

REGISTRY ファイル スクリーン辞書の使い方



JAICI 社団法人 化学情報協会

情報事業部 ヘルプデスク

〒113-0021 東京都文京区本駒込6-25-4 中居ビル

TEL: 0120-003-462 FAX: 03-5978-3600

URL: www.jaici.or.jp

E-mail: support@jaici.or.jp

スクリーンの有効性

- 構造検索のサンプル検索において、フルファイル検索、バッチ検索の予想として INCOMPLETE が表示された場合に利用する。

スクリーン番号から作成された L 番号と構造質問式の L 番号を演算することによって、INCOMPLETE を COMPLETE にすることができる。

サンプル検索とは？

サンプル検索は全ファイルの 5% を対象とする、無料の試し検索である。

- フルファイル検索の前に必ずサンプル検索を行い、以下の内容を確認する。
 - フルファイル検索の予想 (FULL FILE PROJECTIONS) が COMPLETE であるか
 - フルファイル検索をしたときの予想回答数
 - 回答が検索の目的にあっているか
- フルファイル検索が完了しないと予想されている場合 (INCOMPLETE) やノイズ (目的でない回答) が得られているような場合は、構造質問式を修正して、サンプル検索をやり直す。

```
=> S L1
SAMPLE SEARCH INITIATED 09:47:20 FILE 'REGISTRY'
SAMPLE SCREEN SEARCH COMPLETED - 10 TO ITERATE
100.0% PROCESSED 10 ITERATIONS
SEARCH TIME: 00.00.01

FULL FILE PROJECTIONS: ONLINE **COMPLETE**
                      BATCH **COMPLETE**

PROJECTED ITERATIONS: 11 TO 389
PROJECTED ANSWERS:    5 TO 234

L2          5 SEA SSS SAM L1
```

●	5 ANSWERS	サンプル検索の候補化合物数
●	FULL FILE PROJECTIONS: ONLINE **COMPLETE**	フルファイル検索の予想
●	BATCH **COMPLETE**	バッチ検索*の予想
●	PROJECTED ANSWERS: 5 TO 234	予想回答数
●	L2 5 SEA SSS SAM L1	サンプル検索の回答

スクリーン検索

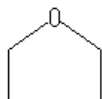
■ スクリーン番号の利用により、サンプル検索の INCOMPLETE の記載が、COMPLETE になる理由

構造検索は、スクリーン検索、イタレーション検索の 2 段階で行われる。サンプル検索の COMPLETE の記載は、サンプル検索時の構造質問式でフルファイル検索した場合に、構造検索が完了することを表す。

