

---

# MARPAT

マッチレベル, 元素数レベルを  
マスターしよう!



**JAICI**  
化学情報協会

---

## 本日の内容

- MARPAT ファイルの概要
- マッチレベルをマスターしよう!
- 元素数レベルをマスターしよう!

---

## 本日の内容

- **MARPAT ファイルの概要**
- マッチレベルをマスターしよう！
- 元素数レベルをマスターしよう！

---

## MARPAT ファイルの概要

- **MARPAT ファイルはマルクーシュ構造を収録するデータベース**

収録対象特許	CA ファイルの収録特許でマルクーシュ構造があるもの (ポリマー, 無機化合物は除く)
収録期間	1961 年～ (CAS : 1985 年～, INPI : 1961～1987 年)
収録内容	CA ファイル由来の書誌情報, 抄録, マルクーシュ構造
レコード単位	文献単位 (特許ファミリー単位)
更新頻度	毎日

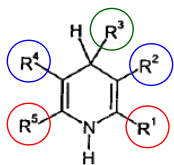
# MARPAT ファイルのデータ

## 請求の範囲

【特許請求の範囲】

【請求項1】式(II)

【化1】

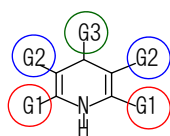


(II)

式中、**R<sup>1</sup>及びR<sup>5</sup>**は同一であるか又は相異なり、そして各々**C<sub>1</sub>~C<sub>10</sub>-アルキル**又は**C<sub>6</sub>~C<sub>10</sub>-アリアル**を表し、**R<sup>2</sup>及びR<sup>4</sup>**は同一であるか又は相異なり、そして各々**水素**、**C<sub>1</sub>~C<sub>10</sub>-アルキル**、**CN**又は**COOR<sup>6</sup>**(ここで**R<sup>6</sup>**は**C<sub>1</sub>~C<sub>10</sub>-アルキル**である)を表し、**R<sup>3</sup>**は**水素**、**C<sub>1</sub>~C<sub>10</sub>-アルキル**を表すか、又は場合により**ハロゲン**、**ニトロ**、**COOR<sup>6</sup>**(**R<sup>6</sup>**は上記で定義したとおりである)、**CN**もしくは**C<sub>1</sub>~C<sub>10</sub>-アルキル**により置換されていてもよい**C<sub>6</sub>~C<sub>10</sub>-アリアル**を表す…

**JAICI**  
化学情報協会

MSTR 1



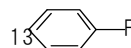
基本骨格

置換基

G1 = alkyl <containing 1-10 C> / aryl <containing 6-10 C> / (Example: Pr-i)

G2 = H / alkyl <containing 1-10 C> / CN / alkoxycarbonyl <containing 1-10 C> / (Examples: CO2Me / CO2Et)

G3 = H / alkyl <containing 1-10 C> / aryl <containing 6-10 C> (opt. substd. by 1 or more G4) / (Example: 13)



G4 = halo / NO2 / alkoxycarbonyl <containing 1-10 C> / CN / alkyl <containing 1-10 C>

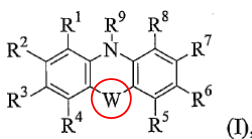
Patent location: claim 1

4

# REGISTRY/CAplus との関係

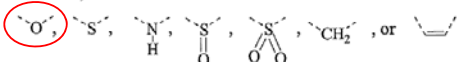
(例) WO200502782

クレーム

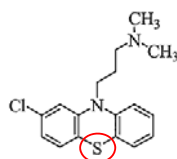


マルク-シュ  
構造を索引

W is NO,



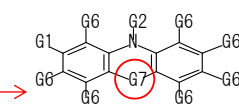
実施例



特定化合物の  
CAS 登録番号  
を索引

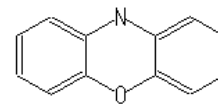
W が S である化合物のみを記載

MARPAT



G7 = S / SO2 / S(O) / NH / CH2 / CH=CH / 106

構造質問式



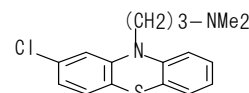
↓ X

CAplus/CA

IT 50-53-3, biological studies  
RL: PAC (Pharmacological activity);  
THU (Therapeutic use); BIOL  
(Biological study); USES (Uses)  
(chlorpromazine compd.-antiproliferative  
drug antitumor combination)

