MARPAT®
USER GUIDE

NOVEMBER 2013

CAS® | STN®

CAS is a division of the American Chemical Society

www.cas.org

TABLE OF CONTENTS

CHAPTER 1: OVERVIEW OF MARPAT

PURPOSE OF THIS GUIDE

This user guide is intended to serve as a learning aid and a reference tool to conduct Markush structure searches. The MARPAT User Guide is organized into chapters on specific topics. This first chapter offers an overview of the database.

What Does MARPAT Mean?

The term, MARPAT, is derived from **MAR**kush or generic structures from **PAT**ents. MARPAT is the CAS database of Markush structures found in the world's patent literature. The database was designed as an extension of the substance searching provided in the CAS REGISTRY^{SM} and CAplus^{SM} databases to perform comprehensive patent substance searching.

What is MARPAT?

MARPAT is a complex and rich database of chemical Markush structures from patents. Structure searching in MARPAT allows retrieval of patent records that might be missed by relying solely on other types of patent searching. The more experience you can gain in the use of this database, the more you will benefit from its valuable content. LMARPAT is a Markush learning database that you can use to practice search techniques.

WHAT ARE MARKUSH STRUCTURES?

Patent documents often contain many specific chemical substances as well as generic structures. Generic structures are named "Markush" because of a dye chemist, Eugene Markush, who was the first person to successfully obtain a patent based on generic structures in 1923. A Markush structure is a single structure used to define a set of implied structures specified on the "atom" level (methyl, phenyl) and/or "generic" level (alkyl, aryl). It occurs in the claims and/or the disclosure portion of patent documents. This allows the patent applicant to reasonably broaden or extend the scope of their invention without specifying every single substance to be covered by the patent. MARPAT on STN offers valuable access points to the patent literature for chemical substances by making these structures searchable.

VALUE OF MARPAT

MARPAT is a database of Markush structures built through intellectual analysis of the entire patent document by highly trained experts. The MARPAT structures mirror the Markush structures and the language found in the patent record itself. The specific claims, examples, and special definitions of terminology are covered to contain complete Markush structures representing many possible substances.

Searching MARPAT for the many substances represented by Markush structures is an invaluable companion to patent searches for specific substances in REGISTRY and CAplus or other patent databases. A patent substance search without MARPAT cannot be considered comprehensive and complete.

Multiple Markush Structures

A patent document can contain more than one Markush structure – for example, the claimed product from claim 1 and starting materials and catalysts from later claims and/or the patent disclosure.

MARPAT COVERAGE

MARPAT contains the Markush structures found in patent records indexed in CAplus. Each record represents a patent document. All subject areas of chemistry are covered; however, the database is heavily organic in nature. MARPAT includes: 1) Patents from all countries covered by CAS with patent publication years of 1988 to the present. Exceptions: Korean patent coverage begins with 2008 and Russian patent coverage began January 10, 2000; 2) Markush structures acquired from INPI, the French National Patent Office, dating from 1961 to 1987; 3) English, French, and German patents covered in CAplus from 1985-1987; 4) Japanese patents covered in CAplus from 1987.

How Many Records?

As of February 2013, MARPAT included more than 398,400 patent records containing more than 969,400 Markush structures. For the latest information enter NEWS FILE at the command prompt in MARPAT or visit www.cas.org.

Structures MARPAT Includes

MARPAT includes Markush structures that are representations of organic or organometallic molecules. Usually, the Markush structure in a patent consists of a structure diagram with real atoms (C, N, P, etc.) and generic representations and textual descriptions defining the extent of the generic representations. The textual descriptions can be specific or generic.

What is Not Included

- Alloys
- Metal oxides
- Inorganic salts
- Intermetallics
- Polymers

Document Selection

Patents are covered from all 80 subject area sections of *Chemical Abstracts*. For inclusion in MARPAT, a complete Markush structure must be shown in the patent with defined variability and organic character.

MARPAT was designed to cover sets of related organic molecules. To qualify for MARPAT coverage, there must be at least one carbon-carbon bond present in the structure. Many organometallics, coordination compounds, and biological molecules such as peptides are covered in MARPAT. Some of these require special indexing and search techniques (refer to Chapter 3).

Structures Not Included

These types of structures *are not* included:

- Specific compounds (Covered in REGISTRY/CAplus.)
- Incompletely defined Markush structures, e.g. "---NR-$(CH_2)_n$-C(=O)---"
- Table structures (with R-groups used as shorthand to list a set of specific compounds)
- "Text" Markush structures, e.g., "halogenated alkenediols"
- Polymers (However, MARPAT does cover low-range structural repeating units and claimed monomers.)

Additional Sources

Two exceptions to the "patent only" coverage of MARPAT are structures from the intellectual property journals *Research Disclosure* and *IP.com*.

Claims and Disclosure

All Markush structures from the patent claims are indexed, and selected "high-value" structures from the patent disclosure are also covered. In most cases, Markush structures are selected from the disclosure because they represent a more broadly defined version of the claimed structure.

For example, if a Markush structure were described as follows, the disclosure version would also be indexed:

Claims version   R = O, S, NH
Disclosure version   R = O, S, Se, NH, NR_1

Document Analysis

In all cases, selection and indexing are based on examination of the whole patent document. Each Markush structure indexed in MARPAT includes a reference to the "patent location", i.e., a claim or the disclosure, as the source of the particular Markush structure.

COMPONENTS OF MARPAT STRUCTURES

The indexed Markush structures are modeled on the language of the patent's Markush structure.

The structures consist of:

- Base or core structure (invariant graphics portion)
- Sets of alternatives (variable groups listed as G1, G2, etc.)
- Definitions of the variable groups
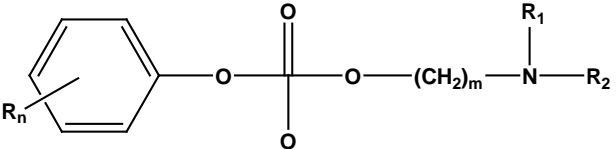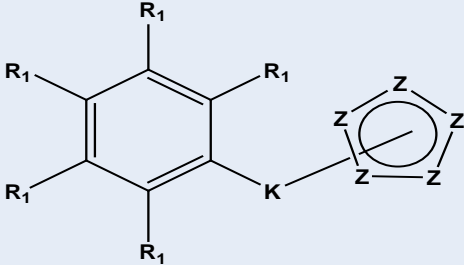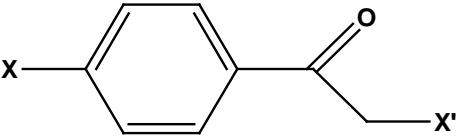
The following also apply:

- Groups can be specified as optionally substituted (opt. substd.) or required substitution (substd. by).
- Groups or substituents can have occurrence counts that indicate numbers of ranges of substitution.

- Definitions can be specific, generic, or textual, e.g., phenyl, alkyl, or "blocking group".
- Location of the Markush structure in the patent is always indicated in MARPAT
- Other information is often included in notes

Examples of Markush Structures

The following table shows some typical Markush structures from patent documents:

MARKUSH STRUCTURE	DESCRIPTION
	in which: • R is optionally substituted alkyl, halo, NO_2, CN, OH, alkoxy, substituted by one or more halo; two adjacent R-groups can combine to form a ring • n = 0-5, m = 1-3 • R_1 and R_2 are H or organic groups such as methyl
	in which: • K is a linking group • R_1 is an H or an organic substituent • Z is independently C, CR, N, NR, O, or S
	in which: • X and X′ are halogens such as F, Cl, or Br

Generic Variability

These are some of the most common techniques used by patent authors to define Markush structures in their patents. A Markush structure is commonly claimed on a hierarchy of generic levels. All of these techniques are indexed in MARPAT.

VARIABILITY	DEFINITION	EXAMPLES
Variation of alternatives	A variable group is defined as a set of different organic and/or inorganic possibilities.	• R is defined as alkyl, halo, NO_2, CN, OH
Frequency variation	Different numbers or occurrences of a substituent are allowed.	• $(CH_2)_m$, in which m is repeated 1-3 times • Substituent R can occur 0-5 times (n = 0-5)
Positional variation	A substituent group can be placed in one or more chemically unique positions in a structure.	• R can occur at the *ortho*-, *meta*-, or *para*- position of the phenylene
Variation of generic level	A variable group can contain a hierarchy of general and specific counterparts for a chemical substituent.	• Halogen (general) such as F (specific) • Organic group such as alkyl, more specifically methyl or ethyl
Optional or required substitution	A substituent is itself optionally substituted or requires further substitution; the optional or required substituents may or may not be explicitly named.	• Optionally substituted alkyl • Alkoxy substituted by one or more halo
Additive variation	Single groups can be taken together to derive new possible combinations.	• Two adjacent R-groups can combine to form a ring

Sample Markush Structures in a Patent Document

The following excerpt is from the claims section of a PCT patent document.

WO 2005/102998   PCT/EP2005/051766

55

CLAIMS

1.   A compound of formula I:

formula I
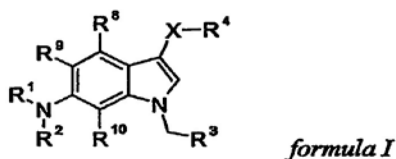
5 wherein

X is	S or SO_2;
R^1 is	(1C-6C)alkyl, (3C-6C)alkenyl, or (3C-6C)alkynyl, each optionally substituted with (3C-6C)cycloalkyl, OH, OC(O)(1C-4C)alkyl, (1C-4C)alkoxy, halogen, cyano, formyl, C(O)(1C-4C)alkyl, CO_2H, CO_2(1C-4C)alkyl, $C(O)NR^5R^6$, S(O)(1C-4C)alkyl or $S(O)_2$(1C-4C)alkyl;
R^2 is	hydrogen, (1C-4C)alkyl or C(O)(1C-4C)alkyl;
R^3 is	a phenyl group optionally substituted with (1C-4C)alkyl, (1C-4C)fluoroalkyl, (1C-4C)alkoxy, (1C-4C)fluoroalkoxy, halogen, cyano or nitro;
	or R^3 is a 5- or 6-membered aromatic heterocyclic ring structure optionally substituted with (1C-4C)alkyl, (1C-4C)fluoroalkyl, (1C-4C)alkoxy, halogen or cyano;
R^4 is	a phenyl group or an aromatic 6-membered heterocycle, substituted at the ortho position with 1-hydroxy(1C-4C)alkyl, (1C-4C)alkoxy, C(O)(1C-4C)alkyl, CO_2(1C-4C)alkyl, $C(O)NH_2$, cyano, nitro, or $CH=NOR^7$, and optionally further substituted with (1C-2C)alkyl, (1C-2C)fluoroalkyl or halogen;
	or R^4 is 2-pyridyl optionally substituted with (1C-2C)alkyl, (1C-2C)fluoroalkyl or halogen;
R^5 and R^6 are	independently hydrogen or (1C-4C)alkyl;
R^7 is	hydrogen or C(O)(1C-4C)alkyl;
R^8, R^9, R^{10} are	independently hydrogen, (1C-2C)alkyl, fluoro or chloro;

or a salt thereof.

30 2. A compound according to claim 1, characterised in that R^8, R^9, R^{10} are hydrogen.

3. A compound according to claim 2, characterised in that:

R^2 is   hydrogen, (1C-2C)alkyl or $C(O)CH_3$;

MARPAT Record

This is a portion of the corresponding MARPAT record for the patent document and an explanation of the various parts of the record.

```
        ● ● ●

AN    143:440257   MARPAT   ①




MSTR 1  ②                            ③


G1    = S / SO2    ④
G2    = alkyl <containing 1-6 C>
        (opt. substd. by 1 or more G3) /
        alkenyl <containing 3-6 C> (opt. substd. by 1 or more G3) /
        alkynyl <containing 3-6 C> (opt. substd. by 1 or more G3) /
        (Specifically claimed: CH3 / CH2CN / 32 / CH2CH2OH / 34)



  H2C———COCH3      H2C———CH2—OMe
     32                34


G3    = cycloalkyl <containing 3-6 C> / OH /   ⑤
        alkylcarbonyloxy <containing 1-4 C> /

        alkoxy <containing 1-4 C> / F / Cl / Br / I / CN / CHO / ⑥
        alkylcarbonyl <containing 1-4 C> / CO2H /
        alkoxycarbonyl <containing 1-4 C> / CONH2 /
        alkylaminocarbonyl <containing 1-4 C> /

        dialkylaminocarbonyl <each alkyl containing 1-4 C> /  ⑦
        alkylsulfinyl <containing 1-4 C> /
        alkylsulfonyl <containing 1-4 C> /
        (Specifically claimed: OCOMe / OCOCH2Me / OMe / OEt / COMe /

        ● ● ●



  G12—OH        HC═══N———OH        HC═══N———O———G13
 22 22            24                 27
G12   = alkylidene <containing 1-4 C>
G13   = alkylcarbonyl <containing 1-4 C> /
        (Specifically claimed: COMe / COCH2Me)
G14   = Me (opt. substd. by 1 or more F) /
        Et (opt. substd. by 1 or more F) / F / Cl / Br / I
G15   = Me (opt. substd. by 1 or more F) /
        Et (opt. substd. by 1 or more F) / F / Cl / Br / I
G16   = H / Me / Et / F / Cl


Patent location:        claim 1    ⑧

Note:                   or salts    ⑨
```

Description of MARPAT Record Fields

RECORD LOCATION	DESCRIPTION
①	The Accession Number (AN) is the CA Abstract Number.
②	The structure displays in the MSTR (Markush STRucture) display field. Each Markush structure in the record has a number. This is the first structure in this record, MSTR 1.
③	The structure diagram contains the base structure and alternatives used in G-group definitions. G-groups, e.g., G1, are used to represent the variable substituent definitions just as R-groups are used in the patent. The node numbers displayed below the structure fragments are those used in G-group definitions.
④	The alternatives for the G-groups are given in their definitions. Each alternative is separated by a slash (/). For example, G1 can be either S or SO2 (G1 corresponds to X in the Markush structure of the patent document).
⑤	The patent specifically claimed the substances formed when G2 is methyl, etc. (R^2 in the Markush structure of the patent document)
⑥	Text groups are used in the alternatives. They may be generic groups, e.g., heterocycle, or they may describe specific real atom groups, e.g., cycloalkyl <containing 3-6 C>.
⑦	If additional information is given about the generic group, it follows the group in angle brackets (< >). In G3, for example, the alkoxycarbonyl group is a chain of one to four carbons. If a group is optionally substituted, this information is included; for example, the alkyl group in G2 can be substituted by one or more G3 groups.
⑧	To help you find the structure in the original patent, its location is given in the Patent location field, e.g., claim 1.
⑨	The Note field includes other pertinent information about the claimed structure. The terms in this field are searchable in the Basic Index (/BI).

Text Descriptions in MARPAT Structures

The text portion of a MSTR display contains generic groups frequently used in the original patent, e.g., *alkyl, aryl, acyl*. The text portion also has statements that define the G-groups found in the structure. This text has two formats. The default (expanded form, GTEXT = ON) shown in the example describes the G-group definitions in detail. The compact form (GTEXT=OFF) uses two-letter codes and compacted descriptions for the definitions. Use the SET command to turn GTEXT OFF. The following shows some of the text from the example with GTEXT=OFF. (The graphic structures have been omitted.)

```
        • • •
G1      = S / SO2
G2      = alkyl<(1-6)> (SO (1-) G3) /
          alkenyl<(3-6)> (SO (1-) G3) / alkynyl<(3-6)> (SO (1-)
          G3) / (SC CH3 / CH2CN / 32 / CH2CH2OH / 34)
G3      = cycloalkyl<(3-6)> / OH / alkylcarbonyloxy<(1-4)> /
          alkoxy<(1-4)> / F / Cl / Br / I / CN / CHO /
          alkylcarbonyl<(1-4)> / CO2H / alkoxycarbonyl<(1-4)> /
          CONH2 / alkylaminocarbonyl<(1-4)> /
          dialkylaminocarbonyl<(1-4)> / alkylsulfinyl<(1-4)> /
          alkylsulfonyl<(1-4)> / (SC OCOMe / OCOCH2Me / OMe /
          OEt / COMe / COCH2Me / CO2Me / CO2Et)
G4      = NH / 19
G5      = alkyl<(1-4)> / alkylcarbonyl<(1-4)> / (SC Me / Et /
          COMe)
G6      = Ph (SO (1-) G7) / heteroaryl<EC (1-) C, AN (1-) C,
          RC (1)> (SO (1-) G8) / (SC 2-pyridyl / 3-pyridyl /
          4-pyridyl / 2-pyrimidinyl / 5-pyrimidinyl / 2-furyl /
          3-furyl / 40 / 44 / 50)
G7      = alkyl<(1-4)> (SO (1-) F) /
          alkoxy<(1-4)> (SO (1-) F) / F / Cl / Br / I / CN /
          NO2 / (SC Me (SO (1-) F) / Et (SO (1-) F) / OMe (SO
          (1-) F) / OEt (SO (1-) F))
G8      = alkyl<(1-4)> (SO (1-) F) / alkoxy<(1-4)> / F / Cl /
          Br / I / CN / (SC Me (SO (1-) F) / Et (SO (1-) F) /
          OMe / OEt)
G9      = 21 / 2-pyridyl (SO (1-) G15)
G10     = o-C6H4 (SO (1-) G14) /
          Hy<EC (6) A (1-) C, AN (1-) C, AR (1-), BD (6-) N, RC
          (1), RS (1) E6> (SO (1-) G14)
G11     = 22 / alkoxy<(1-4)> / alkylcarbonyl<(1-4)> /
          alkoxycarbonyl<(1-4)> / CONH2 / CN / NO2 / 24 / 27 /
          (SC CH2OH / OMe / COMe / CO2Me)
```

CAPLUS PORTION OF THE MARPAT RECORD

The remaining portion of the MARPAT record includes the bibliographic information, abstract, and indexing from the corresponding CAplus record. Only the Accession Number is searchable. The remaining information is only displayable. The CBIB ABS IND display is shown here.

```
143:440257  Preparation of indole derivatives as androgens.
     Van Der Louw, Jaap; Teerhuis, Neeltje Miranda; Lommerse,
     Johannes Petrus Maria; Stock, Herman Thijs; Hermkens,
     Pedro Harold Han (Akzo Nobel N.V., Neth.).   PCT
     Int. Appl. WO 2005102998 A1 20051103, 65 pp.   DESIGNATED
     STATES: W:  AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR,
     BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ,
     • • •
     MR, NE, NL, PT, SE, SN, TD, TG, TR.  (English.  CODEN:
     PIXXD2.  APPLICATION: WO 2005-EP51766 20050421.
     PRIORITY: US 2004-2004/PV56504E 20040423; EP
     2004-101700 20040423.
 GI
```
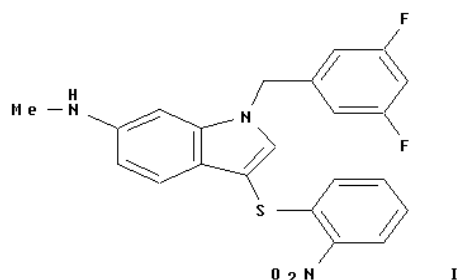
```
AB    Indole derivs. were prepd. and tested for androgenic activity.
      e.g., I was prepd. starting from 6-nitroindole and 1-
bromomethyl-
      3,5-difluorobenzene.  The prepn. of a no. of indole derivs.
Was
      given and extensive androgenic activity data.
IC    ICM  C07D0209-40
      ICS  C07D0401-04; C07D0403-04; C07D0405-04; C07D0413-04;
           C07D0401-12; A61K0031-40; A61P0005-26; A61P0015-16
CC    27-11 (Heterocyclic Compounds (One Hetero Atom))
      Section cross-reference(s): 2
ST    indole deriv prepn androgen
IT    Androgen receptors
      RL: BSU (Biological study, unclassified); BIOL (Biological
           study) (disorders; prepn. of indole derivs. as androgens)
IT    Contraceptives
           male; prepn. of indole derivs. as androgens)
IT    Androgens
      RL: PRP (Properties); SPN (Synthetic preparation); THU
      therapeutic use); BIOL (Biological study); PREP (Preparation);
      USES (Uses)
           (prepn. of indole derivs. as androgens)
IT    868672-39-3P   868672-56-4P   868672-67-7P
      RL: PAC (Pharmacological activity); PRP (Properties); RCT
      (Reactant); SPN (Synthetic preparation); THU (Therapeutic
use);
      BIOL (Biological study);
      PREP (Preparation); RACT (Reactant or reagent); USES (Uses)
           (prepn. of indole derivs. as androgens)
      ● ● ●
```

GENERIC AND SPECIFIC REPRESENTATIONS OF MARPAT STRUCTURES

Unlike other structure search databases on STN, the MARPAT search file contains both generic representations of chemical structures as well as specific representations. The variables in a Markush structure can designate a defined set of specific atoms or groups. They can also designate more generalized structural features using generic terms.

