Obecné principy chemických strukturních bází dat – předmět projektu VaVpI ChemEIZ

Jaroslav Šilhánek

Projekt ChemEIZ, OP VaVpI
CZ.1.05/3.2.00/12.0231

$C_9H_{11}BrN_2$

Benzaldehyde, 2-bromo-
dimethylhydrazone *[148492-28-8]*, 95373m

$C_9H_{11}BrN_2O$

Acetamide, *N*-(5-bromo-4,6-dimethyl-2-pyridinyl)-
[99179-94-9], P 95341z

Butanamide, *N*-(5-bromo-2-pyridinyl)-
[148612-12-8], P 95341z

Propanamide, *N*-(5-bromo-4-methyl-2-pyridinyl)-
[148612-13-9], P 95341z

$C_9H_{11}BrN_2O_2$

Benzenamine, 3-(bromomethyl)-*N*,*N*-dimethyl-=
2-nitro- *[149155-80-6]*, 117176d

prepn. of and stomach acid secretion and ATPas inhibition by, structure in relation to, **112:** 48250q

2*H*-Imidazo[4,5-*f*]quinoline-2-thione

——, 1,3-dihydro- *[107002-97-1]*

prepn.

alkylation, and acylation of, with alkyl or acyl halides, **110:** 212681n

reaction with chloroacetic acid, and cyclization with dibromoethane, **106:** 102161c

prepn. and *N*- and *S*-methylation and trideuteriomethylation of, **109:** 210963m

2*H*-Imidazo[4,5-*h*]quinoline-2-thione

——, 1,3-dihydro-3,7-dimethyl- *[132500-82-4]*

prepn. of, as herbicide, **114:** P 122372r

Strukturní báze dat

=

Grafická representace struktur chemických sloučenin (+ další informace)

Reakční báze dat

=

Grafická representace strukturních změn chemických sloučenin (+ plus další údaje)

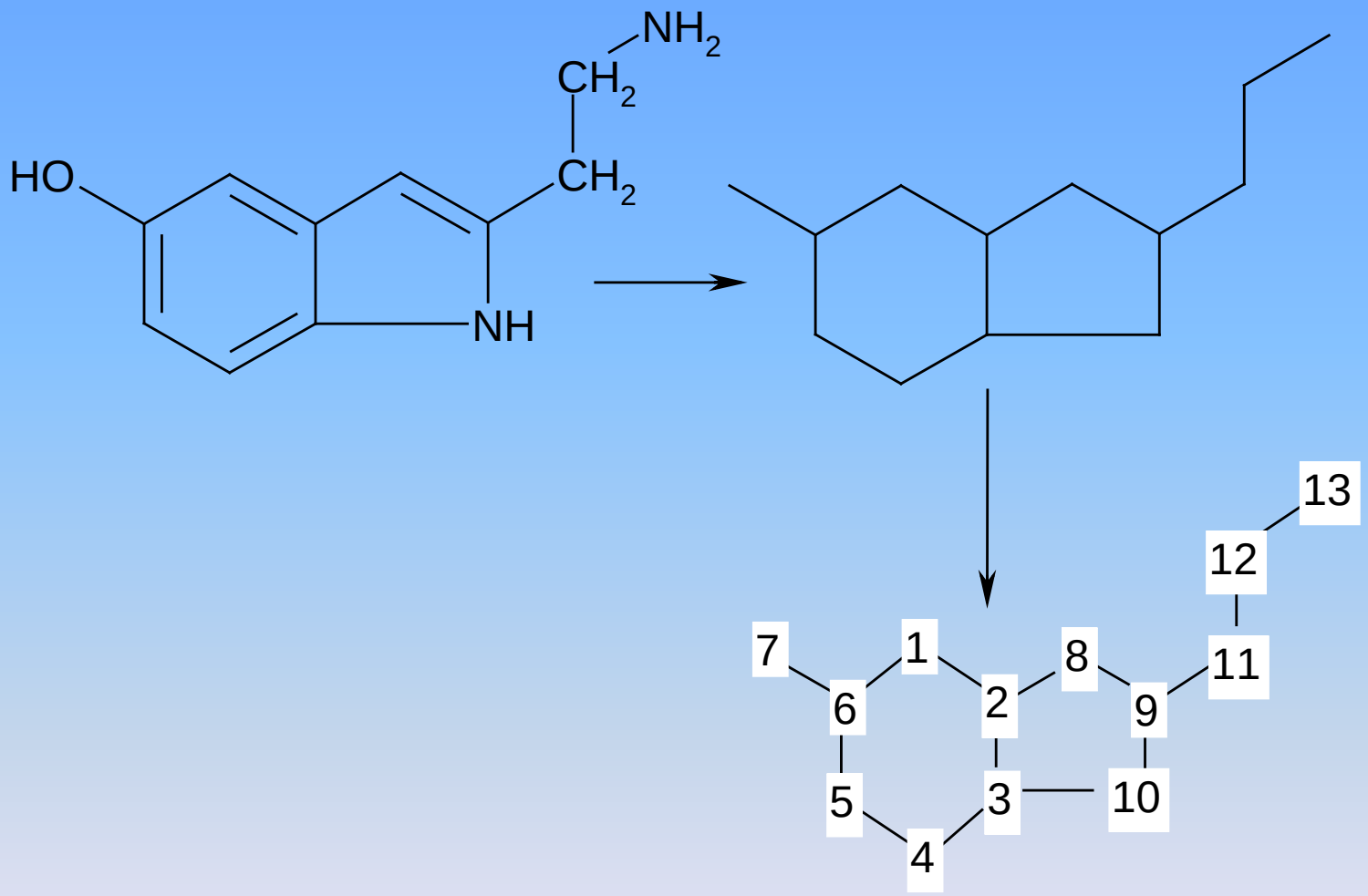
Projekt ChemEIZ, OP VaVpI
CZ.1.05/3.2.00/12.0231

