

GC/MS분석자료의 목표물질 인식에 관한 연구

민 흥 기 · 흥 승 흥*

A Study on the Identification of Target Compounds From the GC/MS Data

Hong Kee Min, Seung Hong Hong*

- Abstract -

In this paper, the computerized interpretation of the analytical chemical data, especially GC/MS data, was performed for the purpose of prescreening of the target compounds. First, the data from the analytical instrument was analyzed to get the information about the retention time of the ISTD and the time interval between the records. Second, the identification of the characteristic ion peaks was performed by calculating the ratio of the heights and the relative slope sensitivity of the characteristic mass abundance.

1. 서 론

현재 의학, 화학, 생물학 등에서 사용되고 있는 분석장치는 전자공학의 발달과 함께 급속히 수량화 및 전산화 되고 있다. 그러나 이들 분석기기에 장치된 컴퓨터에 의해 처리된 출력결과는 단순자료로써 수량화 혹은 그래픽으로 나타내는 것이 가장 최근의 실정이기 때문에 각 응용분야의 전문적인 지식과 숙달된 경험이 없으면 출력결과를 해독할 수 없다.^{1,2)}

따라서 본 연구에서는 가스크로마토그래피/질량스펙트로메트리(Gas Chromatography/Mass Spe-

ctrometry : GC/MS)에서 발생된 자료를 적용 대상으로 하고, 전문가에 의한 분석과정을 알고리즘화하여 이를 이용한 분석의 기초자료를 제공하는 것이 그 목적이다.

본 논문은 자료의 분석과정, 즉 분석기에서 발생된 자료를 해독하여 인식에 필요한 정보를 획득하는 과정과 자료의 인식과정, 즉 자료분석과정을 수행하여 획득한 자료를 이용하여 목표물질을 인식하는 과정으로 구성되었다.

2. 자료분석

우선 파라미터화일을 읽고 해당되는 내부표준물질(Internal Standard : ISTD)을 인식한 후 판정하고저 하는 물질들의 머무름시간(Retention Time)을 구하고, 그 머무름시간 근처의 자료로 부터 입력된 질량수(Mass Number)와 이에 해당하는 질량크기(Mass Abundance)를 찾아내어 피크값과 그 피크의 정의에 필요한 보조자료를 출력한다.

<접수 : 1989년 11월 30일>

한국과학기술연구원 유전공학센터

*인하대학교 공과대학 전자공학과

Genetic Engineering Center, Korea Institute of Science and Technology.

*Dept. of Electronics, Inha Univ.

2-1 알고리즘

입력된 내부표준물질의 질량수와 시간폭자료 (Time Range Data)로 부터 내부표준물질의 정확한 시간과 디스크 구동시간을 줄이기 위한 알고리즘은 다음과 같다.

알고리즘1

- i) 내부표준물질의 탐색
- ii) 정해진 시간범위내의 화일레코드지정

2-1-1 내부표준물질의 탐색

내부표준물질의 질량수와 머무름시간 범위가 주어지면 자료화일의 각 레코드별 머무름시간이 시간창 (Time Window)내에 있는가 여부를 검사한다. 이때 머무름시간이 시간창내에 존재하지 않으면 다음 레코드를 읽고, 존재하면 내부표준물질의 질량수를 레코드의 질량크기쌍에서 찾은 후 해당 크기를 기록하고 이 중 최대크기값을 찾아서 이때의 머무름시간을 기록한 후 이러한 과정을 순차적으로 수행한다.

