



ÚSTAV INFORMAČNÍCH STUDIÍ A KNIHOVNICTVÍ  
FF UK V PRAZE

Lucie Šerá

# Ukázky vyhledávání v chemických databázích

Verze 1.0

Praha

Listopad 2008

## 1 POSTUP

Cílem je prozkoumat jakým způsobem lze vyhledávat v konkrétních databázích a porovnat jejich nástroje a možnosti. Dalším záměrem je ukázat postupy vyhledávání v těch samých databázích, ale přístupných prostřednictvím jiného vyhledávacího prostředí s rozdílným účtováním. Rešerše byly vedeny v těchto databázích:

- báze dostupné v rámci licence pro instituci
  - Registry v rámci SciFinder Scholar 2006
  - Beilstein v prostředí CrossFire Commander V7
- zdarma dostupná databáze
  - PubChem
- báze dostupné prostřednictvím STN International (platba systémem pay-as-you-go, pomocí softwaru STN Express 8.3).
  - Registry
  - Beilstein

V průběhu rešerše se postupuje podle diagramů uvedených v kapitole 4.1 a u každého kroku je uvedeno vysvětlení. Co se týče rešerše vedené v rámci STN International, snahou bylo přiblížit se reálnému vedení rešerše, tzn. snažit se o co nejmenší náklady, proto u rešerší z STN International není postup plně dodržen, protože je potřeba myslet na závěrečnou cenu rešerše, a tudíž se standardní kroky musí obejít či nahradit jiným postupem. Registrační číslo je záměrně opomíjeno, aby bylo možné postup náležitě rozvinout.

Pokud jde o ceny rešerší, jsou uvedeny jen pro zajímavost. Navzájem je porovnávat by nebylo směrodatné vzhledem k tomu, že každá databáze má trochu jinou cenovou strategii vypisovaných formátů. Zatímco Registry poskytuje základní informace o látce zdarma, Beilstein žádný zdarma dostupný formát nenabízí, ale na druhou stranu se ze základního placeného formátu dozvíme více.

## 2 VYBRANÉ DATABÁZE

Následující tři databáze byly vybrány, protože reprezentují každá trochu jinou skupinu informačních zdrojů. Databáze Chemical Abstracts patří k základním zdrojům pro všechna odvětví chemie. Představuje integrovaný systém bibliografických i faktografických informací nebývalého rozsahu. Databáze Beilstein zase představuje skupinu zdrojů zaměřených na faktografické informace o látkách, včetně reakcí. Databáze PubChem se odlišuje faktem, že je dostupná veřejně a zdarma. Snahou je poskytnout alespoň základní přehled o tom, jaké vyhledávací nástroje jmenované databáze poskytují a jak je možné řešit zadané úkoly.

### 2.1 CHEMICAL ABSTRACTS

#### 2.1.1 Historie a obecné informace

Databázi Chemical Abstracts (ovšem stále vydávanou i jako tištěný referátový časopis) produkuje Chemical Abstract Service, divize Americké chemické společnosti. V současnosti zpracovává okolo 9 500 titulů časopisů.

V roce 1895 na Massachusetts Institute of Technology vznikl časopis Review of American Chemical Research, který měl kompenzovat nedostatečný zájem o americký výzkum ze strany evropských referátových časopisů. V roce 1907 byl tento časopis nahrazen širší publikací Chemical Abstracts (CA), kterou sponzorovala Americká chemická společnost (ACS).

Už od počátku měli editoři Chemical Abstracts od Americké chemické společnosti zadán jasný cíl – abstrahovat veškerou světovou literaturu zaměřenou na chemii. Začátky byly skromné, s dvěma editory na půl úvazku a 195 dobrovolnými tvůrci abstraktů s ročním rozpočtem 15 500 dolarů bylo v roce 1907 vydáno 11 847 abstraktů [CAS1].

I přes nedostatek papíru a ztíženou kooperaci s evropským kontinentem vycházely CA i během druhé světové války

V roce 1955 náklady poprvé přesáhly milión dolarů, ACS byla nucena přikročit k radikální změně financování a ceny předplatného výrazně vzrostly. V tomto roce bylo publikováno již 86 322 abstraktů [CAS1]. V roce 1966 se změnil styl práce, výrazně vzrostl počet placených pracovníků na úkor dobrovolníků. [Baker, 1980] V šedesátých letech se také začalo pracovat na automatizaci a tvorbě elektronické databáze. Tato snaha byla podpořena granty National Science Foundation a zároveň se CAS stala v USA jakýmsi prototypem automatizovaných služeb zaměřených na sekundární informace.

V roce 1980 byla spuštěna databáze CAS ONLINE. O tři roky později založila CAS společně s Fachinformationszentrum (FIZ) Karlsruhe dnes již proslulé databázové centrum STN International, ve kterém nejdříve zpřístupňovali svoje databáze, ale postupem času přidali produkty i jiných institucí.

V Československu byly počítačové produkty CAS hojně využívány již od roku 1971 formou dávkového zpracování z magnetických pásek [Vymětal, 2001].

### 2.1.2 Elektronické verze CA

Co se týče uživatelského rozhraní, nejvíce práci s původními tištěnými rejstříky připomíná verze Chemical Abstracts distribuovaná na CD ROM – CA on CD. Předplatitel získává každý měsíc nové CD, kde jsou kumulovány všechny předchozí abstrakty vydané toho roku.

Další možností jak přistupovat k Chemical Abstracts je internetové připojení a nástroj SciFinder (případně SciFinder Scholar určený pro akademické prostředí). Toto velmi přívětivé rozhraní, které se v posledních letech rychle vyvíjí, umožňuje spolu s abstrakty získat naráz i další informace (zejména faktografického charakteru) z databáze Registry a CASREACT.

Naopak pro profesionální řešeršéry je určen přístup v rámci databázového centra STN International. Na rozdíl od předchozích dvou přístupů nespolehá na uživatelskou intuici, ale na znalosti a praxi. Komunikace se odehrává pomocí příkazového řádku a dotazovacího jazyka Messenger.

### 2.1.3 CAS Registry

Systém CAS Registry byl založen v roce 1965 pro potřeby zpracovatelů abstraktů CAS, kteří museli zjišťovat, zda už byla konkrétní látka do Abstrakt zpracována a jaký je její rejstříkový název. V roce 1984 začala CAS zpracovávat i látky indexované v CA před rokem 1965, aby se tak zajistil přístup na základě chemické struktury i do literatury vydané před tímto datem. Dnes (27.9. 2008) obsahuje 38 260 638 záznamů [CAS5] organických a anorganických látek a stala se tak nejrozsáhlejším zdrojem identifikace látek pro vědce, průmysl i regulační orgány [Weisgerber, 1997].

Databáze CAS Registry obsahuje záznamy látek indexovaných do báze Chemical Abstracts (v posledních letech však nejen jich) a je v podstatě vstupní branou do databáze Chemical Abstracts (a pokud pracujeme v rámci databázového centra, pak i do jiných databází obsahující pole pro registrační číslo CAS).

Každý záznam zahrnuje registrační číslo, rejstříkové jméno, synonyma, strukturní a sumární vzorec. U části látek jsou přítomné i informace o vlastnostech, získané buď experimentálně, nebo vypočítané.

Tím jak se pole působnosti registračních čísel rozšiřuje, v databázi Registry se objevují i látky, ke kterým neexistují žádná zpracovaná abstrakta. Jako příklad uvedme některé látky ze seznamu ingrediencí pro kosmetický průmysl vydaného Evropskou komisí jako součást Rozhodnutí 96/335/EC, ve kterém figurují přírodní látky typu extraktu z violky vonné (90147-36-7) nebo chmelu otáčivého (8060-28-4), i produkty potravinářského průmyslu jako je např. pivo (8029-31-0) [Komise ES]. Dalším příkladem může být již zmiňovaný EINECS (European Inventory of Existing Commercial chemical Substances [0]). Dále se mohou v Registry vyskytnout látky bez záznamu v CA na základě žádosti některých organizací typu World Health Organization, US Environment Protection Agency a dokonce i na základě žádosti podané přes CAS Client Service. Také nové látky, které se objeví v některém z obchodních katalogů zařazených v bázi CHEMCATS a od roku 2008 jsou zařazovány i látky z webových kolekcí, např. ze zmiňovaného zdroje ChemSpider. Další speciální případy jsou popsány zde [STN, 2008].

### 2.1.4 CASREACT

Tato databáze zahrnuje záznamy reakcí publikovaných v některém z více než stovky klíčových časopisů od roku 1985. Kromě reaktantů a produktů, záznamy obsahují i informace o reagentech, rozpouštědlech, katalyzátorech, reakčních a bezpečnostních podmínkách. Výtěžek reakce původně zahrnut nebyl kvůli nedostupnosti této informace u některých

reakcí a kvůli nejednotným podmínkám měření, nicméně na nátlak uživatelů nakonec zaveden byl [Blake, 1990]. Databáze obsahuje k dnešnímu dni (27.9.2008) 15 700 474 reakcí [CAS5].

## 2.2 BEILSTEIN

Databáze Beilstein je největší faktografickou databází zaměřenou na organickou chemii. Registruje více jak 9 miliónů sloučenin (listopad 2004) a pro téměř 8 miliónů z nich uchovává data o reakcích. Je aktualizována čtvrtletně.

Historicky vychází z tištěného díla Beilsteins Handbuch der organischen Chemie (krátce Beilsteinovo kompendium), pojmenovaného podle autora Konráda Beilsteina (1883-1906), německého profesora chemie. První soupis 1500 organických sloučenin a jejich vlastností vydal v roce 1881, úspěch byl mezi chemiky obrovský a další rozšířená vydání rychle následovala. Po smrti profesora Beilsteina se vydávání ujala Německá chemická společnost. V 50. letech 20. století byl založen „Beilstein Institut,“ který se celý soustředil pouze na vydávání kompendia. O 30 let později vyšlo již páté vydání čítající úctyhodných půl miliónu stránek. V té době se již také pracuje na převodu kompendia do elektronické podoby, výsledkem bylo zpřístupnění databáze Beilstein v databázovém centru STN International (1988) a Dialog (1989). V roce 1994 byla založena společnost Beilstein Informationssysteme GmbH. O 5 let později bylo ukončeno souběžné vydávání tištěné formy kompendia.

Současným producentem je Molecular Design Ltd., dceřinná společnost vydavatelství Elsevier, které převzalo Beilstein Informationssysteme. Zajímavostí je, že zatímco MDL bylo v říjnu 2007 prodáno dále (Symyx Technologies, Inc.), databáze Beilstein i Gmelin zůstala ve vlastnictví Elsevieru.

Co se týče forem zpřístupnění báze, stále je možnost online přístupu do databázových center (např. v Dialogu pod souborem 381- Beilstein Reactions, 390- Beilstein Facts a 393- Beilstein Abstracts). Na konci 90. let však společnost Beilstein Informationssysteme podnikla revoluční krok, když nabídla licence na pronájem databáze i komunikačního softwaru do lokálních sítí. Vznikl tak systém pod názvem **CrossFire**, který je doplněn o modul CrossFireplus Reactions a databázi CrossFire Abstracts.

Zdroje databáze tvoří původní Beilstein Handbook, pokrývající literaturu od roku 1771, dále jsou to časopisy, knihy a patenty vydané od roku 1960 do 1979, od roku 1980 je zpracováváno 176 titulů časopisů.