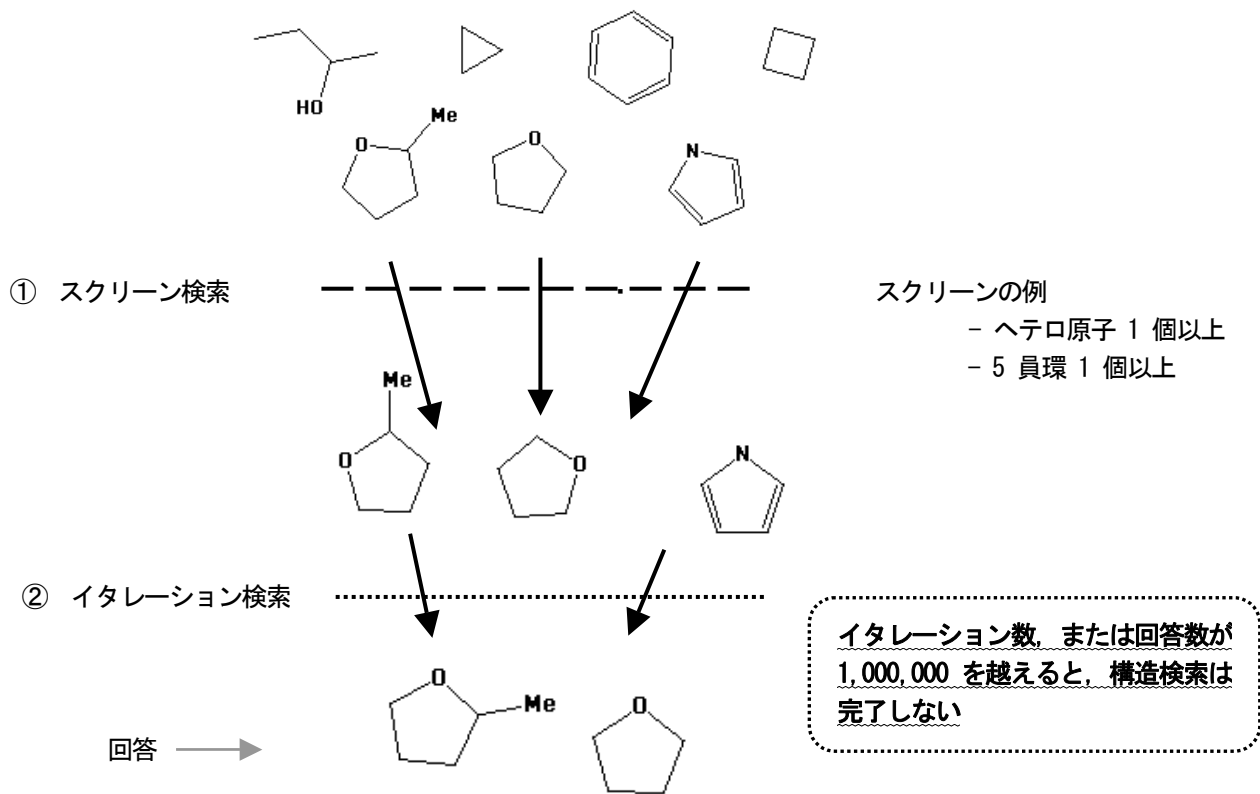
COMPLETE か、INCOMPLETE かはイタレーション検索の予想値がイタレーション検索の制限値内にあるかどうかによって判断される。現在のイタレーション検索の制限値は、1,000,000 件であり、その数を超える物質がイタレーション検索の検索対象になる可能性がある場合に INCOMPLETE が表示される。

スクリーン番号を利用することにより、イタレーション検索対象数を減らすことができるため、INCOMPLETE の予想を COMPLETE にすることができる。

- ・ 構造質問式 L1



- ・ 構造検索の仕組み (⇒ S L1)



- ・ REGISTRY ファイルの構造検索におけるシステム制限値

検索	オンライン検索		バッチ検索*
	サンプル検索	フルファイル検索	フルファイル検索
制限値			
イタレーション数*	2,000 件	1,000,000 件	1,500,000 件
回答数	50 件	1,000,000 件	1,500,000 件

* バッチ検索とは、構造質問式をシステムに登録しておき、コンピュータの利用度の少ない時に一括して行わせる検索のこと。オンライン検索に比較してシステム制限が緩和される。

スクリーン番号の種類

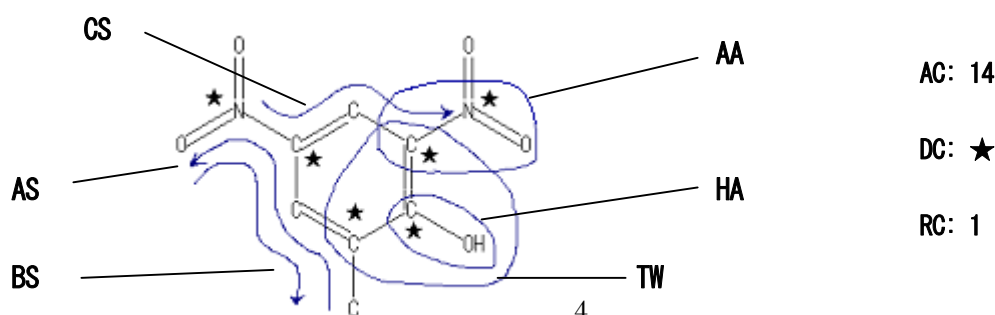
● スクリーン番号

- 12 種類の 2,100 を超えるスクリーンが利用できる。
(ただし、現在、スクリーン番号の 2058 から 2066 までの立体等に関するスクリーンは利用できない。)

