell weight (g/L)
DO	Dissolved oxygen concentration (%)
EGFP	Enhanced green fluorescent protein
GSH	Glutathione
$h_p(i,k)$	Neighborhood size of the winning neuron p at iteration of k
i, j, p	Neuron
$I(s,r)$	Fluorescence intensity of a spectral component in a class at one determined time
IPTG	Isopropylthiogalactoside
k	Training time or number of iterations
LA	Levulinic acid
LB	Luria Bertini
M	Number of classes
n	Number of spectral components in the class
OUR	Oxygen uptake rate (mM/min)
PCA	Principal component analysis
PLS	Partial least square regression analysis
r	Present number in the fluorescence spectrum
s	Spectral components in a class i
SOM	Self-organizing map
T	Number of spectral components in a class
$VAR_i(r)$	Time-dependent variance of the fluorescence intensities of all spectral components in a class
$VAR_m$	Mean variance of the fluorescence intensities for all spectral components of all classes in a distribution card
$w_i, w_p$	Individual weight vector or winning neuron with the corresponding weight vector
xj	Input vector
YF1, YF2	Fermentation processes by <i>S. cerevisiae</i> Greek letters
(k)	Learning rate at iteration of k

## REFERENCES

- Sonnleitner, B. (2000), Instrumentation of biotechnological processes, In: Advances in Biochemical Engineering and Biotechnology, Springer, Berlin, 1-64.
- Harms P., Y. Kostov, G. Rao (2002), Bioprocess monitoring, *Curr. Opin. Biotechnol.* **13**, 124-127.
- Marose S., C. Lindemann, R. Ulber, T. Scheper (1999), Optical sensor systems for bioprocess monitoring, *Trends Biotechnol.* **17**, 30-33.
- Li J. K., E. C. Asali, A. E. Humphrey (1992), Monitoring cell concentration and activity by multiple excitation fluorometry, *Biotechnol. Prog.* **7**, 21-27.
- Schügerl K., C. Lindemann, S. Marose, T. Scheper (1998), Two-dimensional fluorescence spectroscopy for on-line bioprocess monitoring, Course material for the Bioprocess Engineering Course, Supertar, Island of Brac, Croatia, **27**, Sept.-02 Oct..
- Tartakovsky B., M. Scheintuch, J. M. Hilmer, T. Scheper (1996), Application of scanning fluorometry for monitoring of a fermentation process, *Biotechnol. Prog.* **12**, 126-131.
- Mukherjee J., C. Lindemann, T. Scheper (1999), Fluorescence monitoring during cultivation of *Enterobacter aerogenes* at different oxygen levels, *Appl. Microbiol. Biotechnol.* **52**, 489-494.
- Skibsted E., C. Lindemann, C. Roca, L. Olsson (2001), On-line bioprocess monitoring with a multi-wavelength fluorescence sensor using multivariate calibration, *J. Biotechnol.* **88**, 47-57.
- Marose S., C. Lindemann, T. Scheper (1998), Two-dimensional fluorescence spectroscopy: A new tool for on-line bioprocess monitoring, *Biotechnol. Prog.* **14**, 63-74.

10. Lindemann C., S. Marose, H.O. Nielson, T. Scheper, (1998), 2-Dimensional fluorescence spectroscopy for on-line bioprocess monitoring, *Sens. Actuat. B*, **51**, 271-277.
11. Low L. D., S. Kalelkar, E. R. Dow (2004), Self-organizing maps for the analysis of NMR spectra, *BioSilico*, **2**, 157-163.
12. Kolehmainen M., P. Ronkko, O. Raatikainen (2003), Monitoring of yeast fermentation by ion mobility spectrometry measurement and data visualization with Self-Organizing Maps, *Anal. Chim. Acta*, **484**, 93-100.
13. Debeljak Z., M. Strapac, M. Medic-Saric (2001), Application of self-organizing maps for the classification of chromatographic systems and prediction of values of chromatographic quantities, *J. Chromatog. A*, **925**, 31-40.
14. Yang, H., I. R. Lewis, P. R. Griffiths (1999), Raman spectrometry and neural networks for the classification of wood type. 2. Kohonen self-organizing maps, *Spectrochim. Acta. Part A*, **55**, 2783-2791.
15. Chung, S. Y., K. H. Seo, J. I. Rhee (2005), Influence of culture conditions on the production of extra-cellular 5-aminolevulinic acid (ALA) by recombinant *E. coli*, *Process Biochem.* **40**, 385-394.
16. Shimizu, H., K. Araki, S. Shioya, K. I. Suga (1991), Optimal production of glutathione by controlling the specific growth rate of yeast in fed-batch culture, *Biotechnol. Bioeng.* **38**, 196-205.
17. Tietze, F. (1969), Enzymic method for quantitative determination of nanogram amount of total and oxidized glutathione, *Anal. Biochem.* **27**, 502-522.
18. Teshima, N., H. Katsumate, M. Kurihara, T. Sakai, T. Kawashima (1999), Flow-injection determination of copper(II) based on its catalysis on the redox reaction of cysteine with iron(III) in the presence of 1,10-phenanthroline, *Talanta* **50**, 41-47.
19. Kohonen, T. (1998), The self-organizing map, *Neurocomputing* **21**, 1-6.
20. Vesanto, J. (1999), SOM-based data visualization methods, *Intell. Data Anal.* **3**, 111-126.
21. Kohonen, T. (2001), *Self-Organizing Maps*, 3rd ed, Springer, Berlin, Germany.
22. Zupan, J., J. Gasteiger (1999), *Neural Networks in Chemistry and Drug Design*, 2nd ed, Wiley-VCH, Weinheim, USA.