

# 化學文獻의 部分構造表示法 (1)

李 正

## 머 리 말

어떤 化合物이 文獻에 이미 알려진 物質일까? 目的하는 性質을 가진 化合物에 대하여 報告된 文獻이 있을까? 등을 調査, 判定한다는 것은 쉬운 일이 아니다. 대부분의 사람들은 지금까지 使用하여 온 一般的인 調査方法으로 充分하고 完全하다고 判定할 수 있을 때까지는 많은 時間과 努力이 檢索하는데 필요함을 알고 있다.

Schwartz의 報告<sup>1)</sup>에 의하면, 環式脂肪族 아민誘導體를 構造式 索引(Formular Index)을 使用하여, 檢索한 바, C.A.의 索引에는 發見되지 않은 것이 平均 約 30% 가 C.Z.(Chemisches Zentralblatte)의 索引에 收錄되어 있으며, 특히 原報가 特許인 경우에는 出發原料 7種中 2種, 中間體 5種中 2種類의 化合物이 各各 索引에 收錄되어 있지 않음을 實例를 들어 引用하고 索引의 不完全性을 지적한 바 있다.

컴퓨터가 情報處理 檢索道具로 利用되고 있다는 것은 衆知의 事實이지만 이것을 能率的으로 活用하기 위하여는 情報를 어떠한 利用方法으로 運營하여야 할 것인가에 關하여 유의하여야 할 것이다.

化合物을 機械檢索에 適用하는데는 前報<sup>2)</sup>에서 그 大략을 說明한 바와 같이 크게 나누어 프래그멘테이션 코오드, 사이퍼노오테이션, 토폴로지칼 코오드의 3가지 方法이 있다. 이중 프래그멘테이션 코오드에 대하여 以下 說明코자 한다.

### 1. GREMAS 法<sup>3)</sup> (Genealogisches Recherchieren durch Magnetband Speicherung)

1957年, 獨逸의 Hoechst社에서 처음으로 그리고 1962年부터 Bayer社와 Badische Anilin社가 參加하여 發展시킨 프래그멘테이션 코오드의 一種이다. 1967年에, IDC(Internationale Dokumentations-gesellschaft für Chemie m.b.h)가 設立되어 C.Z.의 SRD(Schnellreferatendienst)와 前記 3社의 特許關係 文獻의 抄錄도 同時에 實施하게 되었다.

論文, 特許 등의 GREMAS 法에 의한 處理件數는 年

間 約 35,000정도이다.

IDC는 토폴로지칼 코오드와 병행하여 이 方法을 採用하고 있으며, 優先的으로 利用하고 있다. 그 理由는 部分構造나 反應의 檢索이 보다 쉽고, 經濟的이기 때문이다.

GREMAS 法은 結合을 含有한 構造單位를 알파벳과 對應시킨 3文字로 表現하고, 그 構造單位의 分子內에서 의 配列을 다시 Y 또는 Z로 시작하는 表現法으로 나타내고 있다.

表 1에 表示한 A→R까지는 炭素原子에 關한 것, S→V의 記號는 環을, W는 置換基의 位置 등을 나타내고 있다. 3文字 記號의 最初의 文字는 表 1에 分類된 것, 두번째 文字는 結合하는 상대의 原子의 性格을 分類한 것, 세번째는 炭素原子 自身の 性格을 區別 分類한 것이다. 예를 들면, H의 記號로 나타낸 경우는 할로젠과 結合하는 炭素를 나타내며, 이것이 HA—, HB—, 로 되면 A는 弗素, B는 鹽素와 결합하는 것을 나타낸다.

表 2는 一級 아민 構造에 대하여 약간의 例를 表示한 것으로서, 最初의 B는 아민構造와 結合하는 炭素, BA—는 아민이 一級 아민임을 나타낸다.

表 1.

A C- $\alpha$  (N,O,S, 할로젠, H 이외의 原子)

B C-N-  $\left\{ \begin{array}{l} \text{BA} \text{ 1級아민} \\ \text{BB} \text{ 2級아민} \\ \text{BC} \text{ 3級아민} \end{array} \right.$

C  $\begin{array}{c} \text{O} \\ \parallel \\ \text{C-N-} \end{array}$

D C-N-N

E C-O-

F C-S-

G C-SO-

H C-할로젠 (HA : F, HB : Cl, HC : Br, HD : I)

I >C=O

K >C=S

L -CH=O

M -CH=S

- N —COO—
- O —CSO—
- P —O—CO—O—
- Q 티오키온산(thiocarbonic acid)
- R 알킬기(RA : C—C, RB : C=C, RC : C≡C)
- S 單環
- T 縮合環
- U 보통의 環
- V 縮合點
- W 置換基의 位置, 鎖의 길이
- X 헤테로原子—헤테로原子

表 2.