種類	名称	概要と例	■	自動発生
AA	Augmented Atom	中心原子の回りの環境 (結合している非水素原子とその結合) 1642 AA N -1C -2O ● 中心原子に結合する全置換基を定義する必要はない	×	○
HA	Hydrogen Augmented Atom	AA に水素を付加したもの ● 中心原子に結合した水素の数は固定 1700 HA OH -1C	△	△
TW	Twin Augmented Atoms	結合する二つの原子上の水素について記述 ● 2 つの原子に結合した水素数は固定 1832 TW C *4C *4C -1OH	△	×
AS	Atom Sequence	4 ~ 6 の原子配列 71 AS C -C *C *C -N -O	×	○
BS	Bond Sequence	4 ~ 6 の原子間の結合の配列 (原子は非特定) 882 BS A -2A -1A *4A *4A -1A	×	○
CS	Connectivity Sequence	原子に結合している非水素原子の個数と原子間の結合の配列 (3 ~ 6) 700 CS 3 -3 *2 *3 -3 ● 非水素原子の数は固定	×	△
AC	Atom Count	同一成分中の非水素原子の最少数 1904 AC 14	×	○
DC	Degree of Connectivity	中心原子に結合している非水素原子の最少数 2113 DC 5 3	×	○
RC	Ring Count	同一成分中の最小環の数 (例: ビスフェノールは 2, ナフタレン環は 2) 1838 RC 1	△	○
TR	Type of Ring	最小環の形状と縮合の状態 1867 TR DDDDD	×	△
EC	Element Count	同一成分中の元素 (水素以外) の最少数	△	△
GM	Graph Modifier	同一成分中の特別な構造上の特徴。ただし、成分数と一部の化合物クラスは、物質全体に働く。	△	×

■ STN Express のフィルタ機能を利用して指定することができる。

- 注意 1) 通常、自動発生するスクリーンであっても、一般式記号や可変原子, G グループで定義されたノードに対しては自動発生しない。
- 注意 2) スクリーンは、各成分の構造に対して働く (例外として GM スクリーンの成分数, 化合物クラスに関するスクリーンは物質全体に働く。ただし、ポリマーと配位化合物については成分に対しても働く)。従って、複数のスクリーンを演算した場合、多成分物質では、必ずしも同一成分中での条件になるとは限らない。



スクリーンの指定方法

① スクリーンコマンドの利用

- 複数のスクリーン番号は, AND または OR で結合することができる.

入力例: => SCR 2049
 => SCR 2043 OR 2127

* AND と OR が同時に存在する場合は, まず AND 演算が実行される.

- 構造質問式とスクリーンの L 番号は, 以下のように入力するのが有効である.

=> S ① AND ② NOT ③

- ① 単独または複数の構造質問式の L 番号
- ② 単独または AND 演算で結合したスクリーン番号の L 番号
- ③ 単独または OR 演算で結合したスクリーン番号の L 番号

注意: => S L1 AND L2 NOT L3 NOT L4 (L1 は構造質問式, L2-L4 は, スクリーン) と入力した場合は,
(L2 NOT L3 NOT L4) の条件にあう構造集合が検索対象になるため, 目的とする構造検索が実行されない.

② STN Express のフィルタの利用

- 構造質問式保存時に、下記の「保存しています」ダイアログボックスが表示されたら、「構造フィルタを使う」にチェックを入れ、「保存」をクリックする。「構造フィルタを使って絞り込む」ダイアログボックスが表示される。

