

목표물질 스크리닝을 위한 피이크 인식 알고리즘

민 흥 기 · 흥 승 홍*

= Abstract =

A Peak Recognition Algorithm for the Screening of Target Compounds

Hong Kee Min, Seung Hong Hong*

In this paper, the peak detection algorithm was developed for the purpose of screening of the target compounds. Algorithm is divided into searching the characteristic ion and peak detection. The heuristic knowledge about analytical chemistry was applied for the searching the characteristic ion. Peak detection was accomplished in comparison with the peak identification strings and pattern strings around the retention time. Pattern strings are composed with the number which generated by pattern identification function. The variables of pattern identification function are the codes which represent the difference of two adjacent abundances. Some of the free steroids were selected to demonstrate the proposed algorithm.

1. 서 론

최근, 분석화학 분야에 컴퓨터의 도입으로 분석 기기로부터의 자료획득, 획득한 자료의 보관 및 해석, 분석기기의 제어, 분석기기의 최적상태 결정등에 대한 자동화가 급속히 이루어지고 있다. 이와 같은 기본적인 컴퓨터의 응용 이외에도 분석방법을 결정하거나 복잡한 분석 결과의 해석을 위한 소위 지능이 부여된 프로그램을 개발하는 연구도 활발히 진행되고 있다. 그러나 현 단계에서 자동화

시스템에 의한 처리 결과는 아직까지는 만족할 만한 결과를 얻지 못하고 있으므로 수작업의 해석과정이 병행되고 있다. 또한 요구되는 정밀도와 정확도는 주로 통계적인 방법에 의하여 최적상태를 결정하는 방법을 사용하고 있기 때문에 신뢰성을 충족시키지 못하며, 최적방식의 결정은 분석화학자의 경험을 바탕으로한 전문지식에 크게 의존한다. 즉 이러한 과정의 작업들은 어디까지나 인간이 문제를 해결하는 과정의 보조수단으로 사용될 뿐이지 인간을 대신하여 문제를 해결한다는 것은 불가능하다.^[1-13]

현재, 분석화학분야에서 다루는 목표물질에 대한 스크리닝(screening)은 분석 기기로부터 얻어진 분석 결과를 단순 그래프로 표시한 후 이 출력 결과를 전문 분석 화학자의 목적에 의해 행해지고 있다. 그러나 이러한 방법은 분석화학자의 경험과 지식에 크게 의존하며 주관적인 판단과 목적으로 인

(접수 : 1993년 5월 17일)
인천대학교 정보통신공학과, 인하대학교 전자공학과*
Dept. of Information & Telecommunication Eng.,
Incheon Univ.
Dept. of Electronic Eng., Inha Univ.*

한 오류 발생의 가능성이 존재하고, 또한 모든 출력결과를 확인해야 하는 과정이 필요하다. 따라서 이러한 문제점을 해결하기 위해서는 보다 객관적이고 수량화된 출력이 요구되며, 동시에 분석 결과를 재확인해야 하는 시료수를 최소화하여야 한다. 이런 요건을 충족시키기 위해서는 자동화된 스크리닝 방법이 제공되어야 하며, 이러한 과정을 수행하려면 분석결과 자료의 질량스펙트럼에서 피이크를 인식하는 과정이 필요하다. 즉 분석화학의 한 분야인 목표물질의 스크리닝과정을 자동화하여 스크리닝에 소요되는 시간을 가능한 줄이고, 분석 결과에 대한 객관적인 판단을 수행하기 위한 피이크 인식 알고리즘을 개발하는 것이 본 연구의 목표이다. 물론 이러한 자동화 스크리닝과정은 매우 제한된 경우로서, 연속적인 반복작업, 제한된 시간내에 분석을 완료해야 하는 작업, 분석대상 시료의 수가 매우 많은 경우, 대상 목표물질의 수가 비교적 적은 경우 및 목표물질의 존재 여부를 확인하는 정성분석등에 적용하기 위한 것이다.

이를 위하여 본 연구는 화학 분석기기에서 발생된 분석자료의 특성해석 분야중 제한된 용도의 목표물질 인식을 위한 전처리 과정, 즉 목표물질 스크리닝에 관한 내용으로서, 이에 필요한 피이크인식 알고리즘을 제안하고 이 알고리즘을 사용하여 목표물질을 스크리닝하여 의심스러운 양성 시료를 선별하는 것이다.