Výchozí předpoklady pro tvorbu strukturních bází dat:

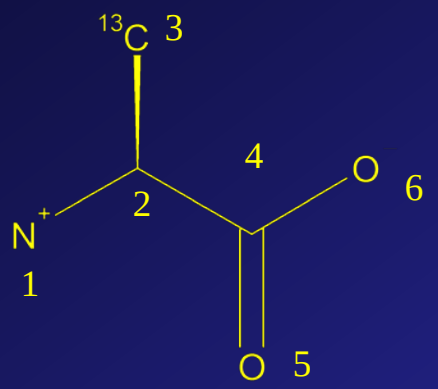
- *Volba základního pracovního principu*
- *Dovedení principu do použitelné podoby*
- *Vytvoření pracovních nástrojů pro tvorbu záznamů*
- *Motivace, smysl vytváření funkčních strukturních bází dat*

Stručný historický exkurs

- První práce o principu grafické representace: Science 126, 814 (1957)
- CAS – 60. léta, práce na *CAS Chemical Registry System*
- Důraz na nalezení schůdného postupu pro substrukturní vyhledávání
- Vývoj v soukromém sektoru, skupina BASIC = Basel Information Center for Chemistry (Ciba-Geigy, Hoffmann-LaRoche, Sandoz)
- 70. léta, spolupráce soukromého sektoru s CAS, funkční řešení



Projekt ChemEIZ, OP VaVpI
CZ.1.05/3.2.00/12.0231



.mol file

```
-ISIS- 02190318482D   6  5  0  0  0  0  0  0  0  0999 V2000
-1.1417   -2.4333    0.0000 N   0  3  0  0  0  0  0  0  0  0  0  0
 0.3917   -1.5333    0.0000 C   0  0  1  0  0  0  0  0  0  0  0  0
 0.3833    0.2458    0.0000 C   1  0  0  0  0  0  0  0  0  0  0  0
 1.9292   -2.4250    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
 1.9208   -4.2042    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
 3.4667   -1.5250    0.0000 O   0  5  0  0  0  0  0  0  0  0  0  0
 1  2  1  0  0  0  0
 2  4  1  0  0  0  0
 4  5  2  0  0  0  0
 2  3  1  1  0  0  0
 4  6  1  0  0  0  0
M  CHG  2   1   1   6  -1
M  ISO  1   3  13
M  END
```

