

What's New in the CAS databases?

STN User Meeting, June 2008

# CAS file enhancements

- Prophetics added to CAS databases
- CAS<sup>SM</sup>/CAPlus<sup>SM</sup>
  - Patent content additions
  - New IPC custom display formats
  - Additional page images
- CAS REGISTRY<sup>SM</sup>
  - 10 digit CAS Registry Number<sup>®</sup> identifiers
  - Additional spectra
- Additional reactions in CASREACT<sup>®</sup>
- MARPAT<sup>®</sup> enhancements

## Substance coverage enhanced with additional *prophetic* substances in CAS files

- **Prophetic** substances are defined by CAS as:
  - **Specific** substances described but not characterised, appearing in the **examples**, but not appearing in the claims section
  - No additional supporting evidence such as a physical property or yield is available
- Substances are indexed if they are identified by
  - Name or structure
  - Appear in a table
  - Additional indexing appears for known substances reporting novel or new uses without substantiation

## Supplemental substance information registered from patents

- Prophetic substances indexed for English basic patents from US, CA, DE, FR, GB and English, French and German language basics from WO and EP
- New prophetic substances are assigned CAS RNs and indexed in CAS files
  - Prophetic substances are flagged with the **Prophetic (PRPH)** super role in CAS files
  - Reactions with prophetic substances as reactants, intermediates, or products are added to CASREACT
- Supplemental coverage of additional substances does not fall under CAS timeliness guarantee

# Prophetic substance examples

## Example 1

Preparation of 4-(3-trifluoromethylphenoxy)-2-(4-trifluoromethylphenyl)pyrimidine

**[0031]** 3,3-Dichloroacrolein (10 mmoles) diluted with acetonitrile (50 ml), is slowly added to a mixture consisting of a 4-trifluoromethylbenzamide (10 mmoles), 3-trifluoromethylphenol (11 mmoles), potassium carbonate (40 mmoles) and acetonitrile (100 ml), which is stirred under reflux. When the addition of 3,3-dichloroacrolein is completed additional 4-trifluoromethylbenzamide (0.5 mmoles) is added. The reaction mixture is stirred for 20 hours under reflux and subsequently cooled down to ambient temperature and filtered through silica. The organic phase is washed with ethyl acetate and concentrated in vacuo. The residue is purified by chromatography on A1203 (petrol ethers / ethyl acetate : 2 / 1) to yield 3.25 g (85 %) of the pure product having a melting point of 66 - 67°C.