1. スクリーンを
ハイライトする

* コントロールキー
を押しながらクリ
ックすれば複数選
択できる

2. 演算子を選択する

3. スクリーン番号が表示される

4. 作成された質問式の確認

5. 保存する

```

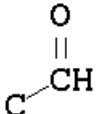
C:\Program Files\stnexp\Queries\samscr.tmp
send "...Testing the current file...↓ screen"
wait
for "THIS COMMAND" goto @errmsg
for "." goto @cleanup
for 5 seconds goto @errmsg1
@cleanup
send "end"
goto @upload
@errmsg
@errmsg
echo "Please change to a suitable file and repeat your upload"
echo ""
exit
@errmsg1
echo
echo "Your upload attempt failed, please try again"
echo ""
exit
@upload
=>screen 2049 OR 2043 OR 1840 OR 2127 ¥)_line1
upload lnum_line2 <C:\Program Files\stnexp\Queries\samscr.str>
=>que_line2 NOT_line1
    
```

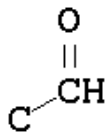
■ SCREEN DICTIONARY の使い方

- AA, HA, TW, AS, BS, CS, RC, TR, AC, DC, EC, GM の順序で記載されている。
- AA, HA, TW, AS, BS の各スクリーンは、元素記号のアルファベット順、次に下記の結合順序に従って表示されている。

結合の表し方	
* 環結合	-鎖結合
*1 環の単結合	-1 鎖の単結合
*2 環の二重結合	-2 鎖の二重結合
*3 環の三重結合	-3 鎖の三重結合
*4 環のノーマライズ結合	-4 鎖のノーマライズ結合
空白 任意の結合	
SCREEN DICTIONARY に表示される記載順位	
なし > * > *1 > *2 > *3 > *4 > - > -1 > -2 > -3 > -4	

原子の表現としては、X はハロゲン、A は水素以外の任意の原子、M は金属を表す。

- 表示例  を 2 個以上持つことを意味するスクリーン番号

Screen Number	Fragment Definition	Freq. %
1305	CH -1C -2O	0.10
↓	↓	↓
スクリーン番号	定義	頻度
↓	↓	↓
スクリーンの種類	下記の構造を示す	
↓	↓	
数 (以上を表す。1 は表記しない)		
		このスクリーン番号は、複数の定義を表している (別の定義にもこの番号が付与されている)。

Screen dictionary の記載順序

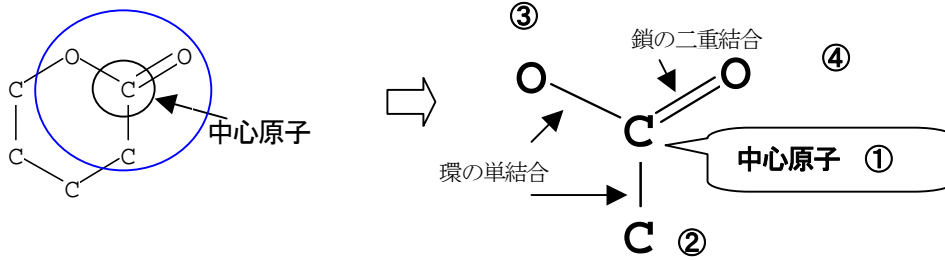
- ① 中心原子と結合している複数の原子の記載順序は、アルファベット順である。
- ② 同じ原子の場合、結合の優先順位に従って記載される。

なし > * > *1 > *2 > *3 > *4 > - > -1 > -2 > -3 > -4

■ 特定の原子の周りの環境を表すスクリーン (3 種)

AA スクリーン

中心原子に結合した非水素原子の種類と、その間の結合を表す。



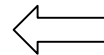
① の中心原子 C は、環の単結合で ② の炭素 (C) と結合、環の単結合で ③ の酸素 (O)、鎖の二重結合で ④ の酸素と結合している。