$\text{—CH}_2\text{—CH}_2\text{—NH}_2$	BAA		BAD
$\text{—CH=CH—NH}_2$	BAF		BAQ
$\text{—C}\equiv\text{C—NH}_2$	BAM		BAS
	BAR		

3번째의 文字는 다시 炭素原子로 돌아가, 炭素原子가 單結合, 二重結合으로 結合하는 炭素인가 아닌가를 나타낸다.

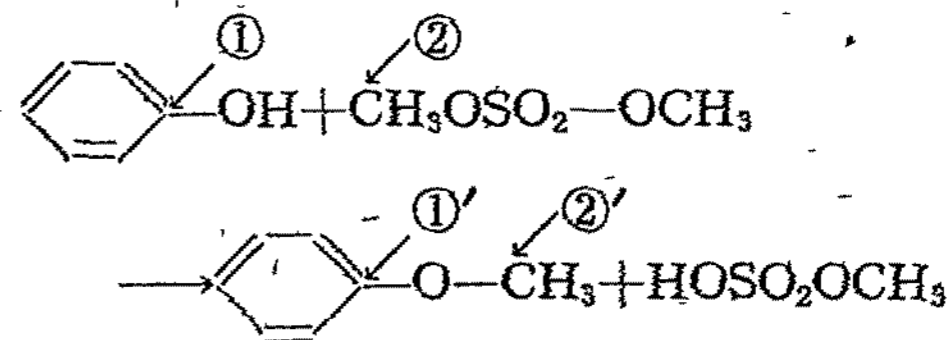
環狀化合物은 S 또는 U로 시작하는 記號로 주어지지만, 이 경우도 一般적으로 頻度가 많은 代表的인 環에 各各 미리 두번째, 세번째의 記號를 정하여 부여한다. 表 3은 이것의 一例이다.

表 3.

	SKC		UDA
	TCC+TKC		UBI

表 4에 나타낸 構造는 2價 페놀의 構造單位를 SAF, EAR로 表示하고 있다. SAF는 벤젠環을 나타내며, EAR은 芳香族環에 붙어 있는 水酸基의 構造單位를 나타낸다. 水酸基가 2個 오소(ortho-) 位置에 結合하고 있는 것은 YTEE 2로 나타낸다.

表 4.



Y로 시작하는 코오드는 YR...이 鑷狀構造, YS...脂環狀, YT...芳香環을 意味한다.

YTEE 2의 EE部分은 E로 表示된 性格의 炭素(表 1)가 2個 存在하고, EE 2의 2는 처음의 E를 1位置로 생각하여, 두번째의 E를 2位置의 炭素로 본 것이다. —OH基가 메타(meta-) 位置의 異性體에서는 SAF, EAR, YTEE 3이 된다.

官能基나 部分構造의 檢索은 이러한 記號를 逆으로 組合하여 質問하면 좋다.

E...는 —C—O 構造의 有無만 이지만, EA...로 質問하면, 水酸基의 有無를 찾는 形이 된다. 마찬가지로 BAA, BAQ는 脂肪族 脂環狀1級아민을 檢索하는데 利用된다.

反應을 코오딩하는 경우도, 目的하는 構造單位만을 고려한다. 上記의 反應에서 反應에 關與하는 ①, ②의 炭素原子에 주목하여 보면 다음 式과 같은 變化를 생각할 수 있다.

- ① EAR → ①' EBR
- ② EIA → ②' EBA

따라서, 反應의 코오딩은 2BE가 되며 이 코오드를 質問으로 利用한다. 이처럼 反應의 檢索, 部分構造의 檢索 등을 위하여, 同時에 125個의 質問이 可能한 프로그램이 이미 開發, 利用되고 있다.

平均 20~25個의 質問을 IBM/360에 인풋(input)하여, 磁氣테이프의 全資料를 處理하여 프린트 아웃(print out)하는 데는 約 60분이 걸린다고 한다.

### 1. Ringdoc

Ringdoc<sup>4), 5)</sup>은 英國의 Derwent社에 의하여 1964년부터 發足한 藥學領域의 情報서어비스 시스템이지만, 그 前身은 歐州의 製藥, 化學系 會社 8個社가 가지고 있는 Documentation Ring에 Hoffman La Roche社와 Sandoz社의 Codeless Scanning方式을 더하여, 1964년에 發足한 組織이다.

