

# 化學文獻의 電子計算化에 있어서

## 化學構造 表示法の 役割

李 正

| <차 례>                               |                                      |
|-------------------------------------|--------------------------------------|
| 1. 머리말                              | 5.3 Chemical Abstracts Service (CAS) |
| 2. 化合物의 表示法                         | A. From attachment 리스트               |
| 3. 프래그멘테이션 코오드 (fragmentation code) | B. Ring closure 리스트                  |
| 4. 사이퍼 노오테이션(cipher notation)       | C. Node value                        |
| 5. 토폴로지칼 코오드(topological code)      | D. Line value                        |
| 5.1 레이法 (Ray method)                | E. Modification 리스트                  |
| 5.2 듀우퐁 시스템                         | 6. 맺는말                               |

### 1. 머리말

電子計算機에 의한 情報檢索의 보급과 함께, 化學構造 表示法(Chemical Notation) 注<sup>(1)</sup>에 대한 관심이 급증하고 있다.

情報檢索의 本質이 데이터의 축적과 탐색임을 고려할 때, 데이터를 어떻게 축적하고 탐색할 것인가 하는 것은 電子計算機의 기억장치의 發達과 함께, 情報檢索分野에 있어서 中心課題가 되어 왔다.

본고에서는 이러한 觀點에서, 化學分野의 케미칼노오테이션 (Chemical Notation)의 몇가지 手法<sup>(1)</sup>과 役割<sup>(2)</sup>에 대하여, 토폴로지칼 코오드를 中心으로 하여 說明코자 한다.

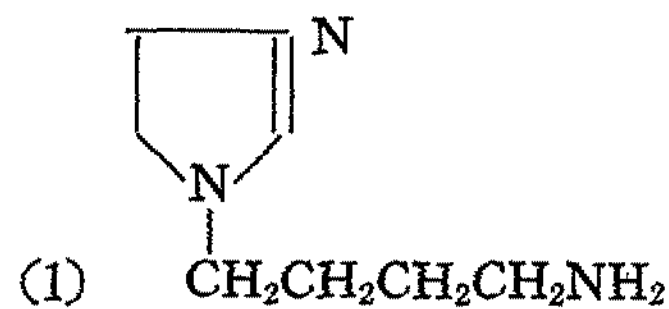
### 2. 化合物의 表示法<sup>(3)</sup>

歴史的으로 化合物의 表示法은 慣例的인 (Conventional)것과 非慣例的인 (non-conventional) 表示法の 두가지 手法으로 分類되고 있다. 慣例的인 表示法은 지금까지 使用되어 온 化合物의 名稱 즉 化合物命名法(nomenclature)에 의한 化學名이나 慣用名, 一般名, 商品名 또는 分子式, 化學構造式 등을 말하지만, 非慣例的인 表示法은 一定한 規則에 따라서, 化學構造 혹은 그의

構成要素를 記號化하는 手法이다. 그림 1은 慣例的인 것을 포함하여 케미칼노오테이션의 展開를 나타낸 것으로 左端에 있는 시스티머틱名(Systematic name)은 IUPAC (International Union of Pure and Applied Chemistry)이나 CA 등의 化合物命名法에 의한 化學名으로, Beilstein (Beilsteins Handbuch der Org. Chemie)이나 CA는 여기에 따라서 化合物을 表示하고 있다. 물론 이 化學名 以外에 一般名, 商品名 등도 補助的으로 使用하고 있지만, 어느 것도 다 語(word)로 表示하고 있다.

여기서 化學名으로부터 化合物을 檢索할 경우, 化合物 檢索의 2가지 形態에 관하여 간단히 說明코자 한다. 化合物의 檢索에는 스페시픽 서어치(specific search)와 제너릭 서어치(generic search)의 두 形態가 있다.

스페시픽 서어치는 全構造檢索으로서, 特定の 化合物 檢索이며, 제너릭 서어치는 部分構造 檢索으로서 化合物파일중에서 特定の 部分構造를 가진 모든 化合物의 檢索을 말한다.



(1)의 化合物을 例로 들어서 말하면, 인돌린(indoline) 骨核을 가진 化合物이든가, 아니면 어떠한 헤테로環이라도 좋기 때문에 헤테로環에 水酸基가 直結된 化合物을 알고 싶을 경우의 檢索을 말한다. 이러한 제너릭 서어치의 部分構造라는 點에 대하여도 化學構造와 生物活性

注 1. Chemical Notation 이란 말은 여기서는 廣義로 理解하여, 非慣例적인 化學構造의 表示法이란 意味로 使用한다. 따라서 노오테이션 (notation) 과 코오드(code)를 區別하지 않는다.