Projekt ChemEIZ, OP VaVpI
CZ.1.05/3.2.00/12.0231

Vyhledávání struktur, resp. substruktur

- 1. krok

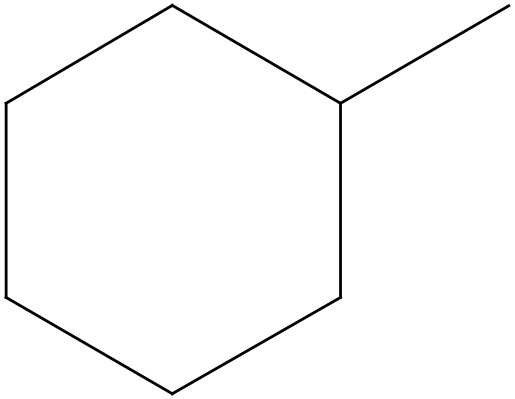
  vytvoření souboru potenciálních kandidátů – screening, „*screens*“
- 2. krok

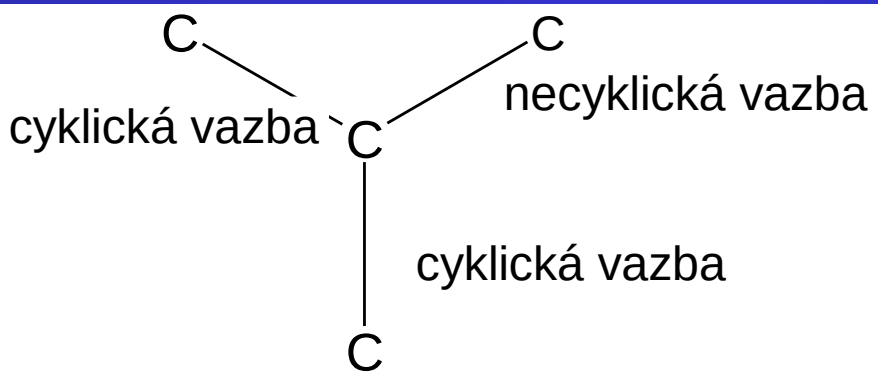
  „*atom-by-atom-search*“

Přehled některých typů „screens“

AA	Augmented Atom	Rozšíření daného atomu danou vazbou
AS	Atom Sequence	Linearní sekvence atomů a vazeb 4 až 6
BS	Bond Sequence	Lineární sekvence vazeb od 3 do 5
CS	Connectivity Sequence	Počet lineárních nevodíkových připojení
RC	Ring Count	Minimální počet kruhů