REGISTRY

RN 50-53-3 REGISTRY

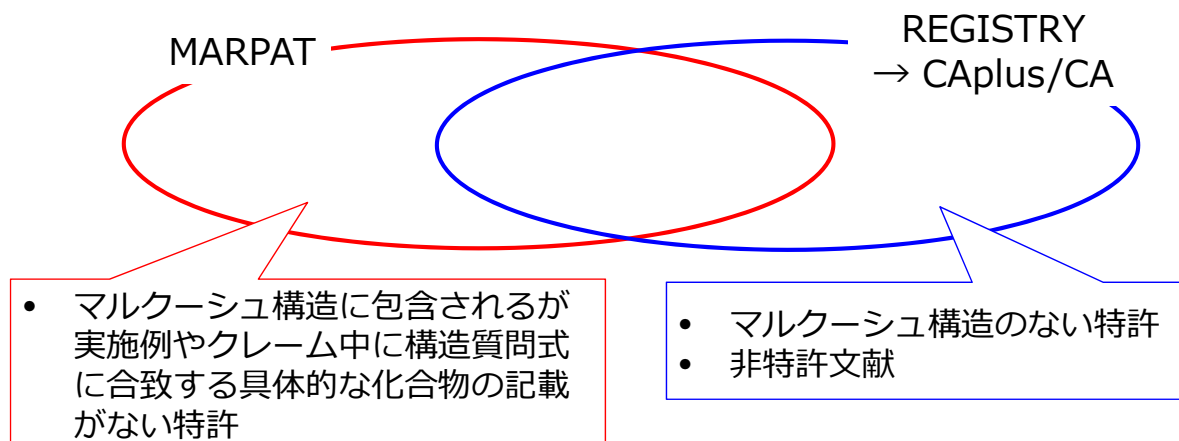


**JAICI**  
化学情報協会

5

# REGISTRY/CAplus との関係

## • 各ファイルに特有の回答



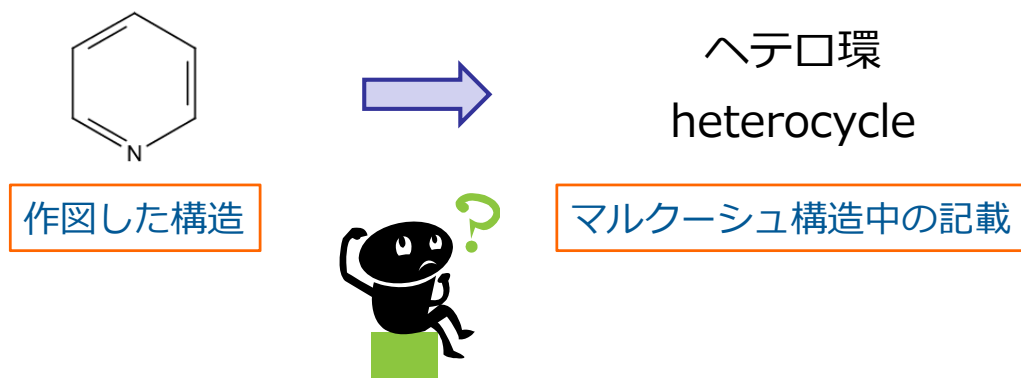
両ファイルを併用することで、より多くの情報が得られます！

## 本日の内容

- MARPAT ファイルの概要
- マッチレベルをマスターしよう！
- 元素数レベルをマスターしよう！

## マッチレベルとは

- 特許中で一般化した表現で書かれた置換基もヒットさせるかを指定するための設定



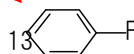
## マルクーシュ構造中の置換基の種類

- 特定原子, 一般式グループ, R グループの3種類がある

具体的 ↑↓ 一般的	<b>特定原子</b>	C, N, O などの原子, 官能基グループ (NO <sub>2</sub> など), 環の固有名称 (pyridyl など), 構造フラグメント
	<b>一般式グループ</b>	ハロゲン, アルキル, ヘテロアリアルなどの一般式
	<b>R グループ</b>	構造図や一般式グループで表現できない置換基 (有機基, 保護基, 脱離基など)