\*KORSTIC 調査檢索部

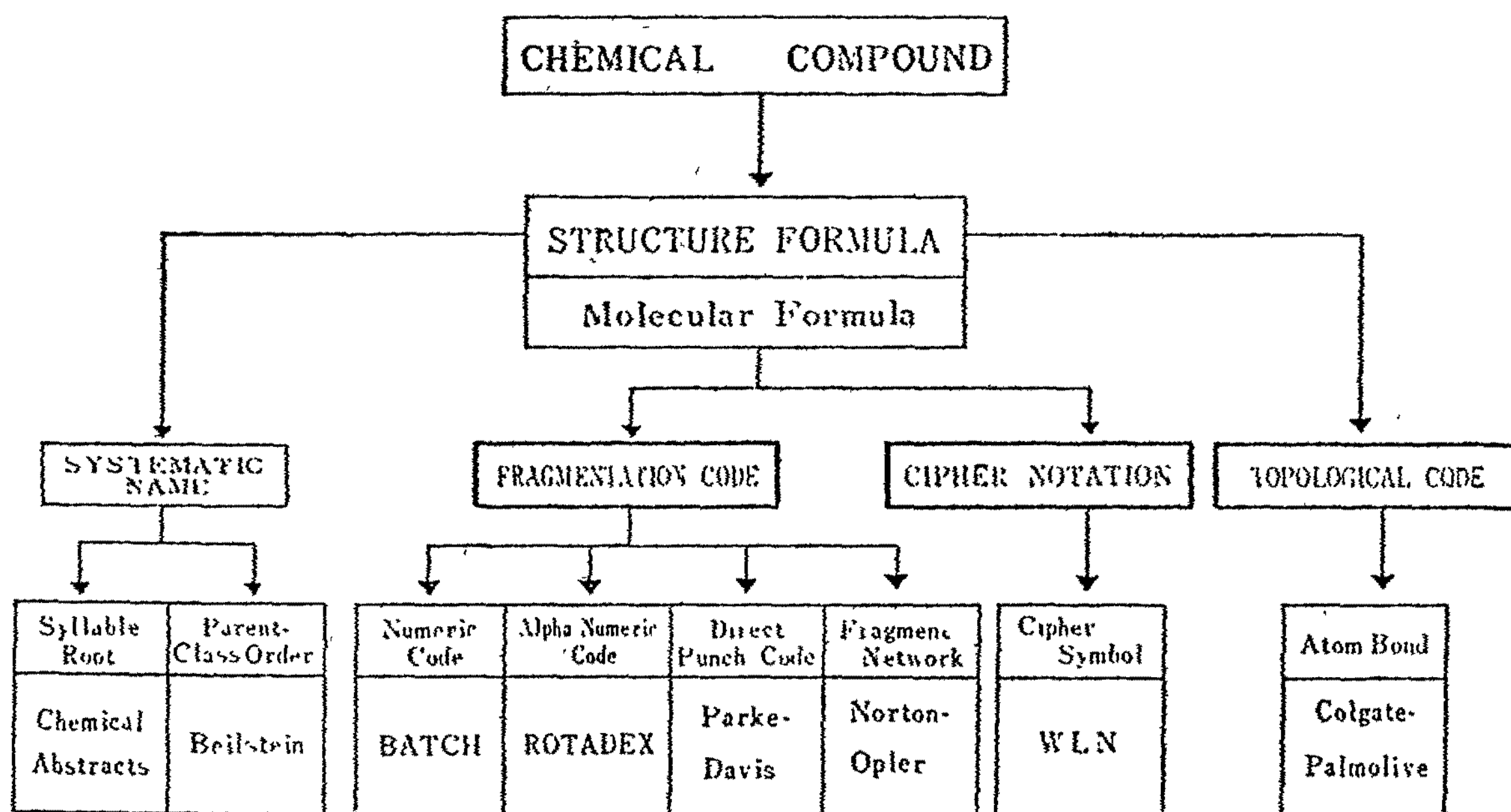


그림 1. \*M.L. Huber의 논문<sup>(4)</sup>中 그림 2를 약간 수정한 것임.

과의 關聯性 즉 構造活性關係라는 觀點에서는 보다 多面性이 있고, 緻密性이 있는 것이 要求되고 있다. 이 化合物에 대해서도 環內의 雜原子 N과 側鎖의 N 사이의 炭素數가 2 혹은 3인 化合物을 탐색코자 하는 質問에도 應해야할 경우가 있는 것이다.

化合物을 化學名으로 檢索하는데는 化學名의 간단한 ABC順 索引만으로는 만족할 수 없다. 化合物命名法의 복잡성, 非單一性을 고려하지 않고, 스페시픽 서어치라는 點에서 생각하여 보면, 化學名의 ABC順 索引만으로는 索引의 價値는 半減하게 된다. 化合物의 索引은 스페시픽 서어치와 함께 어느 정도의 제너릭 서어치에도 應답할 수 있어야 하는 것이다.

제너릭 서어치를 어느 정도의 多樣性(search variety)으로 하느냐에 따라서 索引의 構成이 變化하게 된다. 결국 化學名으로 어느 정도의 제너릭서어치에 適應할 수 있는 索引을 만드는데는 化學構造의 階層的 分類(structure hierarchy)가 必要하다.

CA의 Subject Index에서는 倒置法 [1-(4-amino-butyl)-2-imidazoline은 2-imidazoline의 主標目中의 aminobutyl이란 副標目으로 記載되어 있다]이란 原則外에, 官能基의 優先性이란 또 하나의 基本原則에 따라서, 化合物을 索引上의 어떤 主目標에 分配하고 있다. Beilstein도 마찬가지로, Beilstein으로 가능한 範圍內에서 제너릭 서어치를 하는데는 Beilstein體系로부터 應答해줘야 한다.

以上 說明한 바와 같이 化學名에 의한 化合物檢索의 특징은

(1) 스페시픽 서어치는 完全하다.