**[0032]** Analogously are prepared

3-methyl-4-(3-trifluoromethylphenoxy)-2-(4-trifluoromethylphenyl)-pyrimidine,

5-methyl-4-(3-trifluoromethylphenoxy)-2-(4-trifluoromethylphenyl)-pyrimidine,

4-phenoxy-2-(4-trifluoromethylphenyl)-pyrimidine

# Prophetic substance examples (ii)

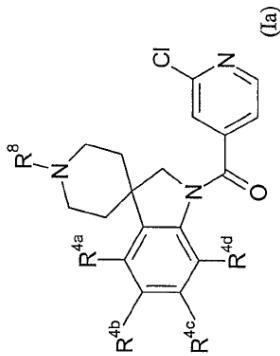


Table 1

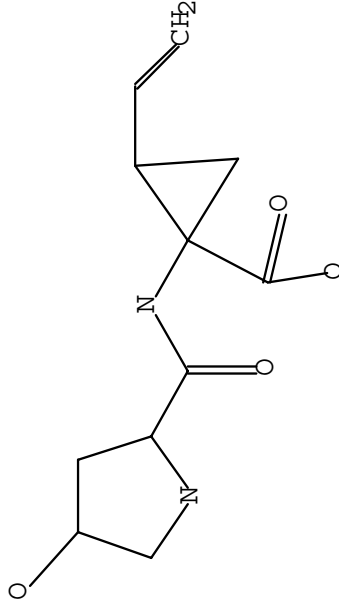
| Compound | R <sup>8</sup>            | R <sup>4a</sup> | R <sup>4b</sup> | R <sup>4c</sup> | R <sup>4d</sup> |
|----------|---------------------------|-----------------|-----------------|-----------------|-----------------|
| I-1      | Cinnamyl                  | H               | H               | H               | H               |
| I-2      | 4-chlorocinnamyl          | H               | H               | H               | H               |
| I-3      | 4-fluorocinnamyl          | H               | H               | H               | H               |
| I-4      | 4-nitrocinnamyl           | H               | H               | H               | H               |
| I-5      | 4-methoxycinnamyl         | H               | H               | H               | H               |
| I-6      | 4-methylcinnamyl          | H               | H               | H               | H               |
| I-7      | 4-trifluoromethylcinnamyl | H               | H               | H               | H               |
| I-8      | 4-cyanocinnamyl           | H               | H               | H               | H               |
| I-9      | 2,4-dichlorocinnamyl      | H               | H               | H               | H               |
| I-10     | 2,4-difluorocinnamyl      | H               | H               | H               | H               |
| I-11     | cinnamyl                  | Cl              | H               | H               | H               |
| I-12     | 4-chlorocinnamyl          | Cl              | H               | H               | H               |

Compounds without MS data

| Table 3     |  |
|-------------|--|
| Compound No | MS data  |
| I-1         | 444<br>(95%),<br>446<br>(100%)                               |
| I-2         | 478<br>(100%),<br>480<br>(70%),<br>482 (15%)                 |
| I-3         | 462<br>(100%),<br>464 (95%)                                  |
| I-4         | 489<br>(100%),<br>491 (70%)                                  |
| I-5         | 147<br>(100%),<br>474<br>(30%),<br>476 (80%)                 |
| I-12        | 512<br>(95%),<br>514<br>(100%),<br>516<br>(35%),<br>518 (5%) |

# Exemplified prophetics: search example

Uploading C:\CASNC\STN Express\Queries\test1.str  
L1 STRUCTURE UPLOADED



Structure attributes must be viewed using STN Express query preparation.

```
=> S L1 SSS SAM  
L2          50 SEA SSS SAM L1  
  
=> S L1 SSS FUL  
L4          1145 SEA SSS FUL L1  
  
=> FILE CAPLUS  
  
=> S L4/PREP  
L5          80 L4/PREP  
  
=> S L4/PREP (NOTL) PRPH/RL  
L6          80 L4/PREP (NOTL) PRPH/RL  
  
=> S L4/PREP NOT L4/PRPH  
L7          78 L4/PREP NOT L4/PRPH
```

The NOT operator would remove those patents where compounds in the structure search result set (L4) would include both exemplified prophetics as well as claimed compounds. Please use (NOTL) to remove exemplified prophetics if necessary.

L8 ANSWER 1 OF 2 CAPLUS COPYRIGHT 2008 ACS on STN  
 AN 2008:42485 CAPLUS Full-text  
 TI Preparation of proline dipeptides and analogs as inhibitors of hepatitis c  
 virus replication

IN Blatt, Lawrence M.; Seiwert, Scott; Beigelman, Leonid; Kercher, Timothy;  
 Kennedy, April L.; Andrews, Steven W.  
 PA Intermune, Inc., USA; Array Biopharma, Inc.  
 SO PCT Int. Appl., 126pp.  
 CODEN: PIXXD2

DT Patent  
 LA English  
 FAN.CNT 1

| PATENT NO.      | KIND | DATE     | APPLICATION NO. | DATE     |
|-----------------|------|----------|-----------------|----------|
| WO 2008005511   | A2   | 20080110 | WO 2007-US15530 | 20070605 |
| US 2008019942   | A1   | 20080124 | US 2007-773912  | 20070705 |
| US 2006-818914P | P    | 20060705 |                 |          |
| US 2006-819128P | P    | 20060706 |                 |          |

o o o

IT 862119-83-3P

RL: PRPH (Prophetic); RCT (Reactant); SPN (Synthetic  
 preparation); PREP (Preparation); RACT (Reactant or reagent)  
 (intermediate; preparation of proline dipeptides and analogs as inhibitors  
 of hepatitis c virus replication)