제안한 알고리즘은 두가지 개념을 사용하였다. 그중 하나는 분석화학자의 분석자료인식에 관한 선행적 지식으로 이것은 내부 표준물질(internal standards)의 존재 시간으로부터 목표물질의 예상 검출 시간의 계산 및 선택적 이온 모니터링(selected ion monitoring) 형식에서 목표물질을 나타내는 특성이온의 질량전하비(mass to charge ratio : m/z) 등을 탐색하는 과정이다. 다른 하나는 목표물질의 존재 가능성을 나타내는 피이크의 인식이다. 이것은 연속된 3개의 정규화된 어번던스(abundance)의 차를 부호화한 후, 이 부호에 의해 질량 어번던스 변화에 따른 패턴을 지정하고, 시간의 변화에 따른 패턴을 나타내는 숫자의 연속적인 나열에 의한 문자열(character string)을 생성하여, 생성된 문자열과 피이크를 나타내는 문자열을 비교함으로써 피이크를 판정한다.

실험과정은 자유형 스테로이드 계통의 약물을 대상 목표물질로 설정하고, 양성으로 알려진 시료

의 분석기기 출력자료와 '88 서울 올림픽 경기 기간중 수거된 1,601개 시료의 분석기기 출력자료중 재생 가능한 850개의 자료를 선택하여 스크리닝을 수행한 후, 양성으로 판정된 결과와 비교하여 이의 유용성을 확인하였다.

2. 피이크인식 알고리즘

우선 자료분석 과정에서 생산된 목표물질의 존재가능 시간창내의 모든 레코드로부터 해당 질량전하비와 이의 어번던스를 탐색하여 목표물질의 절대 머무름 시간 전후의 어번던스의 변화에 따른 패턴형태를 지정한 후 결정된 패턴형태의 연속적인 나열, 즉 패턴형태를 표시한 구문에 의하여 피이크를 인식하는 과정을 하나의 목표물질의 모든 특성이온에 적용하여 양성/음성의 예비판정을 내린다.

목표물질 인식에 관한 알고리즘을 다음과 같다 <알고리즘>

Step 1 (Target m/z Searching)

- : Read Abundance of Target m/z in Time Window
- : Select the Maximum Abundance's RT around the RT of Target Compound

Step 2 (Determination of Characteristic Ion peak)

- : Setting Relative Scaling and Determination of Threshold
- : Production of pattern Identification Number
- : Production of Syntax around the RT of Target compound

Step 3 (Result for One Target M/z)

- : Comparison with the Peak Syntax
- : Production of Screening Result for One Target m/z

Step 4 (Repeation)

- : Repeation from Step 1 to Step 3 for All Target m/z

Step 5 (Results for one target Compound)

- : Production of Screening Results for one Compound

Step 6 (Results for One Sample)

- : Repeation form Step 1 to Step 5 for All Target Compounds
- : Production of Screening Results for One Sample

2. 1 목표 특성이온 탐색

분석대상 시료에서 목표물질의 검출을 위해서는 목표물질을 나타내는 피이크의 인식과정이 필요하다. 목표물질의 존재가능 시간창내에 존재하는 분석자료중 대상 특성이온의 질량 전하비와 해당 어번던스를 탐색한후, 이들 자료중 계산된 목표물질의 절대 머무름 시간 전후의 3 자료를 선택하여 이중 최대값에 해당하는 시간을 수정된 예상 피이크의 머무름 시간으로 결정한다. 또한 목표특성이온의 인식에 필요한 패턴을 지정하기 위하여 수정된 머무름 시간의 어번던스와 수정된 머무름 시간 전후의 각각 3 개의 어번던스, 총 7개의 어번던스를 배열에 기록한다. 이러한 목표특성이온 탐색과정을 나타내는 가상코드는 다음과 같다.

```

RESET Mass-Data File
READ Record
SET Retention Time of Record
IF (Retention Time is in Time Window)
SET Array Masss Abundance
DETERMINE the Predicted Peak Value around
the Calculated RT
SELECT the Abundance for Pattern Indication
    
```

2. 2 특성이온 피이크의 결정

피이크의 인식을 위해서는 우선 대상 자료중 인접한 두 어번던스를 비교하여 증가, 감소 및 평형의 상태를 알아야한다. 그러나 분석자료의 어번던스는 목표 특성이온에 따라서 달라지기 때문에 절대 어번던스는 의미가 없으며 특히 절대 어번던스를 이용하여 문턱값(threshold value)을 지정할 수 없다. 따라서 문턱값을 지정하기 위해서는 시간창안의 최고값을 상대적 최고값으로 정규화하고 상대적인 값 "1"을 문턱값으로 설정하였으며 시간창내의 자료를 그래프로 출력하기 위하여 어번던스 눈금의 표시구간은 상대적 최고 어번던스에 의해 결정되었다. 본 연구에서는 프린터의 출력크기를 고려하여 상대적 최고값을 "120"으로 설정하였다. 다음은 어번던스 눈금의 표시구간을 구식이다.

$$S = a \times b / c \dots\dots\dots (1)$$

여기서, S는 어번던스 눈금의 표시구간이고, a는 설정한 상대 최고값이고, b는 수정된 머무름 시간

에서의 어번던스이며, c는 시간창내의 최대 어번던스이다.