For example, in a Markush structure, R1 can be defined as hydrogen or chlorine or alkyl; R2 can be defined as a carboxy group, cyano group, halogen atom, methyl group, or carbocycle. The point of attachment of a substituent on a ring system can also be variable. In its most generic description, a group of atoms cannot be structurable or even definable in any specific sense, e.g., "blocking group".

Hierarchy of Descriptions

Regardless of the specificity of a structural representation, degrees of specificity are implied and converted through MARPAT indexing. For example, pyridine: the graphic displays varying degrees that are implied by the presence of a pyridine ring.

Pyridine → **6-Membered N-containing Heterocycle** → **Heterocycle** → **Cyclic Organic Substituent** → **Organic Substituent**

Most Specific → **Most Generic**
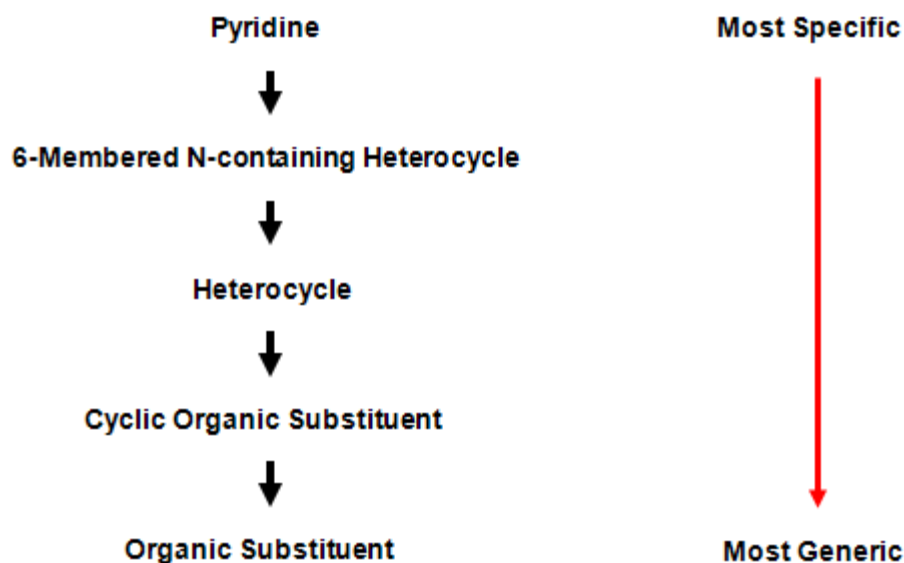
Any structural group can be similarly mapped. Certain characteristics are identifiable. The nodes making up the MARPAT structure are either real atoms or generic groups. They are either carbon or non-carbon. They are either cyclic or acyclic.

You Can Control Precision of Retrieval

MARPAT allows you to control the level of matching between a structure query and the file structure (refer to Chapter 6). Look at pyridine as an example of the scope of the search file. In a standard STN structure database, a pyridine ring retrieves compounds with a pyridine ring or rings fused to pyridine. In MARPAT, a pyridine retrieves those structures, and depending on the match level assigned to the ring in the query, can also retrieve many Markush structures matching the full range of specificity, i.e., pyridine, six-membered N-containing heterocycle, any heterocycle, any ring system, or an organic substituent.

SEARCHABLE TEXT IN MARPAT

Searchable text in MARPAT includes:

- Accession Number (/AN)
- Basic Index (/BI) (made up of the Note, Derivative, Stereochemistry, and Patent Location fields)
- Entry Date (/ED), and Update Date (/UP)

The patent information and other text in a MARPAT record are *not* searchable. To search this information for a MARPAT record, it is necessary to take your MARPAT answer set into CAplus (refer to Chapter 4).

LMARPAT DATABASE

LMARPAT, a learning database, is designed to allow you learn how to search in MARPAT with minimal cost. It is a small, static subset of MARPAT. Use LMARPAT to practice and develop your search skills, so that you can be comfortable and confident in your MARPAT searching.

MARPAT RESOURCES

- In STN use HELP DIRECTORY, HELP COMMANDS, or HELP TIPS
- MARPAT Database Summary Sheet can be found on www.cas.org

CHAPTER 2: DISPLAYING MARPAT RESULTS

Introduction to MARPAT Display

MARPAT displays are more complex than displays in other STN structure search databases. This chapter covers the major aspects of MARPAT answer displays. Further detail is found in the MARPAT Database Summary Sheet, available at www.cas.org.

VARIABILITY IN MARPAT STRUCTURES

Remember that answers to a MARPAT search are document records with Markush structures representing a multitude of specific chemical substances.

To represent this variability, the Markush structure in a MARPAT record uses chemical structure fragments with G-groups (Generic groups) as variable substituents. The G-group definitions are interspersed among the fragments and consist of specific chemical features (NH_2), generic variables (X or Hy), locants (14), or even textual descriptions ("alkoxycarbonyl" or "organic group"). For example:
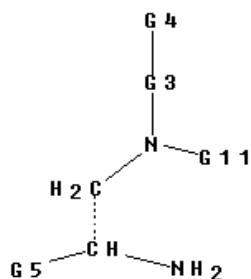

```
G3     = phenylene
G4     = H / F / Cl / Br / I / alkyl <containing 1-4 C> /
         CN / CF3 / OH / alkoxy <containing 1-4 C> / NH2 /
         alkylamino <containing 1-4 C> /
         dialkylamino <each alkyl containing 1-4 C> / CHO /
         alkylcarbonyl <containing 1-4 C> / SH / 29 / 32 / 34
G5     = heterocycle <containing 1-3 heteroatoms,
         1 or more N, zero or more O, zero or more S (no other
         heteroatoms), monocyclic, (0-1) 5-membered,
         (0-1) 6-membered rings only> (opt. substd. by (1-3)
         G8) / 37 / 39 / CHO / alkylcarbonyl <containing 1-4
         C> / CO2H / alkoxycarbonyl <containing 1-4 C> / CONH2
         • • •
```

DISPLAY OPTIONS

Markush structures in patents are often very complex. The resulting display of a Markush structure in MARPAT is often complex as well. The purpose of this chapter is to lessen the intimidation factor of some MARPAT displays.

The default display format (BIB, bibliographic information) does not show any Markush information. Six displays that do show Markush structures are:

COMMAND	DEFINITION
FHIT	The first full Markush structure that matches the query structure and (or) the fields containing hit text terms
FQHIT	Portions of the first Markush structure that match the query structure and (or) fields containing the first query focus hit text terms
FQHITEXG	FQHIT plus the definitions of any unmatched G-groups that are visible in the assembled display
HIT	The full Markush structure that match the query structure and (or) the fields containing hit text terms
QHIT	The portions of each Markush structure that match the query structure and (or) the fields containing hit text terms
QHITEXG	QHIT plus definition of any unmatched G-groups that are in the assembled display
D SCAN	Randomly selected answers showing FQHIT and selected indexing information
ALL	All Markush structures with highlighting and all document record information

SET MPTASSEMBLY

The SET MPTASSEMBLY command allows you to control answer assembly in the QHIT, FQHIT, QHITEXG, and FQHITEXG formats. The options are:

- ON (default)
- OFF
- BOTH

Answer assembly often provides an easier way to interpret the view of the Markush results. The BOTH option will provide an assembled view, as well as the classic MARPAT display for each answer. For example;

```
=> SET MPTASSEMBLY ON
=> D FQHIT

    MSTR 1A Assembled                        cont. on next page
```
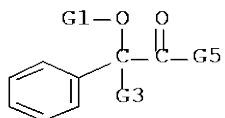
```
    G1—O  O
       |  ||
       C—C—G5
       |
      G3
```

```
Patent location:           claim 1
Note:                      also incorporates claim 10
Note:                      or pharmaceutically acceptable
                           salts, solvates, or prodrugs
Stereochemistry:           or diastereomers or
                           enantiomers or stereoisomers
```

=> **SET MPTASSEMBLY BOTH**
=> **D FQHIT**

    MSTR 1A Assembled
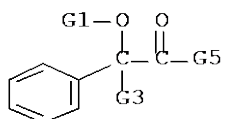
```
    G1—O  O
       |  ||
       C—C—G5
       |
      G3
```

```
Patent location:        claim 1
Note:                   also incorporates claim 10
Note:                   or pharmaceutically
                        acceptable salts, solvates,
                            or prodrugs
Stereochemistry:        or diastereomers or
                            enantiomers or stereoisomers
```
    MSTR 1   *(Classic display)*
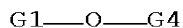
```
 G1——O——G4
```
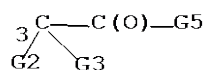
```
G2     = Ph
G4     = 3
```

```
   C——C(O)—G5
  3/
 G2   G3
```

```
Patent location:           claim 1
Note:                      also incorporates claim 10
Note:                   or pharmaceutically
                            acceptable salts, solvates,
                        or prodrugs
Stereochemistry:           or diastereomers or
                        enantiomers or stereoisomers
```

=> **SET MPTASSEMBLY OFF**
=> **D FQHIT**

    MSTR 1

```
 G1——O——G4
```

```
G2     = Ph
G4     = 3
```

```
   3  C——C(O)——G5
    /
  G2     G3
```

• • •
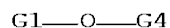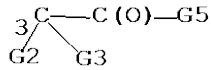
To retain SET MPTASSEMBLY settings for BOTH or OFF beyond the current session you can follow either command with PERM to make it a permanent change under your login ID. For example;

SET MPTASSEMBLY BOTH PERM

-- or --

SET MPTASSEMBLY OFF PERM

To see the current permanent values for your MPTASSEMBLY setting, enter DISPLAY SET at a command prompt.

WHEN SHOULD I CHANGE THE SET MPTASSEMBLY OPTION? – MARPAT HELP TIP 13 (HELP T13)

SET MPTASSEMBLY has three options:

- **ON** (system default for all users) – displays answers in the assembled QHIT/FQHIT format
- **OFF** – QHIT/FQHIT displays in the classic QHIT/FQHIT format which means no assembly will be performed
- **BOTH** – the assembled format and the classic QHIT/FQHIT formats are displayed. You will only be charged once when using the BOTH option.

As with other SET options, this can be set either for the current file session or permanently for your login ID, depending on whether the keyword PERM is used: e.g., SET MPTASSEMBLY BOTH PERM

When MPTASSEMBLY is ON or BOTH, the GTEXT and MARHIGHLIGHT options are considered to be ON for the assembled display. Your GTEXT and MARHIGHLIGHT values are honored for the unassembled QHIT/FQHIT display.

When MPTASSEMBLY is set to OFF, the non-hit G-group definitions available in the QHITEXG and FQHITEXG formats are shown with assembled displays. If MPTASSEMBLY is set to BOTH, an unassembled display will follow. You might want to change SET MPTASSEMBLY to BOTH until you are comfortable using the assembled display format.

STRUCTURE QUERY USED FOR DISPLAY EXAMPLES
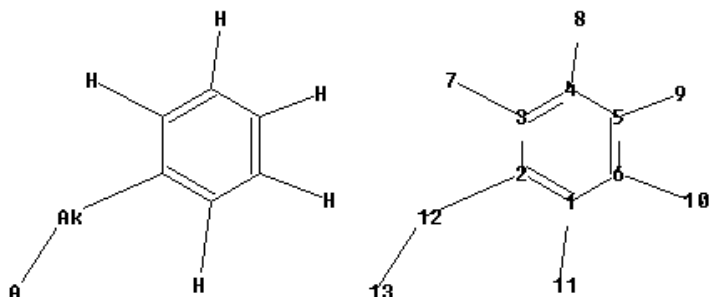
D SCAN DISPLAY

The no-fee D SCAN display shows a randomly selected answer including some bibliographic and indexing data and the FQHIT Markush display. No answer numbers are shown.

```
IPCI  C07D0277-68 [I,A]; A01N0043-78 [I,A]; A01P0007-04
[I,A];A01P0007-02 [I,A];
     A01P0003-00 [I,A]
CC    28-7 (Heterocyclic Compounds (More Than One Hetero Atom))
      Section cross-reference(s): 5
TI    Preparation of 2(3H)-oxobenzothiazol-3-ylacetyl hydrazone
      derivatives as insecticides and fungicides
ST    benzothiazolylacetyl hydrazone prepn insecticide
      Fungicide
IT    Fungicides
      Insecticides
          (preparation of 2-oxobenzothiazol-3-ylacetyl hydrazone
              derivs. as insecticides and fungicides)
● ● ●


MSTR 1 Assembled
```
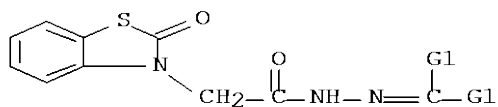


```
G1    = alkyl <containing 1-4 C> /
        Ph (opt. substd. by 1 or more G2)
Patent location:        claim 1
```

HIT DISPLAY FORMAT

The HIT format displays all hit Markush structures with highlighting including hit terms.

```
MSTR 2
```
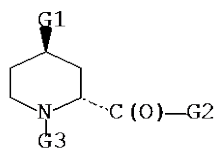


```
G1      = alkyl <containing 1-10 C> /
          any ring <containing zero or more N, zero or more O,
          zero or more S (no other heteroatoms), aromatic, 1-3
          rings, including 5- or 6-membered rings> (opt.
          substd. by G7) /
          (Specifically claimed: Me / Ph (opt. substd. by G7))/
          (Examples: Et / Pr-n / Pr-i / Bu-n / Bu-i / Bu-s /
          Bu-t)
G2      = OH / 12
```
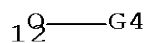


```
G3      = H / R
G4      = alkyl <containing 1-4 C> /
          cycloalkyl <containing 3-6 C> / 14
```
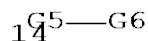


```
G5      = alkylene <containing 1-4 C>
G6      = Ph (opt. substd. by G7)
G7      = alkyl <containing 1-6 C> /
          alkoxy <containing 1-6 C> / halo
Patent location:           claim 6
Note:                      also incorporates claim 7
Note:                      or salts
```

QHIT DISPLAY FORMAT

The QHIT format displays all hit Markush structures showing only query related portions. The example is the same Markush structure as the HIT display above.

```
MSTR 2 Assembled
```
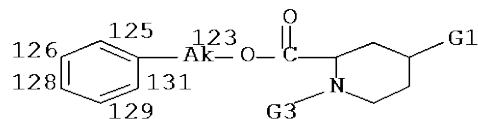

> **QHIT** does not highlight the hit fragments. It only displays them.

```
123: alkylene <containing 1-4 C>
125, 126, 128, 129, 131: opt. substd. by G7
Patent location:           claim 6
Note:                      also incorporates claim 7
Note:                      or salts
```

QHITEXG DISPLAY FORMAT

The QHITEXG format displays the QHIT display plus definition of any unmatched G-groups that are in the assembled display. The example is the same Markush structure as the QHIT display above.
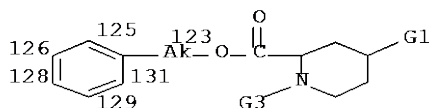
```
 MSTR 2 Assembled

                          O
          125     123     ||       G1
   126  __/\__ __Ak_O_ C_/ \_
   128 |  |   131        | N |
        \__/                \__/
          129       G3

123: alkylene <containing 1-4 C>
125, 126, 128, 129, 131: opt. substd. by G7

Additional displayed G-groups:
G1     = alkyl <containing 1-10 C> /
         any ring <containing zero or more N, zero or more O,
         zero or more S (no other heteroatoms), aromatic, 1-3
         rings, including 5- or 6-membered rings> (opt.
         substd. by G7) /(Specifically claimed: Me / Ph (opt.
         substd. by G7)) /
         (Examples: Et / Pr-n / Pr-i / Bu-n / Bu-i / Bu-s /
         Bu-t)

G3     = H / R

Patent location:           claim 6
Note:                      also incorporates claim 7
Note:                      or salts
```

GENERIC TEXT IN DISPLAYS

Generic Text Used in Displays

MARPAT displays use a number of generic text terms to describe what the patent includes in the Markush structure.

Ak-Based Terms

The Ak-based terms and their definitions are shown.

TERM	DEFINITION
Alkylene	-alkyl- A divalent saturated hydrocarbon radical
Alkylidene	Alkyl= A divalent alkyl radical that is attached to the parent via a double bond or two single bonds from the same carbon, e.g., =CH-CH2-CH3
Alkynyl	Ak with no double bonds, one or more triple bonds, and two or more carbons. An unsaturated monovalent radical chain of two or more carbons, branched or linear, containing one or more carbon-to-carbon triple bonds, but no double bonds. Formed by removal of one hydrogen

	from the corresponding alkyne, e.g., CH≡C-, CH≡C-CH2-
Lower	Any of the "alk" terms can be preceded by the term "lower", which implies a total carbon count of 1 to 6 carbons for alkanes, 2 to 6 carbons for alkenes and alkynes. This term is used only when no carbon count is given in the patent.
Loweralkyl	Ak with all bonds single exact, and 1 to 6 carbons
Perhaloalkyl	Saturated Ak with all hydrogens replaced by halogen atoms

Cb/Hy-Based Terms

The Cb/Hy-based text term descriptions and their definitions are shown below. Cb is any monocyclic or polycyclic group containing all carbon atoms with any bond values between the atoms. Hy is any monocyclic or polycyclic group containing one or more non-carbon atoms with any bond values between the atoms.