中心原子①	環単結合	②	環単結合	③	鎖結合	④
C	* 1	C	* 1	O	- 2	O



● 該当する Screen Dictionary での記載

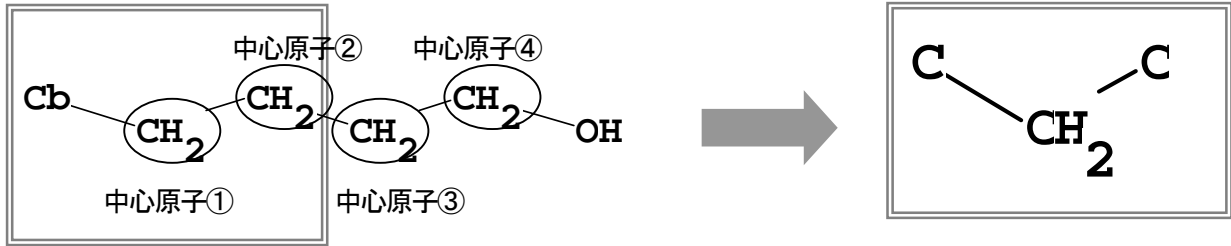
1310 AA C * C * O - O



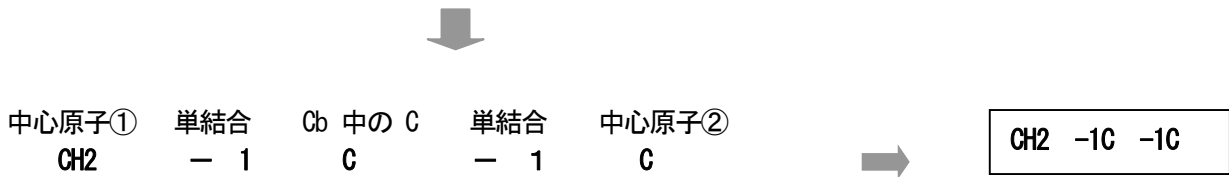
*1C *1O -2O
のスクリーンは
ないため、これ
よりも広い定義
を意味する
*C *O -O を利
用する

HA スクリーン

中心原子に結合する正確な水素の数とその原子に結合した非水素原子の種類とその間の結合を表す。



中心原子①のCH2は、Cb中のC(炭素)と単結合で結合、②のCと単結合で結合している。



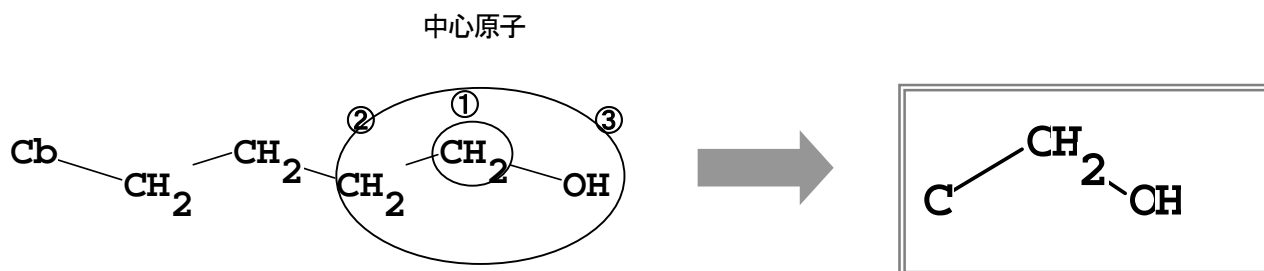
中心原子②, 中心原子③も同様である。

中心原子①	C-CH2-C	}	⇒ スクリーン定義 :	CH2 -1C -1C	が 3 以上存在する
中心原子②	C-CH2-C				
中心原子③	C-CH2-C				
中心原子④	C-CH2-O		⇒ スクリーン定義 :	CH2 -1C -1O	が 1 以上存在する。

● 該当する Screen Dictionary での記載					
1008	HA	3	CH2	-1 C	-1 C
1297	HA		CH2	-1 C	-1 O

TW スクリーン

中心原子・炭素に結合する水素の正確な数と、その炭素に結合している CH₃, NH₂, OH, SH とその間の結合を表す。



中心原子①の CH₂ は、②の C と単結合で結合、さらに③の OH と単結合で結合している。

中心原子① 単結合 ② の原子 単結合 ③ の原子

CH₂ - 1 C - 1 OH → CH₂ -1C -1OH

スクリーン定義： CH₂ -1C -1OH が 1 以上存在する。

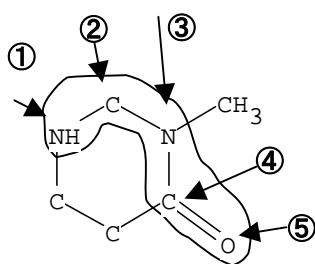
● 該当する Screen Dictionary での記載

1836 TW CH₂ -1 C -1 OH

■ 原子の配列を表すスクリーン (3 種)