Dr. W. Steidle에 의하여 化合物의 構造, 生物學데이터를 表示하는 링코오드(Ring Code)가 開發되었다.

IBM카드를 使用하여 最初의 27칼럼까지는 化合物을 表示하고 나머지 칼럼에 酵素, 糖 및 生物學데이터(解剖學, 代謝, 藥理, 疾病 등)와 抄錄番號를 表示한다. 酵素 및 生物學데이터의 表示에는 英文 數字코오드, 數

字코오드(WHO의 疾病, 副作用코오드를 採用)도 使用하고 있다.

化合物 表示用의 코오드에는 化學一般 코오드, 스테로이드 코오드(Steroid Code), 펩타이드 코오드(Peptide Code)의 3種類 以外에 構造不明物에 대하여는 Clear-text라고 불리는 自然語에 의한 表示法이 있다. 이러한 4種의 코오드는 카아드의 칼럼 1의 편치로 識別하고, 칼럼 2에서 27까지는 共通필드로 使用하고 있다.

原則적으로 한 化合物에 한개의 카아드가 만들어 지도록 하고 있다.

1.1 化學一般코오드

表 5에 化學一般코오드의 IBM카아드에의 配分을 나타내었다.

칼럼別의 詳細한 內容은 다음과 같다.

칼럼

1: 識別코오드(예를 들면, 스테로이드 코오드에는 1/11이 편치된다.)

2: 化合物 全體 및 環시스템別로 環의 數를 각각 指定位置에 편치한다.

3: 芳香族環의 數, 狀態

4-5: 炭素環의 크기, 數, 狀態, 架橋狀態, 스페로化合物的 區別

6: 헤테로環의 크기, 狀態

7: 헤테로原子的 種類, 數, 狀態

8: 헤테로原자의 相對的 位置, 狀態

9-10: 置換基의 位置

11-16: 炭素鎖에 관하여 炭素連鎖, 炭素構成單位의 相對的 位置, 鎖狀 C-C 結合, C=C 結合의 數.

17: 할로젠, NO<sub>2</sub>, NO 置換基, 나이트릴, 이소나이트릴

18: 알콜, 티오알콜, 페놀, 티오페놀, 에텔, 티오에텔

19: 아민類

20-21: 燐酸, 黃酸 및 그의 同族體의 誘導體

22: 아조化合物

23: 알데하이드, 케톤, 炭酸 및 그의 誘導體

表 5.

칼럼 2	10	16	27
識別 코오드	骨 格		置 換 基
	環	鎖	各各의 置換基 個別的으로 表示
	芳香族 脂環式化合物	飽和鎖 不飽和鎖	
	헤테로化合物		

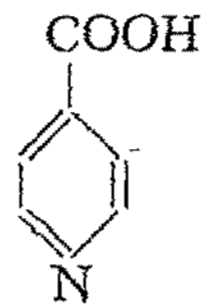
24: 炭素 誘導體

25: 環에 直結한 炭보닐 및 그의 誘導體

26: 칼럼 18~25의 補助코오드

27: 元素, 金屬, 標識 化合物

다음 이소니코틴酸(isonicotinic acid)의 코오드 및 그의 意味를 나타내면 아래와 같다.



穿孔位置 (칼럼/열)      意味

2/12: 環1個

6/12: 單離의 헤테로環

6/3: 의 헤테로環

7/1: 헤테로環內의 N 1個

10/5: 피리딘의 4位置換

23/2, 23/7: 炭酸

1.2 스테로이드 코오드

칼럼 2에서 16까지 스테로이드骨核에 대하여 環의 型, 位置와 配位(α, β), 置換原자의 種類(-OH, =O 등) 및 不飽和結合이 表示되며, 특히 置換頻도가 높은 3位, 5位, 14位, 17位에 대하여는 特別히 코오드가 指定되어 있다.

아자스테로이드(azasteroid), 옥사스테로이드(oxasteroid) 등 헤테로環이나 비타민 D와 같은 세코(seco)型의 것은 스테로이드 코오드에는 없고, 化學一般코오드로 表示한다. 또한 置換基를 나타내는 코오드(칼럼 17~27)는 化學一般코오드와 同一하다.

1.3 펩타이드 코오드

構成아미노酸의 數, 펩타이드의 狀態(環狀, 分枝狀, 架橋狀), 光學活性(d體) 外에 펩타이드構造의 表示에는 22칼럼을 使用하여 個個의 아미노酸의 結合順序를 코오딩한다.

링코오드에 의한 檢索은 基本的으로는 제너릭서어치가 되지만, 프래그먼트(fragment)의 單位가 비교적 적기 때문에 檢索時 코오드化를 잘하면 스페시픽서어치도 可能하다.