다음 단계는 탐색된 수정된 머무름 시간 주위의 7개 어번던스중 앞에서부터 차례로 연속적인 3개의 어번던스를 택하여 3자료에 의한 패턴을 지정하고자 한다. 3개의 어번던스에 의해 발생되는 패턴형태는 그림 1과 같다. 우선 피이크 인식에 필요한 패턴지정을 위한 패턴 지정함수를 유도한다. 패턴형태 지정을 위하여 상태결정 함수 F(g)를 정의한다.

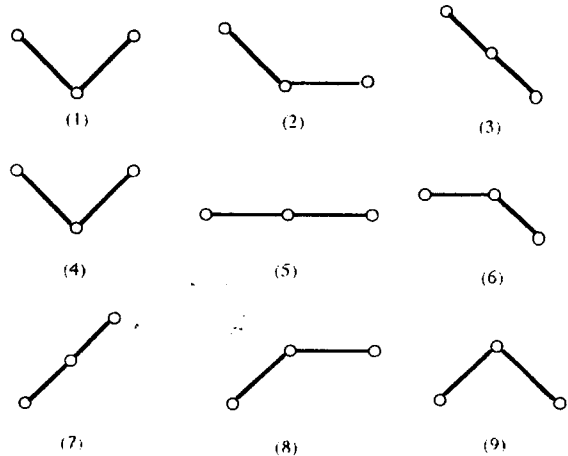


그림 1 3 점에 의한 패턴
Fig. 1 Pattern generated by 3 points

표 1 패턴 지정함수의 출력

Table1 Output results from pattern identification function

dl	d2	PI(d1,d2)	Description
-1	1	1	Beginning of Peak
-1	0	2	Beginning of Saddle Point
-1	-1	3	Going Down
0	1	4	End of Saddle Point
0	0	5	No Change
0	-1	6	End of Saddle Point
1	1	7	Going Up
1	0	8	Beginning of Saddle Point
1	-1	9	Peak Point

정의 1 :

임의의 실수 g와 양수 TH에 대하여 상태결정 함수 F는 다음과 같은 값을 갖는다.

$$F(g) = \begin{cases} 1 & (\text{if } g > TH) \\ 0 & (\text{if } |g| \leq TH) \\ -1 & (\text{if } g < -TH) \end{cases} \dots\dots\dots (2)$$

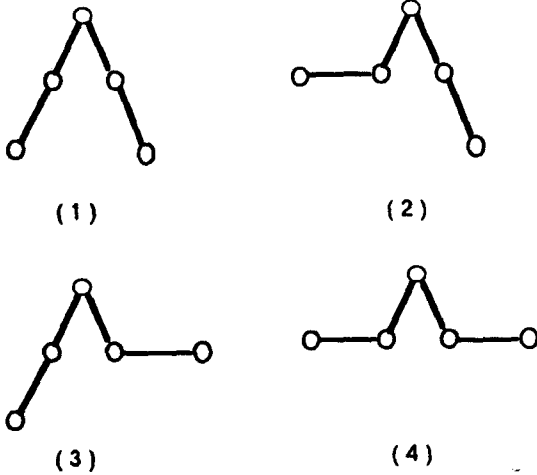


그림 2 피이크 패턴
Fig. 2 Possible peak patterns

다음단계는 어번던스의 최고값을 일정한 값으로 정규화하기 위한 정규화 함수 N(h)를 정의한다.

정의 2 :

임의의 양수 h에 대하여 정규화 함수 N은 다음과 같이 정의한다.

$$N(h) = k \times h / \text{MAX}(h) \dots\dots\dots (3)$$

여기서 MAX(h)는 주어진 집합의 원소중 최고값을 나타내며, $h \in \{h_i\}$ 이고 k는 정규화 상수로서 본 연구에서는 120으로 정하였다.

