A chemical company's view and expectations on Markush structure searching

International Conference on Trends for Scientific Information Professionals

14.-17.10.2012

Dr. Peter Geyer, BASF Group Information Center, BASF SE

BASF
The Chemical Company

Outline

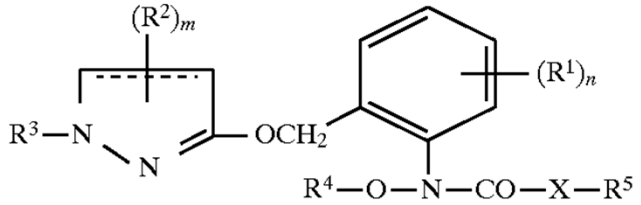
- Introduction
- Why Markush searching is necessary
- Regained interest in Markush data...but
- Our expectations

Introduction
What is a Markush structure

- Generic structural description of chemicals used in patent documents to cover more than one specific chemical
- Named after Dr. Eugene Markush who first successfully used such a type of claim in 1924
- Uses several types of variations (substituent, position, frequency and homology)

We claim:

1. A 2-[(dihydro)pyrazol-3'-yloxymethylene]anilide of the formula I



I

where -- is a single or double bond and the indices and the substituents have the following meanings:

n is 0, 1, 2, 3 or 4, it being possible for the substituents R^1 to be different if n is greater than 1;

m is 0, 1 or 2, it being possible for the substituents R^2 to be different if m is greater than 1;

X is a direct bond, O or NR^a;

R^a is hydrogen, alkyl, alkenyl, alkynyl, cycloalkyl or cycloalkenyl;

R^1 is nitro, cyano, halogen, unsubstituted or substituted alkyl, alkenyl, alkynyl, alkoxy, alkenyloxy, alkynyloxy or

in the case where n is 2, additionally is an unsubstituted or substituted bridge bonded to two adjacent ring

Introduction
How Markush structures are made searchable

Markush structures are indexed in the following systems

- Topological Search Systems (available since the end of the 1980ies)
  Marpat (Chemical Abstracts Service)
  Merged Markush Service (MMS, Thomson-Reuters/INPI/Questel)

- Fragmentation Code Systems (available since the 1960ies)
  Derwent CPI Fragmentation Codes
  IFI Claims Codes
  GREMAS

Recent Review see: Downs GM & Barnard JM, WIREs Comput Mol Sci, 2011, 1, 727–741

Why Markush searching is necessary

- Several databases cover specific chemicals of patents (CAS Registry, Derwent DCR, Reaxys, Chemspider, ...) ;
  also available to end-users (graphical search interfaces)

- Do not even comprehensively cover all specific chemicals described in a single patent document (esp. “prophetics” from tables)

- Searches there will not reveal patents covering chemical of interest ONLY by their Markush claims

- For commercial applications such patents are highly relevant (FTO)

Why Markush searching is necessary
Use case

- Is it possible to market the mixture of the fungicide Pyraclostrobin with the insecticide Fipronil?

- European Patent EP 1696728 B1 covers this mixture

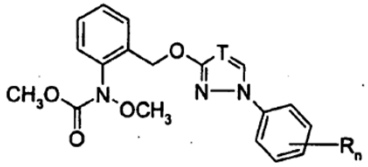
Why Markush searching is necessary
Use case

EP1696728
“Fungicidal mixtures based on carbamate derivatives and insecticides”

Claims

1. A mixture for crop protection, comprising as active components

a) carbamate derivatives of the formula I

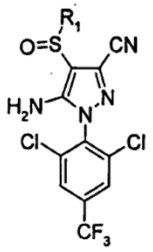


I

in which the substituents and the index have the following meaning:

TisCHorN
n is 0, 1 or 2
R is halogen, C_1-C_4-alkyl or C_1-C_4-haloalkyl, it being possible for the radicals R to be different when n is 2,

and
b) at least one compound of the formulae II



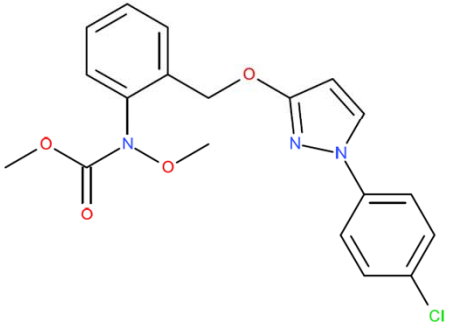
II

in which R_1 is C_1-C_4-alkyl or C_1-C_4-haloalkyl;
in a synergistic effective amount.