IT 98977-34-5P, 1-tert-Butyl 3-ethyl 4-oxopiperidine-1,3-dicarboxylate  
 862119-82-2P 1000994-18-2P 1000994-19-3P 1000994-20-6P  
 1000994-23-9P 1000994-24-0P 1000994-25-1P 1000994-26-2P  
 1000994-27-3P

RL: RCT (Reactant); SPN (Synthetic preparation); PREP  
 (Preparation); RACT (Reactant or reagent)  
 (intermediate; preparation of proline dipeptides and analogs as inhibitors  
 of hepatitis c virus replication)



## Exemplified prophetics: implications for searching

- The PRPH role indexing only applies to the basic patent
- Use caution – later members in a family may claim a substance identified with the PRPH role in the basic
- Useful for identifying substances which are disclosed, but perhaps not claimed
- PRPH role is also available for searching in the REGISTRY file and in CASREACT
- Began 12 Dec 2007 – backfile indexing is planned, but currently with no definitive timeline

## CAS provides more Asian patent information with industry-leading currency

- Chemistry-related **South Korean** patent applications and granted patents are available in CAplus at least 3 months sooner than from other services
  - South Korean utility models from 2006 being added
- English-language bibliographic information and abstracts are available for **Chinese** invention patents and utility models in CAplus within 14 days of publication
  - Full indexed records now available in less than 50 days from receipt on average
- Chinese granted utility models from 2006 being added
- **Japanese** utility models from 2006 are being added
- Bibliographic data and abstracts are provided for **Indian** patents within 14 days of publication
- CAS Company Name Thesaurus includes more Asian company names as a result of the increased coverage of Asian patents

## CA/CAplus enhanced with new custom IPC display formats

- Two new custom display formats for International Patent Classification (IPC) codes have been added to CA/CAplus
  - IPC.UNIQ – displays a set of unique IPC codes for a basic patent and equivalents (if any)
    - Compact display of IPC information
    - CLASS and IPC.TAB formats are available for more extensive IPC information, including IPC metadata
  - IPC.HIT – displays IPC codes matching those specified in the search query
    - Quickly determine why a specific record has been retrieved with a complex search strategy

# Examples of the IPC.UNIQ and IPC.HIT format content

=> **D IPC.UNIQ**

L1 ANSWER 1 OF 480 CAPLUS COPYRIGHT

IPC.UNIQ

A61K0038-00; A61K0039-00; C07K0014-435; C07K0014-705; C07K0016-18;  
C07K0016-28; A61D; C12N0015-18; C12N0015-16; A61P0019-08; A61P0019-00;  
A61K0038-18; C07K0016-22; A61K0039-395; C07K0014-51; C12N0015-09;  
A61K0039-39; A61P0019-10; A61P0043-00; C07K0014-47; C12N0001-10;  
C12N0001-11; C12N0001-15; C12N0001-19; C12N0001-21; C12N0005-10;  
C12P0021-02; C12P0021-08; A61K0038-17; C07H0021-00; C07H0021-04;  
C12P0021-06

=> **D IPC.HIT**

L2 ANSWER 1 OF 23 CAPLUS COPYRIGHT 2008 ACS on STN

IPC.HIT

A61K0038-00; A61K0039-395; C07H0021-00

Displays a set of unique IPC code  
for a patent and its equivalents.

## Additional page images added to CA/CAplus

- Page images from 1967-1998 (Volumes 66-129) are being added to CA files
- Page images provide access to print graphics not available in electronic file
  - Reaction schemes
  - Markush structures
  - Tables of data
- Records with page images are identified with a GI field display in the abstract