어떤 하나의 化學構造에 대하여 하나의 化學名이 주어지고, 또 그 化學名에서 元化學構造에 對應하여 1對1의 關係가 있으므로, 스페시픽 서어치는 完全하다. 그러나 化合物命名法이 單一하지 않거나, 分類規準에 支配되어 檢索은 그렇게 쉬운 것이 아니다.

(2) 제너릭 서어치에는 多樣性이 없고, 좁은 範圍로 限定된다.

이와같이 語(word)에 의한 表示에는 化學構造를 解體하고, 그 解體한 블록部分構造로 檢索을 하는데는 어느 정도의 限界가 있는 것이다. 이러한 限界의 認識과 多元的 檢索을 가능하게 하는 情報檢索機器의 發展(이를테면, Plain Card에서 Hole-sorted Card로, 그리고 IBM Card 즉 PCS에서 EDPS化 등)이 化學構造를 코오드화하는 케미칼노오테이션이란 手法에의 關心을 급속히 促進하고 있는 것이다.

非慣例的인 케미칼 노오테이션의 手法은 프래그멘테이션 코오드, 사이퍼 노오테이션 및 토폴로지칼 코오드로 大別할 수 있다.

### 3. 프래그멘테이션 코오드

케미스트(chemist)는 예를 들면, 나이트로基, 아미노基와 같은 官能基나 피리딘, 인돌과 같은 骨核(ring) 등을 化學構造의 單位로 생각하여, 化合物을 이러한 單位의 組合으로 보고 있다. 化學反應이나 分析化學에 있어서도 이러한 單位의 認識으로 體系化되어 있다. 프래그멘테이션 코오드는 가장 충실하게 有機化學上의 常識적인 생각을 베이스로 하고 있다.

Wiselogle 코오드, Frear 코오드, CBCC 코오드 등의

뒤를 이어, 처음에는 Hole-Sorted Card用 코오드로出發한 프래그멘테이션 코오드도 情報檢索機器의 發展에 따라 PCS, EDPS에도 그 原理가 應用되어, 20여년 동안에 30以上の 變法이 생기게 되었다.

이러한 많은 變法은 각각 化合物 파일의 量, 質 및 檢索에 使用하는 機械의 制約 등을 고려하면서, 어떠한 構造單位를 選定하는 것이 좋을 것인가 하는데 따라서 考案된 것이다.

이러한 프래그멘테이션 코오드를 몇가지 形態로 分類하는데는 상당히 어려운 점이 있지만, 편의상 그림 1과 같이,

- (1) 數字코오드로 나타나는 것<sup>(5~7)</sup>
- (2) 英文數字코오드로 나타나는 것<sup>(8~13)</sup>
- (3) direct punch code 方式<sup>(14~17)</sup>
- (4) fragment network<sup>(18)</sup>의 4가지로 大別할 수 있다.

수많은 프래그멘테이션 코오드를 단순히 특징지운다는 것은 다소의 위험이 있을지 모르나, 다음과 같이 말할 수 있다.

(1) 제너릭 서어치는 有效하다

官能基, 骨核이라는 構造單位에 입각하여, 多元的 檢索이 가능하므로, 제너릭 서어치라는 點에서는 實際로 그 有効性이 證明되어 있다. 파일의 크기나 그 內容에 따라서, 化合物을 檢索할 때의 質問式의 形態에 對應하는 設計를 할 수 있다. 그러나 사이퍼 노오테이션, 토폴로지칼 코오드에 비하면, 構造單位는 固定的으로 情報를 檢索할 때의 任意性이 없다. 바꾸어 말하면, 體系的 分類性格이 불식되어 있지 않다.

(2) 스페시픽 서어치는 不完全하다.

(3) 化學構造와 코오드사이에 化學名처럼 1對1의 關係가 없다.

#### 4. 사이퍼 노오테이션

Dyson, Wiswesser, Silk, Hayward와 같은 사람들이 開發한 사이퍼 노오테이션<sup>(19~27)</sup>은 원래 3次元的인 化學構造를 2次元으로 投影한 化學構造式을 英文字, 數字를 使用하여 線形(linear)으로 表示하는 것으로, 線形노오테이션(linear notation)이라고도 한다.

原子, 官能器, 環 등에 記號(cipher)를 부여하는 點에서는 프래그멘테이션 코오드나 마찬가지로 이지만 構造單位(fragment)의 끝부분 構造사이의 結合關係에도 注意

注 2. Unique란 어떤 하나의 化學構造에 대하여 하나의 노오테이션 밖에 주어지지 않을 경우를 말한다.

注 3. Unambiguous란 어떤 하나의 노오테이션으로부터 하나의 化學構造가 再現될 경우를 말한다.

하면서 코오드化하고 있기 때문에, Unique 注<sup>2)</sup>하면서도 unambiguous 注<sup>3)</sup>한 關係가 있다. 따라서 그 특징은 다음과 같다.

- (1) 스페시픽 서어치는 完全하다.
- (2) 제너릭 서어치도 물론 가능하지만, 토폴로지칼코오드에 비하면, 역시 記號의 定義에 支配되고 있다.
- (3) 리스트化 할 수 있다.