TERM	DEFINITION
Aryl	Cb with one or more aromatic rings, six or more normalized bonds, and one or more six-membered rings. A monovalent radical derived from an aromatic monocyclic or polycyclic molecule by the removal of one hydrogen. Included are six-membered carbocyclic rings with six normalized bonds, e.g., benzene and rings containing benzene as part of a larger system.
Arylene	- aryl - A divalent aromatic radical formed by the removal of two hydrogens from two different carbon atoms on the aromatic molecule, e.g., phenylene.
Cycloalkenyl	Cb with one or more double bonds and no triple bonds. An unsaturated monovalent monocyclic or polycyclic radical containing one or more carbon-to-carbon double bonds. Formed by removal of one hydrogen from the corresponding cycloalkene, e.g., cyclopentadienyl or cyclohexenyl.
Cycloalkyl	Cb with all bonds single exact. A saturated monovalent alicyclic radical. Formed by the removal of one hydrogen from the corresponding cycloalkane, e.g., cyclopropyl or decahydronaphthyl.
Heteroaryl	Hy with one or more aromatic rings with six or more normalized bonds and one or more six-membered rings. Hy with one or more aromatic rings with two or more double bonds and one or more five-membered rings. A monovalent radical derived from an aromatic molecule that contains at least one heteroatom. Formed by removal of a hydrogen from the corresponding heteroaromatic molecule, e.g., furyl, pyridyl or benzopyranyl.

Hybrid Terms

Hybrid terms are those that combine multiple functionalities.

TERM	DEFINITION
Acyl	• Carbonyl bonded to Ak (which may have R's on it), bonded at the carbonyl carbon • Carbonyl bonded to R, bonded at the carbonyl carbon • Formyl
Aralkyl	Alkyl bonded to 1 – 3 aryls, bonded at the alkyl
Hydrocarbyl	• Ak • Cb • Ak bonded to Cb, bonded at Ak • Cb bonded to Cb, bonded at one Cb • Cb bonded to Ak, bonded at Cb

SET GTEXT FOR G-GROUP DISPLAY

MARPAT Structure Display Options (SET GTEXT)

There are two display options for G-group text in Markush displays in MARPAT:

- Expanded G-group definitions (default)
- Compact G-group definitions

The default can be changed using the SET GTEXT command. The addition of the PERM keyword will retain the setting for future online sessions.

```
=>  SET GTEXT OFF     (compact, uses codes to define the G-
                       group)
=>  SET GTEXT ON      (expanded text definitions, restores the
                       default display)
```

Example with SET GTEXT ON

```
G6    = H / loweralkyl (opt. substd. by 1 or more halo) /
        halo / CN / (Examples: CF3 / Me)
G7    = loweralkyl / halo / loweralkylthio
G8    = Ph (opt. substd. by 1 or more halo) /
        Ph (substd. by G13) / Ph (substd. by G14) / thienyl / SO2Ph
        / OH
G9    = halo / loweralkyl / (Example: Me)
G10   = H / loweralkyl (opt. substd. by Ph) /
        Ph (opt. substd. by 1 or more halo) / CN / CONH2 / NCS /
        (Example: Me)
G11        = loweralkyl / halo / (Example: Me)
Patent location:          claim 1
```

Example with SET GTEXT OFF

```
G6     = H / loweralkyl (SO (1-) X) / X / CN / (EX CF3 / Me)
G7     = loweralkyl / X / loweralkylthio
G8     = Ph (SO (1-) X) / Ph (SR G13) / Ph (SR G14) /
         thienyl / SO2Ph / OH
G9     = X / loweralkyl / (EX Me)
G10    = H / loweralkyl (SO Ph) / Ph (SO (1-) X) / CN /
         CONH2 / NCS / (EX Me)
G11    = loweralkyl / X / (EX Me)
Patent location:           claim 1
```

The compact form shortcuts are defined later in this chapter.

HELP SET GTEXT

=> **HELP SET GTEXT**

The SET GTEXT command allows you to specify the form of the Generic group (G-group) definitions in the MARPAT File. SET GTEXT ON displays a verbose form of the G-group definitions. SET GTEXT OFF displays a compacted form of the G-group definitions.

Example:

```
SET GTEXT ON
G1     = alkyl <each alkyl containing 1-6 C> /
         (Specifically claimed: Me)
 SET GTEXT OFF
G1     = alkyl<(1-6)> / (SC Me)
```

The SET GTEXT default setting is ON. If you wish to retain a SET GTEXT setting of OFF beyond your current session, enter SET GTEXT OFF PERM.

To see the current and permanent values for your GTEXT setting, enter DISPLAY SET GTEXT at a command prompt.

COMPACT MARKUSH DISPLAY CODES

Codes in G-group Definitions

The following codes are used in the G-group definitions in compact displays (SET GTEXT OFF). In almost all cases, the expanded display form uses an expanded G-group definition rather than the compact code.

CODE	DEFINITION
G# =	Defines the alternatives for the G# or defines the repeating group and the number of times it repeats. *Example*: **G1 = O / S / 16 / NULL** *Translation*: G1 is O or S or node 16 or G1 is a direct bond. Node 16 appears in that fragment of the structure diagram. *Example*: **G3 = (0 – 7) CH2** *Translation*: CH2 repeats 0 to 7 times
CVA	The conditional variable statement, ignored at search time. *Example*: **G1 = H / OH / X** 　　　　　**G2 = NO2 / Me / Et** 　　　　　**CVA = IF G1 = OH THEN G2 = Me** *Translation*: If G1 is the OH alternative, then G2 must be Me and not NO2 or Et. Searched as G1 = H / OH / X 　　　　　　　 G2 = NO2 / Me / Et

NULL in G-groups

NULL is used in the file structure G-group definition to indicate that the G-group might not be present and a bond takes its place. For example, the following file structure:

$G1 = NULL / CH_2$
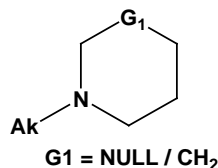
represents two ring systems:
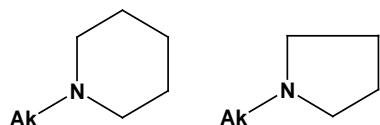
Shared Variables

Sometimes a patent will specify that, for example, R1 and R2 are hydrogen or alkyl or together are trimethylene. Shared variables are used to cover the alternative that combines two R's. The MARPAT file structure would be, for example:
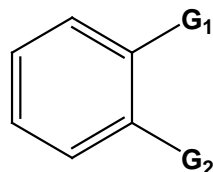
G1 = H / alkyl
G2 = H / alkyl
G1 + G2 = CH2CH2CH2

Node Number Pairs

Sometimes chemical fragments in multivalent G-groups need to be fitted into the base structure (or other fragments) using "node number pairs". In the classic display example below, G2 contains a ring alternative indexed as fragment 190/192 to be inserted with node 190 attaching to base structure node 3, node 192 to node 5. The node number pairs reflect the connectivity and directionality defined in the patent. In node numbers, the first value (190 or 192) represents the constituent fragment; the second value (3 or 5), the point of attachment on the base. This example shows two node number pairs, 190-3 and 192-5. n an analogous fashion, 3, 4, or 5 sets of node number pairs can also occur in MARPAT records.
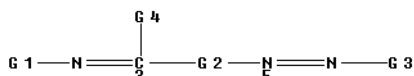
```
MSTR 1

              G 4
              │
 G 1 ─── N ═══ C ─────── G 2 ─── N ═══ N ─── G 3
               3                 5

G1      = heterocycle (opt. substd.)
G2      = 190-3 192-5

         G 2 3
        ╱      ╲
 1 9 0 │        │ 1 9 2
        ╲      ╱
          G 2 5
      G 2 7


G23     = S
G25     = N
Patent location:           claim 1
```
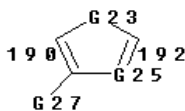
Generic Group Attributes

The generic groups are:

- Ak – any carbon chain
- Cy – any cyclic group
- Hy – any cyclic group with one or more non-carbon atoms
- Cb - any cyclic group with all carbon atoms

The attributes are found in < > following the generic group they modify. The starred attributes in the table are searched as Generic Group Definitions. All others are ignored in searching.

ATTRIBUTE	COMPACT DISPLAY	TRANSLATION
AN (Attachment Nodes)	• AN (3) A • AN (2) N	• Attached through three nodes of any kind • Attached through two N nodes
AR (ARyl)	• AR (1-) • AR (0)	• One or more aromatic rings • Is not aromatic
BD* (BonDs)	• BD (0) T • BD (1-) D (2) SE	• No triple bonds • One or more double bonds and two single exact bonds
CH (CHarge)	• CH (2) + • CH (1) +-	• Total of two positive charges • One positive or negative charge
DC** (Degree of Connectivity)	• DC (0) M3 • DC (1-) M3	• No branching • Branched
EC*** (Element Count)	• EC (1-8) C • EC (2-3) N	• 1 to 8 C atoms • Two or three N atoms
FA**** (Fusion Atoms)	• FA (2-4) C • FA (2-) C	• 2 to 4 C atoms fuse rings • Two or more C atoms fuse rings
RC**** (Ring Count)	• RC (2-) • RC (1)	• Polycyclic with 2 or more rings • Monocyclic
RS (Ring Size)	• RS (2-3) E6 • RS (1) MS (1) X6 and RC (1)	• 2 to 3 six-membered rings • A five- to six-membered ring
TX (TeXt Qualifiers)	• R <TX "protecting group"> • R <TX "residue">	• The patent said "protecting group" • The patent said "residue"

* Searchable as Generic Definition of saturated/unsaturated.
** Searchable as Generic Definition of linear/branched.
*** Searchable as Generic Definition of high carbon/low carbon and high heteroatom/low heteroatom.
**** Searchable as Generic Definition of monocyclic/polycyclic.

Precision Qualifiers – EX and SC

These qualifiers are found in parentheses at the end of the variable definition.

QUALIFIER	DEFINITION
EX	Alternatives are found in the EXamples (in the disclosure). *Example*: G1 = R<TX "leaving group"> / **(EX Cl / Br /I)** *Translation*: Cl, Br, and I are alternatives for G1 and were found in the disclosure.
SC	Alternatives are Specifically Claimed. *Example*: G3 = H / alkyl /alkoxy / **(SC Me / OMe)** *Translation*: Methyl and methoxy are alternatives for G3 and are specifically claimed.

Substitution Qualifiers – SO and SR

These qualifiers are found in parentheses immediately after the group they modify. In the current implementation, SO and SR are both searched as SO. They are searched as if the substituent is present on the group any number of times, including zero.

QUALIFIER	DEFINITION
SO	Alternative is Substituted Optionally by this group. *Example*: G1 = Ph **(SO X)** / alkoxycarbonyl **(SO CO2H)** *Translation*: Phenyl is substituted with 0 to 5 halogens Alkoxycarbonyl is substituted with zero or more carboxy groups.
SR	Alternative has Substitution Required by 1 or more of this group. Searched as Substitution Optional. *Example*: G1 = Ph **(SR X)** / alkoxycarbonyl **(SR CO2H)** *Translation*: Phenyl is substituted with 1 to 5 halogens. Alkoxycarbonyl is substituted with one or more carboxy groups. Searched as 0 to 5 halogens and zero or more carboxys.

Occurrence Counts

Occurrence Counts precede the alternative that they modify. They are used to limit the substitution by specifying how many times an alternative is present. Occurrence counts are not searchable.

COUNT	DEFINITION
(n)	Alternative occurs exactly "n" times. *Example*: G1 = **(1) X** / Me *Translation*: Exactly one of the G1's is X. Searched as G1 = X / Me.
(n-)	Alternative occurs "n" or more times. *Example*: G3 = **(2-)H** / loweralkyl / X *Translation*: Two or more of the G3's are H. Searched as G3 = H / loweralkyl / X.
(-n)	Alternative occurs 0 to "n" times. *Example*: G6 = **(-1) H** / alkyl / CH2Ph *Translation*: 0 or 1 of the G6's are H. Searched as G6 = H / alkyl / CH2Ph.
(m-n)	Alternative occurs "m" to "n" times. *Example*: G2 = X / CF3 / Ph **(SO (1-2) G3)** *Translation*: The phenyl is optionally substituted by one or two of the variable G3's. Searched as the phenyl substituted by 1 to 2 G3's.

CHAPTER 3: MARKUSH STRUCTURE SEARCHING STRATEGIES

PLANNING A SEARCH IN MARPAT

User Guide Focus

The remainder of this user guide focuses on the options and strategies for structure searching in MARPAT. Structure searching for Markush structures utilizes the same structure search skills required for any structure searching in STN. The search process consists of the same two steps, screening and iteration, as in other STN structure databases.

Structure Query Building Experience

This user guide assumes that you are experienced in building STN structure queries using STN Express®, the structure plug-in for STN® on the Web[SM], or the online STRUCTURE command. Only the query building techniques that are important for MARPAT will be discussed in detail.

MARPAT vs. Other Structure Search Files

The nature of the Markush structure search file differs from that of other STN structure search databases in that the structures being searched do NOT represent specific substances but rather Markush structures. There are some special techniques that allow you to exercise more control over the search process in MARPAT. The search examples that follow in this chapter are designed to illustrate the search options and help you make decisions about which options to choose for a given Markush search question.

Search Process

A more detailed discussion of the Markush search process is found in the latter portion of this chapter. It is not necessary to understand the process in depth to obtain very good search results.

Questions When Planning a Search

Before undertaking a structure search in MARPAT, there are two questions you need to consider:

- How comprehensive does your search need to be?

- How precise should your answers be?

The answers to these two questions will affect the type of query that you build and type of search that you run.

How Comprehensive?

The degree of comprehensiveness is related to the degree of precision you desire. In general, no patent substance search can be considered comprehensive if it does not include MARPAT as part of the search strategy. If you require the most

comprehensive search results possible, then you will need to extend your search to all patent databases that permit substance searching by structure and/or name. At a minimum, we suggest the use of CASLINK (refer to Chapter 4).

How Precise?

Techniques for controlling the precision of your answers are discussed in Chapters 5 – 7. You can control the number and type of answers you retrieve by the query structure you create. Specific techniques include:

- Creating a very specific query with no generic character
- Making generic groups more precise using Generic Definitions
- Making generic groups more precise using Element Count
- Controlling the degree of query matching using Match Level

For example, in MARPAT, a pyridine ring in your query may retrieve file structures that match the entire range of specificity implied by the structure, i.e., file structures that are pyridine, or a six-membered N-containing heterocycle, or a heterocycle, or a cyclic organic substituent, or even an organic substituent. All of these representations are in MARPAT file structures, and you can specify the level of matching in your query to control the number and type of answers retrieved.

Search Fields

The primary search index in MARPAT is:

- Markush structure search index

In addition, you can search:
- Basic Index (/BI) – contains single words from the textual information fields
- Accession Number (/AN)
- Entry Date (/ED)
- Update Date (/UP)

For further details, refer to the MARPAT Database Summary Sheet.

MARKUSH STRUCTURE QUERY AND SEARCHING

Structure Queries

MARPAT is a document-based structure-searchable database. In most cases, the structure query used to search in MARPAT can also be used to search in other STN structure databases. The query includes MARPAT-specific features that are ignored when the query is searched in other STN structure databases.

Structure queries can be built by using STN Express, the STN on the Web structure plug-in, or the online STRUCTURE command.

Resources for Query Building

For complete information on how to build structures using STN Express, refer to the *STN Express User Guide* (www.cas.org). You can also use online help messages for assistance. For online information on how to use the STRUCTURE command, enter HELP STRUCTURE at the command prompt.

Query Variability

On STN, structure queries may be specific or allow for generic or variable representations of the search strategy. For example:

- Bond types may be represented specifically or undefined.

- Nodes can be specific elements or represented by variables:
    - A = any atom except hydrogen
    - Q = any atom except carbon or hydrogen (heteroatom)
    - M = any metal atom
    - X = any halogen atom

- Nodes can be generic groups of atoms:
    - Ak = any carbon chain
    - Cy = any ring system
    - Cb = any carbocyclic system
    - Hy = any heterocyclic system (contains at least one non-carbon atom)
    - User-defined generic group (G-group)

- Generic groups can be made more precise by using Generic Definitions:
    - Saturated or unsaturated
    - Number of carbon atoms (less than 7 or 7 or more)
    - Straight chain or branched
    - Number of heteroatoms
    - Monocyclic or polycyclic

Substituents (specific or generic) can be variably attached to ring systems.

RING/CHAIN BOND Not Allowed

All query building features found in other structure databases are allowed in MARPAT *except* the RING/CHAIN Bond Characteristic. If your query contains ring/chain bonds, it cannot be searched in MARPAT and the following error message is received:

```
RC BOND NOT ALLOWED IN A GENERIC STRUCTURE FILE
Searching a Markush structure file requires that all bonds in
a query structure have definite specifications as ring bonds
or chain bonds. The bond type 'RC', representing Ring or
Chain, is not allowed for searching a Markush structure file.
You can re-specify the 'RC' bond in the structure as 'R' or
'C' by using the STRUCTURE command to RECALL this structure
and then using the BOND command at the colon prompt (:). For
more information on using these commands, enter "STRUCTURE" at
```

```
a command prompt (=>) and then at the colon prompt (:) enter
"HELP RECALL" or "HELP BOND".
```

To include both possibilities in your search, two queries and two searches are required.

Assigning Attributes to G-groups

Often, you might need to assign an attribute such as Match Level to a G-group in your query. Attributes cannot be assigned directly to a G-group. However, this can be done indirectly as follows:

STEP	ACTION
1	Define the G-group as separate fragments drawn in the query.
2	Assign the attributes to the separate fragments.
3	Define the G-group in terms of the fragments.

Multi-Component Structures

MARPAT includes multicomponent structures, e.g., salts of acids in the search file. Since your query must be a single L-number, you must build all components in the single L-number. For example:
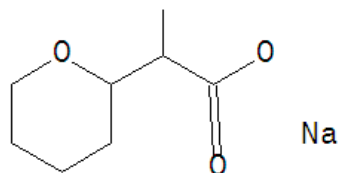
Note: This is a different strategy than a search in REGISTRY.

Query Requirements

The MARPAT query must consist of *only a single structure L-number*. The query cannot include screens (filters) or Boolean operators.

Two-Step Structure Search

As in the other structure databases, structure searches in MARPAT consist of a *screen search* followed by an *iterative search*.

- A screen search is an initial pass through the search index to quickly eliminate all Markush structures that cannot be answers to the query and forward a set of candidate answers to iteration.
- Iteration compares candidate answer structures with the search query and eliminates those that do not match.

Structures that pass both the screen and iterative steps are added to the answer set L-number. An answer record for the document includes the Markush structure(s), bibliographic, abstract, and indexing information. All information in the record can be displayed.

System Limits

All structure searches are subject to system limits. The limits (as of June 2011) are shown below. To see current limits, enter HELP SLIMITS in MARPAT.

SEARCH SCOPE	ITERATIONS	ANSWERS	MINUTES
Online SAMPLE	2,000	50	5
Subset SAMPLE	2,000	50	5
Online FULL	500,000	500,000	30
Subset FULL	500,000	500,000	30
BATCH FULL	500,000	500,000	360
Online RANGE	500,000	500,000	30
Subset RANGE	500,000	500,000	30
BATCH RANGE	500,000	500,000	360

Controlling Search Results

Search results can be controlled in several ways by using the various MARPAT features in the desired combinations.

- Search Types
  - o SSS (less restrictive)
  - o CSS (more restrictive)

- Ring Isolation
  - o Isolated/embedded (less restrictive)
  - o Isolated (more restrictive)

- Match Level
  - ATOM (most restrictive)
  - CLASS (less restrictive)
  - ANY (least restrictive)

- Generic Definitions (more restrictive)

- Element Count
  - Limited (more restrictive)
  - Unlimited (less restrictive)

- SUBSET Searching
  - Structure search L-number
  - Text search L-number

Controlling Hit Displays

In MARPAT, the order of variables defined within a query affects the display of hit terms. For example, G1 defined as X, OH, SH can display different hit terms than G1 defined as OH, SH, X. Variable definition order does not affect the number of answers retrieved, only the display of hit terms. FQHIT or FHIT formats can also be affected.