### 2.2.1 PubChem

PubChem je databáze poskytující informace o biologické aktivitě malých molekul. Je částí iniciativy Molecular Libraries Roadmap a byl spuštěn 16. září 2004. K vyhledávání využívá systém Entrez. Prostřednictvím tohoto rozhraní jsou dostupné i další databáze bibliografického a faktografického charakteru. Celý tento systém zajišťuje National Center for Biotechnology Information, který je součástí National Institutes of Health (NIH). Databáze je dostupná zdarma, což je příčinou sporu mezi NIH a Americkou chemickou společností. Ta tvrdí, že je PubChem pro jejich placené produkty konkurentem (a to je podle nich plýtvání penězi daňových poplatníků) a požaduje, aby byly do databáze přijímány pouze molekuly zkoumané v centrech NIH [Marris, 2005].

Součástí je např. databáze PubMed, která pokrývá databázi Medline (14 mil. záznamů s retrospektivou do poloviny 60.let), OldMedline (2 miliony záznamů z let 1950-1956) a navíc poskytuje další služby jako jsou odkazy na další vhodné zdroje atd. Je to největší biomedicínská databáze dostupná zdarma [PubMed]. Příkladem faktografické databáze je Molecular Modeling DataBase, která shromažďuje 3D struktury bílkovin [MMDB]. Každá databáze poskytuje odkazy na své původní zdroje.

Samotný systém PubChem je organizován jako tři navzájem propojené databáze PubChem Substance, PubChem Compound a PubChem BioAssay. Odkazy z chemických struktur v rámci PubChem se lze dostat na biologické vlastnosti látky, které jsou zaznamenány v jiných databázích (PubMed, zdroje 3D struktur...). Data jsou získávána od producentů různých typů (výzkumné instituce atd.)

Databáze PubChem Compound (5,5 miliónů záznamů) obsahuje validované chemické vzorce sloučenin zařazených do PubChem Substance. Lze vyhledávat podle názvů, synonym, klíčových slov, ale také pomocí strukturního vyhledávání. Strukturní editor je přímo implementován do webového rozhraní, ovšem soubory s nakreslenými strukturami (samořejmě v podporovaných formátech) lze nahrát i z vlastního počítače.

Databáze PubChem Substance (11 miliónů záznamů) obsahuje popis vlastností látek s odkazy na zdroje. Opět lze vyhledávat prostřednictvím názvů, synonym nebo klíčových slov.

Databáze PubChem BioAssay (téměř 300 záznamů) obsahuje stručné popisy analýz bioaktivity látek indexovaných v databázi PubChem Substance.

Co se týče vyhledávání, pokud uživatel chce informace od konkrétního zdroje, je výhodnější zadat dotaz do databáze PubChem Substance. Platí to i pro požadavky ohledně přírodních látek, ke kterým nebývá přiřazena struktura. V ostatních případech je výhodnější hledat v databázi PubChem Compound [PubChem-FAQ].

### 3 REŠERŠE Č. 1: VYHLEDEJTE ZÁZNAM SLOUČENINY S CHEMICKÝM NÁZVEM BENZALDEHYD

#### SciFinder Scholar

Q1 Je známo CAS RN? NE

Q2 Je znám název? ANO

Q5 Je název dlouhý nebo nejednoznačný? NE

Poznámka: Existují i jiné názvy pro benzaldehyd (např. fenylmethanal), ale jde o běžnou sloučeninu, lze se tedy domnívat, že je indexována pod všemi svými názvy.

Q6 Nalezeno pomocí relevantního pole? ANO

Počet výsledků: 1 - v Registry (58028 - v CAPLUS)

The image illustrates the step-by-step process of searching for benzaldehyde in SciFinder Scholar. It starts with the 'New Task' dialog box where 'Locate' is selected. This leads to the 'Locate' dialog box where 'Locate Substances' is chosen. The next step is the 'Locate by Substance Identifier' dialog box, where 'benzaldehyde' is entered into the text field. Finally, the search results are displayed, showing the chemical structure of benzaldehyde and a list of 55,297 references from the Registry.

Dále můžeme pokračovat:

- zobrazením bibliografických záznamů
- získáním přehledu reakcí, kde figuruje benzaldehyd
- zúžením dotazu (např. podle názvů časopisů) nebo přeformulováním dotazu
- získáním záznamu sloučeniny v databázi Registry

## Beilstein

Q1 Je známo CAS RN? NE

Q2 Je znám název? ANO

Q5 Je název dlouhý nebo nejednoznačný? NE

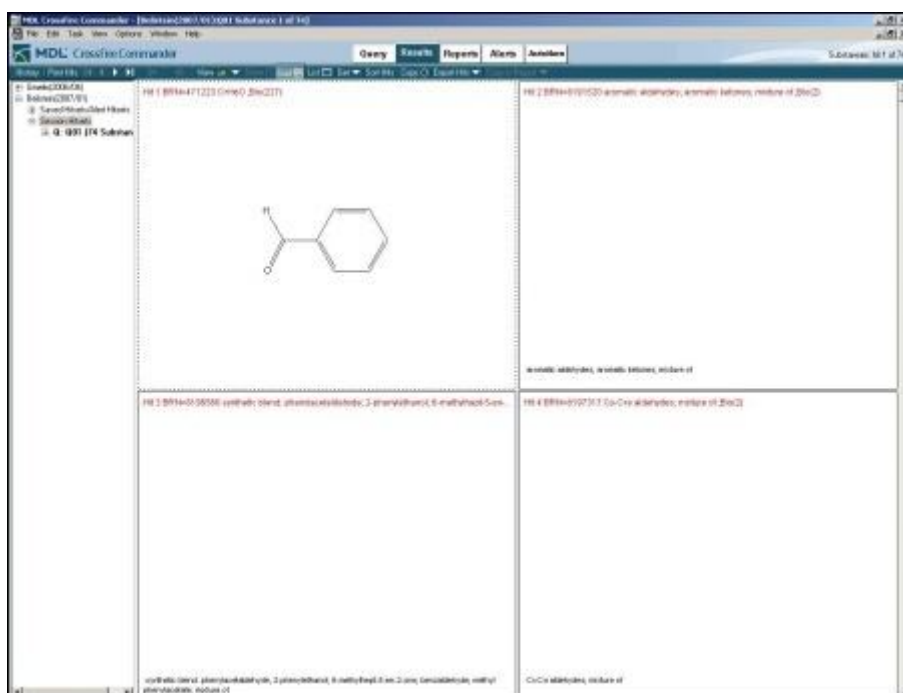
Poznámka: Název lze vyplnit do příslušného pole, případně použít připravený formulář pro vyhledávání látek na základě jejich identifikačních dat.

Q6 Nalezeno pomocí relevantního pole? ANO

Počet výsledků: 74

The screenshot shows the MDL CrossFire Commander interface. On the left, a tree view of search fields is visible, with 'Chemical Name' highlighted. A red circle and arrow point to this field, which is also selected in the 'Search Fields' table at the bottom. The table has columns for 'Operator', 'Field name', 'Position', 'Field content', and 'List'. The first row shows 'and', 'Chemical Name(DN)', '6', and 'benzaldehyde'. The 'Search Context' is set to 'Substances' and the 'Start Search' button is visible.

Operator	Field name	Position	Field content	List
and	Chemical Name(DN)	6	benzaldehyde	
and				
and				



**Alternativně:** Strukturální hledání umožňuje odfiltrout směsi, izotopy atd. (zruší se zaškrtnutí v nabídce). Při tomto postupu jsou výsledkem jen dva záznamy - benzaldehyd a jeho radikál.

**Počet výsledků:** 2

### PubChem

Q1 Je známo CAS RN? NE

Q2 Je znám název? ANO

Q5 Je název dlouhý nebo nejjednoznačný? NE

Q6 Nalezeno pomocí relevantního pole? ANO

**Počet výsledků:** 11046





## STN International

### Registry

Q1 Je známo CAS RN? NE

Q2 Je znám název? ANO

Q5 Je název dlouhý nebo nejednoznačný? NE

Q6 Nalezeno pomocí relevantního pole? ANO

Počet výsledků: 1

```
***** STN Karlsruhe *****
```

```
FILE 'HOME' ENTERED AT 11:16:14 ON 04 JUL 2008
```

```
=> fil zreg
```

COST IN EUROS	SINCE FILE ENTRY	TOTAL SESSION
FULL ESTIMATED COST	0,38	0,38

```
FILE 'ZREGISTRY' ENTERED AT 11:16:31 ON 04 JUL 2008
```

```
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
```

```
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
```

```
COPYRIGHT (C) 2008 American Chemical Society (ACS)
```



Nejprve vyhledáme správný název v rejstříku názvů sloučenin (pole COMPOUND NAME) pomocí příkazu EXPAND („e“) – na rozdíl od příkazu SEARCH („s“) se neplatí za vyhledaný termín a slouží k ověření, zda se konkrétní termín v bázi nachází

```
=> e benzaldehyd/cn
```

```
E1          1  BENZALDAZINE-1,3-BIS(3-ISOCYANATO-4-METHYLPHENYLENE) PARABA  
              NIC ACID COPOLYMER/CN  
E2          1  BENZALDAZINECOBALT IODIDE/CN  
E3          0 --> BENZALDEHYD/CN  
E4          1  BENZALDEHYDE/CN  
E5          1  BENZALDEHYDE ((2,4,6-TRIIISOPROPYLPHENYL)SULFONYL)HYDRAZONE  
              /CN  
E6          1  BENZALDEHYDE (2-CARBOXY-5-SULFOPHENYL)HYDRAZONE/CN  
E7          1  BENZALDEHYDE (2-METHYLALLYL)HYDRAZONE/CN  
E8          1  BENZALDEHYDE (4-CHLOROBENZOYL)HYDRAZONE/CN  
E9          1  BENZALDEHYDE (4-METHYLBENZOYL)HYDRAZONE/CN  
E10         1  BENZALDEHYDE (4-PYRIDYLCARBONYL)HYDRAZONE/CN  
E11         1  BENZALDEHYDE (5,6-DIPHENYL-1,2,4-TRIAZIN-3-YL)HYDRAZONE/CN  
E12         1  BENZALDEHYDE (5-PHENYL-1,2,4-TRIAZOL-3-YL)HYDRAZONE/CN
```

Nyní je potřeba vyhledat zvolený termín v bázi

```
=> s e4
```

```
L1          1  BENZALDEHYDE/CN
```

Zvolíme bezplatný formát „scan“ pouze se základními informacemi

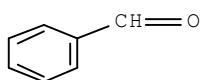
=> d scan

L1 1 ANSWERS REGISTRY COPYRIGHT 2008 ACS on STN

IN **Benzaldehyde**

MF C7 H6 O

CI COM



\*\*PROPERTY DATA AVAILABLE IN THE 'PROP' FORMAT\*\*

ALL ANSWERS HAVE BEEN SCANNED

\*\*\*\*\*

Cena: 5 €

### Beilstein

Q1 Je známo CAS RN? NE

Q2 Je znám název? ANO

Q5 Je název dlouhý nebo nejjednoznačný? NE

Q6 Nalezeno pomocí relevantního pole? ANO

Počet výsledků: 4

=> fil beilstein

COST IN EUROS	SINCE FILE	TOTAL
	ENTRY	SESSION
FULL ESTIMATED COST	9,88	22,16

FILE 'BEILSTEIN' ENTERED AT 13:50:54 ON 04 JUL 2008

COPYRIGHT (c) 2008 Beilstein-Institut zur Foerderung der Chemischen Wissenschaften licensed to Beilstein GmbH and MDL Information Systems GmbH

FILE LAST UPDATED ON April 1, 2008

>>> FOR SEARCHING PREPARATIONS SEE HELP PRE <<<

```

*****
* PLEASE NOTE THAT THERE ARE NO FORMATS FREE OF COST. *
* SET NOTICE FEATURE: THE COST ESTIMATES CALCULATED FOR SET NOTICE *
* ARE BASED ON THE HIGHEST PRICE CATEGORY. THEREFORE; THESE *
* ESTIMATES MAY NOT REFLECT THE ACTUAL COSTS. *
* FOR PRICE INFORMATION SEE HELP COST *
*****

```

Navážeme rovnou číslem dotazu z předchozí databáze

```

=> s l1
L5          4 L1

```

Výsledkem jsou čtyři záznamy, zkusíme, zda zadání registračního čísla benzaldehydu, hledanou sloučeninu upřesní.

```

=> s 100-52-7/rn
L6          4 100-52-7/RN

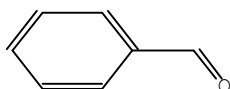
=> s l1 and l6
          4 L1
L7          4 L1 AND L6

```

Zjistili jsme, že všechny 4 záznamy mají to stejné registrační číslo CAS a nezbývá než vypsát jejich identifikační pole.