사이퍼 노오테이션의 큰 특징은 리스트化하는데 있다. 리스트 중의 노오테이션을 ABC 順으로 整理하면, 化學構造가 유사한 化合物이 서로 近接하게 되며 記號를 잘 정하면 電子計算機로 노오테이션으로부터 化學構造式을 再現시킬 수가 있다.

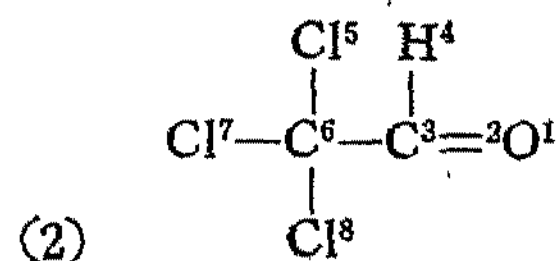
#### 5. 토폴로지칼 코오드

사이퍼 노오테이션은 立體的인 化學構造를 線形으로 表示하기 때문에 엄격한 規則이 있다. 또한 제너릭 서어치라 하더라도 檢索은 어떤 構造單位라도 좋다는 것은 아니다. 이러한 認識과 동시에, 化學構造의 最小 構造單位는 原子로, 그 結合도, 單結合, 二重結合, 三重結合밖에 없기 때문에 原子레벨의 코오드가 當然히 기대되었다. 따라서 1957년 L.C. Ray 등이 Science誌에 發表<sup>(28)</sup>한 이래, 주로 美國에서 많이 研究되어 온 토폴로지칼 코오드는 構造式을 位相的 圖形으로 생각하여, 構成原子와 그 結合關係를 數學的手法으로 表現하기 때문에 이른바 化學構造의 數學的 스냅사진이라고 말할 수 있다. CAS에서도 이 토폴로지칼 코오드法에 의한 檢索시스템을 化合物登錄시스템(CAS Registry system)과 結合시켜, 化學構造式 專用의 電動타이프라이터와 電子計算機를 作動시켜서, 構造式을 타이프라이터로 쳐, 自動적으로 인풋(input)하도록 하고 있다. 그러나 이처럼 入力이 쉬운 토폴로지칼 코오드法도 현재는 아직 一般的인 實用段階에는 이르지 못하고 있다. 그 原因은 한 化合物에 대한 情報量이 막대하여 檢索時間이 길어지기 때문이다.

##### 5.1 레 이 法<sup>(28)</sup>

위에서 잠깐 言及한 바와 같이 토폴로지칼 코오드法을 처음으로 開發한 사람은 L.C. Ray와 R.A. Kirsch이다.

NBS (National Bureau of Standards)의 Data Processing System 部에 所屬하여 있던 L.C. Ray와 R.A. Kirsch는 美國의 特許廳과 協力하여, 特許調査 특히 化合物의 蓄積과 檢索의 自動化에 대하여 研究하고, 1957년 그 成果를 Science誌에 發表하였다. 이 시스템은





| Component No. | 結合    | 記號 | Component No. | 結合      | 記號 |
|---------------|-------|----|---------------|---------|----|
| 1             | 2     | O  | 5             | 6       | Cl |
| 2             | 1,3   | =  | 6             | 3,7,8,5 | C  |
| 3             | 2,4,6 | C  | 7             | 6       | Cl |
| 4             | 3     | H  | 8             | 6       | Cl |

表 1.

클로랄(Chloral) (2)를 例로 들어, 構造原子 및 多重結合에 번호 (이것을 Component No. 라고 함)를 붙이고 (번호를 붙이는 방법은 任意임), 이 번호순에 따라서 표 1과 같이 컨넥션테이블(Connection table)을 만들고

各 原子에 대하여 이러한 情報를 인풋하였다.

檢索프로그램은 蓄積파일중의 化學構造와 質問式의 化學構造사이에, 各 原子間 照會를 할 수 있도록 設計하였다. 스크린 處理法에 의하여 質問式은 分子式 혹은 環數로 미리 選別하여 檢索時間을 단축하였다.

5.2 듀우퐁 시스템<sup>(29)</sup>

듀우퐁社는 1964년 D.J. Gluck 에 의하여 化學構造 檢索시스템을 開發한 이래, 이 시스템을 改良 補充하여, CS<sup>4</sup> (Chemical Structure Storage and Search System) 라는 시스템을 確立하였다. 이 시스템은 化合物의 表示

그림 2.

|   |          |             |
|---|----------|-------------|
| M.F.- C <sub>9</sub> H <sub>8</sub> F <sub>4</sub> O <sub>1</sub> | C- 41638 | TID- 123456 |
| STRUCTURE-  |          |             |
|   |          |             |
| NAME-   |          |             |
| PAGE ____ OF ____   |          |             |