Structure Search Types

There are two structure search types in MARPAT: substructure (SSS) and closed substructure (CSS). These have the same definition as in other STN structure databases.

- **SSS** (the default) search is the broader of the two search types. It allows the query to retrieve Markush structures that match it exactly and those that have substitution at any available positions.
- **CSS** searches block all positions in the query from further substitution unless otherwise specified. CSS is much more restrictive than SSS. CSS must be specified on the command line or in response to a command type prompt.

EXACT or FAMILY searches are not available in MARPAT because there are no specific structures in MARPAT. Every file structure is a variable Markush representing two or more specific substances.

Structure Search Scopes

Like the other STN structure databases, MARPAT has four search scopes:

- **SAMPLE** – (default; a fixed sample percentage of the file) SAMPLE searches incur no search fees and are used to test whether your query will run to completion in the full file and to get sample answers. SAMPLE searches should always be run to see if the search gives acceptable results and will complete within system limits.
- **FULL** – (100% of the file)
- **RANGE** – (user-defined portion of the file) RANGE in this document-based database is usually specified using CA volume ranges or years.
- **SUBSET** – (subset of the file as defined by a previous search) SUBSET allows you to run a structure query against a portion of MARPAT defined by an answer

set previously created by a MARPAT search or an L-number from the CAplus database.

SAMPLE SEARCH

A typical SAMPLE search result is shown below:

```
=> S L1

SAMPLE SEARCH INITIATED 13:32:14
SAMPLE SCREEN SEARCH COMPLETED -          1605 TO ITERATE
100.0% PROCESSED   1605 ITERATIONS               50 ANSWERS
INCOMPLETE SEARCH (SYSTEM LIMIT EXCEEDED)
SEARCH TIME: 00.00.03
FULL FILE PROJECTIONS:          ONLINE  **COMPLETE**
                         BATCH **COMPLETE**
PROJECTED ITERATIONS:         29840 TO          34360
PROJECTED ANSWERS:                   1232 TO     2368


L2          50 SEA SSS SAM L1
```

SAMPLE Search Results

SAMPLE searches often exceed system limits (2000 iterations and/or 50 answers). *This is not important*. Review "SAMPLE search results" and the "Full File Projections" to determine if the search will run to completion (COMPLETE) and that the answers retrieved meet your expectations.

RANGE SEARCH

The primary reasons for a range search are either to limit the results by time or to bring the search within system limits. If L1 were the search query, then enter:

```
=> S L1 RAN=(2000,2005)
```

Enter HELP RANGE at the command prompt in MARPAT for information and examples of how to enter ranges.

Range Search Costs

The cost of a range search is usually the same as the cost of a full file search. Searches in MARPAT may be restricted to a range of CA Collective Periods, CA Volumes, CA publications years, or Accession Numbers.

SUBSET SEARCH

A SUBSET search can refine an answer set created by a structure or text search in MARPAT or by bringing the results of a CAplus search over to MARPAT. This capability allows you to do a complex text search in CAplus, carry the results into MARPAT and use a SUBSET search to find the desired results.

For example, if L1 is your MARPAT query, and L3 is an answer set in CAplus that reflects your desired search strategy, then you enter MARPAT and type:

```
=> S L3
```
L4 (an answer set in MARPAT of your CAplus records[1])

```
=> S L1 SUB=L4
```

The subset search may be either SAMPLE or FULL, depending on whether you need to see example answers to test the query prior to running a FULL search.

In the above example, the subset information is entered with the search command in the "expert" mode (using S or SEA). In "novice" mode (using SEARCH), the SUBSET keyword would be supplied in answer to the prompt for the search type.

```
=> S L3
L4

=> SEARCH L1
ENTER TYPE OF SEARCH (SSS) OR CSS:SSS
ENTER SCOPE OF SEARCH (SAMPLE), FULL, RANGE, OR
SUBSET:SUBSET
ENTER SUBSET L# OR (END) :L4

● ● ●
```

[1] Note that only those CAplus answers in L3 that have corresponding records in MARPAT will be crossed over in the search of L3 and included in L4.

EXTEND SEARCH

The EXTEND feature creates an L-number answer set from the candidate answers found in the screening step of a structure search in addition to the usual answer set following the iteration step. The feature is activated by entering EXTEND on the search command line or using the SET EXTEND ON command. EXTEND is not active with SAMPLE searches.

EXTEND is useful when a search provides few or no results. Note that many, if not most, of the candidate answers might not be relevant for the search.

EXTEND is Free

There is no fee associated with the EXTEND feature. However, you may decide to use it after having completed a full structure search that resulted in few answers. For this reason, you can repeat your search with SET EXTEND ON for a much reduced fee. To take advantage of the reduced fee second search, do the following:

STEP	ACTION
1	Complete a Markush structure search.
2	Enter SET EXTEND ON or add the EXTEND keyword on the search command line.

STEP	ACTION
3	Use the answer set L-number from the initial search as the search query L-number for the second search.

Iteration Incomplete

MARPAT search results can include Markush structures for which the iteration process is not completed in the time allowed. The candidate is passed to the answer set L-number and marked "ITERATION INCOMPLETE." The occurrence of incomplete iterations is reported as the search proceeds and in the final search report.

Separating Iteration Incompletes

The Markush structures for which iteration was not completed need to be examined manually to determine if they meet the search requirements. To create an L-number with only records containing those structures that did not complete the iteration process, search the L-number answer set with /INCOMPLETE (/INC) appended. For example (L6 is the answer set):

```
=> S L6/INC
```

To restrict the answer set to those Markush structures with completed iteration, search the L-number answer set with /COMPLETE (/COM) appended. For example (L6 is the answer set):

```
=> S L6/COM
```

STRUCTURE SEARCH EXAMPLES

Example 1: SAMPLE Search

Find patents on substances with the following Markush structure:
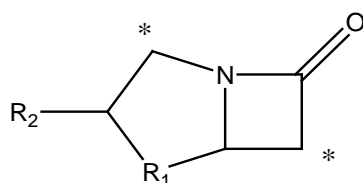
R1 = any atom
R2 = O or S or N
R3 = C(O)O at either
position marked by an *

Special Options with SEARCH Command

When you use the SEARCH command in MARPAT, you get prompts that give you options to change the attributes of Match Level (MLEVEL) and Element Count Level (ECLEVEL) and to isolate all rings in the structure. These changes are for the current search only and are not saved with the query.

These prompts are not provided when you use the S or SEA forms of the search command. Changing Match Level and/or Element Count Level allows you to

broaden generic structure searching in MARPAT. The options for assigning Match Level and Element Count Level and their implications for answer retrieval are discussed in Chapters 6 and 7, respectively.

Enter MARPAT, Upload, and DISPLAY the Query
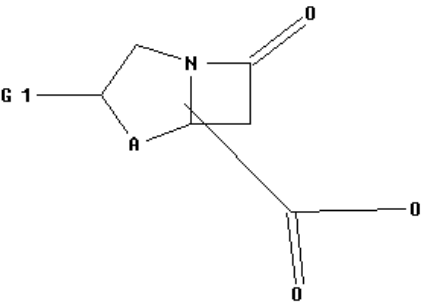
Always display an uploaded query prior to search as a final verification of the structure.

```
=> FILE MARPAT
=>
Uploading ...
=>
L3    STRUCTURE UPLOADED

=> D L3
L3 HAS NO ANSWERS
L3          STR
```



```
G 1  O , S , N
```

Conduct the SAMPLE Search

The command defaults are shown in parentheses with each prompt.

```
=> SEARCH L3

ENTER TYPE OF SEARCH (SSS) OR CSS: SSS
ENTER SCOPE OF SEARCH (SAMPLE), FULL, RANGE, OR SUBSET: SAMPLE
CHANGE MLEVELFOR ANY NODE?  Y/(N)/?: N
CHANGE ECLEVEL FOR ALL NODES AND GROUPS? (LIMITED)/UNLIMITED/N/?: N
ISOLATE ALL RINDS IN THE STRUCTURE? (Y)/N/?: N
SAMPLE SEARCH INITIATED 09:32:56
SAMPLE SCREEN SEARCH COMPLETED -          633 TO ITERATE
100.0% PROCESSED             633 ITERATIONS
ANSWERS
SEARCH TIME: 00.00.01


FULL FILE PROJECTIONS:   ONLINE              **COMPLETE**
                         BATCH        **COMPLETE**
PROJECTED ITERATIONS:                   11172 TO 14148
PROJECTED ANSWERS:                          5 TO 234


L4              5 SEA SSS SAM L3
```

Search Statistics

Notice that, while MARPAT is a document-based database, the structure search statistics are similar to those of other STN structure-based databases. However, there is an important difference. The number to iterate (e.g., 633 TO ITERATE) represents the number of documents that contain potential hit structures. The number of answers (e.g., 5 ANSWERS) represents the number of documents that contain at least one structure that matched the query.

Displaying Results

Use D SCAN to review the answers. This is a free display option for evaluating search results. Answers are displayed randomly with no answer numbers given. D SCAN displays the classification, title, and indexing fields (IC, CC, TI, ST, IT) of the patent and the first query-focused, i.e., partial Markush "hit" structure (FQHIT display format). You can customize D SCAN by requesting one or more fields comprising the full D SCAN, e.g., D SCAN TI, or D SCAN TI FQHIT.

```
=> D SCAN


L2    5 ANSWERS    MARPAT   COPYRIGHT 2011 ACS on STN
INCL 540304000
IPCI C07D0499-00 [ICM,4]; A61K0031-425 [ICS,4]

NCL    540/304.000; 540/313.000; 540/360.000
CC     26-5 (Biomolecules and Their Synthetic Analogs)
       Section cross-reference(s): 1
TI     Process for the synthesis and formulation of penems
       and penams as antibacterials
ST     penam prepn antibacterial; penem prepn
       antibacterial; antibacterial penam penem prepn
IT     81779-73-9
       RL:RCT (Reactant); RACT (Reactant or reagent)
          (alkylation by, of azetidinone deriv.)


● ● ●

MSTR 1 Assembled
```
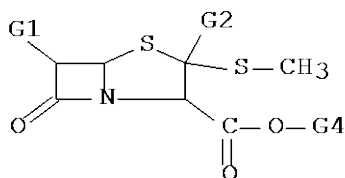


```
Patent location:                       claims
Note:                          record may include structures
                               from disclosure
```

Conducting the FULL Search

When you are satisfied with the results of the SAMPLE search, run a FULL file search. If you do not wish to be prompted for information, use the S or SEA version of the command. Since SSS is the default search scope, you need to only enter the L-number and FULL.

FULL Search (SSS)

```
=> S L3 FULL


FULL SEARCH INITIATED   09:33:18
FULL SCREEN SEARCH COMPLETED – 13025 TO ITERATE

100.0% PROCESSED   13025 ITERATIONS 48 ANSWERS
SEARCH TIME:  00.00.04

L4           48 SEA SSS FUL L3
```

Example 2: FULL Search

Retrieve MARPAT records relating to the structure shown. Isolate those patents most closely related to the structure.

G1 = O, S, N

```
=> FILE MARPAT
=>
Uploading ...
=>
L1    STRUCTURE UPLOADED

●  ●  ●

=> S L1 SSS FULL
FULL SEARCH INITIATED  09:59:18
FULL SCREEN SEARCH COMPLETED -  47235 TO ITERATE

100.0% PROCESSED        47235 ITERATIONS  5045 ANSWERS
SEARCH TIME:  00.00.35

L2          5045  SEA SSS FUL L1
```
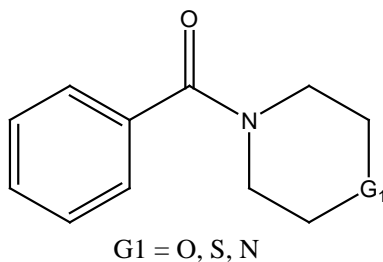
Closed Substructure Search (CSS)

To conduct a closed substructure search in MARPAT, enter CSS with the SEARCH (or S) command. A closed substructure search blocks all available positions in the query from further substitution, unless you have specified some type of substitution for that position in your query.

A CSS search can be used when the SSS search will not run to completion within system limits. A CSS search results in fewer more precise answers than are obtained with an SSS search.

SUBSET CSS Search

One way to refine an answer set for the most closely related answers is to use a CSS SUBSET search on the initial answer set. Novice commands are shown below:

```
=> SEARCH

ENTER LOGIC EXPRESSION, QUERY NAME, OR (END):L1
ENTER TYPE OF SEARCH (SSS) OR CSS:CSS
ENTER SCOPE OF SEARCH (SAMPLE), FULL, RANGE, OR

      SUBSET:SUBSET
 ENTER SUBSET L# OR (END):L2
 ENTER SUBSET SEARCH SCOPE - SAMPLE, FULL, RANGE, OR
      (END):FULL
CHANGE MLEVEL FOR ANY NODES? Y/(N)/?:N
CHANGE ECLEVEL FOR ALL NODES AND GROUPS?
(LIMITED)/UNLIMITED/N/?:N
ISOLATE ALL RINGS IN THE STRUCTURE? (Y)/N/?:N
FULL SUBSET SEARCH INITIATED 10:08:38
FULL SUBSET SCREEN SEARCH COMPLETED -   4943 TO ITERATE
100.0% PROCESSED    4943 ITERATIONS      280 ANSWERS
SEARCH TIME: 00.00.03
L3          280 SEA SUB=L2 CSS FUL L1
```
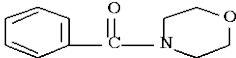
Free D SCAN FQHIT Display

Display a hit Markush structure at no cost using D SCAN FQHIT. FQHIT is the abbreviated display of the first hit Markush structure.

```
=> D SCAN FQHIT
L3   280 ANSWERS   MARPAT   COPYRIGHT 2011 ACS on STN

MSTR 9 Assembled
```



```
Patent Location:        claim 5

HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1):0
=>
```

Example 3: Refining Structure Search Results

Bibliographic data, abstracts, and indexing are available in MARPAT for display only. To refine structure search results by bibliographic or subject terms, crossover MARPAT answer sets into CAplus. In the following example, L3 is an answer set from MARPAT.

STEP	ACTION
1	Enter CAplus.
2	Search the MARPAT L-number answer set in CAplus.
3	Refine the CAplus l-number answer set with desired bibliographic and/or subject search terms.

Note: The limit for l-number crossover between MARPAT and CAPLUS is 100,000 records.

Crossover to CAplus

```
=> FILE CAPLUS

=> S L3
L4 280 L3

=> S L4 AND (ANTI INFLAM? OR ANTITUMO? OR ANTIPARK or
ANTI TUMO? OR ANTI PARK?) AND PY>=2007
        ● ● ●
L5   26 L4 AND (ANTI INFLAM? OR ANTITUMO? OR ANTIPARK OR
     ANTI TUMO? OR ANTI PARK?) AND PY>=2007
```

Beyond MARPAT

From CAplus, you can carry your results into other STN databases by using standard STN patent search techniques.

FOR MORE COMPREHENSIVE SEARCHING

For comprehensive results, you typically need to search your structure query not only in MARPAT, but also in REGISTRY/CAplus.

- MARPAT contains Markush structures representing the chemical substances covered by the patent.
- MARPAT contains Markush structures from patents published since 1961.
- REGISTRY/CAplus covers the specific substances that are prepared or claimed by the patent.
- REGISTRY/CAplus covers the exemplified substances from English language basic patents (coverage dates can be found at www.cas.org)
- REGISTRY/CAplus includes specific substances indexed from patents and other references bask to the early 1900s.
- REGISTRY answers can be linked to the patent literature in CAplus.

The Markush structures cover specific substances that were not actually prepared or specifically claimed, referred to as generic substances and structures that contain generic terms, e.g., aryl, alkyl, heterocycle.

Combining Generic and Specific Search Results

MARPAT permits searching for generic substances in patent documents. REGISTRY and CAplus permit searching for specific substances indexed from patent documents. Chapter 4 discusses automatic tools (CASLINK or HCASLINK) to do both and combine the results in a single set of unique answers.

Merging Search Results from Multiple Databases

You can conduct separate structure searches in MARPAT and in REGISTRY, followed by crossover of the REGISTRY answer set into CAplus and then merge the results from MARPAT and CAplus by using the DUPLICATE command.

Alternatively, you can conveniently conduct the same searches in one step using the CASLINK feature. Chapter 4 shows how you can use CASLINK (or its companion HCASLINK) for one-step comprehensive structure searching of MARPAT, REGISTRY, and CAplus (or HCAplus).

REPEATING GROUPS

In search file structures, there are repeating groups, just as you may have repeating groups in your structure query. The repeating group definition in the file structure is expanded into all possible structures in the search file. If the query were…
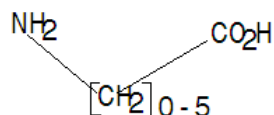
… then answers such as the following would be retrieved.
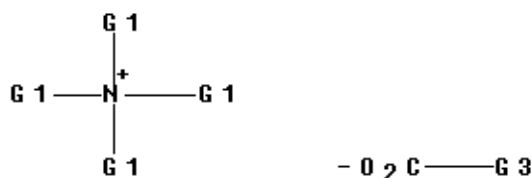

```
G3    = carbon chain <containing 1-18 C,
        0 or more double bonds, 0 or more triple bonds>
        (opt. substd. by 1 or more G4)
G4          = NH2 (opt. substd.)
Patent location:  claim 1
Note:  additional ring formation also claimed

L4 280 L3
```

COORDINATION COMPOUNDS WITH A RING BOND BETWEEN A METAL AND LIGAND

Special techniques are required for searching Markush coordination structures in which one or more of the metal-ligand bonds are ring bonds. A special structure query must be used to obtain the desired results in MARPAT. These special queries should not be used in REGISTRY, and therefore should not be used in CASLINK.

Note: Coordination queries with no metal-ligand ring bonds are drawn in the usual manner and can be used in both MARPAT and REGISTRY.
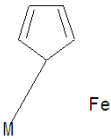
Coordination Query

To create a coordination structure query where a metal-ligand bond is a ring bond, use the following procedure:

STEP	ACTION
1	Draw the coordination structure.
2	Use the variable "M" node for the metal atom(s).
3	Draw the specific metal(s) as a separate fragment(s). Leave the Node Characteristic as Chain.
4	Assign the "M" node(s) as Match Level CLASS.
5	Assign the separate fragment [specific metal(s)] as Match Level ATOM.