이때 TH를 문턱값 즉, 시간창내의 상대적값 "1"로 설정하고, g를 두개의 연속된 검출시간의 어번던스를 정규화한 값의 차 즉,

$$g = N\{y(i+1)\} - N\{y(i)\} \dots\dots\dots (4)$$

라 하면, F(g) 값에 따라서 증가, 감소, 평형이 구별될 수 있다. 또한 연속된 3개의 어번던스 즉, $y(i), y(i+1), y(i+2)$ 에 대하여 $g1 = N\{y(i+1)\} - N\{y(i)\}$, $g2 = N\{y(i+2)\} - N\{y(i+1)\}$ 이라 하고, $d1 = F(g1)$, $d2 = F(g2)$ 라 하면, d1, d2는 두 어번던스 사이의 변화를 나타내는 부호가 된다. 즉 증가이면 1, 감소이면 -1, 평형이면 0의 값을 갖는다. 따라서 정의된 d1, d2를 이용하면 패턴지정함수 (pattern identification function) PI(d1, d2)는 다음 식으로 결정된다.

$$PI(d1, d2) = 3d1 - d2 + 5 \dots\dots\dots (5)$$

표 2 피이크 패턴의 문자열
Table 2 Character strings of possible peak pattern

패턴형태			그림 2의 번호
7	9	3	1
4	9	3	2
7	9	2	3
4	9	2	4

여기서 d1, d2는 인접한 두 어번던스의 변화를 나타내며 d1, d2값에 의해 결정되는 패턴지정함수의 출력값은 그림 1의 각 패턴을 나타내는 번호와 같으며 이 값은 연속된 3개의 어번던스중 중앙의 어번던스 전후의 변화상태를 나타낸다. 각각의 d1, d2에 대한 PI(d1,d2)의 출력결과와 변화상태를 표 1에 나타냈다. 또한 수정된 머무름 시간 주위에서 선택된 5개의 어번던스의 변화에 의해 피이크로 인식될 수 있는 패턴이 그림 2에 나타나 있다.

이 경우 연속된 5개의 어번던스중 앞에서부터 순차적으로 3개의 값을 취하여 정규화한 후 발생된 d1,d2를 이용하여 식(5)로 표시된 패턴지정함수의 값을 구하면, 3개의 값이 생산된다. 이 값을 연속적으로 나열하면, 수정된 머무름 시간 주위에서 패턴을 나타내는 문자열이 된다. 조합 가능한 $9 \times 3^2(81)$ 개의 패턴중 그림 2에 나타나 있는 피이크 패턴에 대한 문장은 표 2와 같다. 따라서 대상 목표물질의 수정된 머무름의 시간 주위에서 모든 특성이온의 패턴이 표 2의 패턴열과 같은가를 비교한 후 기본 피이크를 포함하여 전체 특성이온중 2/3개이상의 특성이온에서 피이크 검출되고, 동시에 기본 피이크 어번던스가 다른 피이크의 어번던스보다 크면 의심스러운 양성으로 스크리닝한다. 또한 이러한 과정을 모든 목표물질에 적용한 후 힌개의 시료에 대한 스크리닝의 최종판정을 내린다. 이러한 특성이온 피이크 결정과정의 가상코드는 다음과 같으며 전체적인 흐름도를 그림 3에 나타냈다.

```

RESET Mass Data File
REPEAT (All Target Compounds for One Sample)
REPEAT (All Target Characteristic Ions for One Target Compound)
DO (One Target Characteristic Ion)
DEAD Abundance
    
```

DETERMINE Modified RT
 PRODUCT Pattern Identification Number
 and Pattern Syntax
 COMPARE with Peak Syntax
 RESULT
 IDENTIFY for One Target Compound
 RESULT
 IDENTIFY for One Sample
 RESULT