| Atom No | Elem | Group | BOND |     | BOND |     | BOND |     | BOND |     | BOND |     | BOND |     | CHARGE | U-OXST | U-MASS | U-CONN | HYG | Atom No |    |
|---------|------|-------|------|-----|------|-----|------|-----|------|-----|------|-----|------|-----|--------|--------|--------|--------|-----|---------|----|
|         |      |       | D    | Att | D    | Att | D    | Att | D    | Att | D    | Att | D    | Att |        |        |        |        |     |         |    |
| 1       |      |       | 1    | 2   | 2    | 2   | 3    | 3   | 4    | 4   | 5    | 5   | 6    | 6   | 7      | 7      |        |        |     |         | 1  |
| 2       |      |       | 1    | 1   | 1    | 3   | 2    | 7   |      |     |      |     |      |     |        |        |        |        |     |         | 2  |
| 3       |      |       | 1    | 2   | 2    | 4   |      |     |      |     |      |     |      |     |        |        |        |        |     |         | 3  |
| 4       |      |       | 2    | 3   | 1    | 5   |      |     |      |     |      |     |      |     |        |        |        |        |     |         | 4  |
| 5       |      |       | 1    | 4   | 2    | 6   |      |     |      |     |      |     |      |     |        |        |        |        |     |         | 5  |
| 6       |      |       | 2    | 5   | 1    | 7   | 1    | 8   |      |     |      |     |      |     |        |        |        |        |     |         | 6  |
| 7       |      |       | 1    | 6   | 2    | 2   |      |     |      |     |      |     |      |     |        |        |        |        |     |         | 7  |
| 8       | O    |       | 1    | 6   | 1    | 9   |      |     |      |     |      |     |      |     |        |        |        |        |     |         | 8  |
| 9       |      |       | 1    | 8   | 1    | 10  | 1    | 11  | 1    | 12  |      |     |      |     |        |        |        |        |     |         | 9  |
| 10      |      |       | 1    | 9   | 1    | 13  | 1    | 14  |      |     |      |     |      |     |        |        |        |        |     |         | 10 |
| 11      | F    |       | 1    | 9   |      |     |      |     |      |     |      |     |      |     |        |        |        |        |     |         | 11 |
| 12      | F    |       | 1    | 9   |      |     |      |     |      |     |      |     |      |     |        |        |        |        |     |         | 12 |
| 13      | F    |       | 1    | 10  |      |     |      |     |      |     |      |     |      |     |        |        |        |        |     |         | 13 |
| 14      | F    |       | 1    | 10  |      |     |      |     |      |     |      |     |      |     |        |        |        |        |     |         | 14 |
| 15      |      |       |      |     |      |     |      |     |      |     |      |     |      |     |        |        |        |        |     |         | 15 |
| 16      |      |       |      |     |      |     |      |     |      |     |      |     |      |     |        |        |        |        |     |         | 16 |
| 17      |      |       |      |     |      |     |      |     |      |     |      |     |      |     |        |        |        |        |     |         | 17 |
| 18      |      |       |      |     |      |     |      |     |      |     |      |     |      |     |        |        |        |        |     |         | 18 |
| 19      |      |       |      |     |      |     |      |     |      |     |      |     |      |     |        |        |        |        |     |         | 19 |
| 20      |      |       |      |     |      |     |      |     |      |     |      |     |      |     |        |        |        |        |     |         | 20 |
| 21      |      |       |      |     |      |     |      |     |      |     |      |     |      |     |        |        |        |        |     |         | 21 |
| 22      |      |       |      |     |      |     |      |     |      |     |      |     |      |     |        |        |        |        |     |         | 22 |
| 23      |      |       |      |     |      |     |      |     |      |     |      |     |      |     |        |        |        |        |     |         | 23 |
| 24      |      |       |      |     |      |     |      |     |      |     |      |     |      |     |        |        |        |        |     |         | 24 |
| 25      |      |       |      |     |      |     |      |     |      |     |      |     |      |     |        |        |        |        |     |         | 25 |

方法으로 토폴로지컬 코오드를採用하고, 그 方式 및 檢索프로그램은 CAS 시스템을 變形, 擴張한 것이다. 原子對 原子(atom by atom) 方式인 이 CS<sup>4</sup>法은 化合物에 대하여 原子의 종류 및 結合樣式을 컨넥션테이블의 形態로 인뜻하고 있다. 그림 2는 그의 코오드 사이트로서 上欄의 構造式에 대하여 水素를 除外한 原子에 任意的 1번번호를 붙이고 있다. 이것이 Atom No. 이다. Elem. 의 項은 炭素 및 水素以外의 原子를 表示한다. 그림 2에서는 酸素와 弗素만이 記入되어 있다. Bond 는 Att.에 쓰여진 Atom No. 의 原子와의 結合樣式으로서, 單結合이면 1, 二重結合이면 2, 三重結合이면 3으로 表示한다. 예를 들어 Atom No. 2의 原子를 보면, 1 및 3의 原子는 單結合이고, 7의 原子와는 二重結合으로 結合되어 있음을 알 수 있다. 原子의 結合關係와 함께, 다음 5가지의 데이터가 인뜻된다.

- (1) CHARGE : 電荷
- (2) UMASS : 同位元素
- (3) HYG : 炭素以外의 原子에 結合하는 水素의 數
- (4) UCONN (Connection number) : 異常原子價數의 表示
- (5) UOXST (Oxidation state) : 配位價電子數

(4), (5)는 錯化合物, 配位化合物 등으로 보통의 原子價가 아닌 경우의 原子價數와 그 價電子數를 表示한다. (표 2.)