G1 = halo / CF<sub>3</sub> / OMe / CN / 13 / aryl / alkoxy

G2 = R <"organic group">



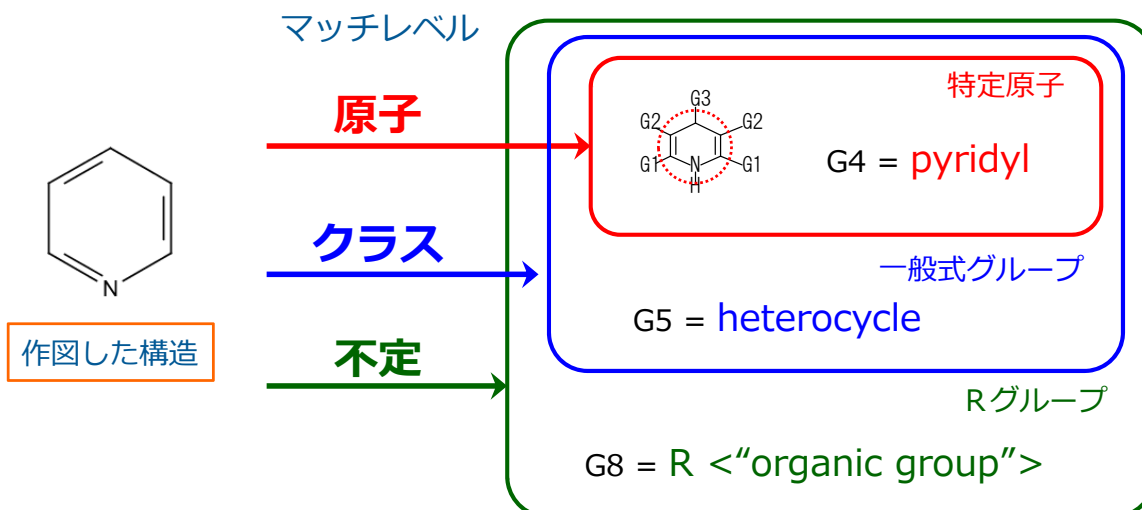
## マッチレベルとヒットする置換基

- マッチレベルは、**原子・クラス・不定**の3種類
  - 同じ構造でも、マッチレベルの指定によってヒットする置換基の範囲を調節できる

マッチレベル	ヒットする置換基の種類
原子	特定原子
クラス	特定原子 + 一般式グループ
不定	特定原子 + 一般式グループ + Rグループ

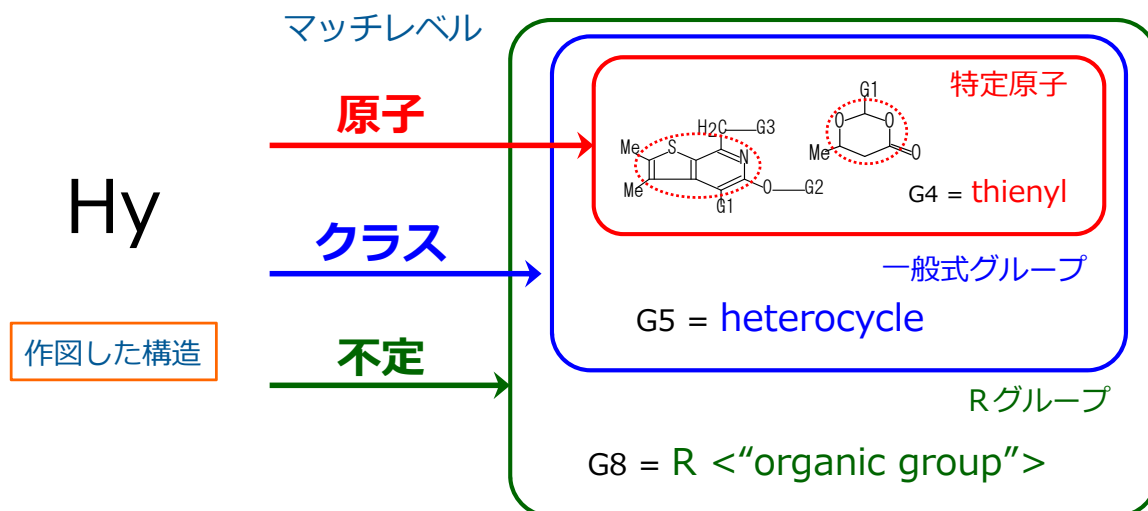
## マッチレベルと得られる回答の関係

- 特定の構造を作図した場合



# マッチレベルと得られる回答の関係

## ● 一般式グループ記号を作図した場合



# マッチレベルの指定方法

## ● デフォルトは 環→原子, 鎖→クラス

原子

クラス

右クリック

Markush 属性

マッチレベル:

クラス

原子

不定

混合

元素数レベル:

限定

OK キャンセル

Markush 属性(M)...

元素数(E)...

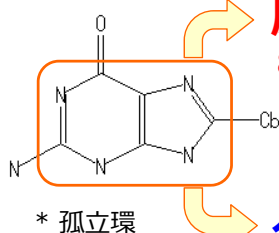
一般式属性(G)...