다음은 내부표준물질탐색의 슈도코드(Pseude Code)이다.

```

RESET MASS - Data File
DO WHILE (Retention Time < Upper Boundary
           of Time)
    READ Record
    SET Retention Time of Record
    IF(Retenion Time is in the Time window)
        SET Number of Peak,
        SET Array Mass Abundance Pair
        PUSH Retention Time to Retention
        Time List
    
```

Linear Search for Maximum Abundance in Time Window

```

IF(Maximum Abundanc)
    SET Larrow = Maximum Abundance'
    s Array order
    
```

Larrow Points to the LSTD's Retention Time in Retention Time List

2-1-2 정해진 시간내의 레코드 지정

일반적으로 화학분석기기로 부터 획득한 자료는 조작자에 의해 다양한 형태로 출력된다. 특히 GS/

MS는 초기 설정상태에 따라서 목표물질이 검출되는 머무름시간이 일정하지 않다. 따라서 각 머무름시간 별로 질량크기를 살펴보려면 처음부터 선형탐색을 해야 하지만 화일전부를 탐색하는것은 처리시간문제에서 매우 비효율적이기 때문에 시간창을 결정하고, 레코드간의 시간간격을 인식하여 원하는 레코드를 직접 해독할 수 있는 방법이 확립되어야 한다.

다음은 시간창을 결정하는 식이다.

$$a - b < TW < a + b$$

여기서 TW는 시간창(Time Window)

a는 해당 물질의 상대머무름 시간 × 내부표준 물질의 머무름시간

b는 시간창의 폭

이다.

다음은 레코드탐색의 슈도코드이다.

```

RESET MASS Data File
READ 1st Record of Record type 4
READ 11th Record of Record type 4
Time Interval between Records =
    (11th Record's Retention Time - 1st
    Record's Retention Time) / 10
TARGET Record = (해당 Compound의 Reten-
    tion Time - 1st Record's
    Retention Time) / Time
    Interval
Starting Record = (Target Record 의 Retention
    Time - b) / Time Interval
    
```

2-2 처리과정

입력된 자료로 부터 내부표준물질의 머무름시간과 레코드간의 시간간격을 구하는 과정이 그림1에 나타나 있다.

3. 자료인식

이 과정의 주목적은 분석기자료를 읽어서 각 목표물질별로 양성/음성의 예비판정을 하는것 이므로

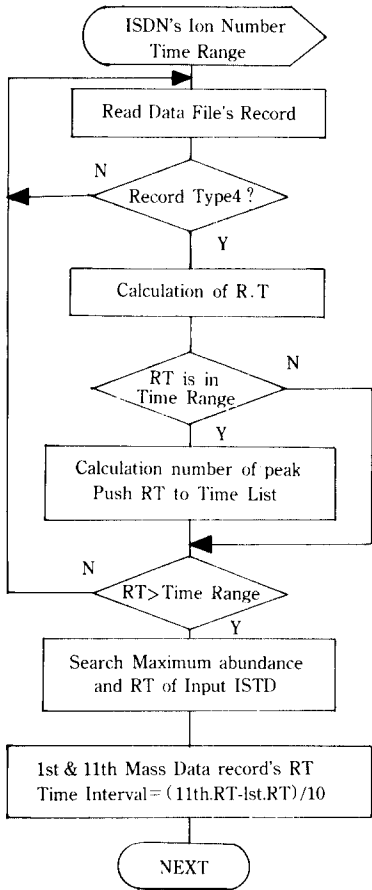


그림 1 머무름시간탐색과 시간간격계산 과정의 흐름도

우선 피이크의 유무를 검출하는 방법이 정립되어야 한다. 즉 내부표준물질을 검출하는 방법은 단순히 주어진 시간범위내에서의 최대크기값을 찾는 일로 가능하지만 목표물질의 특성이온별 피이크유무를 확인하는 것은 외부잡음, 넓게 퍼진 피이크의 판정문제, 상대적 크기값의 차이등 때문에 정량적인 산출에 의한 적정수준을 정하기가 매우 어렵다.

3-1 알고리즘

자료분석과정을 거쳐서 해독된 내용으로부터 목표물질의 인식을 위한 처리알고리즘은 다음과 같다.