Augmented Atoms



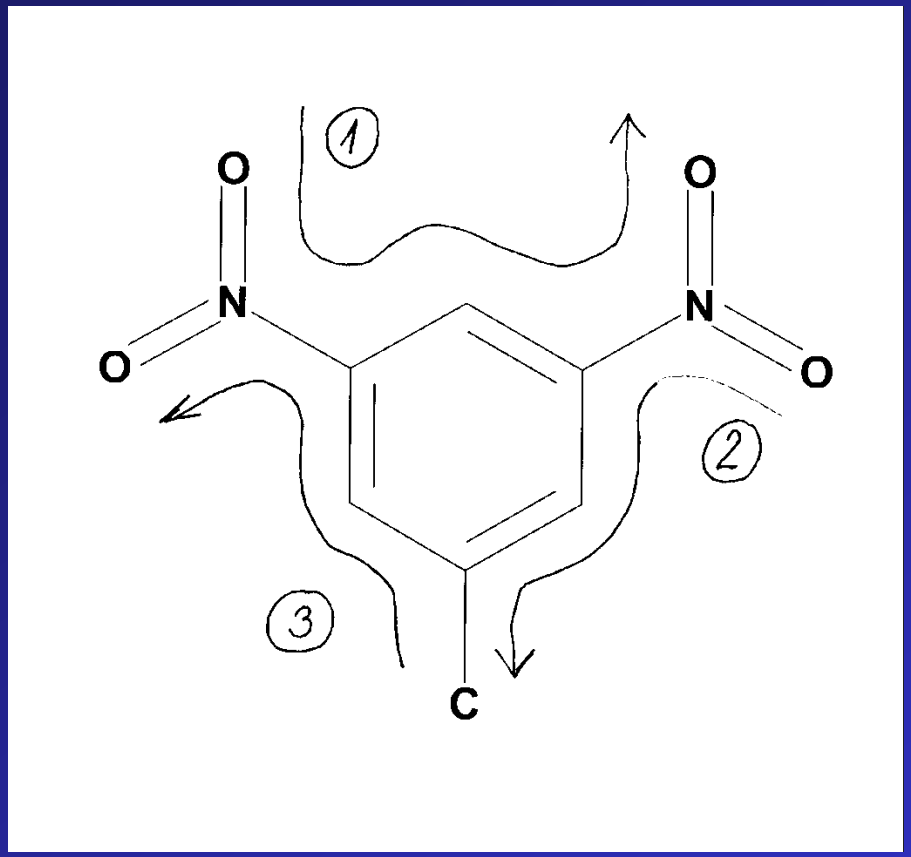


1031 AA C *1C *1C -1C 15.38 %

Aromatická varianta

1044 AA C *4C *4C -1C 38.04%

Atom Sequence, Bond Sequence



1

367 AS O – N – C * C * C - N

1,56 %

2

882 BS A – 2A – 1A * 4A * 4A – 1A

5,66 %

3

71 AS C – C * C * C – N – O

0,93 %

Strukturní a reakční báze dat jako pracovní nástroj

- Prázdné strukturní a reakční báze dat k disposici chemikům

  systém ISIS (ISIS/Base), systém ChemOffice (ChemFinder) aj.
- Uzavřené strukturní a reakční báze dat obsahující data

  báze CAS REGISTRY aj.
- Systémy s možností otevření pro vlastní data nebo kombinující obě možnosti