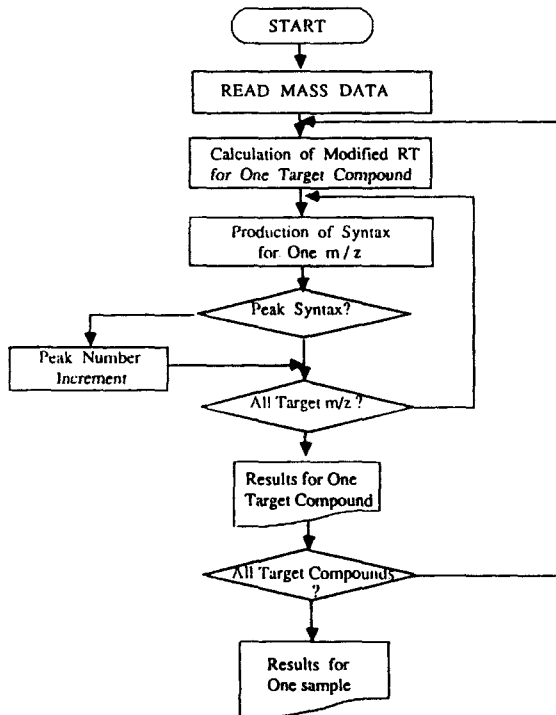


그림 3 알고리즘의 흐름도
 Fig. 3 Flow chart of algorithm

3. 실험결과 및 고찰

제안된 알고리즘의 타당성 및 유용성을 판정하기 위한 실험대상 목표물질을 자유형 스테로이드 (Free steroids) 계통으로 선택하여 자료처리를 수행하였다. 수행방법은 기존의 알려진 분석자료에 제안된 알고리즘을 적용하여 타당성을 확인한 후 '88 서울 올림픽 경기 기간중 수거된 1601개의 실제 시료의 분석자료중 재생 가능한 850개의 자료를 대상으로 목표물질의 양성/음성을 확인하여 제

안된 알고리즘의 유용성 여부를 확인하였다.

우선 자유형 스테로이드의 표준분석자료를 이용하여 타당성을 확인하였다. 화일명 >FSE06은 메산디에논이 양성인 자료이고 화일명 >FSE08은 오랄 튠리나볼이 양성인 자료이다. 그림 4에 화일명 >FSE06에 대한 기존의 화학분석 방식에 의한 스크리닝 과정에 관한 출력 결과이며 그림 5는 제안된 알고리즘을 사용하여 화일명 >FSE06을 처리한 결과로서 스크리닝 대상물질의 특성이온별로 계산된 머무름 시간 주위에서 어번던스의 변화를 코드화 하여 피이크를 인식한 후 판정기준에 의하여 의심스러운 양성/음성의 판정결과를 나타냈다, 그림 6은 양성으로 스크리닝된 물질의 특성이온들의 질량전하비에 대한 어번던스의 변화를 나타낸 그래프 출력이다. 그림 5에서 보듯이 메산디에논과 오랄 튠리나볼은 머무름시간이 4.880 분과 5.610 분에서 각각 4개의 특성이온의 질량전하비에 대하여 베이스 피이크를 포함하여 3 개의 피이크를 갖고 있기 때문에 모두 의심스러운 양성으로 판정되었지만 실제로는 메산디에논만이 양성이다. 그 이유는 그림 6에 나타나 있듯이 특성이온의 질량전하비 143 에서 메산디에논의 피이크 어번던스가 오랄 튠리나볼의 피이크 어번던스보다 약 100배의 크기를 갖는다. 즉, 오랄 튠리나볼의 베이스 피이크를 포함한 다른 특성이온의 피이크 어번던스가 매우 작다. 또한 오랄 튠리나볼의 두번째 특성이온의 질량전하비 243에서 피이크 어번던스가 베이스 피이크 어번던스의 약 3배이기 때문에 양성 판정 조건에 포함되지 않는다. 따라서 재검정 및 확인 과정을 거친 후 오랄 튠리나볼은 음성으로 판명되었다. 이러한 스크리닝 결과를 가상의 양성이라 부르며 스크리닝 과정에서 발생 가능하다. 또한 화일명 >FSE08 은 머무름 시간 5.601 분에서 오랄 튠리나볼의 4 개의 모든 특성이온의 질량전하비에대하여 피이크가 존재하고 양성 판정 기준에 적합하기 때문에 의심스러운 양성으로 판명되었다. 이 경우는 화일명 >FSE08의 자료가 오랄 튠리나볼이 검출된 표준분석자료의 결과와 일치함을 보였다.