Example 4: Coordination Search (SAMPLE)

The M node is Match level CLASS and the M-ligand bond is RING. The Fe atom is Match level ATOM and the Node Characteristic is CHAIN.

```
L1      STRUCTURE UPLOADED
=> D L1
L1 HAS NO ANSWERS
L1                  STR




        Fe
 M

Structure attributes must be viewed using STN Express query preparation.

=> S L1 SSS SAM
SAMPLE SEARCH INITIATED 12:36:39
SAMPLE SCREEN SEARCH COMPLETED -     853 TO ITERATE
100.0% PROCESSED      853 ITERATIONS      50 ANSWERS
FULL FILE PROJECTIONS:            ONLINE   **COMPLETE**
                        BATCH    **COMPLETE**
 PROJECTED ITERATIONS:          15343 TO   18777
 PROJECTED ANSWERS:              1063 TO    2137
L2            50 SEA SSS SAM L1
```

Coordination Search (FULL)

```
=> S L1 SSS FULL
FULL SEARCH INITIATED 12:36:57
FULL SCREEN SEARCH COMPLETED – 17253 TO ITERATE
100.0% PROCESSED    17253 ITERATIONS      1284 ANSWERS
SEARCH TIME: 00.00.03


L3        1284 SEA SSS FUL L1
```
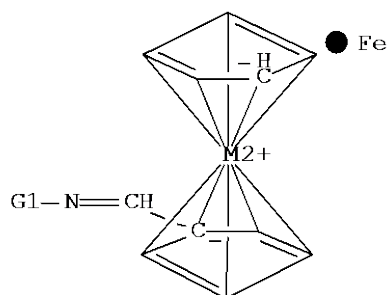
Coordination Search Answer Display

```
=> D 1 FHIT


L3    ANSWER 5 OF 1284  MARPAT  COPYRIGHT 2011 ACS on STN
TI    Ferrocene Schiff base derivatives, preparation method and
      application thereof
```

MSTR 1 Assembled


```
Additional displayed G-groups:
G1    = Ph (opt. substd. by G3) /
      naphthyl (opt. substd. by G3) /
      thiazolyl (opt. substd. by G3) /
      benzothiazolyl (opt. substd. by G3)

Patent location:          claim 1
```

Coordination Compounds

It should be noted that a generic coordination search query would omit the separate fragment for specific metals.

MORE DETAILED SEARCH INFORMATION

Useful (Not Essential) Information

This portion of Chapter 3 explains the MARPAT search process in more detail. An understanding of this detail is not necessary for successful MARPAT searching, but is provided to allow you to have a better vision of what your search is doing.

Specificity in Markush Structures

To gain a deeper understanding of the MARPAT search process, it is necessary to discuss the various degrees of specificity used to represent chemical substances. A single chemical substance, such as pyridine, is represented in an STN structure database such as REGISTRY as a specific set of atoms and bonds. A Markush or generic structure represents a set of specific substances. The number of substances the set includes may vary from a few to many hundreds of thousands or even millions. For example:
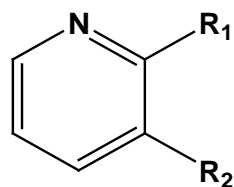
$$R_1 = H \ / \ Cl$$

$$R_2 = CO_2H \ / \ CN \ / \ X \ / \ Me$$

X represents any halogen. This Markush structure is quite specific. Many are much more generic. Sometimes, the options represented in the Markush structure cannot be defined with a structure at all and can only be described in textual terms, e.g., "electron withdrawing group" or "blocking group".

Levels of Specificity

Every structure, whether specific or generic, implies some level of specificity. We will use pyridine as our starting point and go stepwise from the specific to the generic.

Most specific	Pyridine
	Six-membered N-containing heterocycle
	Heterocycle
	Cyclic organic substituent
Most Generic	Organic substituent

In MARPAT, all levels of specificity and all structure characteristics are found in the structures and are allowed in the query structure. Assigning Match Level lets you control the degree to which your query matches the specificity of a particular candidate answer in the search file.

G-group Specificity

It is important to be able to visualize both your query and file structures in terms of real atoms and the generic groups they represent. If you do not realize that the pyridine set at match level CLASS in your query might retrieve a Hy in the file

structure, it becomes difficult to recognize the query structure in your retrieved answers.

The MARPAT system is able to recognize both the specific (ATOM) level of your query as well as the generic (CLASS) level of your query. For example, if your query included an ethyl (CH2CH3) substituent, the system would recognize the ethyl group, but it would also recognize a saturated, linear Ak with two carbon atoms.

Spinoff Generic Groups

The generic representation of the query has two types of generic groups: those actually contained in the query and those "spun off" from real atoms in the query. In the table, an asterisk (*) is appended to the spun off generic nodes to differentiate them from the generic nodes that were actually in the query.
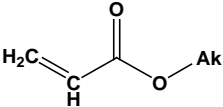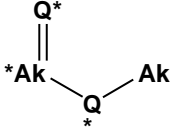
Spin off generic groups cannot be displayed.
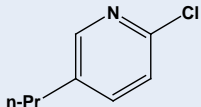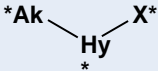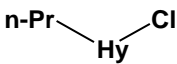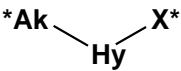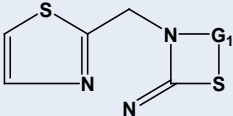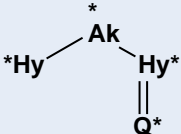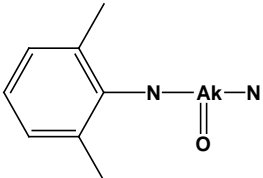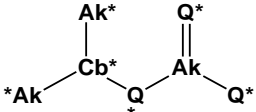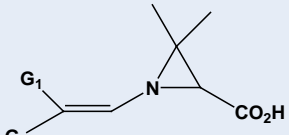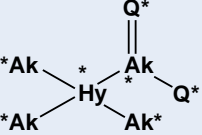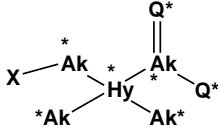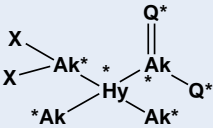
Spinoff Nodes in Searches

Not only does the system recognize the various levels of specificity but it also knows which generic groups are "spinoffs", derived from real atoms, and which were part of the original structure. *This is very important!* Query spinoff generic groups are *never* matched to structure file spinoff generic groups. This prevents false matches. For example, pyridine has a spinoff generic group Hy*. Thiophene also has a spinoff generic group Hy*. Pyridine never matches thiophene even though both generate Hy* spinoff nodes.

Query		
Generic Representation		
Hit		
No Hit		
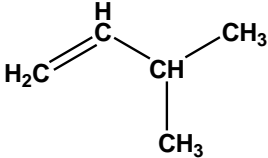
Generic Representations

In the table, nodes in the generic representation that have an asterisk (*) indicate that the generic representation is spun off from real query atoms.

QUERY STRUCTURE	GENERIC REPRESENTATION
$H_2C=CH-C(=O)-O-Ak$	Q*, *Ak, Q*, Ak
pyridine with CI, N, n-Pr	*Ak, Hy*, X*
n-Pr–Hy–CI	*Ak, Hy, X*
thiazole-CH2-N ring, G1, S, N; G1 = 2-3 C	*Ak, *Hy, Hy*, Q*
dimethylphenyl–N–Ak–N, O	Ak*, Q*, Cb*, Ak, *Ak, Q, Q*
G1, G2 ring with N, CO2H; G1 = C/X G2 = C/X	Q*, *Ak, Hy, Ak, Q*, *Ak, Ak*; when G1 =G2 = C
	Q*, X–Ak, Hy, Ak, Q*, *Ak, Ak*; when G1 = X G2 = C
	Q*, X, X–Ak*, Hy, Ak, Q*, *Ak, Ak*; when G1 = G2 = X

Spinoff Generic Groups

Real atoms not only give rise to spinoff generic groups but to a set of appropriate generic definitions for the spinoff groups. The exact definition may vary with the search type.

REAL ATOMS	SSS	CSS
	• Unsaturated • Branched	• Six or fewer • Unsaturated • Branched

Ring Isolation

Whether a ring is isolated or isolated/embedded determines the type of spinoff nodes the ring generates:

- Isolated carbocycles generate Cb* spinoff generic nodes.
- Isolated/embedded carbocycles generate Cy* spinoff generic nodes for SSS searches, but Cb* for CSS searches.
- Heterocycles generate Hy* spinoff generic nodes in all cases.

Benefits of Spinoff Nodes

Spinoff node matching allows MARPAT to retrieve answers that might otherwise be missed.

SEARCH PROCESS SUMMARY

To summarize:

- Real atoms in the query:
    - Match real atoms in the search file
    - Match original generic nodes in the search file
- Original generic nodes in the query:
    - Match real atoms in the search file
    - Match original generic nodes in the search file
    - Match spinoff generic nodes in the search file
- Spinoff generic nodes in the query:
    - Match *only* original generic nodes in the search file

An in-depth understanding of the search process is not necessary to use MARPAT effectively, but it can aid you making your search more precise.

CHAPTER 4: CASLINK COMPREHENSIVE SUBSTANCE PATENT SEARCH

Complete Patent Substance Searching

A comprehensive substance search of the patent literature covered in *Chemical Abstracts* must include not only a search of the generic substances in MARPAT but also a specific substance search in REGISTRY and CAplus. A single set of unique answers can then be created from the search results by using the DUPLICATE REMOVE command.

CASLINK OVERVIEW

CASLINK is an automatic one-step structure search tool that:

- Searches the query in MARPAT
- Searches the query in REGISTRY
- Searches the REGISTRY answer set in CAplus
- Merges the results into a single answer set of unique answers
- Provides a reduced price for the MARPAT search
- Allows for further refinement with text terms in both CAplus and MARPAT

HCASLINK uses HCAplus instead of CAplus. CAplus and HCAplus differ only in how they are priced. Enter HELP COST at a command prompt for both databases to see the different pricing.

Search Discount

Using CASLINK automatically provides a reduced fee for the MARPAT structure search. Enter HELP COST at the command prompt in either database.

You can carry out the same steps manually, as well and still get the reduced MARPAT search fee if the MARPAT search uses the REGISTRY L-number answer set as the query.

Excluded from CASLINK

Certain types of structure searches cannot be used with CASLINK (or HCASLINK) because different structure queries must be used in REGISTRY and MARPAT. These special cases (discussed in Chapter 3) do not get a reduced fee MARPAT search because they cannot use the REGISTRY L-number answer set as their query.

STARTING CASLINK

To initiate CASLINK, enter:

```
=> FILE CASLINK
```

(Substitute HCASLINK to use HCAplus rather than CAplus.)

Example 1: CASLINK

Find patents for specific or generic substances containing the following structure:
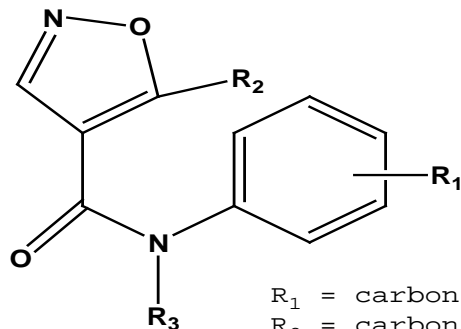

```
R₁ = carbon chain with 3 halogens
R₂ = carbon chain
R₃ = any atom
```

CASLINK Cluster

Enter FILE CASLINK to enter all the databases in the cluster: REGISTRY, MARPAT, and CAplus.

As a final verification, display your uploaded query structure before beginning the search process.

```
=> FILE CASLINK

L1       STRUCTURE UPLOADED

=> D L1
L1 HAS NO ANSWERS
L1              STR
```
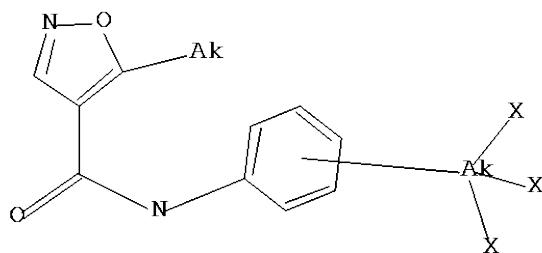


```
Structure attributes must be viewed using STN Express query
preparation.
```

Sample Search in CASLINK

The sample search is first conducted automatically in REGISTRY. An L-number answer set (L2) is created for answers from the sample portion of REGISTRY. MARPAT is then automatically entered. The sample search is conducted in MARPAT and full file projections are given. An L-number answer set (L3) is created for answers from the sample search of MARPAT.

```
=> S L1

S L1 SSS SAM FILE=REGISTRY
```

```
FILE 'REGISTRY'
SAMPLE SEARCH INITIATED 14:51:18
SAMPLE SCREEN SEARCH COMPLETED -    846 TO ITERATE

100.0% PROCESSED     846 ITERATIONS       12 ANSWERS
SEARCH TIME: 00.00.01

FULL FILE PROJECTIONS:        ONLINE  **COMPLETE**
                       BATCH   **COMPLETE**
PROJECTED ITERATIONS:        15175 TO     18665
PROJECTED ANSWERS:              33 TO       447

L2              12 SEA SSS SAM L1

S L2 SSS SAM FILE=MARPAT
FILE 'MARPAT'
SAMPLE SEARCH INITIATED 14:51:19
SAMPLE SCREEN SEARCH COMPLETED -    512 TO ITERATE

100.0% PROCESSED     512 ITERATIONS        4 ANSWERS
SEARCH TIME: 00.00.01

FULL FILE PROJECTIONS:        ONLINE  **COMPLETE**
                       BATCH   **COMPLETE**
PROJECTED ITERATIONS:         8897 TO     11583
PROJECTED ANSWERS:               4 TO       200

L3               4 SEA SSS SAM L1
```

The full-file projections for both databases indicate full searches will run to completion. Next examine the answers to confirm the validity of the query.

S or SEA Command

In the example above, the S shortcut search command is used along with the L-number for the structure query. The default SSS SAMPLE search is assumed. The sample search results in two answer sets (L2 from REGISTRY and L3 from MARPAT).

D SCAN to See Search Results

DISPLAY SCAN is the no-fee display option for CASLINK that allows you to examine the results of your sample search. Display an answer from REGISTRY for evaluation. After evaluating the sample answers, proceed with the full-file search.

```
=> D SCAN L2

L2    12 ANSWERS   REGISTRY  COPYRIGHT 2011 ACS on STN
IN    4-Isoxazolecarboxamide, N-[4-chloro-3-
      (trifluoromethyl)phenyl]-5-ethyl-3-methyl-
MF    C14 H12 Cl F3 N2 O2
```
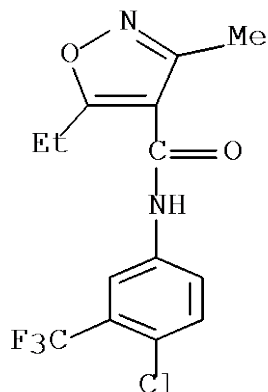
```
**PROPERTY DATA AVAILABLE IN THE 'PROP' FORMAT**

HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1):0

=> D SCAN TI FQHIT L3

L3    4 ANSWERS   MARPAT  COPYRIGHT 2011 ACS on STN
TI    Synthesis of isoxazoles and usage as anti HPV agent
```

MSTR 1 Assembled
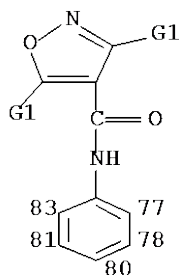

```
77, 78, 80, 81, 83: opt. substd. by (1-3) CF3
G1    = alkyl <containing 1-8 C>
      (opt. substd. by 1 or more G10)
Patent location:          claim 1
Note:                     substitution is restricted

HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1):0
```

FULL-FILE SEARCH IN CASLINK

The next step is to conduct a FULL substructure search, i.e., to search the entirety of REGISTRY, MARPAT, and CAplus. In the SEARCH command, enter the L-number for the structure query and the keyword FULL.

```
=> S L1 FULL
```

When you use the "S" command, the structure is searched as it was created. There are no prompts to override the structure search defaults.

```
=> S L1 FULL

S L1 SSS FUL FILE=REGISTRY
FILE 'REGISTRY'
FULL SEARCH INITIATED  11:02:50
FULL SCREEN SEARCH COMPLETED – 17091 TO ITERATE
L4     194 SEA SSS FUL L1

S L4 SSS FUL FILE=MARPAT
FILE 'MARPAT'
FULL SEARCH INITIATED  11:02:51
FULL SCREEN SEARCH COMPLETED – 10655 TO ITERATE
L5      74 SEA SSS FUL L1

S L4 FILE=CAPLUS
L6     1625 FILE CAPLUS
DUP REM L5 L6
PROCESSING COMPLETED FOR L5
PROCESSING COMPLETED FOR L6
L7     1670 DUP REM L6 L5 (29 DUPLICATES REMOVED)
               ANSWERS '1-74' FROM FILE MARPAT
               ANSWERS '75-1670 FROM FILE CAPLUS
```

Review of Steps in CASLINK

The following steps are automatically performed in CASLINK when you request a substructure search (SSS) in the FULL files:

STEP	ACTION
1	Search the full REGISTRY database (answer set L4).
2	Search the full MARPAT database (answer set L5).
3	Search for references in CAplus (L6) citing the substances found in REGISTRY.
4	Remove duplicate CAplus records.

The final answer set (L7) is a set of unique documents containing the substances or Markush structures with the specified structure fragment. Answers are automatically sorted in the following database order: MARPAT, CAplus. If you prefer to keep the CAplus records and remove duplicates from MARPAT, re-enter DUP REM L6 L5 at a command prompt.

Added Value of MARPAT

Note that only 29 answers in MARPAT were also retrieved by the REGISTRY/CAplus search. MARPAT found 45 answers that might have been missed without CASLINK.