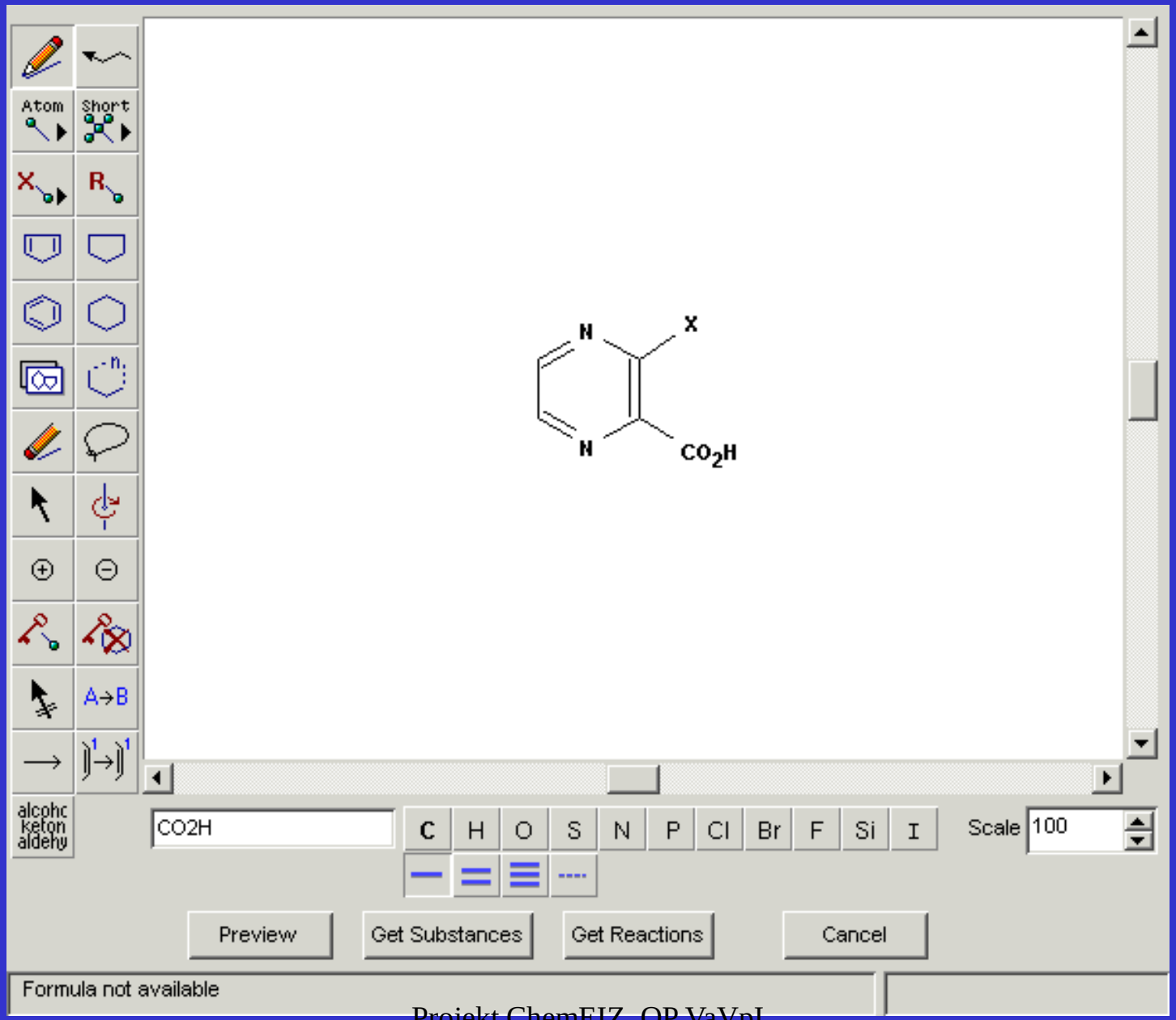
  systém CrossFire pro chemický průmysl, systém ISIS jako prostředí pro některé reakční a strukturní báze dat

Motivace pro vytváření a využívání strukturní representace chemických sloučenin

- Registrace a evidence chemických sloučenin
- Nástroj pro studium vztahů mezi strukturou a vlastnostmi v nejširším slova smyslu
- Reakční báze dat

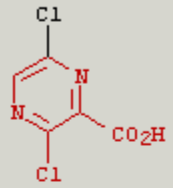
CAS Chemical Registry System

- do r. 1967
  - Author Index
  - Patent Index
  - Subject Index
  - General Subject Index
  - Formula Index
  - Chemical Substance Index
- od r. 1967
  - Bibliographic File
  - báze **CA, CAplus**
  - Registry File,
  - báze **REGISTRY**


Atom
Short
X
R
A→B
alcohc
keton
aldehy
CO2H
C
H
O
S
N
P
Cl
Br
F
Si
I
Scale
100
Preview
Get Substances
Get Reactions
Cancel
Formula not available

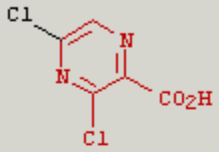
New Task | Back | Forward | Print | Save As | Full Text | Prefs | Database | History | Internet | Help | Exit