L7 ANSWER 1 OF 4 BEILSTEIN COPYRIGHT 2008 BEILSTEIN MDL on STN

Beilstein Records (BRN): 5919529  
Beilstein Pref. RN (BPR): 100-52-7  
CAS Reg. No. (RN): 100-52-7  
Chemical Name (CN): **benzaldehyde cation**  
Lin. Struct. Formula (LSF): C7H6O(1+)  
Molec. Formula (MF): C7 H6 O  
Molecular Weight (MW): 106.12  
Lawson Number (LN): 7132  
Compound Type (CTYPE): isocyclic  
Constitution ID (CONSID): 5159954  
Tautomer ID (TAUTID): 5612920  
Beilstein Citation (BSO): 6-07  
Entry Date (DED): 1993/07/22  
Update Date (DUPD): 1994/02/18



Fragment Notes:

Unknown location for Localized Charge of (+1)

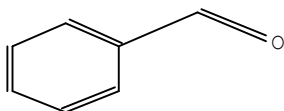
Field Availability:

Code	Name	Occurrence
BRN	Beilstein Records	1
BPR	Beilstein Preferred RN	1
RN	CAS Registry Number	1
CN	Chemical Name	1
LSF	Linearized Structure Formula	1
MF	Molecular Formula	1
FW	Formular Weight	1
LN	Lawson Number	1
CTYPE	Compound Type	1

CONSID	Constitution ID	1
TAUTID	Tautomer ID	1
BSO	Beilstein Citation	1
ED	Entry Date	1
UPD	Update Date	1
ESR	ESR Data	4

L7 ANSWER 2 OF 4 BEILSTEIN COPYRIGHT 2008 BEILSTEIN MDL on STN

Beilstein Records (BRN): 4176385  
 Beilstein Pref. RN (BPR): 100-52-7  
 CAS Reg. No. (RN): 100-52-7  
 Chemical Name (CN): **benzaldehyde anion radical**  
 Lin. Struct. Formula (LSF): C7H6O(1-)  
 Molec. Formula (MF): C7 H6 O  
 [...]



Fragment Notes:

Unknown location for Localized Charge of (-1)

Field Availability:

Code	Name	Occurrence
=====		
BRN	Beilstein Records	1

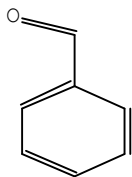
[...]

L7 ANSWER 3 OF 4 BEILSTEIN COPYRIGHT 2008 BEILSTEIN MDL on STN

Beilstein Records (BRN): 3931460  
 Beilstein Pref. RN (BPR): 100-52-7  
 CAS Reg. No. (RN): 100-52-7

Chemical Name (CN): **benzaldehyde radical**

[...]



Field Availability:

Code	Name	Occurrence
------	------	------------

=====

BRN	Beilstein Records	1
-----	-------------------	---

[...]

**L7 ANSWER 4 OF 4 BEILSTEIN COPYRIGHT 2008 BEILSTEIN MDL on STN**

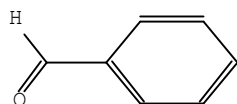
**Beilstein Records (BRN): 471223**

**Beilstein Pref. RN (BPR): 100-52-7**

**CAS Reg. No. (RN): 100-52-7**

**Chemical Name (CN): benzaldehyde**

[...]



Field Availability:

Code	Name	Occurrence
------	------	------------

=====

BRN	Beilstein Records	1
-----	-------------------	---

[...]

This substance also occurs in Reaction Documents:

Code	Name	Occurrence
------	------	------------

=====

RX	Reaction Documents	5567
----	--------------------	------

RXREA	Substance is Reaction Reactant	5133
RXPRO	Substance is Reaction Product	4338

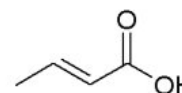
Záznam, který hledáme, je v seznamu výsledků poslední. Kromě chemického názvu i počet dostupných polí (zde zkráceno) a frekvence s jakou se vyskytuje v dokumentech o reakcích napovídá, že jde o hledanou sloučeninu.

**Cena:** 23,58 €

### 3.1 ZÁVĚR

Vyhledávání látky podle chemického názvu v SciFinder Scholar lokalizovalo pouze jeden záznam benzaldehydu (ve srovnání se strukturním hledáním - při volbě přesného hledání, jednosložkové látky - 69 záznamů). Naproti tomu v databázi Beilstein se hledáním názvu objevilo více záznamů, než při strukturním hledáním (s příslušnými volbami). Algoritmy databáze PubChem prohledávají plné znění všech názvů, seznam výsledků je proto dost dlouhý, nicméně požadovaná sloučenina je na prvním místě. Stejně jako Beilstein i PubChem zpřesní své výsledky použitím strukturního hledání. V rámci STN v databázi Registry proběhlo hledání podle názvu totožně s jedním výsledkem. Databáze Beilstein lokalizovala podobně jako její protějšek v rámci CrossFire více záznamů – šlo o radikály. Nepříjemná je cena, kterou je potřeba zaplatit za zjištění, který záznam je ten vhodný. Z praxe je známo, že obvykle ten poslední, takže se začíná s ním a lze tak vylučovací metodou něco ušetřit.

## 4 REŠERŠE Č. 2: VYHLEDEJTE ZÁZNAM SLOUČENINY S TOUTO STRUKTUROU:



### SciFinder Scholar

**Q1 Je známo CAS RN?** NE

**Q2 Je znám název?** ANO

**Q5 Je název dlouhý nebo nejjednoznačný?** ANO

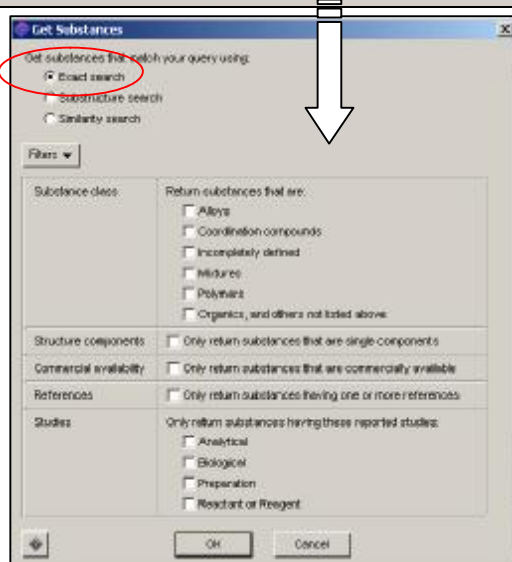
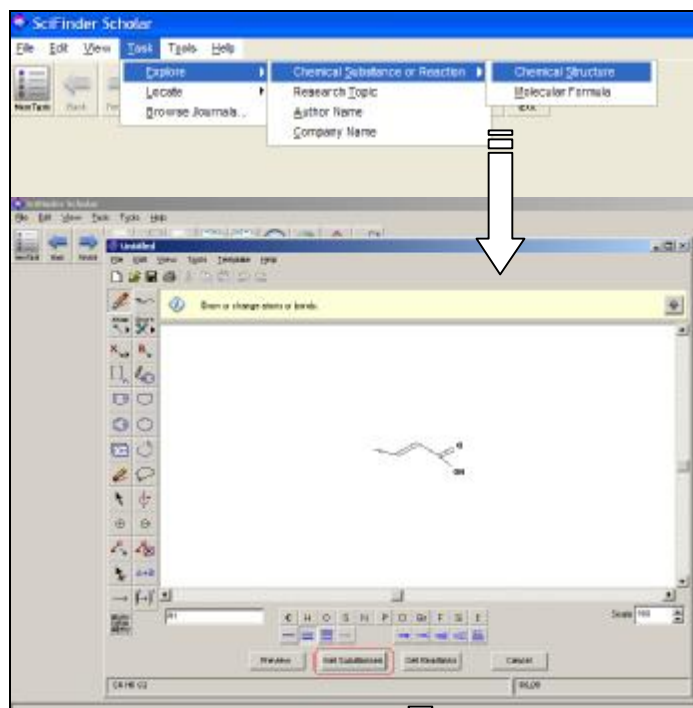
Poznámka: Chemický název bude patrně obsahovat symbol pro stereoizomerní formu a polohu dvojně vazby, které se mohou značit různě.

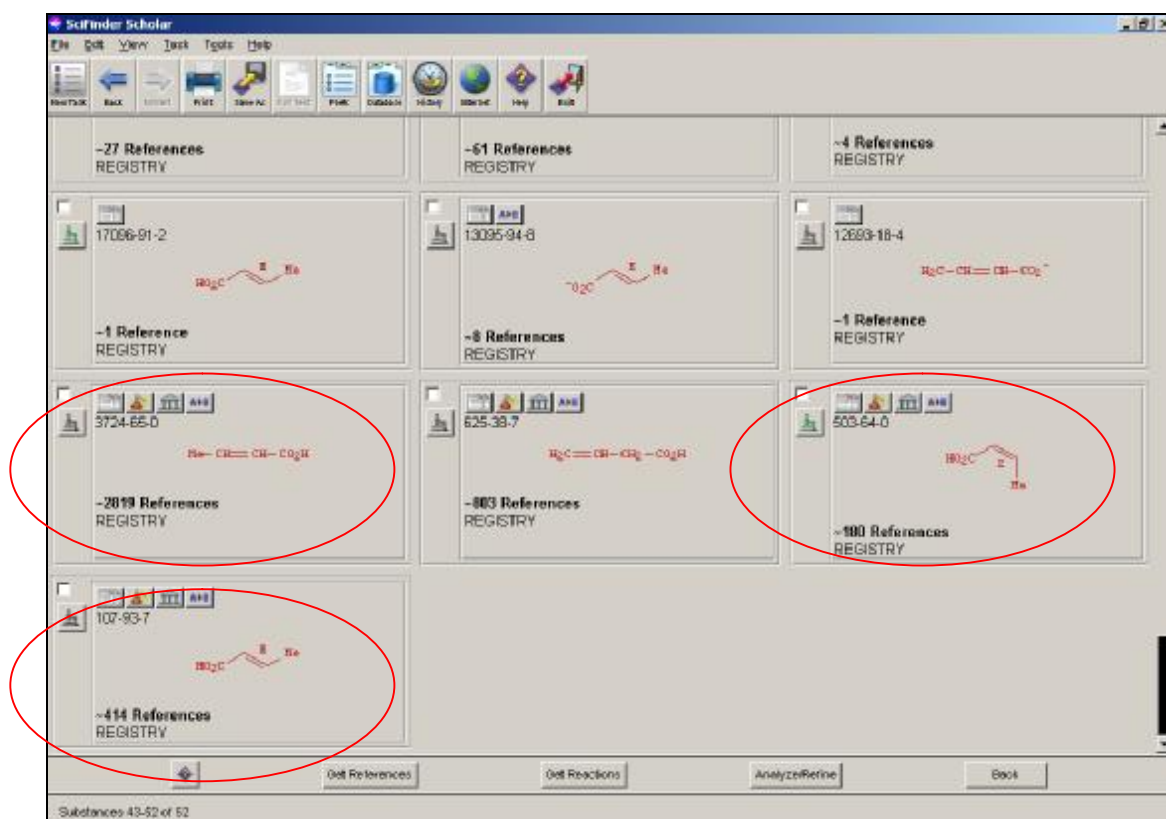
**Q7 Nalezeno strukturním vyhledáváním?** ANO

Poznámka: Je třeba nakreslit vzorec ve strukturním editoru a zadat přesné hledání.