AS スクリーン

配列中の 4-6 の原子と結合 (環または鎖) を表す。



構造の中で特徴的な原子配列を切り出し、SCREEN DICTIONARY に対応する配列を探す。必ずしもすべての配列が存在しているわけではない。ここでは、N をスタート地点とし、C-N-C-O の配列を探す。

構造図より	→	① N	② 環結合 C	③ 環結合 N	④ 環結合 C	⑤ 鎖結合 O
スクリーン定義	→	N	* C	* N	* C	- O

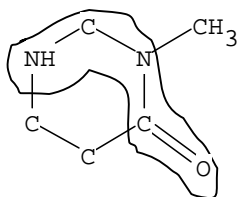
● 該当する Screen Dictionary での記載

426 AS N * C * N * C - O

BS スクリーン

配列中の結合についてその次数（単結合，二重結合，三重結合，ノーマライズ）と結合タイプ（環または鎖）を表す。

原子は，任意の元素（水素以外）として A で表す。



構造図より →
スクリーン定義 →

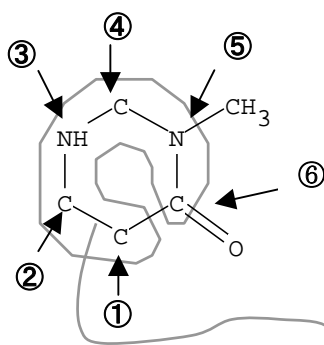
①	②	③	④	⑤
A	環結合 A	環結合 A	環結合 A	鎖結合 A
A	*1 A	*1 A	*1 A	-2 A

● 該当する Screen Dictionary での記載

764 BS A *1 A *1 A *1 A -2A

CS スクリーン

配列中の原子の結合手数(非水素原子との結合手数)とその間の結合のタイプ
(3 ~ 6 の原子の正確な結合手数の列)



結合手数とは，水素以外の結合している原子の個数で表す。
結合次数は関係ない。
例：③ の N は，② と④ の 2 つの相手と結合している
ので，結合手数は，2 となる。

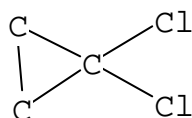
構造図より → ① 手数 環結合 ② 手数 環結合 ③ 手数 環結合 ④ 手数 環結合 ⑤ 手数 環結合 ⑥ 手数

スクリーン定義 → 2 * 2 * 2 * 2 * 3 * 3

● 該当する Screen Dictionary での記載

600 CS 2 * 2 * 2 * 2 * 3 * 3

注) ヘテロ原子を含む長い経路か，環と鎖を結合している長い経路を探す。次に各経路を 4-6 個の長さの配列に分解して調べる。その時に，任意の原子が同一の経路の中に 2 度でてくることはない。



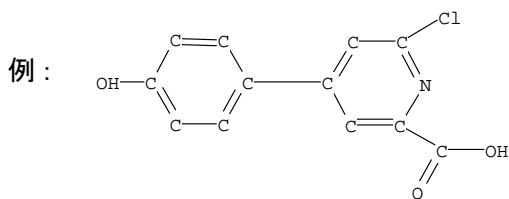
1,1-ジクロロシクロプロパンは

AS -Cl * C * C * C -Cl を用いることはできない

■ 構造的な特徴を表すスクリーン (6 種)

AC スクリーン

同一成分中の非水素原子の最少数を表す



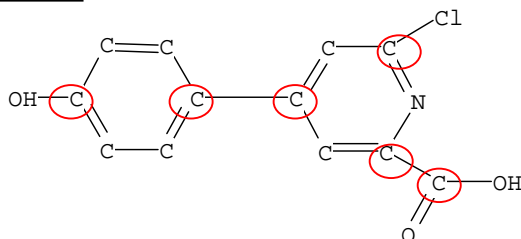
上記の構造の非水素原子数（水素を除く原子の数）は、17 である。
AC スクリーンには、17 以上がないため、16 以上を使用する。