356783-15-8



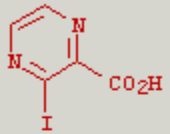
~1 Reference
REGISTRY

312736-49-5



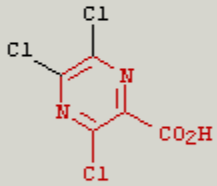
~5 References
REGISTRY

212471-40-4



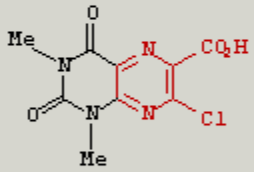
~2 References
REGISTRY

179754-49-5



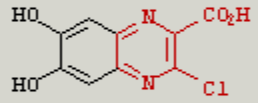
~1 Reference
REGISTRY

126889-70-1



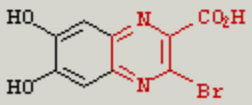
~1 Reference
REGISTRY

122234-94-0

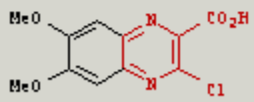


~1 Reference
REGISTRY

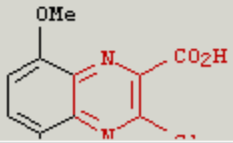
122234-93-9



122234-54-2



99852-30-9



Get References | Analyze or Refine Substances | Back

Substances 1-9 of 16

Registry Number: 312736-49-5

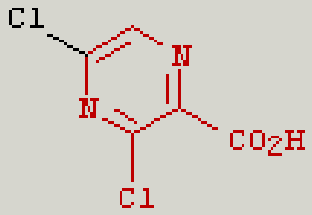
CA Index Name: Pyrazinecarboxylic acid, 3,5-dichloro- (9CI)

Other Names: 2-Carboxy-3,5-dichloropyrazine; 3,5-Dichloropyrazine-2-carboxylic acid

Formula: C5 H2 Cl2 N2 O2

STN Files: CAPLUS, CA, TOXCENTER

(Additional Information is available through STN International. Contact your information specialist, a local CAS representative, or the CAS Help Desk for Assistance)



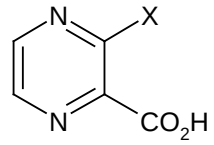
FILE 'REGISTRY' ENTERED AT 10:23:14 ON 18 FEB 2003

COPYRIGHT (C) 2003 American Chemical Society (ACS)

Uploading Halopyr2.str

L1 STRUCTURE UPLOADED

=> Display Query



=> S L1

SAMPLE SEARCH INITIATED 10:25:35 FILE 'REGISTRY'

SAMPLE SCREEN SEARCH COMPLETED - 19 TO ITERATE

100.0% PROCESSED 19 ITERATIONS 2 ANSWERS

SEARCH TIME: 00.00.01

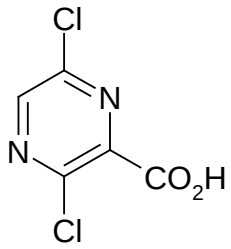
FULL FILE PROJECTIONS: ONLINE **COMPLETE**