**Počet výsledků:** 2889 - v CAPLUS, 1741 - v Registry (kromě požadovaných sloučenin, výsledky zahrnují i směsi, homopolymery, anionty, sloučeniny značené uhlíkem 14 atd. Výběr může být zúžen nastavením filtrů.)







**Alternativně:** Nastavme přesné vyhledávání v jednosložkových látkách. Výsledkem jsou izomery Z i E a nekompletně definovaná látka.

**Počet výsledků:** 52

**Alternativně:** Nastavme přesné vyhledávání v jednosložkových látkách a zadejme navíc ještě volbu hledání v komerčně dostupných látkách. Tato možnost zajímavě omezí množinu na běžněji se vyskytující látky - ve výsledku se objeví Z, E izomery i nekompletně definovaná sloučenina. Tento postup ale nelze doporučit jako standardní.

**Počet výsledků:** 4

### Beilstein

**Q1 Je známo CAS RN?** NE

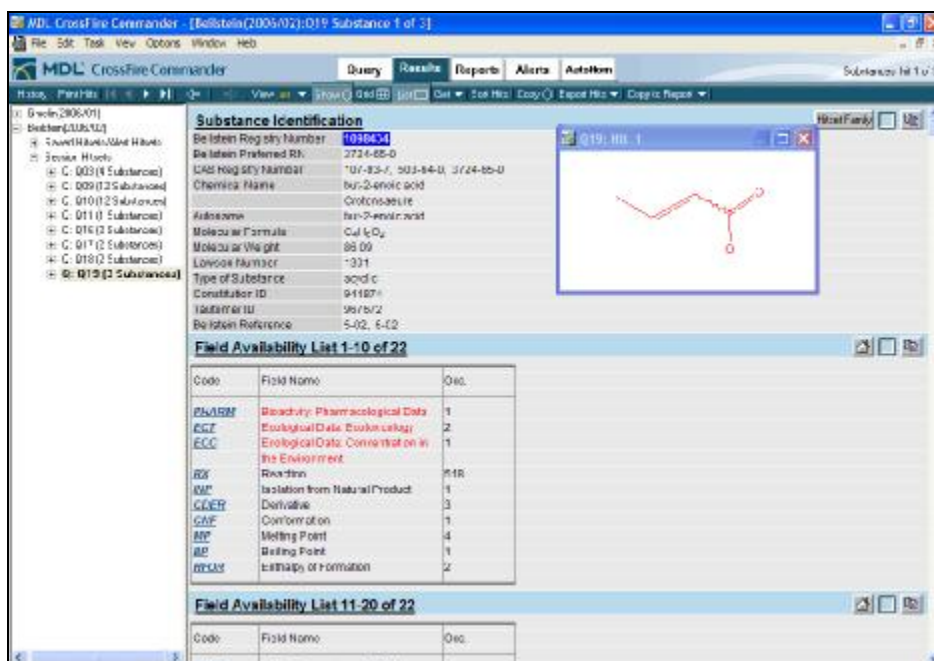
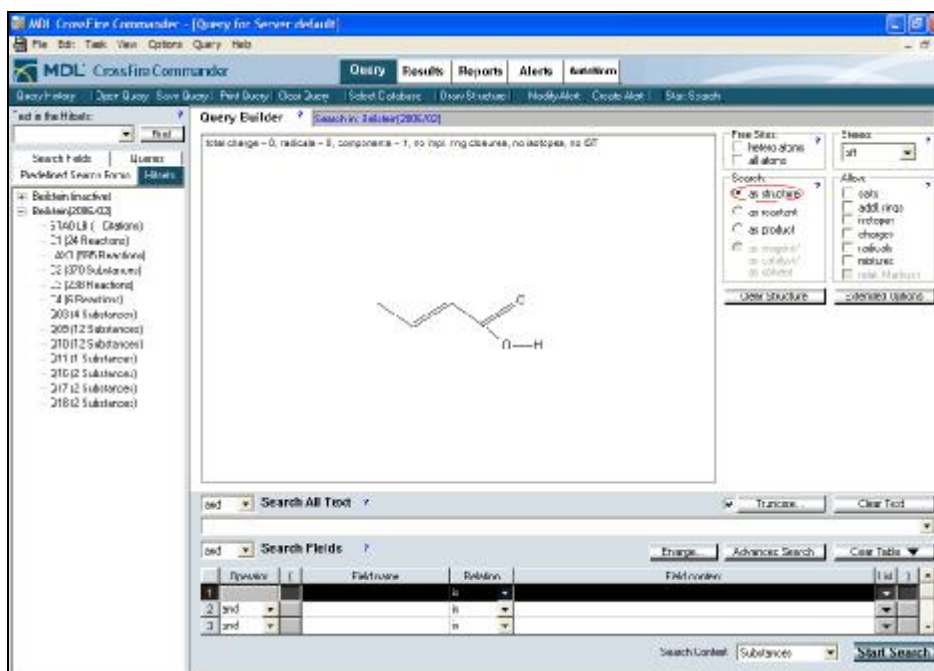
**Q2 Je znám název?** ANO

**Q5 Je název dlouhý nebo nejednoznačný?** ANO

**Q7 Nalezeno strukturním vyhledáváním?** ANO

Poznámka: Strukturní dotaz je zadán do editoru (máme tři možnosti: MDL ISIS/Draw, MDL Draw a MDL CrossFire Structure Editor).

**Počet výsledků:** 3 (Systém našel obě izomerní formy i nekompletně definovanou sloučeninu. Vazba, jejíž orientace není jasná, je naznačena vlnovkou.)



## PubChem

Q1 Je známo CAS RN? NE

Q2 Je znám název? ANO

Q5 Je název dlouhý nebo nejednoznačný? ANO

Q7 Nalezeno strukturním vyhledáváním? ANO

Poznámka: Struktura může být vložena jednak nakreslením přímo ve strukturním editoru (pod odkazem "Sketch") a jednak nahráním z připraveného souboru (v tomto případě vytvořeného v ISIS Draw).

Počet výsledků: 1 (Pouze izomer, který byl nakreslen v rámci strukturního dotazu..)




- PubChem Compound:** Search unique chemical structures using names, synonyms or keywords. Links to available biological property information are provided for each compound.
- PubChem Substance:** Search deposited chemical substance records using names, synonyms or keywords. Links to biological property information and depositor web sites are provided.
- PubChem BioAssay:** Search bioassay records using terms from the bioassay description, for example "cancer cell line". Links to active compounds and bioassay results are provided.
- Structure Search:** Search PubChem's Compound database using a chemical structure as the query. Structures may be sketched or specified by SMILES, MOL files, or other formats.

The screenshot shows the PubChem website interface. At the top, there are logos for NCBI, PubChem, and the National Library of Medicine (NLM). Below the logos is a navigation bar with links to PubMed, Entrez, Structure, GenBank, PubChem, and Help. A search bar labeled "Compound Text Search" is visible. The main section is titled "Structure Search" with radio buttons for "Basic" (selected) and "Advanced". There are "Search" and "Clear" buttons. Below this, the "Search Input" section has radio buttons for "SMILES/SMARTS, InChI, CID or Formula" (selected) and "Structure File". A text input field contains the SMILES string E/Bakalarka/kysolina.skic and a "Procházet..." button. The "Search Type" section has a dropdown menu set to "Identical Structures". Overlaid on the right side is a "SIBIDraw" window, which is a chemical structure drawing tool. It has a toolbar with various icons and a central area showing a skeletal structure of a branched alkane chain. An arrow points from the drawing tool's toolbar to the SMILES input field.

NCBI PubChem National Library of Medicine NLM

Search: PubChem Compound

Compound Summary:



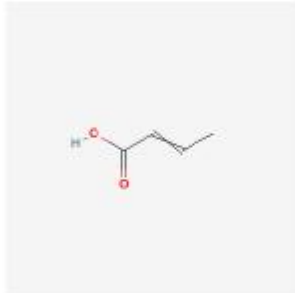
CID: 643792

- BiActivity: Summary (Inactive: 7 Links)
- NLM Toxicology: Link
- Substances: All: 12 Links, Same: 9 Links, Mixture: 3 Links
- Related Compounds: Same, Connectivity: 4 Links, Same, Isotopes: 3 Links
- Similar Compounds: 23 Links
- Structure Search

NCBI PubChem National Library of Medicine NLM

Search: PubChem Compound

Compound Summary:



CID: 19499

- Substances: All: 8 Links, Same: 3 Links, Mixture: 5 Links
- Related Compounds: Same, Connectivity: 5 Links, Same, Isotopes: 3 Links
- Similar Compounds: 23 Links
- Structure Search

Synonyms Properties Descriptors Category Experts

to

to

to

to

to

to

partially specified

**Alternativně:** Pokud užijeme další volby pro strukturní hledání (částečně specifikované E/Z izomery), najdeme nekompletně definovanou látku (tj. but-2-enovou kyselinu) a pomocí odkazu z jejího záznamu se snadno dostaneme k oběma izomerům.