2. Wislogle 코오드

이 코오드는 F.Y. Wislogle에 의하여 第二次世界大



戰後發表<sup>6)</sup>된 것으로서 1950年代末에는 이 코오드를 改良하여 Warner-Lambert Research Institute가 利用하여 왔다.

Wiselogle 코오드<sup>7)</sup>의 特徵은 化合物의 functional unit(官能基와 環內 雜原子를 포함하는 概念)를 Division, Class, Descriptor의 3個의 범주로 分類하고 英文數字코오드로 나타내는데 있다.

各 범주의 內容은 다음과 같다.

2.1 Division

化合物中の 主要原子를 英文字로 나타내며, N, O, S, 할로겐에 대하여는 各各의 原子記號가 使用된다. 其他 原子는 foreign element로서 單一 記號로 表示한다.

2.2 Class

functional unit를 다음 6가지로 分類하고 있다.

- 1: 無機의 functional Unit(주로 加水分解에 의하여 생기는 것)
- 2: 脂肪族 鎖狀化合物에 結合하는 functional unit
- 3: 非芳香族環에 結合하는 functional unit
- 4: 芳香族環에 結合하는 functional unit
- 5: 非芳香族環 및 5員環中에 雜原子를 含有하는 경우
- 6: 芳香族環中에 雜原子를 含有하는 경우.

5와 6중 특히 N을 含有하는 環에 대하여는 다시 A 와 B로 區分하고 N5A와 N6A는 N 1個 또는 O와 NS 와 N을 含有하는 雜環을 나타내며, N5B와 N6B 는 N을 2個 以上을 含有하는 環을 나타낼 때 使用한다.

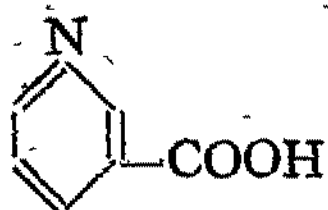
2.3 Descriptor

functional unit를 다시 자세하게 記述하는 코오드로서, 表 6은 functional unit에 使用되는 O, S, N,의 descriptor를 例示한 것이다.

表 6. Functional Unit Descriptors.

O 誘導體	S 誘導體	N 誘導體
1 ROH	1 RSH	1 R <sub>3</sub> N
7 RCHO	7 RSOH	7 R <sub>2</sub> NOH
10 RCOOH	11 RSO <sub>3</sub> H	11 RNO <sub>2</sub>
15 ROOR	14 RSSR	15 RN=NR

니코틴酸(Nicotinic acid)를 코오드化하여 보면,



N6A-1 피리딘(pyridine)

04-10 芳香族 카복실酸(aromatic carboxylic acid) 各各의 記號의 意味는 다음과 같다.

Division	Class	Descriptor
N	6A	1
O	4	10

N: N原子  
 6A: 피리딘環內의 N  
 1: R<sub>3</sub>N 즉 3級의 N(表 6)  
 0: O原子  
 4: 카복실基가 피리딘環에 直結  
 10: RCOOH(表 6)

3. Warner-Lambert法

Warner-Lambert法<sup>8), 9), 10)</sup>은 제너릭서어치에 特徵이 있으며, 피-커-부(Peek-a-boo)카아드 方式<sup>11)</sup>을 취하고 있지만, 인풋에는 IBM카아드를 使用하고 있다. 이것은 장래 電子計算機베이스로서의 檢索을 고려하고 있기 때문이다.

化合物 코오드에는 1) 分子式 2) 프래그먼트코오드 3) 單環式(monocycle)코오드 4) fused ring코오드의 4 種類의 IBM카아드가 使用되고 있다.

3.1 分子式

分子式카아드에는 分子式外에 化合物 番號, 제너릭코오드가 코오드化된다.

제너릭코오드에는 스테로이드, 펩타이드, 糖 등의 化合物의 型이나 시스, 트란스, 스피로(cis, trans, spiro) 등의 化合物의 立體的 性質을 表示한다.

3.2 프래그먼트코오드

官能基를 表示하는 프래그먼트코오드에는 다음과 같은 3가지 手法이 使用되고 있다.

A) 6行코오드

6行코오드의 第 1行은 英文字를 使用하여 化合物을 몇개의 官能基群으로 分類한다. 카아드는 다음의 3분류에 포함되는 官能基群마다 製作한다.