알고리즘 2

i) 좌표계산

ii) 각 물질별 특성이온들의 최대피이크비율계산

좌표계산은 피이크를 출력할때 해당 시간창내의 최대피이크를 출력영역의 최대값으로 설정하고 크기값의 눈금을 최대피이크 크기에 맞추어 표시하는 과정이다. 또한 특성이온별 피이크비의 계산은 판정의 기준을 구하기 위해 해당 물질의 머무름시간에서 각 이온별 크기의 상대비를 구하는 과정이다. 이때 첫번째 특성이온의 크기값을 1로 정하고 다른 특성이온 크기값의 상대비를 구한다. 상대비를 구하는 식은 다음과 같다.

$$\text{상대비} = (a-b)/(c-d)$$

여기서

a는 해당 특성이온의 크기값

b는 해당 특성이온이 존재하는 시간창내의 최소크기값

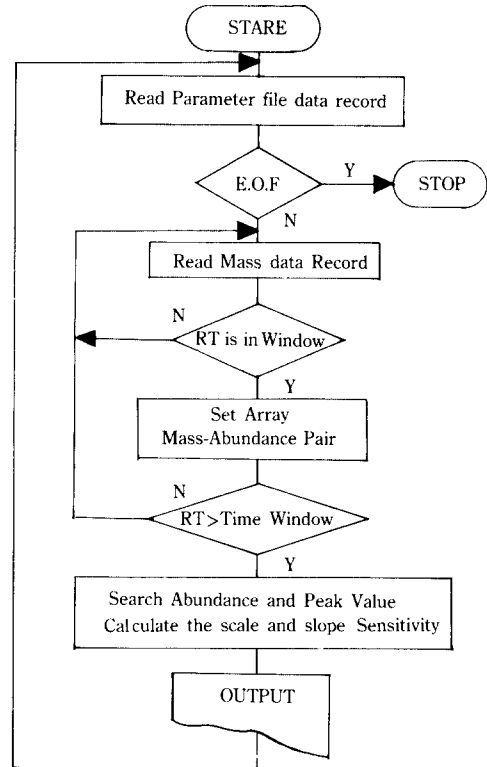


그림 2 인식과정의 흐름도

c는 첫번째 입력된 특성이온의 크기값
d는 첫번째 입력된 특성이온이 존재하는 시간창내의
최소크기값이다.

3-2 처리과정

자료분석과정에서 획득한 자료로 부터 목표물질인식
과 출력에 관한 과정이 그림2에 나타나 있다.

4. 처리 결과

4-1 자료화일의 구조

자료화일의 1개의 레코드당 첫번째워드부터 여
섯번째 워드 까지는 자료화일의 특성을 나타내며
그 이후의 워어드는 실제 자료이다.

즉 첫번째 워어드는 해당 레코드를 구성하는 총
워드수(N)과 레코드의 형태, 두번째 워어드는 분
석방식과 질량크기쌍의 갯수, 세번째와 네번째 워어
드는 해당 레코드의 머무름시간이 millisecond 단위
로 표시되고, 다섯번째 워어드는 최대피크값을 갖
는 질량수에 20을 곱한 값이고, 여섯번째 워어드는
최대피크의 값을 표시하며 그 값을 6×8^a 이다.
여기서 a는 지수로써 15, 14비트 값을, b는 상수로
써 13비트에서 0비트까지의 값을 각각 10진수로
환산한 값이다. 또한 일곱번째 워어드부터는 총(N-6)
/6개의 질량수와 질량크기쌍이 2 워어드로 구성되
어 있다.

4-2 피이크 결정

시간창내의 크기값의 변화를 추적하여 피이크와
잡음의 구별, 넓은 피이크의 인식등을 위하여 몇가지
방법을 시도하였다. 첫번째는 초기값을 설정하여 피
이크와 잡음의 구분을 시도하였으나 기준선의 변동
이 심하여 의미가 없었다. 두번째는 기울기를 구하여
피이크의 인식을 시도하였으나 크기값의 차이가 각
질량별로 크게 나타나기 때문에 절대기울기는 의미
가 없었다. 따라서 상대적인 기울기를 구하여 피이크
의 유무를 알아 본 후 상대기울기의 비를 근거로 양
성/음성의 판정을 행하였다.