# BIB ABS display includes GI information

L2 ANSWER 116 of 1023 HCAPLUS COPYRIGHT 2008 ACS on STN  
AN 1968:106050 HCAPLUS  
DN 68:106050  
OREF 68:20503a,20506a  
TI Utilization of 3-hydroxy-2-naphthoic acid. I. Azo dyes derived from  
3-(2-benzoxazolyl)-2-naphthol  
AU Matsui, Koji; Kuroda, Toshio  
CS Gumma Univ., Kiryu, Japan  
SO Kogyo Kagaku Zasshi (1967), 70(11)  
CODEN: KGKZA7; ISSN: 0368-5462  
DT Journal  
LA Japanese  
GI For diagram(s), see printed CA Issue.  
AB Condensation of 3,2-HOC10H6CO2H with o-C6H4(NH2)2 in PhCl in the  
presence of SO2Cl2 gives 3-(2-benzimidazolyl)-2-naphthol (I) in 80-5%  
yield. Condensation with o-H2NC6H4OH in PhCl in the presence of PCl3  
gives 3-(2-benzoxazolyl)-2-naphthol (II) in 85-90% yield. The  
tabulated azo dyes of the general formula III were synthesized by the  
coupling of diazonium or tetrazonium salts derived from aromatic  
primary amines containing no SO3H or CO2H groups with I or II.  
[TABLE OMITTED] III (X = NH) dye polyacrylonitrile fibers brilliant  
shades of good fastness.

ABS display includes GI information if  
a record has a page image available.

=> D L2 116 PAGE

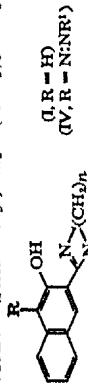
START LOCAL KERMIT RECEIVE PROCESS

BINARY DATA HAS BEEN DOWNLOADED TO MULTIPLE FILES 'IMAGE...TIF'

## 40—DYES, FLUORESCENT BRIGHTENING AGENTS, AND PHOTSENSITIZERS

J. J. LEAVITT

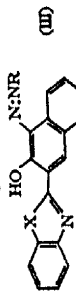
106050k Utilization of 3-hydroxy-2-naphthoic acid. I. Azo dyes derived from 3-(2-benzoxazolyl)-2-naphthol. Koji Matsui and Toshio Kuroda (Gunma Univ., Kiryu, Japan). *Kogyo Kagaku Zasshi* 70(12), 2325-9(1967)(Japan); cf. following abstr. Condensation of 3,2-HOC<sub>10</sub>H<sub>6</sub>CO<sub>2</sub>H with *o*-C<sub>6</sub>H<sub>4</sub>(NH<sub>2</sub>)<sub>2</sub> in PhCl in the presence of SO<sub>2</sub>Cl<sub>2</sub> gives 3-(2-benzimidazolyl)-2-naphthol (I) in 80-5% yield. Condensation with *o*-H<sub>2</sub>N<sub>2</sub>C<sub>6</sub>H<sub>4</sub>OH in PhCl in the presence of PCl<sub>5</sub> gives 3-(2-benzoxazolyl)-2-naphthol (II) in 85-90% yield. The tabulated azo dyes of the general formula