結合非水素数(T)...

## マッチレベル選択の指針

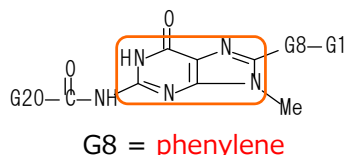
- 主要な構造はマッチレベルを**原子**に指定
  - 化合物の新規性, 生物活性や物性の発現に必須の構造など

構造質問式

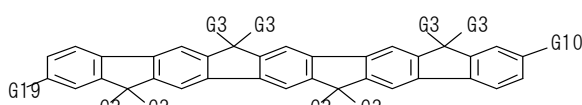


**原子**  
82 件

**クラス**  
5816 件



主要部分の具体的な構造が存在

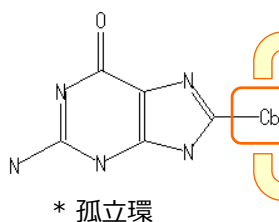


主要部分が置換基の一般式でもヒット

## マッチレベル選択の指針

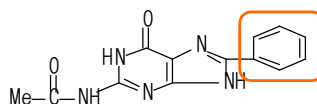
- 許容範囲の広い置換基は**クラス**に指定
  - 一般式グループ記号で作図した側鎖など

構造質問式

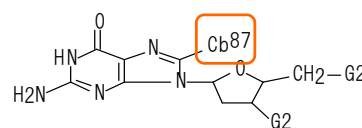


**原子**  
82 件

**クラス**  
231 件



具体的な構造で記載された置換基のみがヒット



一般式グループで記載された置換基もヒット



決められない場合はまず**クラス**で広めに検索して,  
ノイズが多ければ**原子**でサブセット検索するとよい

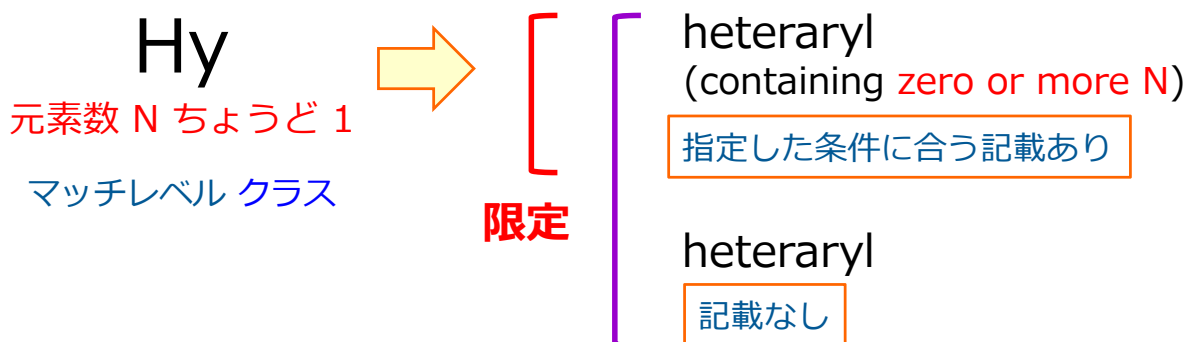


## 本日の内容

- MARPAT ファイルの概要
- マッチレベルをマスターしよう！
- 元素数レベルをマスターしよう！

## 元素数レベルとは

- 指定した条件に合う元素数が記載された一般式グループに限定する設定



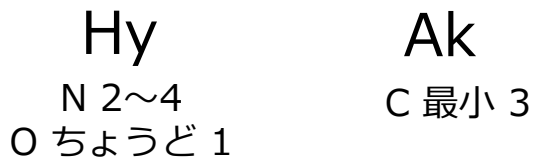
限定しない



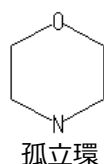
元素数レベル

# 元素数の指定方法

## ● 一般式グループ記号に元素数を指定する




## ● 具体的な構造を作図する



N ちょうど 1  
0 ちょうど 1



# 元素数レベルと得られる回答の関係

作図内容	マッチレベル	元素数レベル	pyridyl	heteroaryl	heteroaryl (0-1 N)	Heteroaryl (2-4 N)
 孤立環	クラス	限定	○	×	○	×
		非限定	○	○	○	×
Hy 元素数 N 1	クラス	限定	○	×	○	×
		非限定	○	○	○	×
Hy 元素数 指定なし	クラス	限定	○	○	○	○
		非限定	○	○	○	○
Hy 元素数 N 1	原子	限定	○	×	×	×
		非限定	○	×	×	×