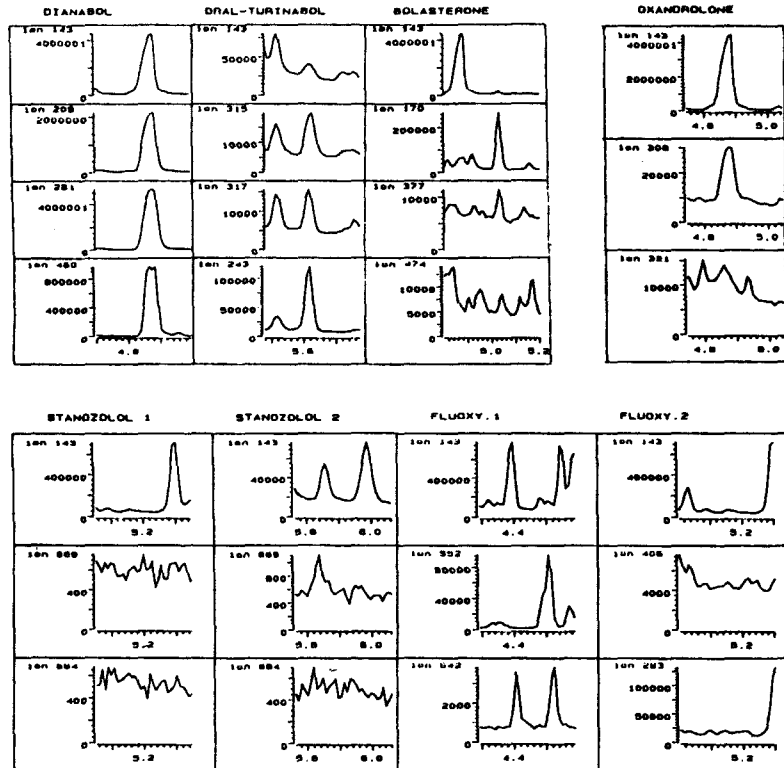


그림 4 GC/MS에 의한 >FSE06자료의 스크리닝 결과
Fig. 4 Screening results for >FSE06 by GC/MS

다음은 실제의 850개 시료의 분석기기 출력자료를 제안된 알고리즘을 적용하여 유용성을 알아 보았다. 이때는 대부분의 자료가 음성이기 때문에 출력시간 단축을 위하여 음성으로 판정된 경우는 그라프 출력을 생략하고 각 목표물질에 대한 분석결과만 출력하며, 양성으로 의심스러운 경우는 분석결과와 그라프를 함께 출력하였다. 실험을 수행한 결과 850개의 자료중 67 개의 자료가 의심스러운 양성으로 판정되었으며 이 중에는 실제로 양성인 3자료를 포함하고 있었다. 따라서 의심스러운 양성으로 판정된 비율은 약 7.9%로서 자료의 양이 방대한 경우등 본 연구에서 적용하고자 하는 제한된 조건에서는 실험의 결과로부터 피이크인식 알고리즘의 유용성을 확인하였다. 의심스러운 양성으로 판정된 67자료중 실제 양성인 자료에 대한 스크리닝 결과를 표 3에 나타냈다.

표 3 실험자료의 양성결과

Table 3 Positive results for experimental data

파일명	시료명	양성물질	여무름시간(분)	피이크 m/z
>DA41A::B2	#18-26	Oxandrolone	5.717/6.372	143,308,321
>EA07A::B2	RE#18-26	Oxandrolone	5.735/6.392	143,308,321
>GA50A::B3	#22-112	Stanozolol	6.098/6.390	143,669,684
			6.973/6.390	669,684

4. 결 과

분석화학자의 선형적지식을 이용하여 스크리닝 알고리즘을 작성한 후 운동선수의 복용 금지약물 중 자유허스테로이드 계통의 불질을 대상물질로 선정하여 분석기기의 원시자료에 대한 스크리닝을 수행하였다.

우선 제안된 알고리즘의 타당성 여부를 확인하기 위하여 기존의 알려진 자료를 처리하였다. 이들 자료중 파일명 >FSE06은 메산디에논과 오랄 튠리나볼이 의심스러운 양성으로 인식되었으며 화일

```

***MASS DATA === >FSE06 : : B6
RETENTION TIME ---4.880MIN.
10N NUMBER --- 143 PATTERN = 77933 PEAK DETECTED
10N NUMBER --- 209 PATTERN = 77933 PEAK DETECTED
10N NUMBER --- 281 PATTERN = 77933 PEAK DETECTED
10N NUMBER --- 460 PATTERN = 91933 NO PEAK
      POSITIVE : Mathandienone  SUSPICIOUS
RETENTION TIME ---5.610MIN.
10N NUMBER --- 143 PATTERN = 77933 PEAK DETECTED
10N NUMBER --- 243 PATTERN = 77933 PEAK DETECTED
10N NUMBER --- 315 PATTERN = 77793 NO PEAK
10N NUMBER --- 317 PATTERN = 77933 PEAK DETECTED
      POSITIVE : Oral -turinabol  SUSPICIOUS
RETENTION TIME ---5.070MIN.
10N NUMBER --- 143 PATTERN = 31793 NO PEAK
10N NUMBER --- 170 PATTERN = 33333 NO PEAK
10N NUMBER --- 377 PATTERN = 33331 NO PEAK
10N NUMBER --- 474 PATTERN = 93333 NO PEAK
      NEGATIVE
RETENTION TIME ---5.260MIN.
10N NUMBER --- 143 PATTERN = 77777 NO PEAK
10N NUMBER --- 669 PATTERN = 17917 NO PEAK
10N NUMBER --- 634 PATTERN = 77917 NO PEAK
      NEGATIVE
RETENTION TIME ---5.820MIN.
10N NUMBER --- 143 PATTERN = 47777 NO PEAK
10N NUMBER --- 669 PATTERN = 77793 NO PEAK
10N NUMBER --- 684 PATTERN = 17917 NO PEAK
      NEGATIVE
RETENTION TIME ---4.990MIN.
10N NUMBER --- 143 PATTERN = 32631 NO PEAK
10N NUMBER --- 308 PATTERN = 31931 NO PEAK
10N NUMBER --- 321 PATTERN = 31931 NO PEAK
      NEGATIVE
RETENTION TIME ---4.500MIN.
10N NUMBER --- 143 PATTERN = 79319 NO PEAK
10N NUMBER --- 552 PATTERN = 77933 PEAK DETECTED
10N NUMBER --- 642 PATTERN = 24779 NO PEAK
      NEGATIVE
RETENTION TIME ---5.289MIN.
10N NUMBER --- 143 PATTERN = 55624 NO PEAK
10N NUMBER --- 283 PATTERN = 47933 PEAK DETECTED
10N NUMBER --- 406 PATTERN = 77917 NO PEAK
      NEGATIVE
    
```