BATCH **COMPLETE**

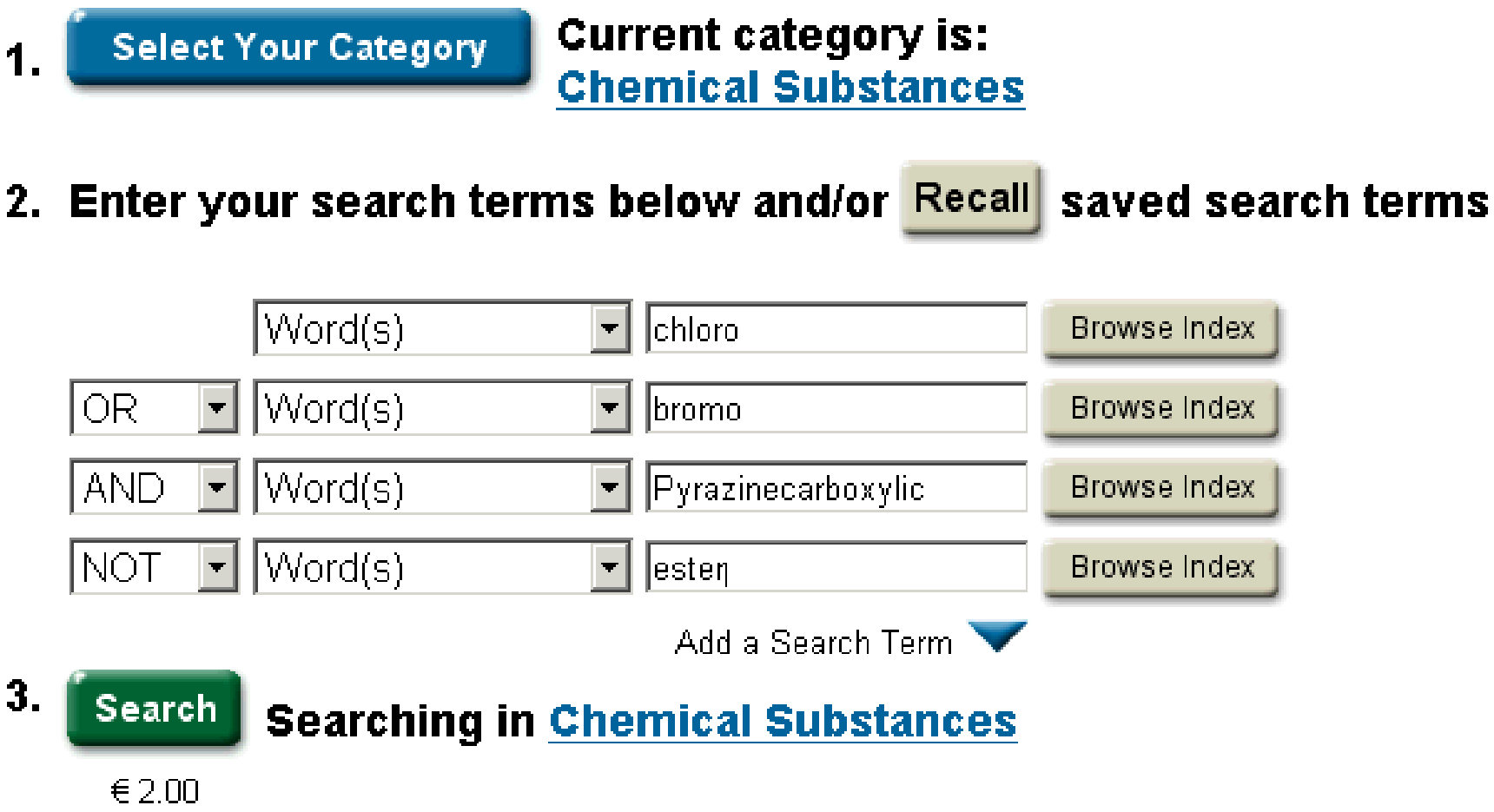
PROJECTED ITERATIONS: 119 TO 641

PROJECTED ANSWERS: 2 TO 124

```
FILE 'REGISTRY' ENTERED AT 12:03:45 ON 18 FEB 2003
COPYRIGHT (C) 2003 American Chemical Society (ACS)

L1      STRUCTURE UPLOADED
=> S L1 Ful
FULL SEARCH INITIATED 12:04:52 FILE 'REGISTRY'
FULL SCREEN SEARCH COMPLETED -    313 TO ITERATE

100.0% PROCESSED     313 ITERATIONS                              15 ANSWERS
SEARCH TIME: 00.00.01

L2           15 SEA SSS FUL L1

L2   ANSWER 1 OF 15  REGISTRY  COPYRIGHT 2003 ACS
RN   356783-15-8  REGISTRY
CN   Pyrazinecarboxylic acid, 3,6-dichloro- (9CI)  (CA INDEX NAME)
OTHER NAMES:
CN   3,6-Dichloro-2-pyrazinecarboxylic acid
FS   3D CONCORD
MF   C5 H2 Cl2 N2 O2
LC   STN Files:   CA, CAPLUS
```



```
=> d cost
COST IN EUROS                                          SINCE FILE      TOTAL
                                                            ENTRY    SESSION
CONNECT CHARGES                                              1,02       1,27
NETWORK CHARGES                                              0,15       0,20
SEARCH CHARGES                                             148,30     148,30
DISPLAY CHARGES                                              8,45       8,45
                                                           -------    -------
FULL ESTIMATED COST                                        157,92     158,22

IN FILE 'REGISTRY' AT 12:05:24 ON 18 FEB 2003


=> logoff
ALL L# QUERIES AND ANSWER SETS ARE DELETED AT LOGOFF
LOGOFF? (Y)/N/HOLD:.
COST IN EUROS                                          SINCE FILE      TOTAL
                                                            ENTRY    SESSION
FULL ESTIMATED COST                                        157,92     158,22
```