Refining Structure Search Results in CASLINK

You can refine the structure search results in CASLINK by combining the final answer set (L7) with other terms such as bibliographic or subject terms. In this example, search results are limited by patent documents and by publication years of 2005 to the present. The searches are automatically conducted in appropriate databases, and duplicates are removed.

```
=> S L7 AND PATENT/DT AND PY>=2005


S L6 AND PATENT/DT AND PY>=2005 FILE=CAPLUS
L8          546 FILE CAPLUS


S L5 AND PATENT/DT AND PY>=2005 FILE=CAPLUS
L9           46 FILE CAPLUS


S L9 AND L5 FILE=MARPAT
L10          46 FILE MARPAT


DUP REM L10 L8
PROCESSING COMPLETED FOR L10
PROCESSING COMPLETED FOR L8
L11          579 DUP REM L10 L8 (13 DUPLICATES REMOVED)
                 ANSWERS '1-46' FROM FILE MARPAT
                 ANSWERS '47-579' FROM FILE CAPLUS
```

Displaying Answers in CASLINK

Answers can be displayed from any of the L-number answer sets created. When displaying results from a combined answer set (L14), use a format that fits the answer selected for display. Here, we choose IBIB QHIT for a MARPAT answer and IBIB HITSTR for a CAplus answer.

```
=> D 8 IBIB QHITEXG;D 50 IBIB HITSTR


L11  ANSWER 8 OF 579  MARPAT  COPYRIGHT 2011 ACS on STN
DUPLICATE 9
ACCESSION NUMBER:       123:218417  MARPAT
TITLE:                  Method of treating hyperproliferative
                        vascular disease with
                        fluorophenylisoxazolecarbooxamide and
                        flurorophenylcyanocrotonamide
                        derivatives
INVENTOR(S):            Bartlett, Robert Ryder; Morris,
                        Randall E.
PATENT ASSIGNEE(S):     Hoechst A.-G., Germany
SOURCE:                 Eur. Pat. Appl., 12 pp.
                        CODEN: EPXXDW
DOCUMENT TYPE:          Patent
LANGUAGE:               English
FAMILY ACC. NUM. COUNT:  1
PATENT INFORMATION:


PATENT NO.   KIND  DATE           APPLICATION NO.  DATE
-----------  ----  --------       ---------------  --------
```

```
EP 665013     A1    19950802      EP 1995-100194   19950109
EP 665013     B1    20011024
US 5519042    A     19960521      US 1994-181116   19940113
AT 207355     T     20011115      AT 1995-100194   19950109
PT 665013     E     20020328      PT 1995-100194   19950109
```

● ● ●

MSTR 1 Assembled
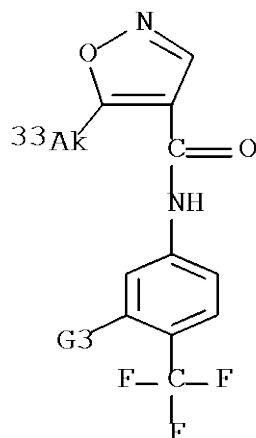
33: alkynyl <containing 2-6 C, 1-2 triple bonds>

Additional displayed G-groups:
G3   = alkyl <containing 1-4 C> / H /
          (Specifically claimed: Me)

Patent location: claim 1

```
L11  ANSWER 50 OF 579  CAPLUS  COPYRIGHT 2011 ACS on STN
ACCESSION NUMBER:       2011:596168   CAPLUS
DOCUMENT NUMBER:        154:599077
TITLE:                  New therapeutic approaches for
                        treating Alzheimer's
                        disease
INVENTOR(S):            Cohen, Daniel; Chumakov, Ilya;
                        Nabirochkin, Serguei
PATENT ASSIGNEE(S):     Pharnext, Fr.
SOURCE:                 PCT Int. Appl., 110pp.
                        CODEN: PIXXD2
DOCUMENT TYPE:          Patent
LANGUAGE:               English
FAMILY ACC. NUM. COUNT: 2
PATENT INFORMATION:

PATENT NO.      KIND   DATE        APPLICATION NO.    DATE
-----------     ----   --------    ---------------    --------
WO 2011054759   A1     20110512    WO 2010-EP66510    20101029
EP 2322163      A1     20110518    EP 2009-306048     20091103

PRIORITY APPLN. INFO.:       EP 2009-306048     A  20091103
IT    75706-12-6, Leflunomide
```
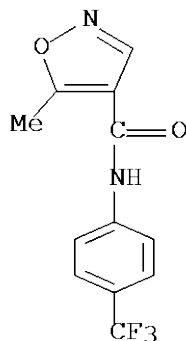
```
      RL: PAC (Pharmacological activity); THU (Therapeutic
      use); BIOL (Biological study); USES (Uses)
            (combination therapy of Alzheimer's disease in
            sustained-release formulations
RN   75706-12-6  CAPLUS
CN   4-Isoxazolecarboxamide, 5-methyl-N-[4-
            (trifluoromethyl)phenyl]-  (CA INDEX NAME)
```


CLOSED SUBSTRUCTURE SEARCH IN CASLINK

To conduct a closed substructure search in CASLINK, simply enter the following with the SEARCH command: the L-number for the structure, CSS for closed substructure search, and SAM or FULL for the scope of the search. The same sequence of steps is automatically performed as in substructure searches.

```
=> S L1 CSS FUL
```

Remember that, since a closed (CSS) search suppresses substitution unless otherwise required, the search usually retrieves fewer and more precise answers than open (SSS) substructure searches. For example, the previous search as a CSS instead of an SSS would result in the following:

```
L15        1614 DUP REM L13 L14 (19 DUPLICATES REMOVED)
               ANSWERS '1-24' FROM FILE MARPAT
               ANSWERS '25-1614' FROM FILE CAPLUS
```

In this case, while only 30 fewer answers were obtained overall, there were less than half as many MARPAT answers as were retrieved from the SSS search.

TYPE OF CASLINK SUBSET SEARCH

There are two types of SUBSET searches in CASLINK, depending on the origin of the subset L-number.

- Using a subset formed by a text search in CASLINK or an L-number brought in from another database environment
- Using a subset created by a prior CASLINK structure search

Text Subset L-Number

To conduct a subset search in CASLINK by using an L-number created by a text search or an L-number crossed from another database, follow these steps:

STEP	ACTION
1	Create an L-number for the subset answer set by conducting a search in any of the databases in CASLINK.
2	Enter the SEARCH command, query L-number, and FUL followed by SUB= and the L-number answer set from the searched in CASLINK. The system automatically selects the appropriate answer set for the subset definition, performs appropriate subset searches, and eliminates duplicates to create a set of unique answers.

Subset Search Using Text Terms

Find references relating to anticancer substances with this substructure:
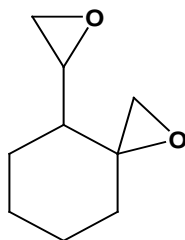
CASLINK SUBSET Search

```
=> FILE CASLINK
FILES 'CAPLUS, MARPAT, REGISTRY' ENTERED AT …
3 FILES IN THE FILE LIST     CLUSTER 'CASLINK' ENTERED


=> S NEOPLASM INHIBITOR OR ANTICANCER? OR ANTI CANCER?


S NEOPLASM INHIBITOR OR ANTICANCER? OR ANTI CANCER?
FILE=CAPLUS
L1      171504 FILE CAPLUS



=>
Uploading ...
L2      STRUCTURE UPLOADED
=> D L2
L2 HAS NO ANSWERS
L2              STR
```
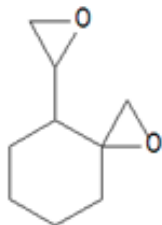
Structure attributes must be viewed
using STN Express query preparation.

```
=> S L2 SUB=L1 FULL

S L1 AND MARPAT/OS FILE=CAPLUS
L8          9046 FILE CAPLUS

S L1 AND NONINDEXED/FS FILE=CAPLUS
L9          9133 FILE CAPLUS

S L8 FILE=MARPAT
L10          9046 FILE MARPAT

S L2 SSS FUL FILE=REGISTRY
FILE 'REGISTRY'
FULL SEARCH INITIATED 16:52:43
FULL SCREEN SEARCH COMPLETED -     19639 TO ITERATE

100.0% PROCESSED     19639 ITERATIONS
928 ANSWERS
SEARCH TIME: 00.00.01

L11          928 SEA SSS FUL L2

S L11 SUBSET=L10 SSS FUL FILE=MARPAT
FILE 'MARPAT'
FULL SUBSET SEARCH INITIATED 16:52:44
FULL SUBSET SCREEN SEARCH COMPLETED -       57 TO ITERATE

100.0% PROCESSED       57 ITERATIONS
11 ANSWERS
SEARCH TIME: 00.00.01

L12           11 SEA SUB=L10 SSS FUL L2

S L11 AND L1 FILE=CAPLUS
L13          220 FILE CAPLUS

DUP REM L12 L13
PROCESSING COMPLETED FOR L12
PROCESSING COMPLETED FOR L13
L14          219 DUP REM L12 L13 (12 DUPLICATES REMOVED)
                  ANSWERS '1-11' FROM FILE MARPAT
                  ANSWERS '12-219' FROM FILE CAPLUS
```

Displaying SUBSET Search Results

The Markush structures in the MARPAT answers (1-11) match the substructure L1, and one or more of the specific compounds indexed in the CAplus answers (12-219) also match on the query substructure L1.
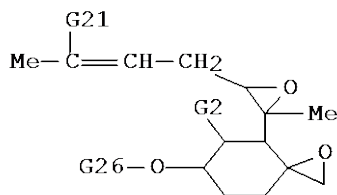
```
=> D L14 TI PI FQHIT


L14   ANSWER 1 OF 219  MARPAT  COPYRIGHT 2011 ACS on STN
DUPLICATE 1
TI    Preparation of 5-demethoxyfumagillol and derivatives for
      the treatment of cancer and other diseases

PATENT NO.    KIND  DATE            APPLICATION NO.  DATE
----------    ----  --------        ---------------  --------
WO 2009117902 A1    20091001        WO 2009-CN328    20090327
```

MSTR 1 Assembled


```
Patent location:             claim 1
Note:                        also incorporates claim 3,
                        formulas II and III;
                        claim 7, formula IV; claim 8,
                        formula V; claim 19,
                        formula VIII ; claim 22, formula
                        IX ; claim 29,
                        formulas XII and XIII
Stereochemistry:        or stereoisomers

=> D 12 TI HITSTR


L14   ANSWER 12 OF 219  CAPLUS  COPYRIGHT 2011 ACS on STN
DUPLICATE 4
TI    Pharmacokinetics and metabolism of O-(chloroacetyl-
      carbamoyl) fumagillol(TNP-470, AGM-1470) in rhesus
      monkeys
IT    129298-91-5, TNP-470   129298-91-5D, TNP-470,
      Metabolites
      RL: BPR (Biological process); BSU (Biological study,
      unclassified); BIOL
      (Biological study); PROC (Process)
      (pharmacokinetics and metab. of O-(chloroacetyl-
      carbamoyl) fumagillol(TNP-470, AGM-1470) in rhesus monkeys)
RN    129298-91-5  CAPLUS
CN    Carbamic acid, N-(2-chloroacetyl)-,
      (3R,4S,5S,6R)-5-methoxy-4-[(2R,3R)-2-methyl-3-(3-methyl-
      2-buten-1-yl)-2-oxiranyl]-1-oxaspiro[2.5]oct-6-yl ester
      (CA INDEX NAME)
```

```
Absolute stereochemistry.
```


• • •
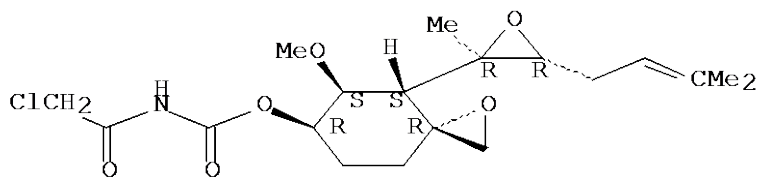
Using a Prior CASLINK Search L-Number

To conduct a subset search in CASLINK by using an L-number created by a prior CASLINK structure search, follow these steps:

STEP	ACTION
1	Carry out a CASLINK structure search.
2	Use the final L-number answer set from the search as the subset search file for the structure search with a new query.
3	If necessary, find the difference between the two CASLINK searches.

Example 2: CASLINK Structure Subset Search

A client wants patents with Markush structures matching the structure on the left. The client does not want patents with Markush structures that match the structure on the right. Retrieve relevant records.

R_1 = Any ring system

R_2 = any monocyclic heterocycle with 2 N
The phenyl ring is isolated
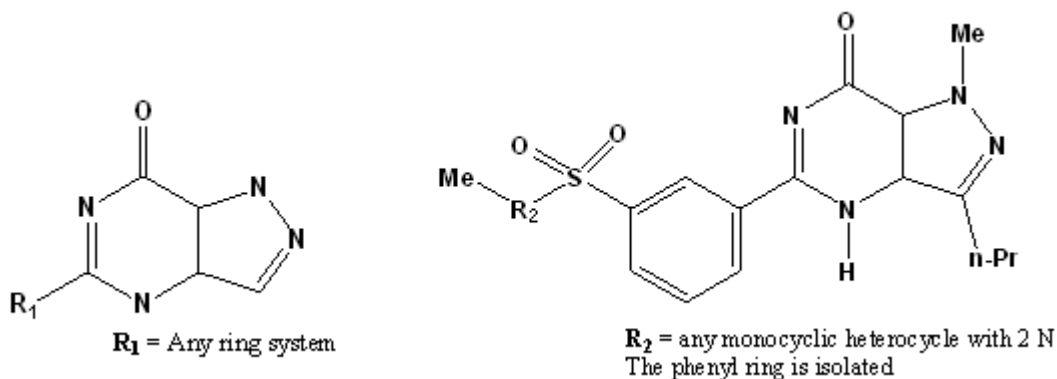
The SAMPLE searches have been omitted from the example.

Exact/Normalized Bonds

By default, the fusion bond between the five- and six-membered rings in both queries is assigned "exact/normalized", so answers with a normalized bond at the fusion will be retrieved. The example used the default assignment. To prevent such answers being retrieved, the query must have an "exact only" bond at the ring fusion.
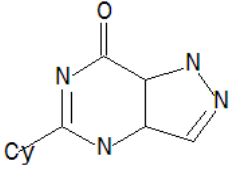
Prior Structure Search SUBSET

```
=> FILE CASLINK
FILES 'CAPLUS, MARPAT, REGISTRY' ENTERED AT 14:43:18

3 FILES IN THE FILE LIST CLUSTER 'CASLINK' ENTERED

Predefined command sequences will be executed in
REGISTRY, MARPAT, and CAPLUS.
=>
Uploading ...
L1      STRUCTURE UPLOADED

=> D L1
L1 HAS NO ANSWERS
L1              STR
```



```
Structure attributes must be viewed using STN Express query
preparation.
=> S L1 FULL

S L1 SSS FUL FILE=REGISTRY
FILE 'REGISTRY'
FULL SEARCH INITIATED 14:44:57
FULL SCREEN SEARCH COMPLETED -    54310 TO ITERATE

100.0% PROCESSED    54310 ITERATIONS
1809 ANSWERS
SEARCH TIME: 00.00.01

L2          1809 SEA SSS FUL L1

S L2 SSS FUL FILE=MARPAT
FILE 'MARPAT'
FULL SEARCH INITIATED 14:44:58
FULL SCREEN SEARCH COMPLETED -     5168 TO ITERATE

100.0% PROCESSED     5168 ITERATIONS              106 ANSWERS
SEARCH TIME: 00.00.03

L3          106 SEA SSS FUL L1
```

```
S L2 FILE=CAPLUS
L4           3307 FILE CAPLUS



DUP REM L3 L4
PROCESSING COMPLETED FOR L3
PROCESSING COMPLETED FOR L4
L5           3323 DUP REM L3 L4 (90 DUPLICATES REMOVED)
             ANSWERS '1-106' FROM FILE MARPAT
             ANSWERS '107-3323' FROM FILE CAPLUS

=>
Uploading ...
L6       STRUCTURE UPLOADED

=> D L6
L6 HAS NO ANSWERS
L6               STR
```

The subset query (L6) is uploaded.
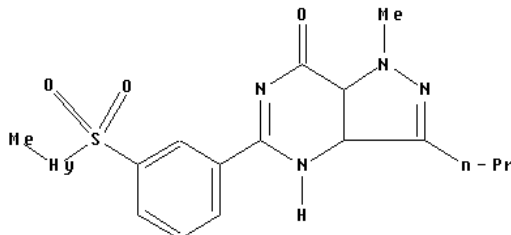

```
Structure attributes must be viewed using STN Express query
preparation.

=> S L6 FUL SUB=L3
```

The subset L-number chosen (L3) is the MARPAT answer set. The DUPLICATE REMOVE answer set (L5) could also have been used.

```
S L6 SSS FUL FILE=REGISTRY
FILE 'REGISTRY'
FULL SEARCH INITIATED 14:47:59
FULL SCREEN SEARCH COMPLETED -     1797 TO ITERATE

100.0% PROCESSED     1797 ITERATIONS
190 ANSWERS
SEARCH TIME: 00.00.01
L7           190 SEA SSS FUL L6

S L7 SUBSET=L3 SSS FUL FILE=MARPAT
FILE 'MARPAT'
FULL SUBSET SEARCH INITIATED 14:48:00
FULL SUBSET SCREEN SEARCH COMPLETED -     77 TO ITERATE

100.0% PROCESSED     77 ITERATIONS          42 ANSWERS
SEARCH TIME: 00.00.01

L8           42 SEA SUB=L3 SSS FUL L6

S L7 FILE=CAPLUS
L9         3094 FILE CAPLUS
DUP REM L8 L9
PROCESSING COMPLETED FOR L8
```

```
PROCESSING COMPLETED FOR L9
L10       3098 DUP REM L8 L9 (38 DUPLICATES REMOVED)
                ANSWERS '1-42' FROM FILE MARPAT
                ANSWERS '43-3098' FROM FILE CAPLUS

=> S L5 NOT L10

S L4 NOT L9 FILE=CAPLUS
L11        213 FILE CAPLUS

S L3 NOT L8 FILE=MARPAT
L12         64 FILE MARPAT

DUP REM L12 L11
PROCESSING COMPLETED FOR L12
PROCESSING COMPLETED FOR L11
L13        234 DUP REM L12 L11 (43 DUPLICATES REMOVED)
                ANSWERS '1-64' FROM FILE MARPAT
                ANSWERS '65-234' FROM FILE CAPLUS
```

> Exclude the subset answers from the prior CASLINK answer set to see those not retrieved by the subset query.

Display an Answer

Use the D SCAN display to see one answer picked at random.

```
=> D SCAN

L13   234 ANSWERS   MARPAT   COPYRIGHT 2011 ACS on STN
IPCI  C07D0487-04 [I,A]; A61K0031-519 [I,A]; A61P0043-00 [I,A]
IPCR  C07D0487-04 [I,A]
CC    28-16 (Heterocyclic Compounds (More Than One Hetero
      Atom))
      Section cross-reference(s): 1, 63
TI    Preparation of pyrazolopyrimidone-containing phenyl
      guanidine derivatives, their medical composition, and
      application in treating male erectile  dysfunction
ST    guanidine phenyl pyrazolopyrimidone prepn PDE5 inhibition
      sexual disorder impotence
      IT   Asthma
              (allergic; prepn. of pyrazolopyrimidone-contg. Ph
              guanidine derivs., their medical compn., and
              application in treating male erectile dysfunction)
IT    Antiarteriosclerotics
      (antiatherosclerotics; prepn. of pyrazolopyrimidone-
      contg. Ph guanidine derivs., their medical compn., and
      application in treating male erectile dysfunction)
IT    Immediate hypersensitivity
              (atopic asthma; prepn. of pyrazolopyrimidone-contg. Ph
              guanidine derivs., their medical compn., and
              application in treating male erectile dysfunction)

• • •

MSTR 1 Assembled
```
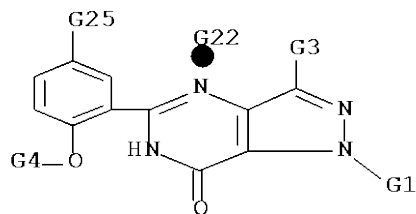
```
Patent location:              claim 1
Note:                         or pharmaceutically acceptable
                              salts, or solvates
Stereochemistry:              or stereoisomers

HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1):0
```

CASLINK SUMMARY

CASLINK offers an easy and efficient method for obtaining comprehensive Markush and specific patent substance search results.

CASLINK can be used with almost any structure query without regard for the special features that will strengthen the power of your substance search query (metal-containing queries constitute an exception, see Chapter 3). These features (discussed in the following chapters) can also be applied to your CASLINK searches to obtain even better results.

CHAPTER 5: GENERIC GROUPS AND GENERIC DEFINITIONS

GENERIC GROUPS IN QUERIES AND ANSWERS

There are two types of generic groups (G-groups) related to Markush structure searching: those in the structure query and those in the retrieved answers.

- Query G-groups
    - System-defined variables: Cy, Cb, Hy, and Ak
    - User-defined variables: G1, G2 etc.

- Answer G-groups
    - G1, G2, etc., defining variables contained in the answer Markush structure
    - Can also contain Cy, Cb, Hy, and Ak

The query generic groups can be system-defined variables and/or user-created variables. The generic groups in the retrieved Markush structures are defined within the answer structure including the Cy, Cb, Hy, and Ak variables.