**Počet výsledků: 1**

### STN International

#### Registry

**Q1 Je známo CAS RN? NE**

**Q2 Je znám název? ANO**

**Q5 Je název dlouhý nebo nejednoznačný? ANO**

**Q7 Nalezeno strukturním vyhledáváním? NE**

Poznámka: Je potřeba se strukturnímu dotazu vyhnout, budeme tedy vyhledávat podle názvu

**Počet výsledků: 2**

```
=> e butenoic acid/cn
E1          1      BUTENETRICARBOXYLIC ACID, TRIS(OXIRANYLMETHYL) ESTER, POLY
              MER WITH ETHENE/CN
E2          1      BUTENETRIOL/CN
E3          1 --> BUTENOIC ACID/CN
E4          1      BUTENOIC ACID, ((TRIMETHYLSILYL)OXY)-, TRIMETHYLSILYL ESTE
              R/CN
E5          1      BUTENOIC ACID, (6E)-3,7,11-TRIMETHYL-6,10-DODECADIENYL EST
              ER/CN

=> s e3
L2          1 "BUTENOIC ACID"/CN

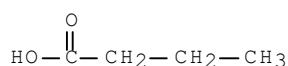
=> d l2

L2  ANSWER 1 OF 1  REGISTRY  COPYRIGHT 2008 ACS on STN
RN  97858-17-8   ZREGISTRY
ED  Entered STN: 31 Aug 1985
CN  Butenoic acid (CA INDEX NAME)
MF  C4 H6 O2
CI  IDS, COM
SR  CA
LC  STN Files:  AGRICOLA, BEILSTEIN*, BIOSIS, CA, CAPLUS, GMELIN*,
              TOXCENTER, USPAT2, USPATFULL
              (*File contains numerically searchable property data)

CM  1

CRN  107-92-6
CMF  C4 H8 O2
```





40 REFERENCES IN FILE CA (1907 TO DATE)

11 REFERENCES TO NON-SPECIFIC DERIVATIVES IN FILE CA

40 REFERENCES IN FILE CAPLUS (1907 TO DATE)

Výsledkem je nekompletně definovaná látka (není jasná ani poloha dvojné vazby)

Následuje pokus o levostranné rozšíření (bohužel nefunguje)

```
=> e left butenoic acid/cn
LEFT TRUNCATION NOT AVAILABLE FOR 'CN' FIELD IN FILE 'ZREGISTRY'
Left truncation cannot be used on search terms in this field of this
file.
```

Pokračujeme přímo vyhledáním názvu v rejstříku

```
=> e 2-butenic/cn
E1          1      2-BUTENIMIDOYL FLUORIDE, 4,4,4-TRIFLUORO-N-(TRIFLUOROMETHY
              L)-, (Z,E)-/CN
E2          1      2-BUTENISOCYANATE, 3-ETHOXY-/CN
E3          0 --> 2-BUTENOIC/CN
E4          1      2-BUTENOIC ACID/CN
E5          1      2-BUTENOIC ACID TROPENOL ESTER/CN
E6          1      2-BUTENOIC ACID, (((3-CHLORO-4-(2-METHYLBENZOYL) PHENYL) (4
              -FLUORO-2-METHYLPHENYL) AMINO) CARBONYL) OXY) METHYL ESTER, (2
              E)-/CN
E45         1      2-BUTENOIC ACID, ((4-(BENZOYLOXY)-2,2,6,6-TETRAMETHYL-1-PI
              PERIDINYL) OXY) METHYL ESTER, (Z)-/CN
E46         1      2-BUTENOIC ACID, ((4-CHLOROPHENYL) (4-(((TRIFLUOROMETHYL) TH
              IO) METHYL) PHENYL) METHYLENE) HYDRAZIDE/CN
```

V seznamu je možno si všimnout způsobu zápisu cis (Z) a trans (E) formy, název tedy podle toho přímo zapíšeme

=> e 2-butenic acid, (E)/cn

E1 1 2-BUTENOIC ACID, (DIMETHYL(2-METHYLENE-3-BUTEN-1-YL)SILYL)  
METHYL ESTER, (2E)-/CN

E2 1 2-BUTENOIC ACID, (DIMETHYL(2-METHYLENE-3-BUTENYL)SILYL)MET  
HYL ESTER, (2E)-/CN

E3 0 --> 2-BUTENOIC ACID, (E)/CN

**E4 1 2-BUTENOIC ACID, (E)-/CN**

E5 1 2-BUTENOIC ACID, (E)-, COMPD. WITH (3B)-CHOLEST-5-EN-  
3-OL (1:2)/CN

=> s e4

**L3 1 "2-BUTENOIC ACID, (E)-"/CN**

=> d scan

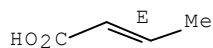
L3 1 ANSWERS REGISTRY COPYRIGHT 2008 ACS on STN

**IN 2-Butenoic acid, (2E)-**

MF C4 H6 O2

CI COM

Double bond geometry as shown.



A analogicky postupujeme i u druhé formy

=> e 2-butenic acid, (Z)/cn

E1 1 2-BUTENOIC ACID, (TETRAHYDRO-2-FURANYL)METHYL ESTER/CN

E2 1 2-BUTENOIC ACID, (TETRAHYDRO-2-FURANYL)METHYL ESTER, (E)-/  
CN

E3 0 --> 2-BUTENOIC ACID, (Z)/CN

**E4 1 2-BUTENOIC ACID, (Z)-/CN**

E5 1 2-BUTENOIC ACID, (Z)-, COMPD. WITH A-METHYLBENZENEME  
THANAMINE (1:1)/CN

E6 1 2-BUTENOIC ACID, (Z)-, COMPD. WITH 1,4-DIAZABICYCLO(2.2.2)

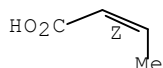
=> s e4

**L4 1 "2-BUTENOIC ACID, (Z)-"/CN**

=> d scan

L4 1 ANSWERS REGISTRY COPYRIGHT 2008 ACS on STN  
IN 2-Butenoic acid, (2Z)-  
MF C4 H6 O2  
CI COM

Double bond geometry as shown.



Cena: 9,88 €

### Beilstein

Q1 Je známo CAS RN? NE

Q2 Je znám název? ANO

Q5 Je název dlouhý nebo nejjednoznačný? ANO

Q7 Nalezeno strukturním vyhledáváním? NE

Poznámka: Pokusíme se postupovat obdobným způsobem zadáním názvu.

Počet výsledků: 2

```
=> e 2-butenic acid/cn
E1 1 2-BUTENE-1-SULFONIC ACID/CN
E2 1 2-BUTENEDIOIC ACID DIBUTYL ESTER/CN
E3 0 --> 2-BUTENOIC ACID/CN
E4 1 2-BUTENOIC ACID, (1R-<1A<RS (RS (E))>, 2B>)-4-<(2-
<<2-(1H-INDOL-3-YL)-2-METHYL-2-(<<(2-METHYLCYCLOHEXYL) OXY>
CARBONYL>AMINO)-1-OXOPROPYL>AMINO>/CN
```

Název z Registry v Beilsteinu nefunguje, zkusíme jiný tvar (podle pravidel IUPAC)

```
=> e but-2-enoic/cn
E1 3 BUT-2-ENETHIOYL CHLORIDE/CN
E2 1 BUT-2-ENIMIDIC ACID ETHYL ESTER/CN
E3 0 --> BUT-2-ENOIC/CN
E4 5 BUT-2-ENOIC ACID/CN
E5 1 BUT-2-ENOIC ACID (1,1-DIMETHYL-2-OXO-2-PHENYL-ETHYL)-AMIDE
/CN
E6 2 BUT-2-ENOIC ACID (1,1-DIMETHYL-2-PHENYL-2-THIOXO-ETHYL)-AM
IDE/CN
E7 1 BUT-2-ENOIC ACID (1,2,2,2-TETRACHLORO-ETHYL)-AMIDE/CN
E8 1 BUT-2-ENOIC ACID (1,2,3,10-TETRAMETHOXY-9-OXO-5,6,7,9-TETR
```

		AHYDRO-BENZO<A>HEPTALEN-7-YL) -AMIDE/CN
E9	1	BUT-2-ENOIC ACID (1,2,3,4-TETRAHYDRO-NAPHTHALEN-2-YL) -AMID E/CN
E10	1	BUT-2-ENOIC ACID (1,2-DIHYDRO-CYCLOBUTA<A>NAPHTHALEN-2-YLM ETHYL) -AMIDE/CN
E11	1	BUT-2-ENOIC ACID (1,3-DIPHENYL-IMIDAZOLIDIN-2-YL) -AMIDE/CN
E12	1	BUT-2-ENOIC ACID (1,4-DIOXY-BENZO<E><1,2,4>TRIAZIN-3-YL) -A MIDE/CN

Pokus o konkretizaci izomeru nevyšel.

=> e but-2-enoic acid, (E)/cn		
E1	1	BUT-2-ENOIC ACID TRITYL-AMIDE/CN
E2	3	BUT-2-ENOIC ACID VINYL ESTER/CN
E3	0 -->	BUT-2-ENOIC ACID, (E)/CN
E4	1	BUT-2-ENOYL/CN
E5	3	BUT-2-ENOYL AZIDE/CN
E6	2	BUT-2-ENOYL BROMIDE/CN
E7	3	BUT-2-ENOYL CHLORIDE/CN
E8	2	BUT-2-ENOYL FLUORIDE/CN
E9	1	BUT-2-ENOYL IODIDE/CN
E10	1	BUT-2-ENOYL ISOCYANATE/CN
E11	1	BUT-2-ENOYL ISOSELENOCYANATE/CN
E12	2	BUT-2-ENOYL ISOTHIOCYANATE/CN

Necháme tedy vypsát záznamy z předchozího hledání v rejstříku (podle zkušenosti začínáme od konce)

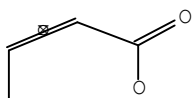
=> s e4

L12 5 "BUT-2-ENOIC ACID"/CN

=> d l12 ide 4-5

L12 ANSWER 4 OF 5 BEILSTEIN COPYRIGHT 2008 BEILSTEIN MDL on STN

Beilstein Records (BRN): 1719942  
Beilstein Pref. RN (BPR): 503-64-0  
CAS Reg. No. (RN): 107-93-7, 503-64-0, 3724-65-0  
Chemical Name (CN): but-2c-enoic acid, **cis-Crotonsaeure**,  
isocrotonic acid  
Autonom Name (AUN): but-2-enoic acid  
Molec. Formula (MF): C4 H6 O2  
Molecular Weight (MW): 86.09  
Lawson Number (LN): 1301  
File Segment (FS): Stereo compound  
Compound Type (CTYPE): acyclic  
Constitution ID (CONSID): 944874  
Tautomer ID (TAUTID): 967672  
Beilstein Citation (BSO): 0-02-00-00412, 1-02-00-00189,  
2-02-00-00394, 2-27-00-00797,  
Entry Date (DED): 1989/02/27  
Update Date (DUPD): 2007/10/11



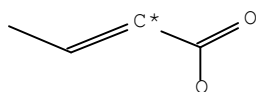
Field Availability:

Code	Name	Occurrence
BRN	Beilstein Records	1

[...]

L12 ANSWER 5 OF 5 BEILSTEIN COPYRIGHT 2008 BEILSTEIN MDL on STN

Beilstein Records (BRN): 1098434  
Beilstein Pref. RN (BPR): 3724-65-0  
CAS Reg. No. (RN): 107-93-7, 503-64-0, 3724-65-0  
Chemical Name (CN): but-2-enoic acid, Crotonsaeure  
Autonom Name (AUN): but-2-enoic acid  
Molec. Formula (MF): C4 H6 O2  
Molecular Weight (MW): 86.09  
Lawson Number (LN): 1301  
Compound Type (CTYPE): acyclic  
Constitution ID (CONSID): 944874  
Tautomer ID (TAUTID): 967672  
Beilstein Citation (BSO): 5-02, 6-02  
Entry Date (DED): 1989/02/27  
Update Date (DUPD): 2008/01/25



[...]

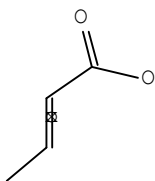
L12 ANSWER 3 OF 5 BEILSTEIN COPYRIGHT 2008 BEILSTEIN MDL on STN

Beilstein Records (BRN): 1719943  
Beilstein Pref. RN (BPR): 107-93-7  
CAS Reg. No. (RN): 107-93-7, 503-64-0, 3724-65-0  
Chemical Name (CN): but-2t-enoic acid, trans-Crotonsaeure,  
**(E)-2-butenic acid**, crotonic acid,  
 $\alpha$ -crotonic acid  
Autonom Name (AUN): but-2-enoic acid  
Molec. Formula (MF): C4 H6 O2  
Molecular Weight (MW): 86.09  
Lawson Number (LN): 1301  
File Segment (FS): Stereo compound  
Compound Type (CTYPE): acyclic  
Constitution ID (CONSID): 944874  
Tautomer ID (TAUTID): 967672

Beilstein Citation (BSO): 0-02-00-00408, 1-02-00-00187,  
2-02-00-00390, 2-27-00-00797,  
3-02-00-01255, 3-10-00-01641,  
4-02-00-01498, 5-02, 6-02

Entry Date (DED): 1989/02/27

Update Date (DUPD): 2008/01/25



[...]