### ● 元素数レベルの指定を変更しても回答が変わらない場合

- 元素数を指定していない ← 限定するための条件がないため
- マッチレベルを原子に指定 ← 一般式グループはヒットしないため

# 元素数レベルの指定方法

- デフォルトは**限定**

右クリック

Markush 属性

マッチレベル

クラス

原子

不定

混合

元素数レベル

限定

OK キャンセル

元素数

限定

追加 削除 キャンセル OK

# 元素数レベル 選択の指針

- ヒットする一般式グループをどこまで許容するかを検討する

元素数 N 1

限定

限定しない

G3 = heteroaryl <containing 1 or more heteroatoms, zero or more N, zero or more O, zero or more S, 1-9 C>

G2 = heteroaryl

N=1 の条件を満たす記載のあるヘテロ環のみがヒット

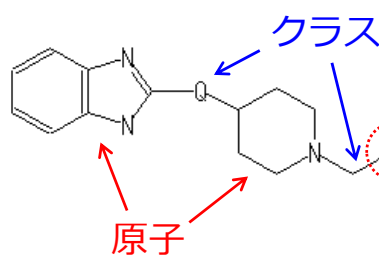
「ヘテロ環」という記載のみでもヒット

決められない場合はまず**非限定**で広めに検索して、  
ノイズが多ければ**限定**でサブセット検索するとよい

# デモンストレーション

- 下記の構造を含むマルクーフ構造を検索する

– Hy のマッチレベル, 元素数レベルの指定を変えた場合の回答を比較する



\* 2つの環は孤立

元素数 N 1~2

- ① マッチレベル: 原子
- ② マッチレベル: クラス  
- 元素数レベル: 限定
- ③ マッチレベル: クラス  
- 元素数レベル: 非限定

## まとめ

- マッチレベル, 元素数レベルの指定によって得られる回答の範囲が変わる

– 調査目的やヒット件数に応じて適切な指定を検討する

- 決められない場合は広めに検索して, ノイズが多ければ絞り込み検索する



## 検索例

=> FILE MARPAT

=>

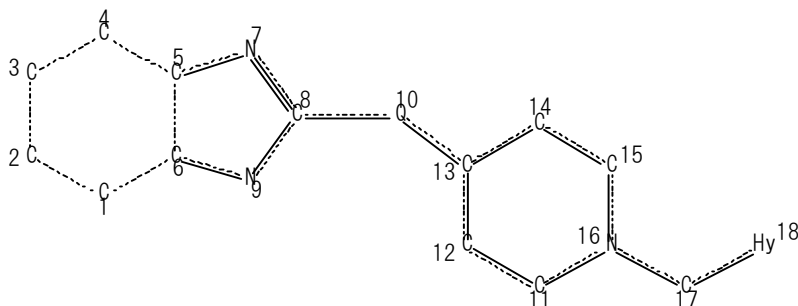
Uploading C:\Users\JAICI\Documents\STN Express 8.5\Queries\ATOM.str

← Hy のマッチレベルが原子の  
構造質問式をアップロード

L1 STRUCTURE UPLOADED

=> D QUE

L1 STR



NODE ATTRIBUTES:

NSPEC IS R AT 1

NSPEC IS R AT 2

:

NSPEC IS C AT 17

NSPEC IS C AT 18

DEFAULT MLEVEL IS ATOM

MLEVEL IS CLASS AT 10 17

DEFAULT ECLEVEL IS LIMITED

ECOUNT IS M1-X2 N AT 18

← デフォルトのマッチレベルは ATOM (原子)

← ノード 10, 17 のマッチレベルは CLASS (クラス)

← デフォルトの元素数レベルは LIMITED (限定)

← 18 (Hy) の元素数は N が最小 1, 最大 2

GRAPH ATTRIBUTES:

RSPEC I

NUMBER OF NODES IS 18

STEREO ATTRIBUTES: NONE

=> S L1 ← サンプル検索

SAMPLE SEARCH INITIATED 10:11:32 FILE 'MARPAT'

SAMPLE SCREEN SEARCH COMPLETED - 1535 TO ITERATE

100.0% PROCESSED 1535 ITERATIONS

1 ANSWERS

SEARCH TIME: 00.00.01

FULL FILE PROJECTIONS: ONLINE \*\*COMPLETE\*\*

BATCH \*\*COMPLETE\*\*

PROJECTED ITERATIONS: 28430 TO 32970

PROJECTED ANSWERS: 1 TO 80

L2 1 SEA SSS SAM L1

=> S L1\_FUL ← フルファイル検索

FULL SEARCH INITIATED 10:11:37 FILE 'MARPAT'