수행예를 그림3에 나타내었다. 이 경우는 양성으로
판정되었고 준비된 파라미터화일로 부터 Methan-

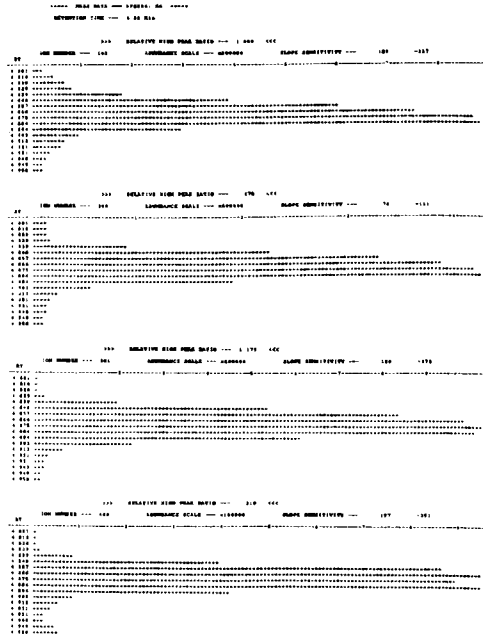


그림 3 수행예

dienone임을 알 수 있었다.

5. 결과 및 고찰

화학분석기기 특히 GC/MS로 부터 획득한 자료
는 미량에 대한 분석자료이기 때문에 전 단계의 화
학적처리, 기기의 상태, 조작자의 능력및 결과의 해
독능력에 따라 상이한 결과를 초래할 수 있다. 따라서
본 연구에서는 분석결과에 대한 양성/음성의 예비
판정에 적용가능한 예비스크리닝과정을 자동화하는
것에 그 목적을 두고 있다.

자료의 해석과정은 자료화일이 가변길이레코드로
구성되어 있기 때문에 화일전체를 읽기 위한 특수기
능의 함수를 필요로 하였으며 질량자료의 해석은 비
이트조작을 필요로 하였다. 또한 양성/음성의 예비
판정은 미리 설정한 질량수에 해당하는 질량크기값
을 조사하여 지정된 시간내에서 피이크의 판독과정
을 거친후 각 피이크의 기울기와 특성질량크기의 상
대비 비교과정을 수행하였다. 그러나 질량크기의 심
한 편차, 넓게 퍼진 피이크등의 문제로 인하여 판정의

신로도는 숙달된 전문가보다 높지 않았다.

현재 피이크의 최적판별법을 개발하여 신뢰도를 향상시키고, 처리시간의 단축, 기준자료베이스구축등에 관한 연구를 계속 진행중이다.

참 고 문 헌

- 1) D. B. BALER, "Computers in Analytical Chemistry", Pergamon Press, New York, 1981
- 2) D. L. MASSART, "Display Methods" in The Interpretation of Analytical Chemical Data, John Wiley & Sons, New York, 1983
- 3) B. R. KOWALSKI, "Chemometrics", Wiley-Interscience, New York, 1986
- 4) S. T. BROWN, etc., "Chemometrics", Anal. Chem., Vol. 60, pp 252-273, 1988
- 5) PETER C. JURIS, "Spectroscopic Library Searching and Structure Elucidation" in Computer Software Applications in Chemistry, John Wiley & Sons, New York, 1986
- 6) HEWLETT-PACKARD, "HP5988A GC/MS RTE-6/VM Data System Manual", part No. 59970-90004, Palo Alto, 1985
- 7) HEWLETT-PACKARD, "HD59970B Workstation Operator's Manual", Part No. 59970-90004, Palo Alto, 1985
- 8) L. KRYGER, "Interpretation of Analytical Chemistry Information by Pattern Recognition Methods", Talanta, Vol. 28, PP871-887, 1981