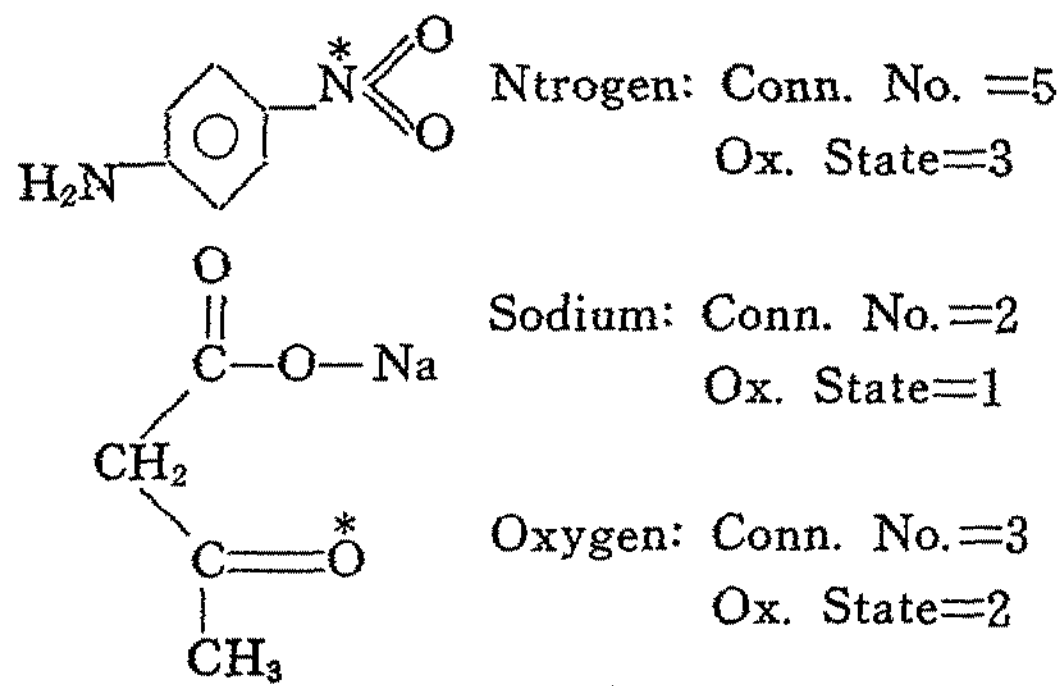


表 2.

CS<sup>4</sup>에서는 벤젠環이나 메틸렌鎖 등의 原子團에는 특수한 記號를 使用하고 있다.

### 5.3 Chemical Abstracts Service(CAS) (30)

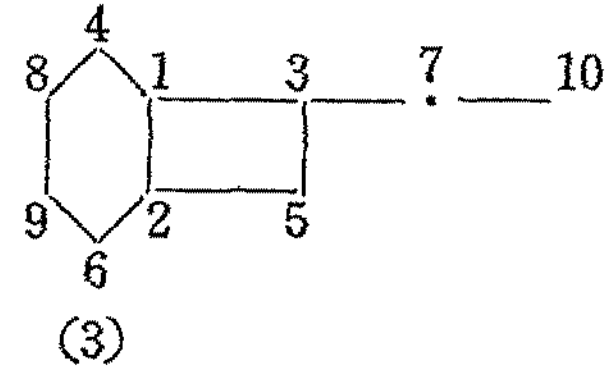
CAS에서는 電子計算機를 베이스로 한 化學情報시스템의 一環으로서 化合物登錄시스템을 開發하고 있지만, 이 시스템을 運用하기 위해서는 어떤 化合物이 이미 登錄되어 있는지 없는지, 蓄積파일중에 있는지 없는지를 먼저 알 必要가 있다.

이를 위하여 化合物의 照會手段으로서 原子對 原子 컨넥션시스템을 採用하고 있다. 化學構造의 表示에는 原子 및 原子사이의 結合關係를 나타내는 컨넥션테이

블을 만들지만, 이에 앞서, 水素를 除外한 原子에 다음과 같은 規則에 따라서 번호(Rank No.)를 붙인다.

(1) 始作點 (Rank No. 1)은 任意로 부여한다.

(2) Rank No.에 直接되어 있는 原子에 2,3.....로 번호를 붙이고, 다음 Rank No. 2의 原子에 直接하는 原子에, 그 다음은 Rank No.3에 直接하는 原子에 順次的으로 번호를 붙여 나간다. 다음의 化合物(3)에 붙인 번호를 보면 알 수 있다.

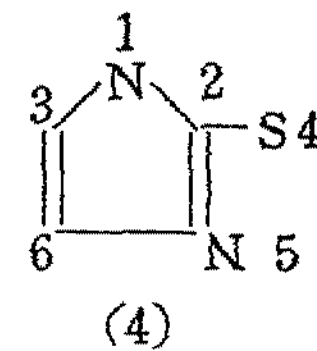


電子計算機에는 컨넥션테이블 製作의 基礎가 되는 原子의 Rank No., 結合關係를 나타내는 From attachment 및 Ring closure, 原子의 種類 (Node value), 結合型 (Line value) 이 인뜻된다.

以下 이러한 리스트에 대하여 간단히 說明코자 한다.

#### A. From attachment 리스트

Rank No.가 주어진 各 原子에 대하여 그 原子와 結合되어 있는 原子中, Rank No.의 最小의 것만을 Rank No.로 表示한다. 化合物(4)에 대하여 나타내면, 표 3라 같이 된다.



| Rank No | From attachment | Ring closure | Node value | Line value |
|---------|-----------------|--------------|------------|------------|
| 1       | Blank           |              | N          | Blank      |
| 2       | 001             |              | C          | 1          |
| 3       | 001             |              | C          | 1          |
| 4       | 002             |              | S          | 1          |
| 5       | 002             |              | N          | 2          |
| 6       | 003             |              | C          | 2          |
|         |                 | 005 006      |            | 1          |

表 3.