Why Markush searching is necessary
Use case

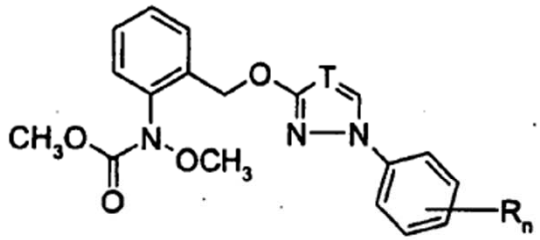
EP1696728
“Fungicidal mixtures based on carbamate derivatives and insecticides”

Pyraclostrobin:



Claims

1. A mixture for crop protection, comprising as active components

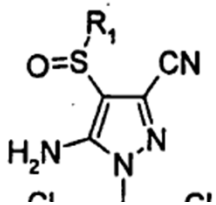


I

in which the substituents and the index have the following meaning:

TisCHorN
n is 0, 1 or 2
R is halogen, C_1-C_4-alkyl or C_1-C_4-haloalkyl, it being possible for the radicals R to be different when n is 2,

and
b) at least one compound of the formulae II



Why Markush searching is necessary
Use case

- Already searches using databases for specific chemicals (Registry/DCR) or using the common names will yield quite some hits, e.g.:

  Family searches Registry
  => CAplus + common names: 192 records (142 patent families)

  Patbase - patent fulltexts
  common names 1208 patent families

- Would such a search, using those databases only, reveal EP1696728?
  => NO!

Why Markush searching is necessary
Use case

- Fipronil mentioned with it's common name but Pyraclostrobin only shown in a table:

Table 1:

No.	T	R_n
I.32	CH	4-Cl

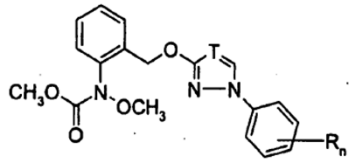
- "In a preferred embodiment the mixture comprises the compound I.32 and fipronil"

- Searches in Markush databases (like MMS and Marpat) will reveal such patents

Claims

1. A mixture for crop protection, comprising as active components

a) carbamate derivatives of the formula I

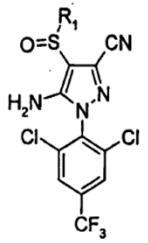


I

in which the substituents and the index have the following meaning:

TisCHorN
n is 0, 1 or 2
R is halogen, C_1-C_4-alkyl or C_1-C_4-haloalkyl, it being possible for the radicals R to be different when n is 2,

and
b) at least one compound of the formulae II



II

in which R_1 is C_1-C_4-alkyl or C_1-C_4-haloalkyl;
in a synergistic effective amount.

Why Markush searching is necessary
Use case

- Especially for questions like Freedom-To-Operate it is essential for chemical companies to also do searches in Markush databases
- Also for “just” filing a new patent application such a previous document would be highly relevant prior art
- Without Markush systems much broader/imprecise additional searches (keywords, codings, citations...) would be necessary (but not necessarily successful...)
- Markush databases are necessary to support important business processes in our companies!

Regained interest in Markush data
Some recent examples

- Thomson-Reuters: New search interface for MMS data together with ChemAxon
- Work of Digital Chemistry on interface for MMS data
- Chemical Abstracts: Marpat data available via SciFinder; New hit display on STN
- InfoChem: Project for automated extraction of Markush data (ChemProspector)
- Roche: Own software for enhanced display of MMS data (MarVis)

  ...

Regained interest in Markush data ...but...

- The Markush databases/indexing systems are all about 25 years old
- Only some minor developments since then
- Just one example: For searching MMS a Windows NT/95 software released 1999 is necessary to prepare the coding (Markush TOPFRAG)

Regained interest in Markush data
...but...