1.
Select Your Category
Current category is:
Chemical Substances
2. Enter your search terms below and/or Recall saved search terms
Word(s)
chloro
Browse Index
OR
Word(s)
bromo
Browse Index
AND
Word(s)
Pyrazinecarboxylic
Browse Index
NOT
Word(s)
ester
Browse Index
Add a Search Term
3.
Search
Searching in Chemical Substances
€ 2.00

419 answers in REGISTRY

419 total hits

[Hide Database Info.]

Too many answers? Refine Your Search

Clear	**CA Index Names from REGISTRY** in **Most Recent Order**	Molecular Formulas
☐ 1	**Pyrazinecarboxylic acid, 2-[(4-chloro-3,5-dimethylphenoxy)acetyl]hydrazide (9CI)** [€ 5.37]	C15 H15 Cl N4 O3
☐ 2	**Pyrazinecarboxylic acid, [(5-chloro-3-methyl-1-phenyl-1H-pyrazol-4-yl)methylene] hydrazide (9CI)** [€ 5.37]	C16 H13 Cl N6 O
☐ 3	**Pyrazinecarboxylic acid, 3-amino-6-bromo- (9CI)** [€ 5.37]	C5 H4 Br N3 O2
☐ 4	**Pyrazinecarboxylic acid, [[4-[(2,4-dichlorobenzoyl)oxy]-3-ethoxyphenyl]methylene] hydrazide (9CI)** [€ 5.37]	C21 H16 Cl2 N4 O4
☐ 5	**Pyrazinecarboxylic acid, [[4-[(3-bromobenzoyl)oxy]-3-methoxyphenyl]methylene] hydrazide (9CI)** [€ 5.37]	C20 H15 Br N4 O4

Results for Search Question:

chloro OR bromo AND Pyrazinecarboxylic NOT ester NOT hydrazide

Save **these search terms for future use**

Total Hits: **122** [Show Answers By Database]

Too many answers? Refine Your Search

Clear	**CA Index Names from REGISTRY** in **Most Recent Order**
☐ 1	Pyrazinecarboxylic acid, 3-amino-6-bromo- (9CI) [€ 5.37]
☐ 2	Pyrazinecarboxylic acid, 3-[(4-chloro-1,6-dihydro-5-hydroxy-6-oxo-2-pyridinyl)azo]- (9CI) [€ 5.37]
☐ 3	Pyrazinecarboxylic acid, 3-[[(4'-chloro[1,1'-biphenyl]-4-yl)amino]carbonyl]- (9CI) [€ 5.37]
☐ 4	Pyrazinecarboxylic acid, 3-[2-(4-chloro-2-methylphenyl)-2,5-dihydro-3,4-dimethyl-5-oxo-1H-pyrrol-1-yl]- (9CI) [€ 5.37]

The Latest CAS Registry Number® and Substance Count

Date	**Tue Feb 25 06:02:06 EST 2003**
Count	**21,131,929** organic and inorganic substances
	25,554,588 sequences
CAS RN	**494745-03-8** is the most recent CAS Registry Number

Specialized Substance Collections Count

CASREACT®	**6,671,749** Single- and multi-step reactions
CHEMLIST®	**228,938** Inventoried/regulated substances
CHEMCATS®	**5,716,289** Commercially available chemicals
MARPAT®	**504,823** Searchable Markush structures