(III) in 21% yield. The tabulated IV were synthesized by

| X   | RNH <sub>2</sub>  | % Yield | m.p. (decomp.)               | λ <sub>max</sub> | λ <sub>max</sub> × 10 <sup>-4</sup> | Shade on          |
|-----|---|---------|------------------------------|------------------|-------------------------------------|-------------------|
| NH  | PhNH <sub>2</sub>   | 100     | 242-2.5° (PhMe)              | 508              | 1.88                                | red               |
| NNH | <i>p</i> -MeOC <sub>6</sub> H <sub>4</sub> NH <sub>2</sub>  | 100     | 194-5° (PhMe)                | 530              | 1.63                                | deep red          |
| NNH | <i>p</i> -MeC <sub>6</sub> H <sub>4</sub> NH <sub>2</sub>   | 58.6    | 263° (CHCl <sub>3</sub> )    | 498              | 2.02                                | violet-red        |
| NNH | 2,4-Me <sub>2</sub> ( <i>o</i> -N)C <sub>6</sub> H <sub>3</sub> NH <sub>2</sub>                   | 70.5    | 273° (HCONMe <sub>2</sub> )  | 508              | 2.02                                | violet-red        |
| NNH | 2,4-Me <sub>2</sub> ( <i>o</i> -N)C <sub>6</sub> H <sub>3</sub> NH <sub>2</sub>                   | 83.6    | 301° (HCONMe <sub>2</sub> )  | 522              | 2.89                                | deep violet-red   |
| NNH | 2,4-Me <sub>2</sub> ( <i>o</i> -N)C <sub>6</sub> H <sub>3</sub> NH <sub>2</sub>                   | 100     | 226° (HCONMe <sub>2</sub> )  | 515              | 1.85                                | deep violet-red   |
| NNH | 5,2,4-Me <sub>3</sub> (MeO)(Bz <sub>2</sub> N)C <sub>6</sub> H <sub>2</sub> NH <sub>2</sub> (VI)  | 76.9    | 264° (PhCl)                  | 548              | 2.50                                | dull violet       |
| NNH | 5,2,4-Me <sub>3</sub> (MeO)(Bz <sub>2</sub> N)C <sub>6</sub> H <sub>2</sub> NH <sub>2</sub> (VII) | 76.8    | 272° (PhMe)                  | 573              | 1.88                                | dark blue         |
| NNH | 1-C <sub>10</sub> H <sub>7</sub> NH <sub>2</sub> (VIII)   | 100     | 257° (HCONMe <sub>2</sub> )  | 538              | 2.37                                | dark violet       |
| NNH | 4-H <sub>2</sub> N <sub>2</sub> C <sub>6</sub> H <sub>4</sub> N <sub>2</sub> (IX)                 | 76.8    | 294° (HCONMe <sub>2</sub> )  | 570              | 3.14                                | dull blue-violet  |
| NZH | PhNH <sub>2</sub> (MeO)X  | 49.7    | >284°                        | 605              | 3.35                                | dull blue         |
| OO  | 2,5-Me <sub>2</sub> ( <i>o</i> -N)C <sub>6</sub> H <sub>3</sub> NH <sub>2</sub>                   | 95.1    | 201° (HCONMe <sub>2</sub> )  | 505              | 2.18                                | red-orange        |
| OO  | V   | 68.1    | 255° (HCONMe <sub>2</sub> )  | 507              | 1.66                                | bright red        |
| OO  | VI  | 75.3    | 285° (HCONMe <sub>2</sub> )  | 520              | 2.09                                | yellow-orange     |
| OO  | VII   | 88.9    | 285° (Me <sub>2</sub> CO)    | 517              | 1.55                                | bright yellow-red |
| OO  | VIII  | 85.5    | 287° (HCONMe <sub>2</sub> )  | 538              | 2.18                                | grayish violet    |
| OO  | IX  | 189.5   | 257° (HCONMe <sub>2</sub> )  | 535              | 1.83                                | deep violet-red   |
| OO  | X   | 50.0    | 265° (HCONMe <sub>2</sub> )  | 530              | 3.22                                | violet-red        |
| OO  | XI  | 100     | >340° (HCONMe <sub>2</sub> ) | 578              | 3.07                                | dull red-violet   |
| OO  | XI  | 100     | >340° (HCONMe <sub>2</sub> ) | 578              | 3.07                                | pale blue violet  |

III were synthesized by the coupling of diazonium or tetrazolium salts derived from aromatic primary amines contg. no SO<sub>3</sub>H or



CO<sub>2</sub>H groups with I or II. III (X = NH) dye polyacrylonitrile fibers brilliant shades of good fastness. S. Inokawa

Data table is available in the print page image.

PAKISTAN. PARISHAD J. Sci. 1706. Res. 10(3), 240-(1957)(Eng.)

## CAS REGISTRY enhanced with additional spectra

- Since the beginning of 2008, ~204,000 experimental spectra have been added to REGISTRY, including:
  - 95,000 proton NMR spectra
  - 31,000 carbon-13 NMR spectra
  - 78,000 mass spectra
- REGISTRY now includes >480,000 experimental spectra