Rank No. 1은 언제나 블랭크(blank)로 되며 Rank No 2는 1,4,5의 原子와 結合하여 있지만 1만이 表示된다.

#### B. Ring closure 리스트

環狀化合物의 경우에 만들어지는 리스트로서, Rank No.의 넘버링(numbering) 規則으로서는 結合하여 있을까, 開環되어 있을까를 判斷할 수 없는 原子사이의 結合을 表示한다. (4)에서는 Rank No. 5와 6이 Ring

closure이다.

C. Node value

原子的 種類를 나타낸다. 水素以外の 原子 (이것을 node라고 함) 즉 Rank No.를 붙인 原子를 原子記號로 나타낸다. (표 3)

D. Line value.

Node 相互間의 結合 즉 Rank No.의 原子와 From attachment의 리스트에 나타낸 原子와의 結合 및 Ring closure 리스트의 原子사이의 結合型을 意味한다.

單結合, 二重結合, 三重結合에는 各各 1, 2, 3의 數字를 使用한다. 표 3에 Rank No. 2, 3, 4의 原子는 各各 Rank No. 1, 1, 2의 原子와 單結合으로 結合되어 있으므로 Line value 는 1이 된다.

Ring closure 리스트의 Rank No. 5와 6의 結合도 單結合이므로 1이 된다.

E. Modification 리스트

이온價, 同位元素, 異常原子價인 경우의 node 나 line 을 表示하는 리스트로서, 듀우퐁시스템의 UMASS 나 UOXST 에 해당되는 것이다.

一般的으로 토폴로지칼 코오드는 Unambiguous 하지만, Non-unique 注<sup>4)</sup>한 것이다.

이것은 인풋데이터를 넣는 方法. 化合物의 번호를 붙이는 方法에 따라서 코오드를 고려하기 때문이다.

CAS 에서는 이 點을 改良하여 原子周圍의 結合數 등 을 고려한 一定한 規則을 採用하고 있어서, 어떠한 넘버링으로 인풋하여도 하나의 化合物에 대하여 단 하나의 코오드가 부여되도록 하고 있다.

立體表示에는 아직 問題가 있지만, 平面構造의 노오테이션에는 Unique 하면서도 Unambiguous 한 토폴로지칼 코오드法을 確立하고 있다.

6. 맺 는 말

限定된 範圍內에서 케미칼 노오테이션을 프래그멘테이션 코오드, 사이퍼 노오테이션, 토폴로지칼 코오드로 나누어서, 그 대략을 說明하였으나 歐美에서는 이미 2, 30년 동안 研究, 發展되어 온 이 領域의 紹介는 그렇게 간단한 것이 아님을 통감하게 되었다.

電子計算化에 있어서 化學構造 表示法이 얼마나 중요한 役割을 하고 있는가에 대하여 조금이라도 理解가 될 수 있다면 다행한 일일 것이다.

차후 기회가 주어지면, 프래그멘테이션 코오드, 사이퍼 노오테이션, 토폴로지칼 코오드에 관하여 좀더 詳細코자 한다. 토폴로지칼 코오드 중에서도 듀우퐁시

注 4. Non-unique란 어떤 하나의 化學構造에 대하여 2개 이상의 노오테이션이 가능한 경우를 말한다.

스템과 CAS 시스템外에도 Chem SEARCH(Colgate-Palmolive) 시스템<sup>(31)</sup>, Penny Connectivity 코오드<sup>(32)</sup>, Tree Structure (Polish Notation)<sup>(33~34)</sup>, MCC (Mechanical Chemical 코오드) 시스템<sup>(35~38)</sup>, Atom Connectivity Matrix (ACM)<sup>(39~40)</sup> 등이 있으나 本稿에서는 생략하였다.

以上과 같이 케미칼 노오테이션의 수 많은 手法中에서 어느 것을 採用할 것인가 또는 어떤 手法를 새로이 開發할 것인가 하는 것은 化合物파일의 量 및 質과 어느 정도의 部分構造檢索(제너릭 서어치)을 求하여야 할 것인가에 따라서 決定하여야 할 것이다.