Attributes Further Define Generic Groups

Generic groups or text descriptions for the generic groups in Markush structures may have attributes that give further definition to the group. The generic attributes follow the generic node description and are set off by **< >**.

A Generic Group Definition:

```
G6      = heterocycle <containing 1 or more heteroatoms,
          zero or more N, zero or more O, zero or more S
          (no other heteroatoms), mono- or
          polycyclic>
```

## GENERIC GROUP DEFINITIONS

The generic groups in the file structures are classified in terms of the following definitions (applicable generic groups shown in parentheses):

- **Saturation (applies to Ak, Cy, Cb, Hy)** – A generic group is *saturated* when all of the bonds between the atoms of the group are single exact bonds. The generic group is *unsaturated* when there is at least one double exact, triple, or normalized bond in the group. G-groups

- **Type of chain (applies to Ak only)** – A generic group is *linear* when all of the atoms within the group have only one or two attachments to other non-hydrogen atoms in the group. A generic group is *branched* when at least one atom within the group has more than two attachments to other atoms in the group.

- **Number of heteroatoms (applies to Cy, Hy only)** – A heteroatom is defined as a non-carbon, non-hydrogen atom. Generic groups are indicated as having *one* heteroatom or having *two or more* heteroatoms.

- **Type of ring system (applies to Cy, Cb, Hy only)** – A ring is *monocyclic* when only one bond must be broken to change the ring to a chain. For *polycyclic* rings, two or more bonds must be broken to change the ring system to a chain.
- **Number of carbon atoms (Ak, Cy, Cb, Hy)** – Generic groups are indicated as having *fewer than 7 carbon atoms* or *7 or more carbon atoms.*

# GENERIC GROUP ATTRIBUTES

Attributes for file structure generic groups that appear in the diagram portion of the structure, rather than in the text portion, display in the Generic Group Attributes field. In the following example, nodes 63, 64, and 65 in the structure diagram are Ak's. The attributes indicate that 63 and 64 contains from 1 to 20 carbons. The Ak at 65 contains 7 to 21 carbons. If no attribute code is displayed within the < >, then any number of carbons is implied for the Ak's.
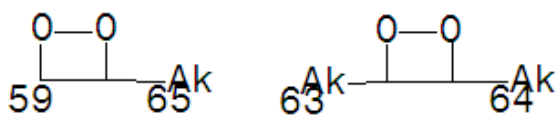


```
Generic group attributes:    63 64 <containing 1-20 C>
Generic group attributes:    65 <containing 7-21 C>
```

Adding attribute information about the generic nodes (Ak, Cb, Cy, and Hy) to your query structure increases precision while preserving recall.

In STN Express, the default settings for the generic nodes in your query are **ANY** (the options are shown below). To change attributes:

- Select the generic node in your structure using the Select Tool or right-click the node.
- Choose **QueryDef > Generic Definition**
- Select the **Generic Definition** option, i.e., **Carbon Atoms, Number of Hetero Atoms,** and **Type of Ring System**.

Choose from available attributes based on the generic group selected.

In this case, a Cy, Cb, or Hy was selected, so the "Type of Chain" option is grayed out.

Impact of Generic Group Definitions

The following example shows the effect of adding generic group attributes to your query. The following query was searched in MARPAT with the attributes for the Ak and Hy set to match level ANY and the benzene ring isolated.
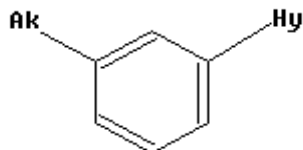
The search for this simple query retrieved over 92,000 documents. The query was restricted by using generic definitions as follows:

- Ak (saturated, linear, and 7 or more C atoms)
- Hy (saturated, monocyclic, and less than 7 C atoms)

The restricted query retrieved 30,000 fewer patent records.

GENERIC GROUP SUMMARY

Adding Generic Group Definitions can be a powerful tool in restricting the type of answers your Markush query can retrieve to those that are desired (Generic Group Definitions also affect your search results in other STN structure databases). Remember, the more definitions you add, the more precise the results.\

CHAPTER 6: MATCH LEVEL

MARKUSH STRUCTURES IN THE SEARCH FILE

Markush structures in the search file consist of:

- Real atoms, e.g., C, O, S, and Fe
- Generic nodes representing 1 or more atoms, e.g., Ak, Cy, X, and M
- R-nodes (text), e.g., "blocking group", "electron-withdrawing group"

The degree to which your structure query matches the Markush structures in the search file is controlled by the Match Level settings used in the query.

WHAT IS MATCH LEVEL?

Structure queries in MARPAT can include generic nodes, i.e., Cy, Cb, Hy, and Ak, as well as specific atoms. Search file structures can also contain both generic nodes, specific atoms, and text descriptions. Match Level specifies the degree of matching between the search query and the Markush structures in the search file.

Three Match Levels

The three Match Levels are:

- ATOM
- CLASS
- ANY

Match Level assignments are powerful retrieval tools. Assigning non-default Match Levels can change both the number and the types of answers retrieved by a structure query.

Default Match Levels

A default Match Level is assigned to your query when you build it.

For queries drawn using STN Express or STN on the Web, the default is **Match Level ATOM** for ring nodes and **Match Level CLASS** for chain nodes, excluding Cb, Cy and Hy.

For queries drawn using the STRUCTURE command, the default is **Match Level ATOM** for all nodes.

You can assign non-default values to specific nodes, rings, or the entire query in order to alter the degree of matching of query nodes to the generic nodes or specific atoms as well as the R nodes in search file structures.

Match Level ATOM

Match Level ATOM is the most restrictive and precise match level. It specifies that real atoms in your query match ONLY real atoms in the search file Markush structures and that generic groups in your query match ONLY real atoms in the answers, and cannot match generic groups.
For example:

- A pyridine ring matches ONLY pyridine, and not any hetero ring Hy.
- Generic groups in your structure match ONLY atom-based groups in the answers, and NOT generic groups.
- Ak in your structure matches ONLY specific carbon chains, e.g., CH3, and cannot match a group defined only as Ak.
- Heteroatoms in your query structure match ONLY the equivalent real atoms in the answers.
- Cl in your structure matches only real Cl, and not X.

M in your query matches ONLY the specific metals, e.g., Fe, and cannot match M.

Match Level CLASS

Broader matching is available with **Match Level CLASS**. Using CLASS, the specific atoms or groups in your structure match not only those specific atoms or groups, but also generic groups in the file structure.
For example:

- Chain carbons match Ak's in the answers.
- Chain atoms also match generic text shortcuts such as acyl, alkyl, alkenyl, alkynyl, alkylidene, aralkyl, alkylidene, and hydrocarbyl in the answers.
- Heteroatoms match Q's, M's, or X's in answers.

Query rings match Hy, Cb, and Cy, as well as text generic shortcuts such as aryl, aralkyl, cycloalkyl, cycloalkenyl, heteroaryl, and hydrocarbyl.
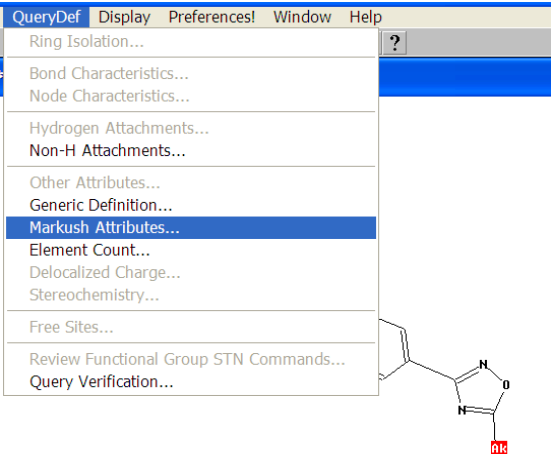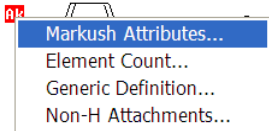
ATOM and CLASS Match Levels
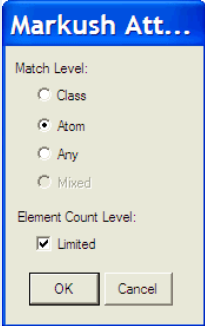
QUERY ATOM OR GROUP	ANSWERS WITH MATCH LEVEL ATOM	ANSWERS WITH MATCH LEVEL CLASS
Cl	Cl	Cl, X
M	Specific metals, e.g., Fe	Specific metals and M
Pyridine	Pyridine	Pyridine, Hy
Ak	Specific carbon chains, e.g., CH3, CH2Ch3, etc.	Specific carbon chains, Ak, or alkyl variables

Match Level ANY

The **Match Level ANY** is the broadest match level available. ANY allows nodes in the query structure to match text R nodes in the file structure as well as specific atoms and generic nodes. Nodes that are searched as CLASS and ATOM match levels do not match R nodes. The R node is found only in file structures. R is not a valid node value in query structures.

Specifying Match Level

You can specify a Match Level other than the default for a query or any selected atom(s) thereof. To change a Match Level assignment, follow these steps:

STEP	ACTION
1	Select the atom(s) for which the Match Level is to be changed. Choose **QueryDef > Markush Attributes** Or right-click the node for which the Match Level is to be changed and select **Markush Attributes**.
2	Select the desired Match Level and click **OK**.

SET MATCH LEVEL

The **SET MATCH LEVEL** feature only affects queries created using the online STRUCTURE command. The options are ATOM (default) or CLASS. Whichever setting is in use when the query is created will be assigned to that query. Queries drawn offline in STN Express or STN on the Web are unaffected by the setting.

Match Level for a G-group

Match Level is an attribute and cannot be assigned to a G-group in a query. To assign a non-default match level to a G-group, first build the G-group as separate atoms or fragments. Then assign the desired Match Level to the nodes in the G-group fragments.

Match Level for a Ring

To avoid unpredictable results, all the nodes in a ring system should be set at the same match level. For example, if you want to change the match level for a ring system from the default level (ATOM) to CLASS, change the level for all the ring system atoms.

Impact of Match Level on Search Results

The table below illustrates the degree to which Match Level can affect your search results. The results shown should not be taken as typical.

MATCH LEVEL ASSIGNED	ANSWERS RETRIEVED
Match Level CLASS for chain nodes Match Level ATOM for ring nodes **(default)**	27,130
Match Level ATOM for all nodes	16,552
Match Level ATOM for chain nodes Match Level CLASS for ring nodes	19,056
Match Level CLASS for all nodes	31,323

R-NODES IN MARPAT

The symbol R is used in MARPAT for groups that cannot be drawn and are described in the patent, e.g., "organic group" or "electron withdrawing group." If a phrase is used in the patent to describe the R-group, then it appears in a text description after the R in the Markush search file structure. The text is in quotes and is set off by < >. *This text is not searchable;* it is there to reflect what the patent says.

```
G4 +G9 = R<"group to form ring">
```

Information from Examples and Other Claims

Other structural information from examples or from other claims is given as an alternative in the G-group and marked as Examples (or EX in the compact display form):

```
G4 = R / (Examples: halo / CO2H / CN / NO2)
```

Or as Specifically Claimed (or SC):

```
G10 = R<"anion"> / (Specifically claimed:
      tetrafluoroborate / hexafluorophosphate /
      45 / hexafluoroantimonate)
```

The terms following "Specifically claimed" or "Examples" are fully searchable in a structure query just as other values defined in a G-group. They are NOT text searchable.

For example, this query with the side chain **C** searched at Match Level CLASS would retrieve the following file structure with an R-node because the side chain **C** would match the alkyl in the **G$_1$** definition.
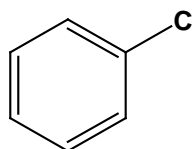
Query: **Retrieved:**
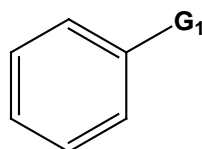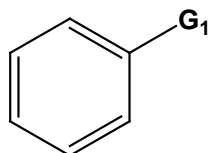

G1 = R / (Specifically claimed: halo /
    alkyl <containing 1-5 C >)

However, the same query searched at Match Level CLASS would not retrieve the following file structure:

G1 = R / (Examples: halo / NO2)

To retrieve this file structure, the side chain on the query must be assigned as Match Level ANY.

R-Nodes that Do Not Display

There are also R nodes in file structures that do not display in the diagram or G-group definitions. If a patent specifies that a group is optionally substituted or has substitution required; that group can be substituted with any number of R's, including none. If the patent states that substitution is optional or substitution is required but did not specify the substituents, "R" is used but not displayed.
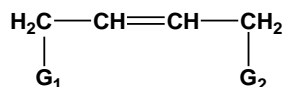
Acyl Includes R-Nodes

The definition of the text shortcut acyl includes R nodes. Acyl is defined as:
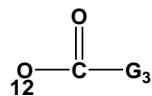

n = (0 or more)
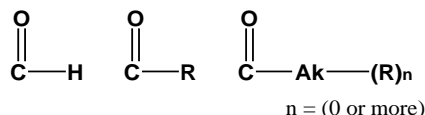
If the Markush structure says "acyloxy" and also lists groups such as benzoyloxy or arylcarbonyloxy as examples of the acyloxy group, then these specifics are also given as alternatives for the G-group.

For example:


```
G1     = alkoxy <containing 1-20 C> (opt. substd. by 1
         or more G4) / aryloxy <containing 6-10 C>
         (opt. substd. by 1 or more G4) / heteroaryloxy
         <containing 4-12 atoms> (opt. substd. by 1 or
         more G4) / acyloxy / 12

G2     = OH / alkoxy <containing 1-20 C> (opt. substd.
         by 1 or more G4) / aryloxy <containing 6-12 C>
         (opt. substd. by 1 or more G4) / heteroaryloxy
         <containing 4-12 atoms> (opt. substd. by 1 or
         more G4) / acyloxy / 12
```



```
G3     = aryl <containing 6-12 C> (opt. substd. by
         1 or more G4) / heteroaryl <containing 4-12
         atoms> (opt. substd. by 1 or more G4)
G4     = R / (Examples: halo / CO2H / CN / NO2)
```

The acyloxy in G1 expands to O-C(O)-H / O-C(O)-R / O-C(O)-Ak-(R)n. The final term indicates a carbon chain bearing optional substituents. In this case, G1 can also be node 12, an additional definition for acyloxy not found in the expansion of the text shortcut. This fragment adds O-C(O)-Cb / O-C(O)-Hy to the list of alternatives for G1. Additionally, G1 or G2 are defined to include benzoyloxy, hence expanding the definition of acyloxy to include other examples cited in the patent.

79

Match Level ANY in the Query

Using Match Level ANY broadens the query beyond CLASS. All of the answers retrieved by Match Level CLASS are also retrieved by ANY. Remember that MARPAT is built knowing that R's are not matched unless the query is assigned to do so. Additional structural information given by the patent is added as alternatives in the G-group. This is done to allow the Match Level CLASS to retrieve the groups specifically mentioned as examples of R. Use Match Level ANY to:

- Broaden the query without deleting nodes
- Have extensive recall
- Match "substitution optional" or "substitution required"
- Match acyl

If a substructure search of the query at Match Level CLASS retrieves zero or just a few answers, consider if any portion of the query might typically be described in patents as, for example, a blocking group. Assign these query nodes as Match Level ANY and see if more answers are found. If a portion of your query could match acyl, consider assigning a portion as Match Level ANY.
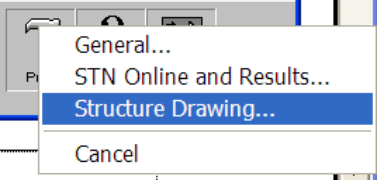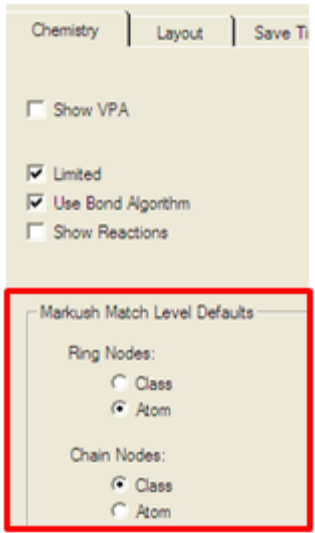
For example, if your query contains…
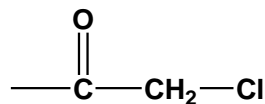
… assign the Cl as Match Level ANY to allow it to match the following definition of acyl:

$$n = (0 \text{ or more})$$
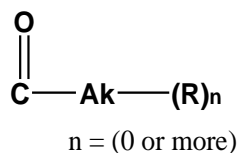
Changing the Default Match Level

The default Match Level can be customized in STN Express.

STEP	ACTION
1	Click the **Prefs** (preferences) button and select **Structure Drawing**. General... STN Online and Results... Structure Drawing... Cancel
2	Select the **Chemistry** tab. Modify **Markush Match Level Defaults** as desired. Chemistry Layout Save Ti ☐ Show VPA ☑ Limited ☑ Use Bond Algorithm ☐ Show Reactions Markush Match Level Defaults Ring Nodes: ○ Class ◉ Atom Chain Nodes: ◉ Class ○ Atom

Changing Match Level for a Single Search

To change Match Levels for the current search only, use the novice SEARCH command.

```
=> SEARCH L1

ENTER TYPE OF SEARCH (SSS) OR CSS:SSS
ENTER SCOPE OF SEARCH (SAMPLE), FULL, RANGE, OR SUBSET:FULL
CHANGE MLEVEL FOR ANY NODES? Y/(N)/?:Y
CHANGE MLEVEL FOR RING NODES AND RING GROUPS?
(ATOM)/CLASS/N/?:CLASS
CHANGE MLEVEL FOR CHAIN NODES AND CHAIN GROUPS?
ATOM/(CLASS)/N/?:N
CHANGE ECLEVEL FOR ALL NODES AND GROUPS?
(LIMITED)/UNLIMITED/N/?:N
ISOLATE ALL RINGS IN THE STRUCTURE? (Y)/N/?:Y
```

CHAPTER 7: ELEMENT COUNT AND ELEMENT COUNT LEVEL

ELEMENT COUNT

Overview

Element Count (EC) is an attribute of generic groups in MARPAT structures. Element Count is also an attribute that you can add to the query structure to more precisely define generic nodes and to refine the types of answers retrieved.

Element Count describes an element and the number of times it occurs in a generic group. In the previous example, the Element Count defines the heterocycle as containing 3-7 atoms with zero or more O, S, and P as well as one or more N.

In MARPAT structures, generic groups can have attributes, such as Element Counts, that further define the group. These attributes follow the generic group and are set off by < >.

```
G53    = heterocycle <containing 3-7 atoms, 1 or more N,
         zero or more O, zero or more S, zero or more P,
         attached through 1 N>
```

Note that hydrogens are not included as part of the Element Count data.

Element Counts for Real Groups

Element Counts are automatically generated for the real (specific) atoms, e.g., C, N, and O, but not for the generic groups, e.g., Ak, Hy, in your query structure. For example, the isolated ring system…
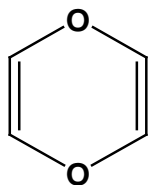
…would have these element counts assigned to it:

- Exactly two oxygen
- Exactly four carbon

If this ring were in your query, these additional element counts would result in a precise set of retrievals, even at Match Level CLASS when the ring would match with Hy nodes in the Markush structures.