그림 5 제안된 알고리즘에 의한 화일명 >FSE06의 스크리닝결과

Fig. 5 Screening results for >FSE06 by proposed algorithm

명 >FSE08에서는 오랄 튠리나볼만이 의심스러운 양성으로 인식되었다. 즉, 양성으로 알려진 자료에 대하여 제안된 스크리닝 알고리즘을 적용한 결과 기존의 화학분석자의 지식과 경험에 의한 판단을 컴퓨터에 의하여 대신할 수 있다는 사실을 알았고 이로부터 제안된 스크리닝 알고리즘의 타당성이 입증되었다. 물론 본 연구에서는 화학분석자들이 사용하는 어번던스의 상대적 크기비에 관한 지식을 사용하지 않고 목표물질을 인식하고자 하였기 때문에 가상의 양성 가능성도 존재함을 알 수 있었다.

또한 제안된 알고리즘의 유용성을 확인하기 위하여 실제 시료의 분석자료를 대상으로 실험하였다. 실제 자료는 '88 서울 올림픽경기시 수거된 1,601 개의 시료에 대한 분석자료중 재생 가능한 850개의 자료를 선택하였다. 제안된 스크리닝 알고리즘에 의하여 자유형스테로이드 물질에 대하여 양성으로 스크리닝된 시료는 총 67 개로 850 개의 자료중 양성으로 확인된 3자료를 모두 포함하고 있었다. 따라서 가상의 양성으로 검출된 비율은 약 7.9% 이다. 따라서 자료의 양이 방대한 경우, 연속적인 분석이 필요한 경우, 제한된 시간내에 결과를 얻어야 하는 경우, 목표물질의 수가 비교적 적은 경우등에는 제안된 알고리즘을 이용하여 목표물질을 스크리닝 할수 있을 것으로 사료되며 다음과 같은 결론을 얻었다.

1) 정규화된 2 값의 차를 부호하여 증감상태와 문턱값을 나타냈고 연속적인 2개 부호에 의해 존재 가능한 9 개의 패턴형태를 설정하였으며 3 값 사이의 패턴 지정함수의 출력은 3 값으로 구성되는 패턴의 고유번호를 지정할 수 있었고 패턴 지정함수의 출력값의 나열에 의해 머무름 시간 전후 7개의 어번던스의 변화를 5문자로 구성된 문자열로 나타내었다.

2) 머무름 시간 전후 5개의 어번던스의 변화에 의해 생성되는 81개의 문자열중 피이크를 나타내는 4개의 문자열을 설정하여 발생된 문자열과 피이크를 나타내는 문자열을 비교하는 방법에 의해 피이크를 인식하였으며 설정된 9개의 패턴과 4 개의 피이크를 나타내는 문자열에 의하여 목표물질을 나타내는 피이크의 인식이 가능 하였다.

3) 제안된 스크리닝 알고리즘을 확인된 양성 자료에 적용하여 목표물질을 정확히 검출하였다.