FULL SCREEN SEARCH COMPLETED - 31674 TO ITERATE

100.0% PROCESSED 31674 ITERATIONS

40 ANSWERS

SEARCH TIME: 00.00.02

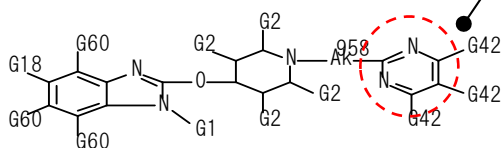
L3 40 SEA SSS FUL L1

=> D L3 SCAN FQHIT

← 最初にヒットしたマルクーシュ構造のヒット部分のみを組み立てて表示

L3 40 ANSWERS MARPAT COPYRIGHT 2013 ACS on STN

**MSTR 1A Assembled**



ヘテロ環の具体的な構造が記載されている  
マルクーシュ構造がヒット

958: alkylene <containing 1-6 C>

Patent location: claim 1

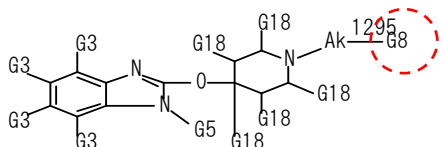
Note: or pharmaceutically acceptable acid addition salts

Stereochemistry: or stereochemically isomeric forms

HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1): 2

L3 40 ANSWERS MARPAT COPYRIGHT 2013 ACS on STN

**MSTR 1C Assembled**



1295: alkylene <containing 1-12 C> (opt. substd.)

G8 = pyridyl (opt. substd. by (1-2))  
alkoxy <containing 1-6 C>

環の固有名称も構造図と同じようにヒット

Patent location: claim 1

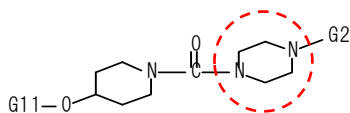
Note: or pharmaceutically acceptable acid addition salts

Note: substitution is restricted

Stereochemistry: or possible stereochemically isomeric forms

L3 40 ANSWERS MARPAT COPYRIGHT 2013 ACS on STN

**MSTR 1 Assembled**



G11 = benzimidazolyl

Patent location: claim 1

Note: = also incorporates claim 9

Note: and pharmaceutically acceptable salts

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

=>

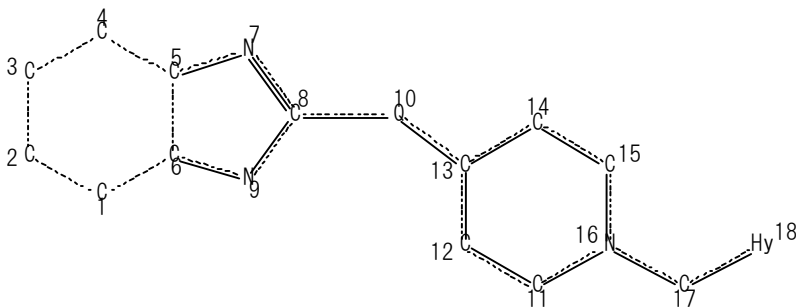
Uploading C:\Users\JAICI\Documents\STN Express 8.5\Queries\CLASS\_LIM.str

← マッチレベル クラス  
元素数レベル 限定

L4 STRUCTURE UPLOADED

=> D QUE

L4 STR



NODE ATTRIBUTES:

NSPEC IS R AT 1

NSPEC IS R AT 2

:

NSPEC IS C AT 17

NSPEC IS C AT 18

DEFAULT MLEVEL IS ATOM

MLEVEL IS CLASS AT 10 17 18

DEFAULT ECLEVEL IS LIMITED

ECOUNT IS M1-X2 N AT 18

← デフォルトのマッチレベルは ATOM (原子)

← ノード 10, 17, 18 のマッチレベルは CLASS (クラス)

← デフォルトの元素数レベルは LIMITED (限定)

← Hy の元素数は N 最小 1, 最大 2

GRAPH ATTRIBUTES:

RSPEC I

NUMBER OF NODES IS 18

STEREO ATTRIBUTES: NONE

=> S L4

SAMPLE SEARCH INITIATED 10:12:00 FILE 'MARPAT'