H: 알콜, 에텔, 티오알콜, 설파이드

O: 카보닐

M: 아민, 나이트릴, 이소나이트릴, 이민

Y: 아마이드, 尿素, 이미드, 시아네이트

L: 할로겐

2~5行은 各各 官能基中の N, S, O, C原子를 나타낸다. 6行은 5行까지의 코오드로서는 識別할 수 없는 경우에 使用한다. 예를 들면, 1級, 2級알콜에 대하여

H00101, H00102로 코오딩한다.

B) attachment 코오드

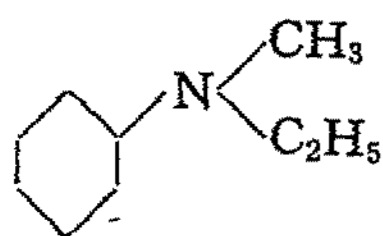
官能基의 殘基(attachment)를 2行의 코오드로 나타내  
고, attachment A에서 D(各各 2行)까지의 4個를 表示  
할 수 있도록 되어 있다.

1行은 제너릭레벨(generic level)을, 2行은 스페시픽  
레벨(specific level)을 表示한다.

表 7. Attachment 코오드

GENERIC LEVEL	SPECIFIC LEVEL
H	(NONE)
A : ALKYL	SPECIFY NO. OF C's (2, 3, 4, ETC.)
L : ALICYCLIC	SPECIFY RING SIZE (3, 4, 5, 6, ETC.)
P : PHENYL	(NONE)
T : HETERO	N : NITROGEN Y : OTHER HETERO S : SULFUR P : NO O : OXYGEN V : NS. ETC.
X : HALOGEN	F : FLUORINE B : BROMINE K : CHLORINE I : IODINE
F : OTHER FUNCTIONS	N : -NR <sub>2</sub> , -NO <sub>2</sub> , ETC. O : -OH, -OR, -O-O-, ETC. S : -SH, -SR, -SO <sub>2</sub> NH <sub>2</sub> , ETC. G : METAL-ORGANIC C : -CN, -COOR, ETC.

예를 들면, 메틸에틸아닐린(methylethylaniline)



attachment A : A1 (CH<sub>3</sub>, 알킬 C 1個)  
B : A2 (C<sub>2</sub>H<sub>5</sub>, 알킬 C 2 )  
C : P (벤젠環, 2行은 없음)

C) 알킬, 알킬렌(alkyl, alkylene)코오드

官能基 혹은 環에 結合하여 있는 部分의 알킬 및 알  
킬렌을 6行코오드로 表示한다.

A)와는 다른 카아드를 使用한다.

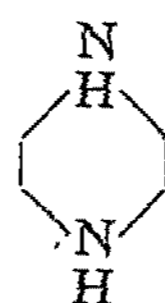
알킬렌코오드는 1行에 A를 表示하고 2~6行에서 各  
各 炭素數, 二重結合, 三重結合, 共役結合의 數, 炭素  
鎖中の 炭素總數(分枝도 포함)를 코오딩한다.

알킬코오드도 2~5行까지는 알킬렌코오드와 同一하지  
만, 1行에는 C를, 6行에는 直鎖이면 O을, 分枝鎖이면  
I을 코오딩한다.

3.3 單環式코오드

單環系 化合物을 環의 크기(5員環, 6員環 등), 2重結  
合의 數, N, S, O原子의 數, 雜테로原子의 相互關係를

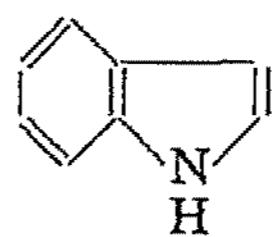
6行으로 코오딩한다.



피페라진(Piperazine) : 602004

- 6 : 6員環
- 0 : 二重結合이 없음
- 2 : N이 2個
- 0 : S無
- 0 : 0無
- 4 : 1,4位의 4
- 4. 4. fused ring코오드

7行의 코오드로서 縮合環을 나타낸다.



인돌(Indole) : 2E11040

- 2 : 環의 數 2個
- E : 2行은 英文字코오드로서 表 8에 의하여 정해진  
것이며 E는 雜테로 原子가 N인 경우이다.
- 1 : 5員環의 數 1個
- 1 : 6員環의 數 1個
- 0 : <5 및 > 6員環 無
- 4 : 二重結合의 數 4個
- 0 : 識別코오드, 이소인돌에는 1을 부여하여 區別  
하고 있다.