Markush Element Counts

The system matches the Element Counts that are recorded with the generic groups in the Markush structures. It matches the heterocycles with two oxygens and eliminates those with two or more heteroatoms that are not oxygen. At Match Level CLASS, the query ring also retrieves the following generic group heterocycle:

```
heterocycle <containing 2 heteroatoms, 1 or more O, zero or more S
(no other hetero atoms)>
```

At Match Level CLASS and Element Count Level (see later in Chapter 7) set to Unlimited, the query ring also retrieves the generic group Hy, without any Element Counts. If no Element Counts are specifically defined for this group, all possible counts for all elements are implied. As with other matching done in MARPAT, when the exact specifications of the query Element Count fall within a range that is part of the file structure, the record is retrieved.

The query does not retrieve generic groups such as:

```
heterocycle <containing 2 heteroatoms, 1 or more N, zero or more O
(no other heteroatoms)>
```

Assigning Element Counts

You can assign Element Counts to generic nodes in structures built with STN Express or STN on the Web or by using the STRUCTURE command. In query structures, Element Counts can be applied only to the Ak, Cy, Cb, and Hy generic nodes. They may not be assigned to real atoms because the system automatically generates a count for real atoms. If no Element Counts are assigned, the generic node retrieves groups matching any number of any elements in the group.

How to Assign Element Counts

To assign Element Count in STN Express, right-click a single generic node and select **Element Count**. Or, select one or more generic nodes in your structure and select **Element Count** from the **QueryDef** menu.

In the element count window, select the element and specify the count values:
- Exact (specify an exact count for an element)
- Range (specify a range of counts from 0-255)
- Minimum (specify a minimum value from 0-)
- Maximum (specify a maximum value from 0-)

*After each specification, click the **Add** button.* Repeat the process for each element for which counts are to be specified. When you are finished, click **OK**.

Note: Only carbon element counts can be specified for Ak and Cb.

Element Count Window

The window below shows assignments of a range of 1-5 for Nitrogen, a minimum of three Carbon, a maximum of one Phosphorous, and exactly zero Sulfur. Oxygen has been selected, and a maximum of two is indicated. Click **Add** to place the definition in the list.

These are minimum criteria. Other atoms can be present, unless specifically set to exactly zero.

Effects of Adding Element Count

By adding Element Counts to the query specification, more precise answers are retrieved. Both ATOM level matches and CLASS level matches better meet the query requirements.

For example, to build the query:

```
Z-NO2      where Z = any heterocycle containing 2 oxygens
```

begin by creating a generic node Hy in the query. Then assign the Element Count of exactly two oxygens to the Hy node.
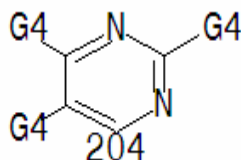
Element Count and Match Level

Some examples of answers retrieved by this query are shown here. Only the matching variable group statements and/or structures are shown. This answer is an ATOM match. There are two oxygen atoms present in the real atom heterocycle retrieved.
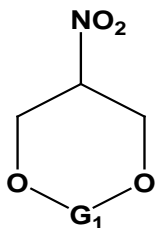
G1 = 8
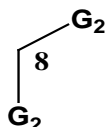
The following file structure is not retrieved by this query because the file element counts do not match those for the query.

G4 = cycloalkyl <containing 3-10 C> (opt. substd.) / NO2

The real atom heterocycle has an Element Count of zero O, as well as exactly two N, and exactly four C, and is therefore not an answer for the query.

Match Level CLASS Answers

If the query Element Count falls within the range of the file structure Element Count, it is considered a match. If the count is outside of the file range, it is not a match. The following example is retrieved for the query with the Match Level set to CLASS. This answer has a Hy with an Element Count of two oxygens, thereby matching the Element Count set in the query.

```
G6 =  aryl (opt. substd. by 1-3 G7) / heterocycle
      <containing 1-2 heteroatoms, zero to 2 O, zero
      to 2 N (no other heteroatoms), monocyclic, 1 or
      more 5-6 membered rings > (opt. substd. 1-3 G7) /
      (Examples: Ph (opt. substd.) / furyl (opt. substd.)
      / 51 / 57 / imidazolyl (opt. substd.) / pyridyl
      (opt. substd.) / pyrimidinyl (opt. substd.)
G7 =  loweralkyl (opt. substd. By OH) / loweralkoxy / X /
      NO2 / NH2 (opt. substd.) / N3 / 48 / alkoxycarbonyl
      / CONH2
```

The following structure would not be retrieved because the Hy in G1 has either exactly zero or exactly one oxygen, which does not match the required two oxygens of the query.

```
G1 =  NH2 / F / Cl / Br / I / CN / H / NO2 / alkyl
      <containing 1-6 C> (opt. substd.) / heterocycle
      <containing 1 or more heteroatoms, 1-3 N, zero
      or 1 S, zero or 1 O (no other heteroatoms),
      monocyclic, 3-6 membered ring> (opt. substd. By
      1 or more G2) / Ph (opt. substd.)
G2 =  F / Cl / Br / I / NO2 / OH / alkyl <containing
      1-6 C> (opt. substd.)
```

The following structure is also retrieved as a CLASS level match if the Element Count Level (discussed later in this chapter) in the query is set to Unlimited. The generic Hy in G9 does not have any elements defined and therefore is retrieved because any Element Count is implied with the level set to unlimited.

```
G5 =  R / (opt. substd. by X / CN / NO2 / alkyl
      <containing 1-5 C> (opt.substd. by 1-9 X) /
      alkoxy <containing 1-5 C> (opt. substd. by 1-9 X)
      / alkylthio <containing 1-5 C> (opt. substd. by 1-
      9 X) / alkoxycarbonyl <containing 1-5 C>
G9 =  heterocycle (opt. substd. by 1 or more G5) /
      (Examples: pyridyl / furyl / thienyl /isoxazolyl /
      isothiazolyl / oxazolyl / thiazolyl / 26 / 32 / 34
      / pyrrolidino / piperidino / hexahydroazepino /
      morpholino / piperazino)
```

The entire definition of the file structure's generic node is examined. When not explicitly blocked, any count of any element is implied with Element Count Level unlimited. In the following example, the file's Hy in G4 specifically states that there are one or more nitrogens present and one or more heteroatoms present. Other heteroatoms are not excluded by the definition. Therefore, this structure is retrieved even though the number of oxygens is not explicitly mentioned in its definition.

```
G4 =  heterocycle <containing 1 or more heteroatoms,
      1 or more N, attached through 1 or more N nodes,
      Zero or more 5-6 membered rings> (opt. Substd. By
      1 or more G9)
G9 =  R / (Specifically claimed: X / OH / NO2 / CN /
      CO2H / SO3H /
```

ELEMENT COUNT AND GENERIC DEFINITIONS

Generic definitions in combination with Element Counts may exclude the elements that are not explicitly excluded. The number of real atoms, the number of explicit heteroatoms, and the number of other heteroatoms are used to calculate the element counts for the G-group. These are considered when determining if the file structure matches the query.

For example, the Element Counts for the search file heterocycle in G14 indicate that there are one or two heteroatoms, one or two nitrogens, and no other heteroatoms.

This eliminates the possibility of any oxygens being part of the group. This structure is not retrieved by the query.

```
G14 = pyridyl (opt. substd.) / pyridazinyl (opt.
      substd.) / pyrimidinyl (opt. substd.) /
      pyrazinyl (opt. substd.) / heterocycle
      <containing 1-2 heteroatoms, 1-2 N (no other
      heteroatoms), 4-5 C, 6-membered monocyclic
      ring> (opt. substd. by 1-2 G27) / 61 / 71 /
      81 / 92 / 237
G27 = alkyl <containing 1-6 C> / F / Cl / Br / I / OH
      / alkoxy <containing 1-6 C> / CHO /
      alkylcarbonyl <containing 1-7 C> /
      arylcarbonyl <containing 6-10 C> / NO2
```

Element Count and SUBSET Searching

Another approach to retrieving structures with the appropriate element counts is to use subset searching. First, search a query that does not have any element counts assigned. This creates a broad answer set that contains structures with all possible element counts. Then modify the query assigning element counts. Search the query by using the broad answer set as a subset. The query with the Element Count attributes is a more precise definition, and all answers to this new query are contained in the broad answer set.

In the following example, a substructure search is first conducted on the following structure, in which **Z** is any carbon chain:
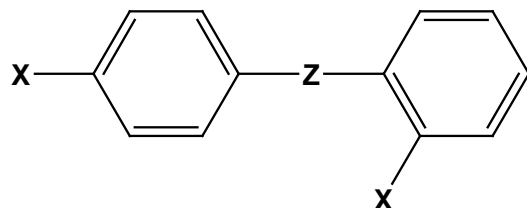
Z = any carbon chain
No further substitution allowed

Because no further substitution is allowed, a closed substructure search (CSS) is conducted on this structure (L1), in which Z is defined as Ak and the Match Level is ATOM at all nodes.

```
=> S L1 FULL CSS
FULL SEARCH INITIATED 15:08:21
FULL SCREEN SEARCH COMPLETED -  117785 TO ITERATE
100.0% PROCESSED  117785 ITERATIONS   353 ANSWERS
SEARCH TIME: 00.02.55
L2 353 SEA CSS FUL L1
```

This query could be modified to retrieve carbon-containing chains with only 5-7 carbons. Modify the original structure (L1) and assign an element count of minimum five and maximum seven carbons at the Ak node. Structure L3 is created. Now conduct a subset closed substructure search on the modified query structure (L3). Define the subset as the previously obtained answer set (L2).

```
=> S L3 SUB=L2 CSS FULL
FULL SUBSET SEARCH INITIATED 15:14:29
FULL SUBSET SCREEN SEARCH COMPLETED - 353 TO ITERATE
100.0% PROCESSED 275 ITERATIONS     113 ANSWERS
SEARCH TIME: 00.01.32
L4 113 SEA SUB=L2 CSS FUL L3
```

The modified query (with element counts specified) retrieves answers such as:

```
TI    Preparation of aryl-substituted acyclic enediyne
      compounds as antitumor agents
```

MSTR 1B Assembled
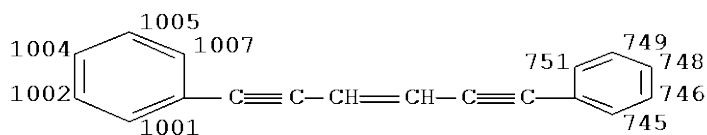

```
745, 746, 748, 749, 751: opt. substd. by (1-3) Cl
1001, 1002, 1004, 1005, 1007: opt. substd. by (1-3) Cl
Patent location:            claim 1
Note:                       substitution is restricted
Note:                       or pharmaceutically acceptable
                            salts or solvates
Note:                       additional substitution also
                            claimed
```

● ● ●

The modified query does not retrieve answers such as:

MSTR 1 Assembled

$$G1 — CH_2 — G1$$

```
G1      = Ph (opt. substd. by 1 or more F)
Patent location:            claim 1
Note:                       or salts, esters, or ethers
Note:                       also incorporates claim
```

ELEMENT COUNT LEVEL

Element Count Level

Element Count Level is the attribute that determines whether a query containing element counts for generic nodes will match against file structures with corresponding nodes that do not have Element Counts.

Element Count Level Options

There are two Element Count Level options: **Limited** and **Unlimited**. The system default is LIMITED, which means that queries with Element Counts will only match file structures with corresponding Element Counts. The **Limited** option is available in the Element Count window. It is selected by default.
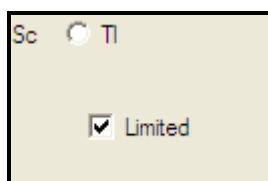

To override the default, you must explicitly assign the value of **Unlimited** to a node or a group of nodes or your entire query. To assign **Unlimited**, deselect the **Limited** check box to remove that default for the specified node(s).

Assigning Element Count Level

You can change the Element Count Level default of **Limited** for the entire query in several ways:

- Deselect the **Limited** option in the **Chemistry** section of the **Structure Drawing Preferences** menu (changes all element count nodes).
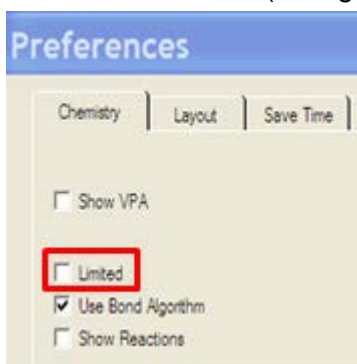
- From the **QueryDef** menu, select **Markush Attributes** and then deselect the **Limited** option for any selected node(s).
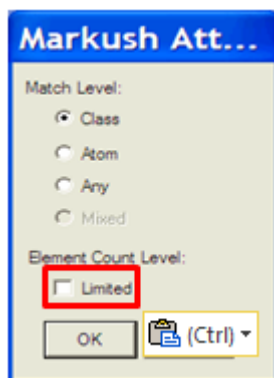
- Right-click a node, select **Markush Attributes**, and then deselect the **Limited** option (see above).

For queries created by using the STRUCTURE command:

- Use the SET command:

```
=> SET ECL UNLIM
SET COMMAND COMPLETED
```

- Use the novice SEARCH command. You receive a prompt with the option to change ECLEVEL. This works even with structure queries which are drawn offline.

```
=> SEARCH L1
ENTER TYPE OF SEARCH (SSS) OR CSS:SSS
ENTER SCOPE OF SEARCH (SAMPLE), FULL, RANGE, OR
 SUBSET:FULL
CHANGE MLEVEL FOR ANY NODES? Y/(N)/?:Y
CHANGE MLEVEL FOR RING NODES AND RING GROUPS?\
 (ATOM)/CLASS/N/?:CLASS
 CHANGE MLEVEL FOR CHAIN NODES AND CHAIN GROUPS?
  ATOM/(CLASS)/N/?:CLASS
 CHANGE ECLEVEL FOR ALL NODES AND GROUPS?
  (LIMITED)/UNLIMITED/N/?:UNLIMITED
```

EFFECT OF ELEMENT COUNT LEVEL LIMITED ON SEARCH RESULTS

When Element Count Level **Limited** is applied to a node, retrievals are limited to generic nodes in the structure that explicitly match the Element Count value in the query. For example, a query Hy node specified with a minimum Element Count value of one sulfur and Element Count Level **Limited** would retrieve a Markush structure containing a heterocyclic node that explicitly permits one or more sulfur atoms. It would not match a heterocyclic node that did not explicitly allow sulfur.

Limited also requires answers to have an Element Count associated with the file structure. It is possible to have a generic node with no Element Count, but **Limited** retrieves only generic nodes with Element Counts.

EFFECT OF ELEMENT COUNT LEVEL UNLIMITED ON SEARCH RESULTS

When Element Count Level **Unlimited** is applied to a node, retrievals include generic groups in the file structures that explicitly match the Element Count value in the query as well as those generic groups that do not explicitly state the Element Count value or allow it in the structure. In other words, with **Unlimited**, the generic group in the MARPAT structure may, but is not required to, have the query's Element Counts.

The difference in retrievals between using of **Limited** or **Unlimited** is illustrated by the following two examples.
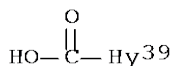
Example 1: Unlimited Element Count

In the first case, a substructure search is conducted on the following structure with Match Level CLASS at all nodes:
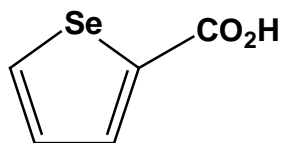
With Element Count Level **Limited**, the query retrieves 302 answers, such as:

```
    MSTR 1 Assembled

         O
         ||
  HO— C— Hy 39


9: heterocycle <containing 1-2 heteroatoms, zero or more B,
      zero or more N, zero or more O, zero or more P,
      zero or more S, zero or more Se, zero or more Si,
      zero or more Te> (opt. substd. by 1 or more G6)
Patent location:          claim 1
```
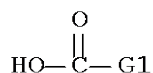
Notice that the definition of the heterocyclic ring includes explicit mention of Se.

With the global Element Count Level of **Unlimited**, 11851 answers are retrieved with Markush structures containing heterocycles that may or may not explicitly have Se in their definition. Additional answers not retrieved with **Limited** include generic structures, such as:

```
    MSTR 1 Assembled

       O
       ||
  HO— C— G1


G1    = heteroaryl (opt. substd.)
Patent location:          claim 1
Note:                     substitution is restricted
```
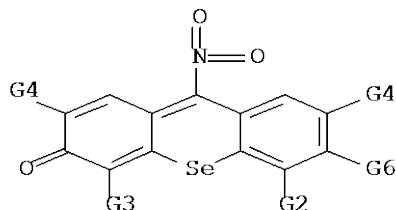
Example 2: More Unlimited Element Count

In this example, Hy-NO2 is the search query with all nodes at Match Level ATOM, the Element Count is exactly one selenium, and the global Element Count Level is **Limited**. The search retrieves 476 answers such as:


```
   MSTR 1 Assembled




Patent location:          claim 10
Note:                     substitution is restricted; also
incorporates claim
```
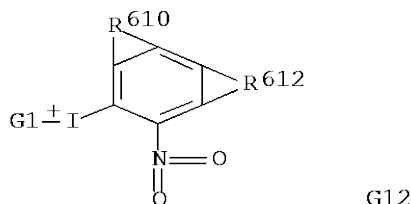
With Element Count Level **Unlimited**, 5146 answers are retrieved, such as:


```
MSTR 1 Assembled




610: R <"group to form ring">
612: R <"group to form ring">
Patent location:          claim 2
Note:                     and protected derivatives
Note:                     also incorporates claim 3
Note:                     substitution is restricted
```

ELEMENT COUNT SUMMARY

Element Counts yield a more precise retrieval for your Markush search. Searching with Element Counts set to unlimited (non-default) allows matching against generic groups with no specified Element Counts while matching only those for which Element Counts match those specified in your query.

SUMMARY

MARPAT, A Useful Database

No substance search of the patent literature can be considered to be successful or complete without including the MARPAT database. Markush structures represent many, many possible substances that have not been specifically reported in the published literature.

CASLINK

Using CASLINK (or HCASLINK) to combine both the specific substance search with the generic substance search yields superior results for your patent searching needs.

ADDITIONAL RESOURCES

RESOURCE	LOCATION	USED FOR
STN Training Resources	www.cas.org	• STN Database Summary Sheets • User documentation

TECHNICAL SUPPORT

CAS Customer Center	Hours:	8 a.m. - 6 p.m. Eastern Time (Monday-Friday)

	Phone:	1-800-753-4227 (North America)
		     Option 2: General information or account-related questions
		     Option 3: Assistance with search strategies, database content, or using a product
		     Option 4: Technical assistance with software setup, installation, and configuration
		 +1 614-447-3700 (outside North America)

	Fax:	1-614-447-3751
	E-mail:	help@cas.org
	Internet:	www.cas.org

FIZ Karlsruhe (Europe)	Hours:	8 a.m. - 6 p.m. Central European Time (Monday-Friday)
	Phone:	+49 7247-808-555
	Fax:	+49 7247-808-259
	E-mail:	helpdesk@fiz-karlsruhe.de
	Internet:	www.stn-international.com

JAICI (Japan Association for - International Chemical Information)	Hours:	9 a.m. - 5:30 p.m. Standard Time (Monday-Friday)
	Phone:	+81 3-5978-3601 (Technical Service) +81 3-5978-3621 (Customer Service)
	Fax:	+81 3-5978-4090
	E-mail:	support@jaici.or.jp (Technical Service) customer@jaici.or.jp (Customer Service)
	Internet:	www.jaici.or.jp