SAMPLE SCREEN SEARCH COMPLETED - 1535 TO ITERATE

100.0% PROCESSED 1535 ITERATIONS

1 ANSWERS

SEARCH TIME: 00.00.01

FULL FILE PROJECTIONS: ONLINE \*\*COMPLETE\*\*

BATCH \*\*COMPLETE\*\*

PROJECTED ITERATIONS: 28430 TO 32970

PROJECTED ANSWERS: 1 TO 80

L5 1 SEA SSS SAM L4

=> S L4 FUL

FULL SEARCH INITIATED 10:12:03 FILE 'MARPAT'

FULL SCREEN SEARCH COMPLETED - 31674 TO ITERATE

100.0% PROCESSED 31674 ITERATIONS

49 ANSWERS

SEARCH TIME: 00.00.02

L6 49 SEA SSS FUL L4

=> S L6 NOT L3

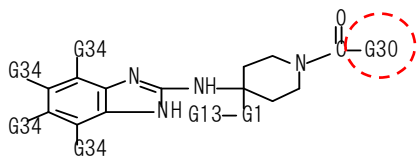
L7 9 L6 NOT L3

← Hy をクラス, 限定に変更すると 9 件増えました  
(L3 : Hy のマッチレベル 原子 の回答)

=> D SCAN FQHIT

L7 9 ANSWERS MARPAT COPYRIGHT 2013 ACS on STN

MSTR 1A Assembled



指定した元素数 (N 1-2) に合う条件が  
記載された一般式グループがヒット

G30 = heteroaryl <containing 5 or more atoms,  
zero or more N, zero or more O, zero or more S>  
(opt. substd.)

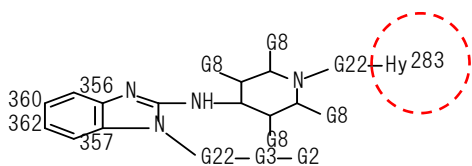
Patent location: claim 1

Note: or pharmaceutically acceptable salts or solvates

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

L7 9 ANSWERS MARPAT COPYRIGHT 2013 ACS on STN

MSTR 1A Assembled



283: heterocycle <monocyclic, 5- or 6-membered rings only,  
containing 1-4 heteroatoms, up to 2 O, up to 2 S,  
zero or more N> (opt. substd.)

356, 357, 360, 362: opt. substd.

G22 = alkylene <containing 1-6 C>

Derivative: or pharmaceutically acceptable acid addition salt

Patent location: claim 1

Note: substitution is restricted

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



=>

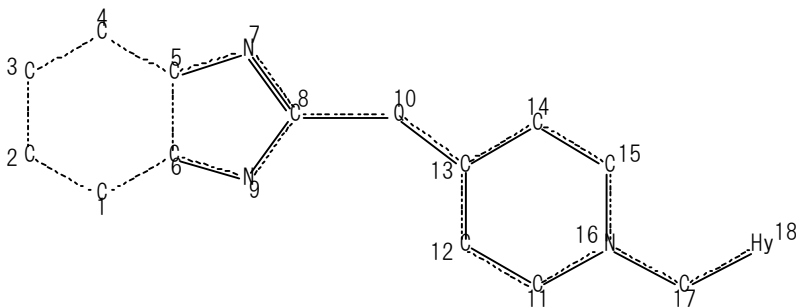
Uploading C:\Users\JAICI\Documents\STN Express 8.5\Queries\CLASS\_UNLIM.str

← クラス, 非限定

L8 STRUCTURE UPLOADED

=> D QUE

L8 STR



NODE ATTRIBUTES:

NSPEC IS R AT 1

NSPEC IS R AT 2

:

NSPEC IS C AT 17

NSPEC IS C AT 18

DEFAULT MLEVEL IS ATOM

← デフォルトのマッチレベルは ATOM (原子)

MLEVEL IS CLASS AT 10 17 18

← ノード 10, 17, 18 のマッチレベルは CLASS (クラス)

DEFAULT ECLEVEL IS LIMITED

← デフォルトの元素数レベルは LIMITED (限定)

ECOUNT IS M1-X2 N UNLIMITED AT 18

← ノード 18 の元素数は 1-2, 元素数レベルは UNLIMITED (非限定)

GRAPH ATTRIBUTES:

RSPEC I

NUMBER OF NODES IS 18

STEREO ATTRIBUTES: NONE

=> S L8

SAMPLE SEARCH INITIATED 10:12:29 FILE 'MARPAT'