表 8. 雜테로原子코오드

NO HET ATOM	-O	-N <sub>3</sub> O <sub>+</sub> S <sub>+</sub> -R
N	-E	N <sub>4</sub> -S
NO	-F	N <sub>4</sub> O <sub>+</sub> S <sub>+</sub> -F
NO <sub>2+</sub>	-G	N <sub>5</sub> <sup>+</sup> -U
NS	-H	N <sub>5</sub> O <sub>+</sub> S <sub>+</sub> -V
NOS	-I	N <sub>6</sub> O <sub>+</sub> S <sub>+</sub> -W
N <sub>2</sub>	-J	N <sub>7</sub> O <sub>+</sub> S <sub>+</sub> -X
N <sub>2</sub> O	-K	O -A
N <sub>2</sub> O <sub>2+</sub>	-L	O <sub>2+</sub> -B
N <sub>2</sub> S	-M	S <sub>+</sub> -C
N <sub>2</sub> S <sub>2+</sub>	-N	OS -D
N <sub>3</sub>	-O	F.E. -Y
N <sub>3</sub> O <sub>+</sub>	-P	Z -PARTLY
N <sub>3</sub> S <sub>+</sub>	-Q	DEFINED

以上, 4種의 카아드중 分子式카아드는 반드시 製作하  
지만 그 外는 化合物의 型에 따라서 選定한다.

IBM카아드로부터 Termatrix카아드가 製作되는 順序  
는 그림 1에서 볼 수 있는 바와 같이 먼저 칼럼 80의 識  
別코오드로 4種類의 카아드로 나누고, J-400 Termat  
rex drill이란 穿孔機로 Termatrix카아드를 만들고  
있다.