**Záznamy 3 a 5 jsou ty hledané.**

Cena: 19 €

#### 4.1 ZÁVĚR

Na rozdíl od databáze Beilstein se hledání přesné struktury ve SciFinder Scholar provádí mezi všemi typy látek (směsi – dají se odfiltrvat, polymery, sloučeniny značené uhlíkem...). Žádný z obou systémů nerozlišuje polohu vazeb nakreslené struktury, mezi výsledky byly jak oba izomery, tak i nekompletně definovaná sloučenina. Naproti tomu databáze PubChem je na grafické znázornění vazeb citlivá, i když určité možnosti rozšíření dotazu tu také fungují. V rámci STN se povedlo formy sloučeniny vystopovat podle názvu (Registry) a výpisem identifikačních údajů pro základní sloučeninu – but-2-enovou kyselinu (Beilstein). Viděli jsme, že pokud rešeršér není dokonale obeznámen s názvoslovnými zásadami v té které databázi, může být hledání poměrně zdlouhavé, a tudíž drahé.



## 5 REŠERŠE Č. 3: NALEZNĚTE VŠECHNY KONSTITUČNÍ IZOMERY\* DIBROMFENANTHRENU

### SciFinder Scholar

**Q1 Je známo CAS RN?** NE

**Q2 Je znám název?** NE

Poznámka: Na molekule fenanthrenu je 10 míst vhodných pro substituci bromu, pokud chceme všechny dibromy, museli bychom zadat celkem 45 různých vzorců. (Např. 1,1-dibromfenanthren, 1,2-dibromfenanthren, 1,3-dibromfenanthren atd. podle polohy bromů v molekule).

**Q3 Je známa struktura?** ČÁSTEČNĚ

Poznámka: Museli bychom zadat 45 vzorců.

**Q8 Nalezeno substrukturním vyhledáváním?** ANO

Poznámka: Použijme speciální funkci pro substrukturní hledání. Vyberme prvek nebo funkční skupinu, které mají být připojeny a označme všechna místa na molekule, kde se mohou vyskytovat. Navíc vyznačme všechny uhlíky – hypotetická místa připojení bromu jako zamčená pro ostatní substituenty. Vyberme substrukturní hledání v jednosložkových látkách.

**Počet výsledků:** 10 - v Registry (všech deset záznamů odpovídá zadání.)

The image displays two screenshots from the SciFinder Scholar interface. The top screenshot shows the 'Get Substances' dialog box with a phenanthrene skeleton and 10 potential bromination sites highlighted by dashed lines. The 'Get Substances' button is circled in red. The bottom screenshot shows the search results, displaying 10 different dibromofenanthrene isomers with their respective CAS numbers and '1 Reference REGISTRY' status.

\* Konstituční izomery se liší polohou některých atomů v molekule. V tomto případě jde o umístění atomů bromu.

## Beilstein

Q1 Je známo CAS RN? NE

Q2 Je znám název? NE

Q3 Je známa struktura? ČÁSTEČNĚ

Q8 Nalezeno substrukturním vyhledáváním? NE

Poznámka: U všech vazebných poloh zadejme výběr ze dvou prvků, bromu a vodíku. Problémem je sice úplný, ale příliš široký seznam výsledků (jsou nalezeny i např. tribromfenathreny). Je třeba hledání zpřesnit.

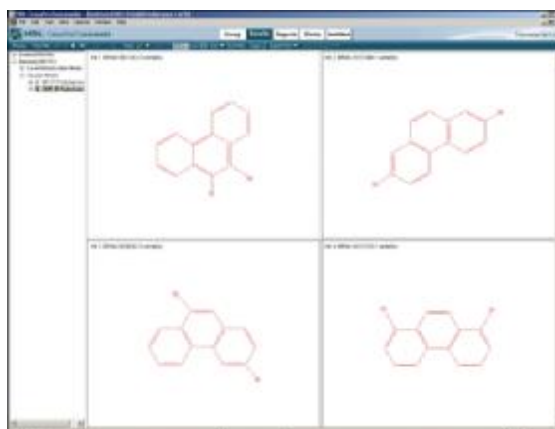
Q4 Je znám sumární vzorec? ANO

Poznámka: Sumární vzorec sám o sobě by byl nepoužitelný vzhledem k počtu sloučenin se stejným vzorcem, je třeba hledat v průniku a oba přístupy (substrukturní hledání a sumární vzorec) spojit.

Q6 Nalezeno pomocí relevantního pole? ANO

Počet výsledků: 9 (všechny výsledky odpovídají zadání.)

Operator	Fieldname	Relation	Field content
1 and	Molecular Formula(MF)	is	C14H8Br2
2 and		is	
3 and		is	



## PubChem

Q1 Je známo CAS RN? NE

Q2 Je znám název? NE

Q3 Je známa struktura? ČÁSTEČNĚ

Q8 Nalezeno substrukturním vyhledáváním? NE

Poznámka: PubChem neumožňuje zadání tohoto typu dotazu najednou. Je třeba provést hledání podle sumárního vzorce a strukturní hledání zvlášť a teprve potom je spojit.

Q4 Je znám sumární vzorec? ANO

Q6 Nalezeno pomocí relevantního pole? ANO

Počet výsledků: 4 (všechny výsledky jsou požadované dibromfenanthreny.)

NCBI PubChem Structure Search

Search By: Name/Text Identity/Similarity Substructure/Superstructure Molecular Formula Saved Search

Molecular Formula Molecular Formula File

Search

Enter single molecular formula

Options

Filters

Clear Save Query

Structure Search Basic Advanced

Search Clear

Search Input: SMILES/SMARTS, InChI

C1=CC=CC2=C1C3=C(C=C2)C=CC=C3

Search Type: Substructure

Search using: Substructure

Compound Subset

PubChem Compound Entrez History: No selection (Refresh)

#2 Select 20 document(s)

#1 Select 19 document(s)

NCBI PubChem Compound National Library of Medicine

Search: PubChem Compound C1=CC=CC2=C1C3=C(C=C2)C=CC=C3 Go Clear Save Search

Lists Preview/Info History Clipboard Details

Display: Summary Show 20 Sort by Send to

All: 4 BioAssay: 0 Pubmed: 0 Pub of S: 0

Items 1 - 4 of 4 One page

<input type="checkbox"/>	<chem>C1=CC=CC2=C1C3=C(C=C2)C=CC=C3</chem>	IUPAC: 9,10-dimethylfluorene MW: 336.021   MF: C14H18	Related Structures
<input type="checkbox"/>	<chem>C1=CC=CC2=C1C3=C(C=C2)C=CC=C3</chem>	IUPAC: 1,9-dimethylfluorene MW: 336.021   MF: C14H18	Related Structures
<input type="checkbox"/>	<chem>C1=CC=CC2=C1C3=C(C=C2)C=CC=C3</chem>	IUPAC: 2,7-dimethylfluorene MW: 336.021   MF: C14H18	Related Structures
<input type="checkbox"/>	<chem>C1=CC=CC2=C1C3=C(C=C2)C=CC=C3</chem>	IUPAC: 3,6-dimethylfluorene MW: 336.021   MF: C14H18	Related Structures

## STN International

### Registry

Q1 Je známo CAS RN? NE

Q2 Je znám název? NE

Q3 Je známa struktura? ČÁSTEČNĚ

Q8 Nalezeno substrukturním vyhledáváním? NE

Q4 Je znám sumární vzorec? ANO

Poznámka: Opět nahradíme substrukturní hledání vyhledáním názvu segmentu sloučeniny

Počet výsledků: 9 (všechny výsledky odpovídají zadání.)

```

=> e phenanthren/cns
E1          15      PHENANTHRAZINE/CNS
E2          1       PHENANTHRAZINYL/CNS
E3          5346 --> PHENANTHREN/CNS
E4          1       PHENANTHRENACYL/CNS
E5          1       PHENANTHRENALDAZINE/CNS
E6          1       PHENANTHRENALDEHYDE/CNS
E7          3       PHENANTHRENAMINATO/CNS
E8          800     PHENANTHRENAMINE/CNS
E9          21      PHENANTHRENAMINIUM/CNS
E10         4       PHENANTHRENAMMONIUM/CNS
E11         40074 PHENANTHRENE/CNS
E12        118     PHENANTHRENEACETALDEHYDE/CNS

```

A zkombinujeme se sumárním vzorcem

```

=> s (e3 or e11) and C14H8Br2/mf
      5346 PHENANTHREN/CNS
      40074 PHENANTHRENE/CNS
      44 C14H8BR2/MF
L11      9 (PHENANTHREN/CNS OR PHENANTHRENE/CNS) AND C14H8BR2/MF

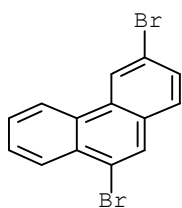
```

=> d scan

```

L11  9 ANSWERS  REGISTRY  COPYRIGHT 2008 ACS on STN
IN   Phenanthrene, 3,9-dibromo-
MF   C14 H8 Br2

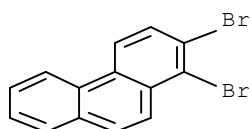
```



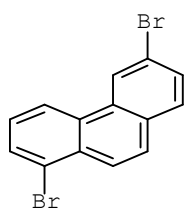
```

L11  9 ANSWERS  REGISTRY  COPYRIGHT 2008 ACS on STN
IN   Phenanthrene, 1,2-dibromo-
MF   C14 H8 Br2

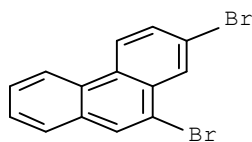
```



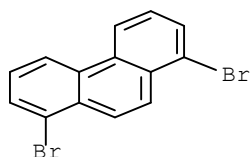
L11 9 ANSWERS REGISTRY COPYRIGHT 2008 ACS on STN  
IN Phenanthrene, 1,6-dibromo-  
MF C14 H8 Br2



L11 9 ANSWERS REGISTRY COPYRIGHT 2008 ACS on STN  
IN Phenanthrene, 2,10-dibromo-  
MF C14 H8 Br2

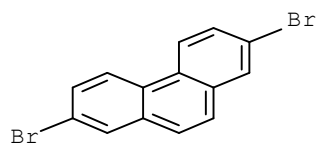


L11 9 ANSWERS REGISTRY COPYRIGHT 2008 ACS on STN  
IN Phenanthrene, 1,8-dibromo-  
MF C14 H8 Br2

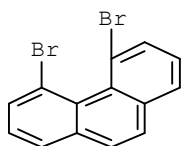


L11 9 ANSWERS REGISTRY COPYRIGHT 2008 ACS on STN  
IN Phenanthrene, 2,7-dibromo-  
MF C14 H8 Br2

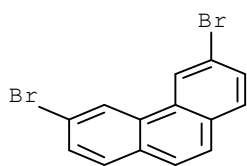
CI COM



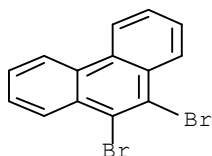
L11 9 ANSWERS REGISTRY COPYRIGHT 2008 ACS on STN  
IN Phenanthrene, 4,5-dibromo-  
MF C14 H8 Br2



L11 9 ANSWERS REGISTRY COPYRIGHT 2008 ACS on STN  
IN Phenanthrene, 3,6-dibromo-  
MF C14 H8 Br2



L11 9 ANSWERS REGISTRY COPYRIGHT 2008 ACS on STN  
IN Phenanthrene, 9,10-dibromo-  
MF C14 H8 Br2  
CI COM