BASF
The Chemical Company

Kopie: Frau Dr. Schoch-Grübler, Herr Dr. Fairhurst 24.2.99
Herr Dr. Kallas, Herr Dr. Geiß
Herr Dr. Kneip, ZDW/C

Derwent Launches *New*

DERWENT
Scientific and Patent Information

Markush TOPFRAG™
version 3.10

ZDW/C-R

Dear TOPFRAG user

We are pleased to announce that we have produced a NEW version of TOPFRAG (version 3.10) which contains the following enhancements:

- New import features which allow structures to be imported from alternative drawing tools such as MACCS and ISIS in SD or MOL file format.
- Year 2000 compliant source code.
- A revised and updated user manual.
- Program modifications to achieve stable operation in Windows NT and Windows 95.
- The incorporation of additional routines to either halt progress or issue a warning when attempting to overlay nodes (to prevent strategy problems due to accidental overlay).

In addition, a number of "bugs" reported in previous versions have now been corrected. These include: the recall of saved strategies using Edit Strategy and the ability to include structures when printing a strategy regenerated from a saved structure. Various other strategy generation faults have also now been fixed.

Regained interest in Markush data ...but...

- The Markush databases/indexing systems are all about 25 years old
- Only some minor developments since then
- Just one example: For searching MMS a Windows NT/95 software released 1999 is necessary to prepare the coding (Markush TOPFRAG)
- TOPFRAG is not running on Windows 7 64-bit, a joint effort by PDG companies was necessary to convince Thomson to re-develop a 64-bit version.

Regained interest in Markush data ...but...

Haloalk(en/yn)yl:

Marpat

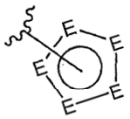
1. A compound having a structure corresponding to Formula (I):

$(A)\text{-}L_A\text{-}(B)\text{-}L_B\text{-}(C)\text{-}L_C\text{-}(D)$

(I)

and pharmaceutically acceptable salts, and prodrugs thereof, wherein:

A is:



wherein each E is independently N, NR, C, C

more than four E's are heteroatoms;

R is hydrogen, substituted or unsubstitute

alkenyl, substituted or unsubstit

unsubstituted alkoxy, substitu

substituted or unsubstituted

cycloalkyl, or substituted

each R^1 is hydrogen, halogen,

unsubstituted alkenyl

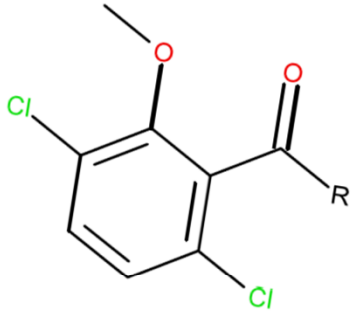
or unsubstituted al

substituted or unsubsti

amino, substituted or unsubsti

```
qt
n
gr
6-1:6-69,1-31:34-37-42-46,2-71,4-7-8,5-70,32-62,33-63
at
G1 7 & ARY 8 & G4 31 & G5 34 & G6 37 & G7 42 & HAL 46 & G8 62 & G9 63
G10 69 & G11 70 & G12 71
bo
no 1:6,1-6 & x 32-62,33-63
fs
1 32,33,3 & 5 8
cr
MON 8
tra
bt 1,2,3,4,5,6,32,33 & nt 8,46
gm
```

Derivatives of Dicamba:



On a generic level:

Any group/molecule – substituted by C=O, substituted by aryl (further substituted by one or more halogens and alkoxy)

Regained interest in Markush data ...but...

There are still areas of possible improvements left, e.g...

- Higher focus on precision (so far main focus on recall), e.g.

  - Require structure parts to match the core structure of hit patents

  - Matching of optional ⇔ mandatory structure parts

- Hit structure visualizing such details (e.g. different colors)

- Roles for Markush structures (claims as structure, process, different uses...)

Regained interest in Markush data ...but...

- Esp. for the traditional MMS search system on Questel we see some question marks on the horizon...

- The recent developments at Thomson and CAS have

  broadened the distribution of Markush data

  added some enhancements to hit highlighting

  ...but not majorly enhanced the use by information professionals

- Automated approaches are still far from replacing traditional indexing systems

Our expectations towards solution providers

- Keep offering Markush search systems
- Actively develop these also for professional use
- Discuss their development plans with us
- Discuss also with us in case they should question the future and/or the economics of these systems or their development

=> Providers are actively driving this topic forward!

Thank you very much
for your attention.

Questions?