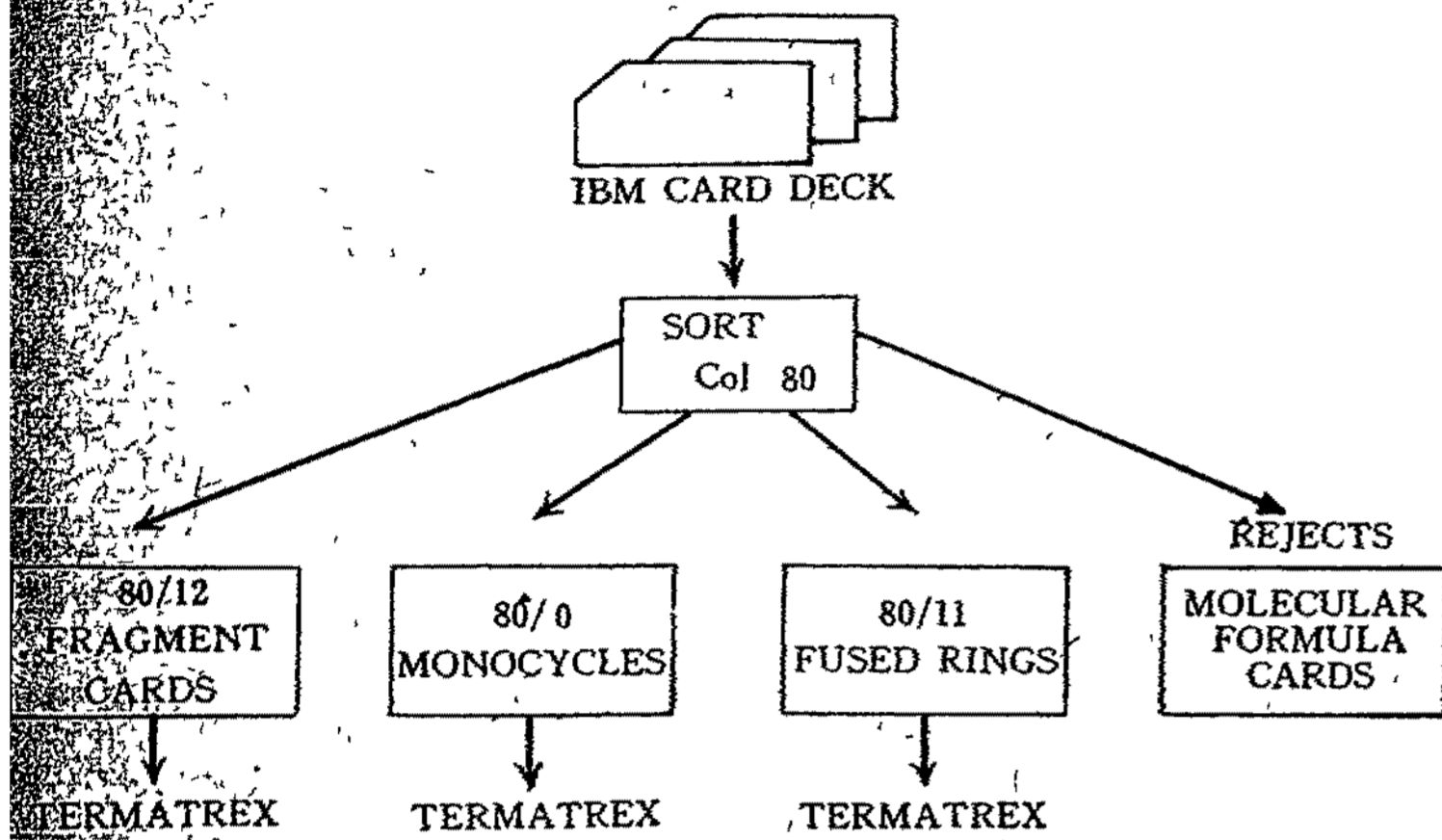


그림 1. Termatex 카아의製作

#### 4. RAMP, CAMP

Random Access Mechanization of Phosphorus (RAMP)와 Card Automatic Mechanization of Phosphorus (CAMP)는 특히 美國 特許廳에서 開發한 磷化合物에 대한 檢索시스템이다. RAMP가 電子計算機에 의한 랜덤 액세스(random access)가 가능한 것임에 비하여 CAMP는 PCS인 점이 다르나, 코오딩의 방식은 거의 같다.

따라서 여기서는 RAMP에 대하여 說明코자 한다.

##### 4.1 RAMP

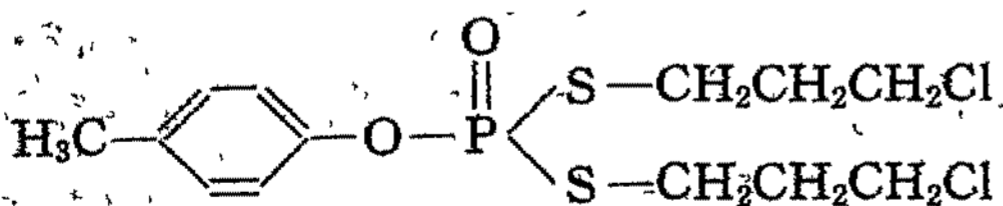
化學構造를 프래그먼트와 node로 分解하여 코오딩한다.

프래그먼트라는 것은 情報檢索의 單位가 되는 原子 또는 原子의 集合을 말한다.

이 프래그먼트가 적어도 2個 以上 結合하여 있고, 그 프래그먼트의 하나에 P를 含有하고 있는 경우에 이것을 node라고 한다.

코오딩의 順序는 다음과 같이 5段階를 거쳐서 이루어진다.

다음의 化合物을 例로 하여 說明코자 한다.



##### 第 1段階

磷原子에 結合하여 있는 原子 즉 node와 그의 數(出現個數)를 코오딩한다.

- (a) P=O-1
- (b) P-S-2
- (c) P-O-1

P-S-2의 2는 P-S node가 2個 있다는 것을 나타낸다.

##### 第 2段階

第 1段階의 node와 여기에 直結하는 프래그먼트 및 그의 數를 表示한다.

- (d) P-S-C<sub>3</sub>-NT-SAT-ST-CH-2

이것은 P-S node에 炭素數 3個(C<sub>3</sub>)의 末端이 아닌 (NT: non-terminal) 飽和(SAT: saturated)의 直鎖狀(ST: straight), 炭素鎖(CH: carbon chain)가 結合하고, 이 그룹의 node가 2個 있음을 나타내고 있다.

##### 第 3段階

第 2段階에서 第 1段階의 node에 結合한 프래그먼트部分만을 코오딩한다.

- (f) C<sub>3</sub>-NT-SAT-ST-CH-2
- (g) phenyl-1

##### 第 4段階

第 2段階의 그룹 node에 直結하여 있는 프래그먼트를 表示한다.

- (h) Cl-CH-2

炭素鎖에 結合하여 있는 鹽素가 2個

- (i) C-T-R-1

環(R: ring)에 結合하는 末端(T: terminal)의 메틸기(C)가 1個 있음을 表示한다.

##### 第 5段階

第 5段階는 環의 置換基의 位置表示(o, m, p)나 프래그먼트의 總數 및 磷의 原子價를 코오딩한다.

- (j) P

파라(para)置換

- (k) 12 Frag

(a)에서 (I)까지 프래그먼트는 12個이다.

- (I) Phos 5.

5價의 磷

(a)에서 (I)까지의 各 node 및 프래그먼트마다 12枚의 IBM카아드를 製作하며, 各 카아드는 特許 No, 入受 No, 化合物 No가 들어 있다.