ALL ANSWERS HAVE BEEN SCANNED

Cena: 14 €



## Beilstein

Q1 Je známo CAS RN? NE

Q2 Je znám název? NE

Q3 Je známa struktura? ČÁSTEČNĚ

Q8 Nalezeno substrukturním vyhledáváním? NE

Q4 Je znám sumární vzorec? ANO

Poznámka: Postup je stejný jako v předchozích případech

Počet výsledků: 9

```
=> e phenanthren/cns
E1          17      PHENANTHRAQUINONE/CNS
E2           2      PHENANTHRAZINE/CNS
E3       193578 --> PHENANTHREN/CNS
E4           1      PHENANTHRENA/CNS
E5           3      PHENANTHRENACETAT/CNS

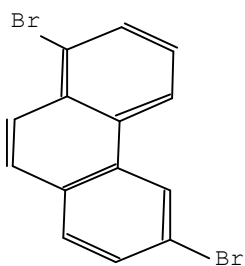
=> s phenanthren#/cns and C14H8Br2/mf
      254257 PHENANTHREN#/CNS
      41 C14H8BR2/MF
L13           9 PHENANTHREN#/CNS AND C14H8BR2/MF

=> d ide 1
```

Na ukázkou vypíšeme pouze jeden záznam.

```
L13 ANSWER 1 OF 9 BEILSTEIN COPYRIGHT 2008 BEILSTEIN MDL on STN

Beilstein Records (BRN):      7703856
Chemical Name (CN):           1,6-dibromophenanthrene
Autonom Name (AUN):           1,6-dibromo-phenanthrene
Molec. Formula (MF):          C14 H8 Br2
Molecular Weight (MW):        336.03
Lawson Number (LN):           4504
Compound Type (CTYPE):        isocyclic
Constitution ID (CONSID):      6579883
Tautomer ID (TAUTID):         7287684
Beilstein Citation (BSO):      6-05
Entry Date (DED):             1997/11/18
Update Date (DUPD):           1998/03/04
```



Field Availability:

Code	Name	Occurrence
BRN	Beilstein Records	1

[...]

**Cena:** 6 € (pokud bychom nechali zobrazit všech 9 záznamů + 44 €)

## 5.1 ZÁVĚR

Nejrychlejší je vyhledání tohoto typu dotazu v databázi SciFinder pomocí speciální funkce ve strukturním editoru. V databázi Beilstein je nutné ošetřit všechny vazebné polohy na rozdíl od substrukturního hledání ve SciFinder. V databázi PubChem se výsledný dotaz skládá ze dvou částí – strukturního a sumárního vzorce, ale i tento způsob vede k uspokojivému výsledku. Dotaz do obou bází v STN se opět skládal z hledání segmentu v kombinaci se sumárním vzorcem. Rozdíl v počtech záznamů v rámci SciFinder (10) a Registry (9) je způsoben tím, že v SciFinderu se objevil navíc i jeden homopolymer.

## 6 REŠERŠE Č. 4: NALEZNĚTE ZÁZNAM KOPOLYMERU STYRENU A ETHYLENU

### SciFinder Scholar

**Q1 Je známo CAS RN?** NE

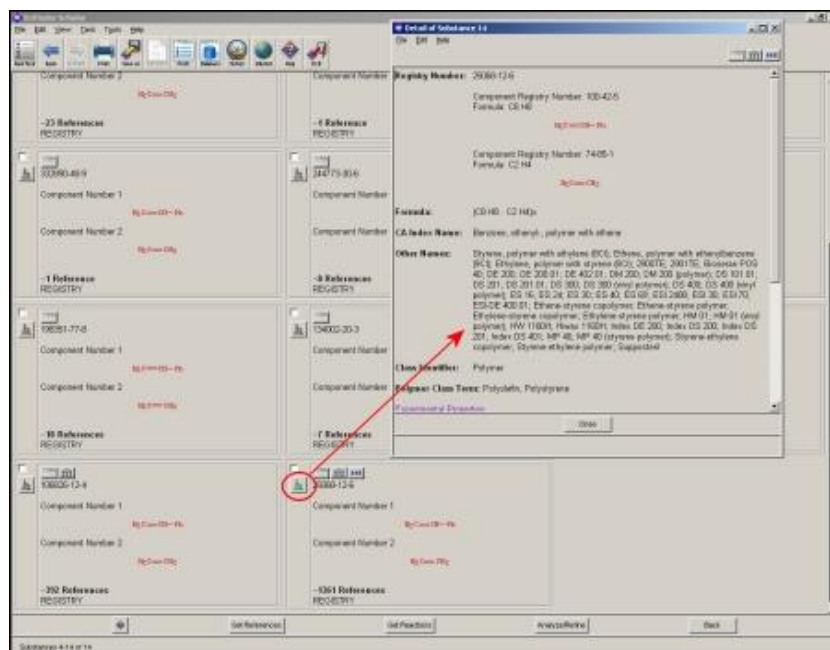
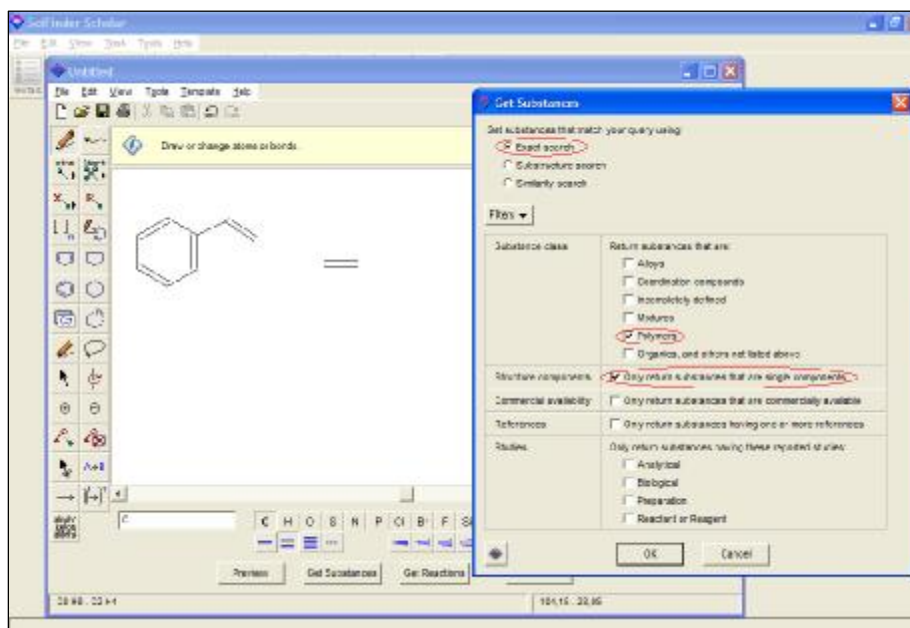
**Q2 Je znám název?** NE

**Q3 Je známa struktura?** ČÁSTEČNĚ

**Q7 Nalezeno strukturním vyhledáváním?** ANO

Poznámka: Zadejme obě struktury najednou, potom vyberme přesné hledání, třídu látek: polymery a hledání v jednosložkových látkách (pokud není zadána poslední volba, systém nalezne polymery sestávající z více různých monomerů).

**Počet výsledků:** 14 - v Registry (1 z výsledků považujeme za správný - má 1361 citací v Chemical Abstracts, to znamená, že půjde o běžnější typ látky, na rozdíl od ostatních nalezených, speciálních typů polymerů.)



## Beilstein

Q1 Je známo CAS RN? NE

Q2 Je znám název? NE

Q3 Je známa struktura? ČÁSTEČNĚ

Q7 Nalezeno strukturním vyhledáváním? NE

Poznámka: Postup jako v SciFinder nefunguje. Vyhledávání je trochu složitější - je potřeba nalézt Beilstein Registry Number (např. strukturním hledáním) obou monomerů a zkombinovat je v poli Composition BRN pomocí „and.“ Poté se do dotazu přidá ještě typ látky „polymer (monomers given)“ – čili všechny polymery, u nichž jsou známy monomery.

**Počet výsledků:** 152 (mezi výsledky jsou ovšem i polymery s další složkou, než jen ethen a styren)

and Search Fields ? Enlarge... Advanced Search Clear Table

	Operator	(	Field name	Relation	Field content	List	)	▲
1			All BRNs(ABRN)	is	1730731			
2	and		All BRNs(ABRN)	is	1071236			
3	and		Type of Substance(STYPE)	is	polymer (monomers given)			

Search Context: Substances Start Search

MDL CrossFire Commander Query Results Reports Alerts AutoMem Substances: Hit 3 of 152

History Prev Hits View all Show Grid List Get Sort Hits Copy Export Hits Copy to Report

**Substance Identification** Hitset Family

Beilstein Registry Number: 9973642  
 Chemical Name: styrene-ethylene copolymer, M<sub>n</sub> ca. 2E5; Monomer(s): ethylene; styrene  
 Type of Substance: [styrene monomers given](#)  
 Composition: Comp. BRN: [1071236](#) phenylethylene  
[1730731](#) ethylene

**Field Availability List 1-2**

Code	Field Name	Occ.
<a href="#">NMR</a>	NMR Spectroscopy	1
<a href="#">GW</a>	Reference	1

## PubChem

Q1 Je známo CAS RN? NE

Q2 Je znám název? NE

Q3 Je známa struktura? ČÁSTEČNĚ

Q7 Nalezeno strukturním vyhledáváním? NE

Poznámka: Strukturní hledání nevedlo k žádným výsledkům, protože k požadované látce není žádná struktura v databázi dostupná. Musíme se vrátit k hledání podle chemického názvu. Ale tady narážíme na problém konkrétního zápisu názvu - místo styrenu, jde zde použít název ethenylbenzen (jak si později pověříme pomocí CAS čísla) ve tvaru: "Benzene, ethenyl-, polymer with ethene". A než přijdeme na ten správný název, hledání vzdáme.

Počet výsledků: 230 (žádná látka neodpovídá zadání.)

NCBI **PubChem Substance** National Library of Medicine My NCBI Sign In Registered

All Databases: Published Residues Proteins Genome Structure PMC PubChem Books

Search: PubChem Substance for 25068-124 Go Clear Save Search

Limits Preview/Index History Clipboard Details

Display: Summary Show: 20 Sort by: Send to:

All: 1 BioAssay: 0 ProteinSD: 0 Rule of 5: 0 OK

SID: [3736063](#) Related Structures

STRUCTURE not available Benzene, ethenyl-, polymer with ethene, 145069-25-6 ...  
 Source: [ChemIDplus\(025068124\)](#)

PubChem Substance Structures supplied by depositors

## STN International

### Registry

Q1 Je známo CAS RN? NE

Q2 Je znám název? NE

Q3 Je známa struktura? ČÁSTEČNĚ

Q7 Nalezeno strukturním vyhledáváním? NE

Poznámka: Tentokrát strukturní hledání obejdeme ne chemickým názvem segmentu, ale polem COMPONENT REGISTRY NUMBER (CRN), které umožňuje identifikovat registrační číslo složky, kterou látka obsahuje. Pole NUMBER OF COMPONENTS (NC) slouží k definování počtu složek, které má daná látka obsahovat. V našem případě se tak podařilo 1332 mezivýsledků zúžit počtem složek na 14.