1905 AC 16 ← 16 以上である 1905 を使用する
1906 AC 18

DC スクリーン

中心原子に結合した非水素原子の最少数と構造中に存在する数を表す。

例：



3 個の非水素原子と結合している原子が 6 個存在する。

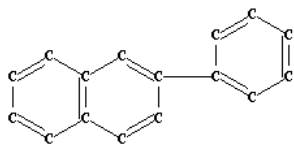
2113 を使用する

2113 DC 5 3 ← 3 個以上の非水素原子と結合している原子が 5 個以上存在する
2114 DC 7 3 ← 3 個以上の非水素原子と結合している原子が 7 個以上存在する

RC スクリーン

成分内の最小環の数を表す。（辞書検索の /CNR フィールドと同じ）

例：



最小環は、合計 3 個なので、1840 を利用する。

1840 RC 3 ← 最小環 3 個以上

最小環の数え方：環系内の環の数は、環結合の数から環内原子の数を差し引き、1 を加えれば求められる。

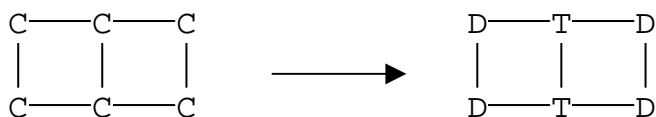
$$7 (\text{結合}) - 6 (\text{原子}) + 1 = 2$$

TR スクリーン

最小環の形状と縮合の状態を表す。

最小環 (3-7 員環) ごとに D と T を利用して環の原子配列を表す。
配列順序は, アルファベット順, DDDDTT > TDDDDT

D : 2 個の環の原子に結合している原子を表す。
T : 3 個以上の環の原子に結合している原子を表す。



最小環ごとにアルファベット順で表すと DDTT と DDTT になる。

1853 TR DDTT ← DDTT の最小環が 1 個以上

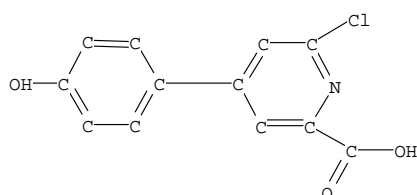
1854 TR 2 DDTT ← DDTT の最小環が 2 個以上

* この場合は 1854 を利用する

EC スクリーン

同一成分中の元素の最少数を表す。

原子 (H, D, T を除く) の最少数を指定する。



作図した炭素数は, 12.