<참 고 문 헌>

- 1) "Survey of chemical systems," Publication No. 1150. National Academy of Sciences-National Research Council, 1964.
- 2) "Chemical structure information handling—A review of the literature 1962—1968," National Academy of Sciences, 1969.
- 3) "Survey of european non-conventional chemical notation systems," Publication No. 1278, National Academy of Sciences-National Research Council, 1965.
- 4) Huber, M L., "Chemical structure codes in perspective" J. Chem. Doc. 5 1 pp.4~8 1965.
- 5) Indris, R.W., "Chemical structure fragmentation for use in a coordinate index retrieval system," J. Chem. Doc. 4 4 pp. 274~278 1964.
- 6) Barnard, A.J.Jr. et al., "Retrieval of organic structures from small-to-medium sized collections," J Chem. Doc. 6 1 pp 41~48 1966.
- 7) Feeman, J.F., "A novel organizational code for organic structures based on functional groups," J. Chem. Doc. 6 3 pp. 184~187 1966.
- 8) Arendell, F.H., "A three-symbol code for searching chemical structures," J. Chem. Doc. 1 3 pp. 47~57 1961.
- 9) Starker, L.N. et al., "A multi-level retrieval system. I. A simple optical coincidence card system," J. Chem. Doc. 8 2 pp. 81~85 1968.
- 10) Starker, L.N. et al., "A multi-level retrieval system. II. Medium-sized collections," J. Chem. Doc. 9 3 pp. 161~167 1969.
- 11) Starker, L.N. et al., "A multi-level retrieval system. III. A generic chemical search system using optical coincidence cards," J. Chem. Doc. 10 3 pp. 206~211 1970.
- 12) Frome, J. et al., "PACIR: practical approach to chemical information retrieval," J. Chem. Doc. 2 4 pp. 248~255 1962.
- 13) Sher, I. H. et al, "Rotadex-a new index for generic searching of chemical compounds," J. chem. Doc. 4 1 pp. 49~53 1964.
- 14) Geer, H.A. et al., "The Parke-Davis code for chemical structures," J. Chem. Doc. 2 2 pp. 110~113 1962.
- 15) Starker, L.N. et al., "The cyanamid organic structure code and search system," J. Chemi Doc. 2 1 pp. 12~15 1962.
- 16) Haefele, C.R. et al., "Some unusual features of a che



- mical retrieval system used in the Eastman Kodak Company," J. Chem. Doc. 4 2 pp. 112~115 1964.
- 17) Willard, J.R. et al., "A working system for retrieval of chemical structure, adaptable to pesticidal screening data," J. Chem. Doc. 4 4 pp. 211~217 1964.
  - 18) Opler, A. et al., "New speed to structural searches," Chem. Eng. News 34, 2812~2816, 1956.
  - 19) Hoffmann, E., "Use of a modified wiswesser notation for the encoding of proteins," J. Chem. Doc. 2 3 pp. 137~140 1969.
  - 20) Sorter, P.F. et al., "Rapid structure searches via permuted chemical line-notations," J. Chem. Doc. 4 1 pp. 56~60 1964.
  - 21) Granito, C.E. et al., "Rapid structure searches via permuted chemical line notations. III. A computer produced index," J. Chem. Doc. 5 4 pp. 229~233 1965.
  - 22) Garfield, E. et al., "Index chemicus registry system: pragmatic approach to substructure chemical retrieval," J. Chem. Doc. 10 1 pp. 54~58 1970.
  - 23) Bonnett, H.T., "Chemical notations—a brief review," J. Chem. Doc. 3 4 pp. 235~242 1963.
  - 24) Sneed, H.M.S. et al., "A line-formula notation system for Markush structures," J. Chem. Doc. 8 3 pp. 173~178 1968.
  - 25) Silk, J.A., "A linear notation for organic compounds," J. Chem. Doc. 3 4 pp. 189~195 1963.
  - 26) Skolnik, H. et al., "A notation system for indexing pesticides," J. Chem. Doc. 4 4 pp. 221~227 1964.
  - 27) Skolnik, H., "A new linear notation system based on combinations of carbon and hydrogen," J. Hetr. Chem. 6 6 pp. 689~695 1969.
  - 28) Ray, L.C. et al., "Finding chemical records by digital computers," Science 126 (3278), 814~819, 1957.
  - 29) Hoffman, W.S., "An integrated chemical structure storage and search system operating at Du Pont," J. Chem. Doc. 8 1 pp. 3~13 1968.
  - 30) Morgan, H.L., "The generation of a unique machine description for chemical structures—a technique developed at Chemical Abstracts Service," J. Chem. Doc. 5 2 pp. 107~113 1965.
  - 31) Gould, D. et al., "Chem SEARCH—an operating computer system for retrieving chemicals selected for equal, analogous, or related character," J. Chem. Doc. 5 1 pp. 24~32 1965.
  - 32) Penny, R.H., "A connectivity code for use in describing chemical structures," J. Chem. Doc. 5 2 pp. 113~117 1965.
  - 33) Eisman, S.H., "A polish-type notation for chemical structures," J. Chem. Doc. 4 3 pp. 186~190 1964.
  - 34) Hiz, H., "A linearization of chemical graphs," J. Chem. Doc. 4 3 pp. 173~180 1964.
  - 35) Lefkovitz, D., "A chemical notation and code for computer manipulation," J. Chem. Doc. 7 4 pp. 186~192 1967.
  - 36) Lefkovitz, D., "Use of a non-unique notation in a large scale chemical information system," J. Chem. Doc. 7 4 pp. 192~200 1967.
  - 37) Lefkovitz, D., "Substructure search in the MCC system," J. Chem. Doc. 8 3 pp. 166~173 1968.
  - 38) Lefkovitz, D., "A utility analysis for the MCC topological screen system," J. Chem. Doc. 10 2 pp. 86~94 1970.
  - 39) Spialter, L., "The atom connectivity matrix (ACM) and its characteristic polynomial (ACMCP)," J. Chem. Doc. 4 4 pp. 261~269 1964.
  - 40) Spialter, L., "The atom connectivity matrix characteristic polynomial (ACMCP) and its physico-geometric (topological) significance," J. Chem. Doc. 4 4 pp. 269~274 1964.