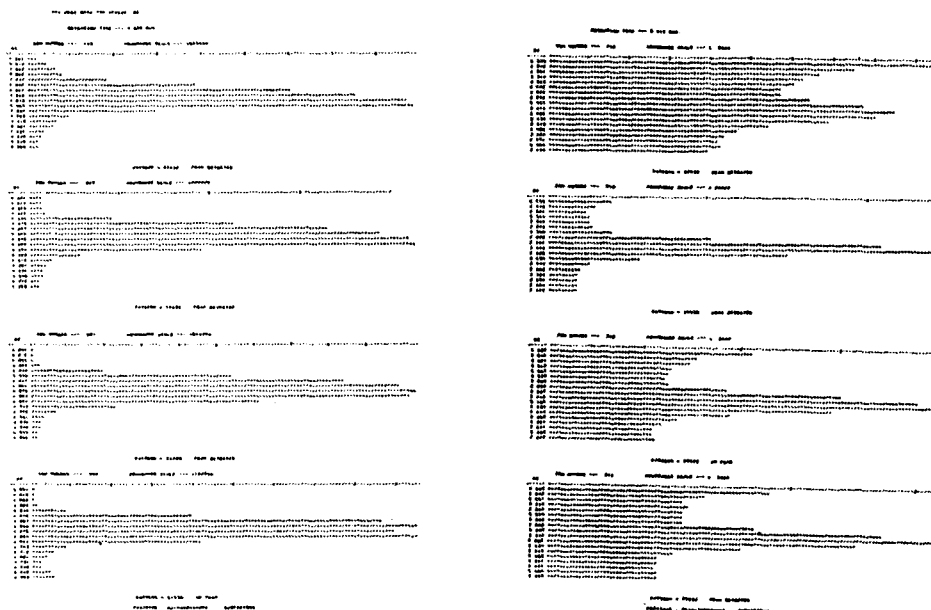


그림 6 스크리닝결과에 대한 그래프 출력
Fig. 6 Graphic output for the screening results

따라서 본 알고리즘의 타당성이 입증되었다.

4) 제안된 스크리닝 알고리즘을 850개의 실제 자료에 적용하여 목표물질의 검출을 수행한 결과 67 개의 자료에서 양성으로 스크리닝되었다. 이는 실제의 양성결과보다 많았지만 스크리닝 과정에서 발생할 수 있는 경우로서 기존의 방식에서도 동일한 현상이 발생된다. 따라서 본 알고리즘의 유용성도 입증되었다.

이상의 결론으로부터 본 연구별과의 유용성이 확인되었으며 제안된 알고리즘을 이용한 목표물질 스크리닝을 성공적으로 수행하였다. 또한 본 연구에서는 목표물질을 자유형스테로이드중 운동선수의 복용 금지약물에 한하였지만 여타 다른 물질의 스크리닝에도 적용 가능하리라 사료된다.

참 고 문 헌

- 1) Kurt Varmuza, 'Pattern recognition in chemistry', Springer-Verlag, 1980.
- 2) C. Armanino, M. Forina, P.H.E. Gardiner, G. Kateman, S. Lanteri, H. Smit and B. G. M. Vandeginste, 'Chemometrics and species identification', Springer-Verlag, 1987.

- 3) O.Strouf, 'Chemical pattern recognition', John Wiley & sons Inc., New York, 1986.
- 4) T.H. Pierce and B.A. Hohne, 'Artificial intelligence application in chemistry', American Chemical Society, Washington, 1986.
- 5) D.L. Massart, B.G. M. Vandeginste, S.N. Deming, Y. Michotte and L Kaufman, 'Chemometrics', Elsevier Science Publishing Co., New York, 1988.
- 6) Philip G. Barker, 'Computer in analytical chemistry', Pergamon Press Inc. New York, 1983.
- 7) B.R. Kowalski, 'Chemometrics', Wiley. Inter-science, New York, 1986.
- 8) 민홍기, 홍승홍, 'GC/MS 분석자료의 목표물질 인식에 관한 연구', 대한 의공학회지, Vol. 10, No. 3, pp 223-227, 1989.
- 9) D. Coomans and D. L. Massart, 'Optimization by statistical linear discriminant analysis in analytical chemistry', Analytica Chimica Acta, Vol. 112, pp97-122, 1979.
- 10) P.S. Shoenfeld and J. R. Devoe, 'statistical and mathematical methods in analytical chemistry', Analytical Chemistry, Vol. 48, No. 5, April, pp

- 403R-411R, 1976.
- 11) K. Varmuza, 'Pattern recognition in analytical chemistry', *Analytica Chimica Acta*, Vol. 122, pp227-240, 1980.
 - 12) M.P. Derde and D.L. massart, 'supervised pattern recognition : the nideal method?', *Analytica Chimica Acta*, Vol. 191, pp1-16, 1986.
 - 13) D. P. Martinsen, 'Survey of computer aided methods for mass spectral interpretation', *Applied Spectroscopy*, Vol. 35, No. 3, pp 255-266, 1981.