#### 5. PACIR

Practical Approach to Chemical Information Retrieval (PACIR)<sup>12)</sup>은 스테로이드, CAMP, RAMP 등의 機械檢索 經驗을 살려서 開發한 美國 特許廳의 情報檢索시스템이다. PACIR는 殺虫劑 關係의 約 5,000件의 特許에 應用하고 있지만, 融通性이 있는 一般의 手法의 하나이다.

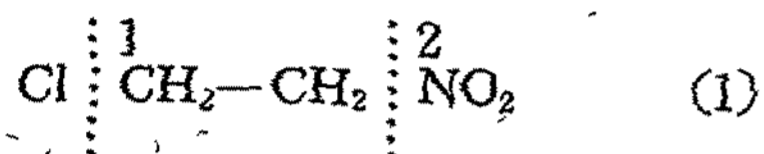
##### 5.1 프래그먼트化



化學構造를 프래그먼트 構造單位로 나누는 規準은

- 1: 環
- 2: 炭素鎖
- 3: 環의 構成原子가 아닌 N, S, O 및 >C=O,  
 $\begin{array}{c} | \\ -N=C-N < \end{array}$ 으로 나타나는 原子의 結合
- 4: P-X (X: O 또는 S)

로서, 다음 化合物의 點線으로 區分한 部分이 프래그먼트이다.



### 5.2 線形으로 코오딩

各 프래그먼트에 다음과 같은 記號(이것을 symbolic code)를 使用하여 코오드를 부여한다.

심볼코오드(symbolic code)의 一部:

- CH: 炭素鎖
- S: 飽和炭化水素
- U: 炭素間의 二重結合
- UT: 炭素間의 三重結合
- ST: 炭素直鎖
- BR: 炭素分枝鎖
- T: 炭素鎖의 末端
- NT: 炭素鎖의 末端이 아닌 部分

頻度記號의 一部:

- W: 프래그먼트가 1個인 경우
- X: 프래그먼트가 2個인 경우
- Y: 프래그먼트가 3個인 경우

심볼코오드는 脂肪族化合物의 例示에 必要한 것이지만, 그 外에 環에 대한 記號가 있다.

(1)은 스페시픽한 化合物의 例로서, 다음과 같이 a)에서 d)까지의 4個의 코오드가 주어진다.

- a) 1-Cl-CH-W/
- b) 1,2-C2-CH-S-ST-NT-W/
- c) 2-NO2-CH-W/
- d) S5-3/.

이것을 說明하면,

鹽素(Cl)가 炭素鎖(CH)와 1個(W) 結合하여 있고 (a), 炭素 2個(C2)의 骨核으로 되어 있다.

이 炭素鎖(CH)는 飽和(S) 直鎖狀(ST)으로, 末端에는 없으며(NT), 1個 存在한다(b). 다시 이 炭素鎖에는 나이트로基(NO2)가 1個 結合하여 있으며 各 코오드의 最後에는 슬래쉬(/)를 넣어 다음 코오드와 區別하고 있다. (d)의 S5-3/는 最後 카아드의 코오드로서 S5-3의 3은 node 및 프래그먼트의 數를 나타낸다. (다음 호에 계속)

### <參考文獻>

- 1) Schwartz, J.M., "Some observations concerning Chemical Abstracts Formula Index," J. Chem. Doc. 9 3 169~171 1969.
- 2) 李正一, "化學文獻의 電子計算化에 있어서 化學構造表表示法의 役割", 情報管理研究 6 6 pp. 137~143, 1973.
- 3) 笹本光雄ら, "IDC의 紹介—GREMAS Code를 中心として—", 情報管理 13 12 pp. 764~773 1971.
- 4) 武田敬一, "Ringdocについて", 情報管理 10 9pp. 488~499 1967.
- 5) 趙載浩, "Ringdoc에 관하여", 情報管理研究 6 4 pp. 94~101, 93 1973.
- 6) Wiselogle, F. Y., et al, "A new system for the classification of compounds," J. Chem. Ed. 23 pp 375~391 1946.
- 7) Arendell, F. H., "A three-symbol code for searching chemical structures", J. chem. Doc. 1 3 pp. 47~57 1961.
- 8) Starker, L.N. et al., "A multi-level retrieval system. I. A simple optical coincidence card system, J. Chem. Doc. 8 2 81~85 1968.
- 9) Starker, L.N., et al., "A multi-level retrieval system. II. Medium-sized collections", J. Chem. Doc 9 3 pp. 161~167 1969.
- 10) 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.
- 11) 司空哲, 도큐멘테이션 概說, 韓國圖書館協會, 1968.
- 12) Frome, J., et al, "PACIR: Practical Approach to Chemical Information Retrieval," J Chem. Doc 2 4 248~255 1962.