C	1	
	4	199
	6	1943
	8	1944
	10	1945
	12	1946 ← 12 個以上の炭素を含む 1946 を利用する

GM スクリーン

特別な構造上の特徴を表す。構造中の SAF の存在を指定できる。

異常原子量, 異常結合価, 電荷を持つ原子の存在を指定できる。

成分数, 化合物クラス (ポリマー, 配位化合物は除く) は, 物質全体に対して機能する。

GM (Graph Modifier Element) スクリーンと EC スクリーンは, SCREEN DICTIONARY の中で次のように記載されている。

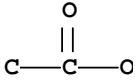
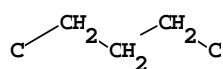
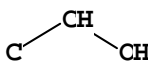
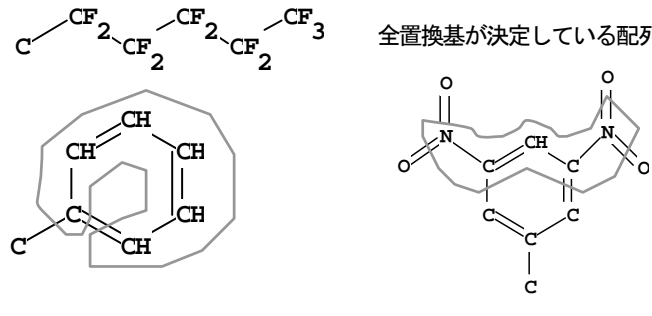
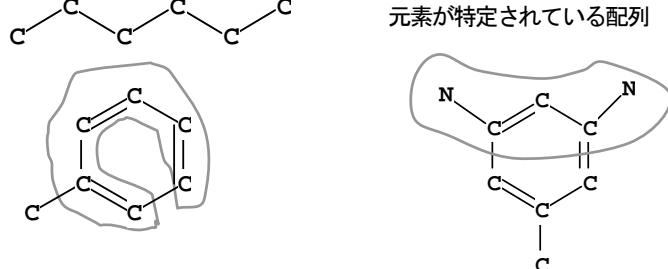
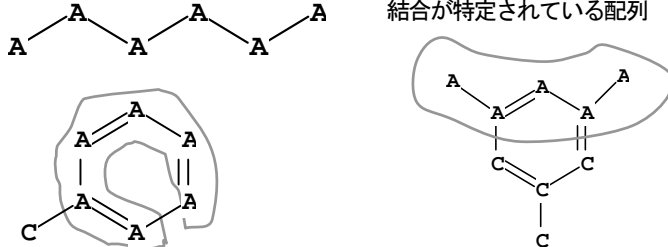
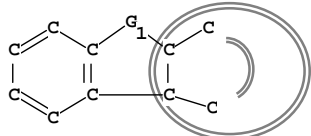
	またはそれ以上	結合表中の原子	単原子フラグメント		
	↓	↓	↓		
	"or-more"	Element Count	Graph Modifier	EC or GM	
	Count	(CT Graph)	(SAF)	(CT + SAF)	
Br	1		2030	0.90 %	1938 5.17 %
	2	1939		1.49 %	

スクリーン利用例

■ スクリーンが自動発生しにくい構造質問式

- ・ 鎖状構造のみの構造
- ・ G グループを利用した構造
- ・ 可変基（可変原子，一般式グループ）を利用した構造

■ 有効なスクリーンの使い方

作図した構造の特徴的な部分	条件と有効なスクリーン
 <p>特徴的な官能基が存在する</p>	AA スクリーン
 <p>水素数が固定されている炭素鎖が存在する</p>	HA スクリーン
 <p>水素数が固定されている炭素鎖に OH, CH3, NH2, SH が存在する (OH の代わりに CH3, NH2, SH)</p>	TW スクリーン
 <p>全置換基が決定している配列</p>	CS スクリーン
 <p>元素が特定されている配列</p>	AS スクリーン
 <p>結合が特定されている配列</p>	BS スクリーン
 <p>特定部位にのみ環が縮合する場合</p>	TR スクリーン CS スクリーン, AS スクリーン, BS スクリーン

特に、環と側鎖から構成されている配列を利用すると有効である

作図した構造の特徴的な部分	条件と有効なスクリーン
構造質問式中に、置換基が完全に定義できるノードが多数存在する	DC スクリーン
構造質問式中のノードに、水素以外の置換基がこれ以上結合しない場合	AC スクリーン (対応する辞書検索機能あり)
構造質問式中に、多数の環が存在している場合	RC スクリーン (対応する辞書検索機能あり)
構造質問式中に、多数の特定元素が存在している場合 例 1. 多数の炭素が作図している場合 (作図している炭素の数を最小数としてスクリーンを利用する) 例 2. C, H 以外の元素を多数作図している場合	EC スクリーン (対応する辞書検索機能あり. ただし、SAF に関する限定は、スクリーンのみ有効)
構造質問式に、異常原子量、異常結合価、電荷を持つ原子が存在している場合	GM スクリーン (対応する辞書検索機能なし)
成分数 (物質全体)	GM スクリーン (辞書検索でも限定可能)
化合物クラス (物質全体、ただし、ポリマーと配位化合物は成分のみに定義されていても検索できる)	GM スクリーン (辞書検索でも限定可能. ただし、ポリマーと配位化合物についてはスクリーンの方が有効.)

注意：スクリーンは、構造全体に有効であるため特定の部分構造に対してのみ有効にすることはできない。
多成分物質の場合、特定の成分にそのスクリーンを限定することはできない。