# CAS begins assigning 10 digit CAS Registry Numbers

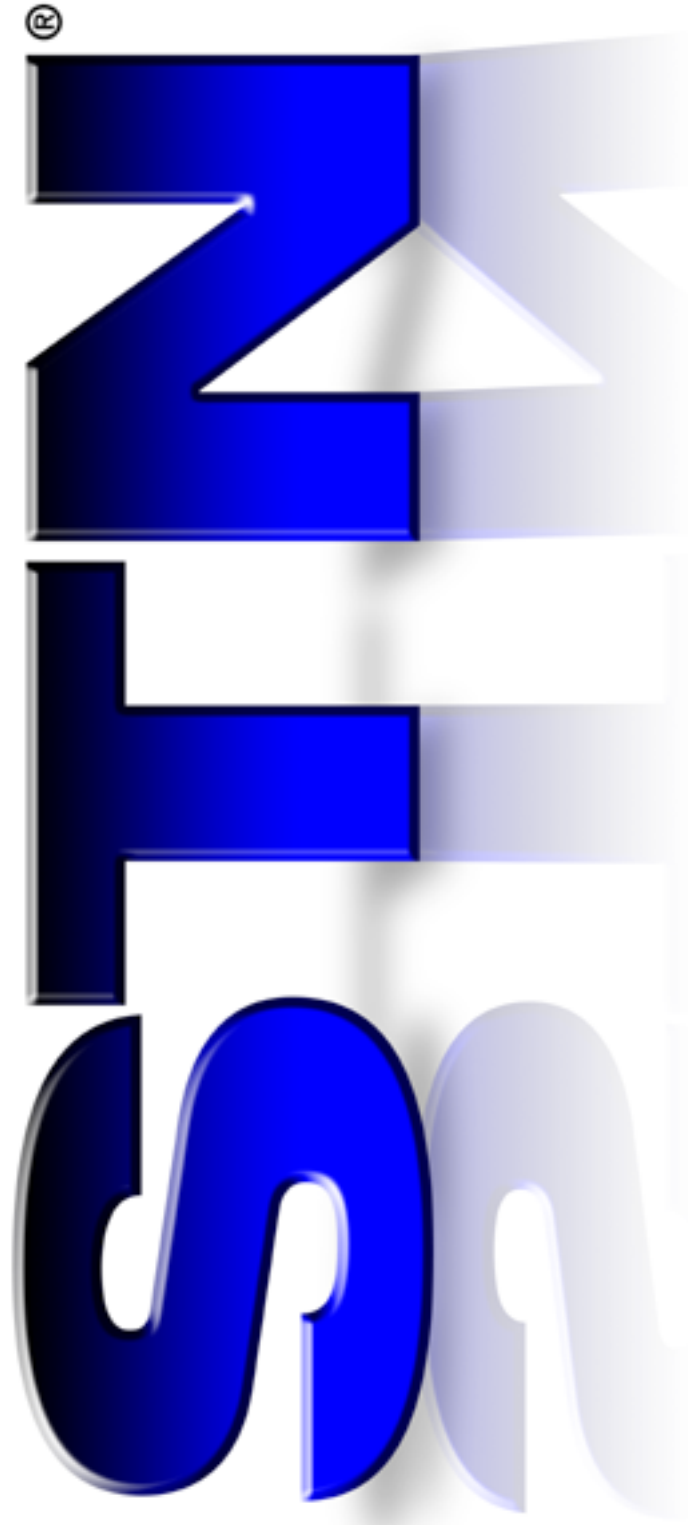
- To accommodate continuing growth of substance information in REGISTRY, 10 digit CAS Registry Numbers are assigned
- The 10th digit appears in the left-most position in the CAS RN
  - eg new CAS RNs have the following form  
XXXXXXXXXX-XX-X
- Upgrade to the latest version of STN Express to take advantage of CAS RN features such as
  - Hyperlinks
  - Custom Tables and Reports
  - BLAST reports
  - Command file scripts

## CASREACT now has over 15 million reactions

- Nearly 18,000 new documents have been added to CASREACT which include
  - Over 270,000 new single step reactions
  - CAplus bibliography, abstract, and indexing
- These reactions were extracted from INFOCHEM data from 1992-1999 and appear in records from Volume 116 to 131

## MARPAT searching has been enhanced

- The following enhancements were recently implemented in MARPAT:
  - The number of searches that return at least one incomplete result has been reduced by more than 60%
  - The iteration and answer limits have been increased from 100,000 to 200,000 for both online and batch searches



What's New in the CAS databases?

STN User Meeting, June 2008