SAMPLE SCREEN SEARCH COMPLETED - 1535 TO ITERATE

100.0% PROCESSED 1535 ITERATIONS

1 ANSWERS

SEARCH TIME: 00.00.01

FULL FILE PROJECTIONS: ONLINE \*\*COMPLETE\*\*

BATCH \*\*COMPLETE\*\*

PROJECTED ITERATIONS: 28430 TO 32970

PROJECTED ANSWERS: 1 TO 80

L9 1 SEA SSS SAM L8

=> S L8 FUL

FULL SEARCH INITIATED 10:12:33 FILE 'MARPAT'

FULL SCREEN SEARCH COMPLETED - 31674 TO ITERATE

100.0% PROCESSED 31674 ITERATIONS

52 ANSWERS

SEARCH TIME: 00.00.02

L10 52 SEA SSS FUL L8

=> S L10 NOT L6

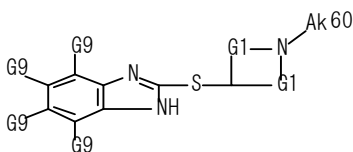
← 元素数レベルを非限定に変更すると 3 件増えました

L11 3 L10 NOT L6 (L6 : Hy がクラス, 限定の回答)

=> D\_SCAN\_FQHIT

L11 3 ANSWERS MARPAT COPYRIGHT 2013 ACS on STN

MSTR 1J Assembled



元素数の記載のない一般式グループが  
ヒット

G0: loweralkyl (opt. substd. by 1 or more heterocycle)

G1 = (1-2) CH<sub>2</sub>

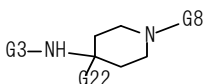
Derivative: or pharmacologically acceptable salts

Patent location: claims

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

L11 3 ANSWERS MARPAT COPYRIGHT 2013 ACS on STN

MSTR 1 Assembled



G3 = benzimidazolyl

G8 = 1190 / 1191

G9-G151190 G10-G151191

G10 = heteroaryl (opt. substd. by (1-5) G11)

G15 = (1-3) CH<sub>2</sub>

Patent location: claim 1

Note: or pharmaceutically acceptable salts

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

■ 広めに検索してから絞り込む場合は、サブセット検索を利用すると経済的です。

- ・ 構造検索料金 (2013 年 10 月現在)
  - フルファイル検索 : 17,400 円
  - サブセット検索 : 4,580 円

L10 : Hy がクラス, 非限定でフルファイル検索した回答セット  
L4 : Hy がクラス, 限定の構造質問式

=> S L4 SUB=L10 SAM ← L10 を母集合として L4 でサンプル検索  
SAMPLE SUBSET SEARCH INITIATED 10:14:04 FILE 'MARPAT'  
SAMPLE SUBSET SCREEN SEARCH COMPLETED - 1 TO ITERATE

100.0% PROCESSED 1 ITERATIONS 1 ANSWERS  
SEARCH TIME: 00.00.01

PROJECTIONS (WITHIN SPECIFIED SUBSET): ONLINE \*\*COMPLETE\*\*  
PROJECTED ITERATIONS (WITHIN SPECIFIED SUBSET): 1 TO 80  
PROJECTED ANSWERS (WITHIN SPECIFIED SUBSET): 1 TO 80

L12 1 SEA SUB=L10 SSS SAM L4

=> S L4 SUB=L10 FUL ← L10 を母集合として L4 でフルファイル検索  
FULL SUBSET SEARCH INITIATED 10:14:09 FILE 'MARPAT'  
FULL SUBSET SCREEN SEARCH COMPLETED - 52 TO ITERATE

100.0% PROCESSED 52 ITERATIONS 49 ANSWERS  
SEARCH TIME: 00.00.01

L13 49 SEA SUB=L10 SSS FUL L4

=> D HIS

(FILE 'HOME' ENTERED AT 10:11:07 ON 08 OCT 2013)

FILE 'MARPAT' ENTERED AT 10:11:14 ON 08 OCT 2013

L1 STRUCTURE UPLOADED  
L2 1 S L1  
L3 40 S L1 FUL  
L4 STRUCTURE UPLOADED  
L5 1 S L4  
**L6 49 S L4 FUL** ← 通常のフルファイル検索  
L7 9 S L6 NOT L3  
L8 STRUCTURE UPLOADED  
L9 1 S L8  
L10 52 S L8 FUL  
L11 3 S L10 NOT L6  
L12 1 S L4 SAM SUB=L10  
**L13 49 S L4 FUL SUB=L10** ← サブセット検索

=> S L6 AND L13  
L14 49 L6 AND L13 ← 回答は同じです