Počet výsledků: 14

```
=> s 100-42-5/crn and 74-85-1/crn
```

```
79317 100-42-5/CRN
```

```
14563 74-85-1/CRN
```

```
L8 1332 100-42-5/CRN AND 74-85-1/CRN
```

```
=> s 2/nc
```

```
L9 3394930 2/NC
```

```
=> s 18 and 19
```

```
L10 14 L8 AND L9
```

```
=> d scan
```

```
L10 14 ANSWERS REGISTRY COPYRIGHT 2008 ACS on STN
```

```
IN Benzene, ethenyl-, polymer with ethene, syndiotactic, diblock
```

```
MF (C8 H8 . C2 H4)x
```

```
CI PMS
```

```
L10 14 ANSWERS REGISTRY COPYRIGHT 2008 ACS on STN
```

```
IN Benzene, ethenyl-, polymer with ethene, diblock
```

```
MF (C8 H8 . C2 H4)x
```

```
CI PMS
```

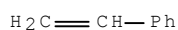
```
L10 14 ANSWERS REGISTRY COPYRIGHT 2008 ACS on STN
```

```
IN Benzene, ethenyl-, polymer with ethene, syndiotactic, graft (9CI)
```

```
MF (C8 H8 . C2 H4)x
```

```
CI PMS
```

L10	14 ANSWERS	REGISTRY	COPYRIGHT 2008 ACS on STN
IN	Benzene, ethenyl-, polymer with ethene, isotactic		
MF	(C8 H8 . C2 H4)x		
CI	PMS		
L10	14 ANSWERS	REGISTRY	COPYRIGHT 2008 ACS on STN
IN	Benzene, ethenyl-, polymer with ethene, graft		
MF	(C8 H8 . C2 H4)x		
CI	PMS		
L10	14 ANSWERS	REGISTRY	COPYRIGHT 2008 ACS on STN
IN	Benzene, ethenyl-, polymer with ethene, syndiotactic		
MF	(C8 H8 . C2 H4)x		
CI	PMS		
L10	14 ANSWERS	REGISTRY	COPYRIGHT 2008 ACS on STN
IN	Benzene, ethenyl-, polymer with ethene, pentablock		
MF	(C8 H8 . C2 H4)x		
CI	PMS		
L10	14 ANSWERS	REGISTRY	COPYRIGHT 2008 ACS on STN
IN	Benzene, ethenyl-, polymer with ethene, syndiotactic, block		
MF	(C8 H8 . C2 H4)x		
CI	PMS		
L10	14 ANSWERS	REGISTRY	COPYRIGHT 2008 ACS on STN
IN	Benzene, ethenyl-, polymer with ethene, alternating (9CI)		
MF	(C8 H8 . C2 H4)x		
CI	PMS		
<b>L10</b>	<b>14 ANSWERS</b>	<b>REGISTRY</b>	<b>COPYRIGHT 2008 ACS on STN</b>
<b>IN</b>	<b>Benzene, ethenyl-, polymer with ethene</b>		
<b>MF</b>	<b>(C8 H8 . C2 H4)x</b>		
<b>CI</b>	<b>PMS, COM</b>		
<b>CM</b>	<b>1</b>		



CM 2



[...]

Hledaný záznam je ztučněn (vzorce jsou u všech výsledků stejné, pro stručnost jsou zde vynechány).

**Cena:** 5 €

### Beilstein\*

**Q1 Je známo CAS RN?** NE

**Q2 Je znám název?** NE

**Q3 Je známa struktura?** ČÁSTEČNĚ

**Q7 Nalezeno strukturním vyhledáváním?** NE

Poznámka: Podobně jako v CA vyhledáme základní monomery podle registračních čísel (tzv. BRN pro Beilstein). Další postup ovšem není tak přímočarý.

**Počet výsledků:** 109

Nejdřív vyhledáme podle jména (na rozdíl od CAS RN není BRN tak snadné zjistit ve zdarma dostupných zdrojích) a pomocí příkazu SEL (select) „vyjmeme“ registrační čísla ze záznamů.

```
=> s styrene/cn
L1          1 STYRENE/CN

=> sel brn
E1 THROUGH E1 ASSIGNED

=> s ethene/cn
L2          2 ETHENE/CN

=> sel brn
E2 THROUGH E3 ASSIGNED
```

\* Za tento návod děkuji paní Renate Hedderich z Helpdesk STN International v Karlsruhe.

Nyní spojíme vyhledávání registračních čísel (COMPBRN = Composition BRN) s typem sloučeniny.

```
=> s e1/compbrn and e2-3/compbrn and polymer?/ctype
      6028 1071236/COMPBRN
      3552 1730731/COMPBRN
           0 5804672/COMPBRN
      66803 POLYMER?/CTYPE
L3 152 1071236/COMPBRN AND (1730731/COMPBRN OR 5804672/COMPBRN) AND POLYMER?/CTYPE
```

Výsledek ovšem zahrnuje i ty polymery, které mají více typů monomerů, ne jen styren a ethen. Je potřeba tyto nadbytečné látky odstranit. V STN to jde poměrně snadno, opět pomocí příkazu SEL vybereme všechny registrační čísla. Pokud výběr zobrazíme, uvidíme, jaký je výskyt registračních čísel.

```
=> sel compbrn
E4 THROUGH E13 ASSIGNED

=> d sel
E1      1      1071236/BRN
E2      1      1730731/BRN
E3      1      5804672/BRN
E4     152     1071236/COMPBRN
E5     152     1730731/COMPBRN
E6      16     3587209/COMPBRN
E7      14     1209317/COMPBRN
E8      14     1696878/COMPBRN
E9       7     3194891/COMPBRN
E10     2     1098262/COMPBRN
E11     1     1731576/COMPBRN
E12     1     1848736/COMPBRN
E13     1     2935169/COMPBRN
```

Zjistili jsme, že BRN s pořadovými čísly E6-E13 jsou monomery, které ve výsledku nechceme. Teď stačí jen předchozí dotaz zopakovat, ovšem s negací všech registračních čísel, která nejsou žádoucí.



```

=> s 13 not e6-13

      427 3587209/COMPBRN
      269 1209317/COMPBRN
     1387 1696878/COMPBRN
      55 3194891/COMPBRN
      63 1098262/COMPBRN
     112 1731576/COMPBRN
     281 1848736/COMPBRN
      53 2935169/COMPBRN

L4      109 L3 NOT (3587209/COMPBRN OR 1209317/COMPBRN OR 1696878/COMPBRN OR
3194891/COMPBRN OR 1098262/COMPBRN OR 1731576/COMPBRN OR 1848736/COMPBRN OR
2935169/COMPBRN)

```

Zobrazíme např. 109. záznam:

```

=> d 109 ide

L4 ANSWER 109 OF 109 BEILSTEIN COPYRIGHT 2008 BEILSTEIN MDL on STN

Beilstein Records (BRN):      8973642
Chemical Name (CN):          styrene-ethylene copolymer, Mn ca. 2E5;
                               Monomer(s): ethylene; styrene
Compound Type (CTYPE):       polymer (monomers given)
Compos.: Comp. Brn (COMPBRN): 1071236, 1730731
Compos.: Comp. Name (COMPN):  vinylbenzene, ethene
Entry Date (DED):            2002/01/24
Update Date (DUPD):          2002/01/24

[...]
```

Cena: 7,40 €

## 6.1 ZÁVĚR

Pokud je výsledkem rešerše velký počet látek se stejnou strukturou, je možné doporučit orientaci podle počtu citací v odborné literatuře. V databázi PubChem i SciFinder bylo možno najít zadaný polymer vyhledáním struktur obou jeho složek najednou. Ovšem v případě databáze PubChem, požadovaná látka neměla přiřazen strukturní vzorec, takže hledání bylo neúspěšné a bylo nutné se pokusit o hledání chemického názvu. V SciFinderu je na rozdíl od druhé testované databáze možnost vyhledávání omezit jen na polymery. Co se týče Registry v rámci STN, pomocí kombinace dotazů jsme dospěli ke 14 výsledkům, stejně jako SciFinder, bohužel se nemůžeme orientovat podle počtu citací v Chemical Abstracts, správný záznam lokalizujeme podle názvu. V databázi Beilstein v rámci STN se podařilo na rozdíl od CrossFire zúžit počet výsledných polymerů jen na ty, které obsahují pouze monomery uvedené v zadání.

## SEZNAM POUŽITÉ LITERATURY

Záznamy odkazující obecně pouze na webové prezentace jsou odděleny a číslovány.

- [Baker, 1980] BAKER, Dale B; HORISZNY, Jean W; METANOMSKI, Wladislaw W. History of Abstracting at Chemical Abstracts Service. *Journal of Chemical Information and Computer Science*. 1980, vol. 20, s. 193-201.
- [Blake, 1990] BLAKE, James E; DANA, Robert C. CASREACT: More than a Million Reactions. *Journal of Chemical Information and Computer Science*. 1990, vol. 30, no. 4, s. 394-399.
- [CAS1] CAS : statistical summary 1907-2004. Columbus, OH : Chemical Abstract Service, 2005. 7 s. Dostupný na WWW: <http://www.cas.org/EO/casstats.pdf>.
- [CAS5] CAS Registry Number and Substance Counts [online]. Chemical Abstract Service, [200?], 2006 [cit.2006-08-13]. Dostupný na WWW: <http://www.cas.org/cgi-bin/regreport.pl>
- [Komise ES] Komise Evropských společenství. Rozhodnutí komise ze dne 8. května 1996, kterým se stanoví soupis a společná nomenklatura přísad používaných v kosmetických prostředcích (96/335/ES). In *Úřední věstník L 132 , 01/06/1996 S. 0001 - 0684*. Dostupný na WWW: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31996D0335:CS:HTML>.
- [Marris, 2005] MARRIS, Emma. American Chemical Society: Chemical reaction. *Nature*. 2005, vol. 437, s. 807-809.
- [MMDB] *MMDB - Entrez's Structure Database* [online]. National Center for Biotechnology Information, [2005], 2005 [cit.2006-07-23]. Dostupný na WWW: <http://www.ncbi.nlm.nih.gov/Structure/MMDB/mmdb.shtml>
- [PubChem - FAQ] *PubChem Help* [online]. National Library of Medicine, [2004], 2006 [cit.2006-08-01]. Dostupný na WWW: <http://pubchem.ncbi.nlm.nih.gov/help.html#faq>
- [PubMed] *What's the Difference Between MEDLINE® and PubMed®?* [online]. National Library of Medicine, 2002, 2004 [cit.2006-07-20]. Dostupný na WWW: [http://www.nlm.nih.gov/pubs/factsheets/dif\\_med\\_pub.html](http://www.nlm.nih.gov/pubs/factsheets/dif_med_pub.html)
- [STN, 2008] *STNews 2/2008 : Finding information about substances that do not have associated references* [online]. STN International, 2008 [cit. 2008-10-10]. Dostupný na WWW: [http://www.stn-international.de/archive/stnews/recent\\_articles/STN%20Ask%20Reggie.pdf](http://www.stn-international.de/archive/stnews/recent_articles/STN%20Ask%20Reggie.pdf)
- [Vymětal, 2001] VYMĚTAL, Jan a kol. *Odborná literatura a informace v chemii*. [Praha] : Orac, 2001. 377 s. ISBN 80-86199-33-9.
- [Weisgerber, 1997] WEISGERBER, David W. Computer- Assisted Topological Analysis and Completion of Chemical Reactions. *Journal of the American Society for Information Science*. 1997, vol. 48, no. 4, s. 349-360.

### Webové prezentace:

PubChem [online]. c2006 [cit.2006-08-10]. Dostupný na WWW <http://pubchem.ncbi.nlm.nih.gov/>.

European chemical Substances Information System : EINECS [online]. c2006 [cit.2006-08-10]. Dostupný na WWW: <http://ecb.jrc.it/esis/>.