|  |  |
| --- | --- |
|  | **PPBRG** |

STN

MANUAL

**This document is electronically controlled. Its accuracy can only be guaranteed when viewed electronically. Effective Date 28/07/2015**

**REVISION HISTORY**

| Date | Name | Revision |
| --- | --- | --- |
| Nov 06 | Marie-Anne Fam | Revision prior to being controlled |
| 31 Mar 10 | Andrew Bryce | Minor checking and updating |
| 31 Mar 13 | Gavin Bartell | Minor checking and updating |
| 28 July 15 | Gavin Bartell | Checking and updating following new fee structure |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

|  |
| --- |
| **CONTENTS** (CTRL+Click on page number to go to that page) |

1. DATABASES AND BACKGROUND 6

1.1 Background 6

1.2 STN Databases 7

1.3 The REGISTRY File 8

A Sample Registry File Display 9

1.4 The CAPLUS File 11

A sample Chemical Abstract display showing different fields 12

1.5 CA Lexicon 13

1.6 BIOSIS 14

A sample BIOSIS display 14

1.7 MEDLINE 15

A sample MEDLINE display 15

1.8 Access to Medline Thesauri 16

A sample FSTA display 18

1.10 WPIDS 19

1.11 AGRICOLA 20

A sample record from AGRICOLA 21

2. BASIC COMMANDS IN STN 22

2.1 The EXPAND Command 22

2.2 Truncation 25

2.3 Boolean Operators 26

2.4 Proximity Operators 28

The (W) operator 29

The (A) operator 31

The (S), (P) and (L) operators 32

2.5 Further Qualification of Proximity Operators nA etc. 34

2.6 Display Options 35

Display Scan 35

Display Abs (D Abs) And Display Bib Abs (D Bib Abs) 35

3. KEYWORD SEARCHING 36

3.1 Background 36

3.2 Searching for Abbreviations, Alternate Spelling and Plurals 37

3.3 Keyword Searching 38

3.4 Multifile Searching Using INDEX 39

3.5 Removing Duplicates in Multifile Searches 40

3.6 Refining answer sets by publication year /PY 43

3.7 Using the RANGE command to search for Publication Year 45

3.8 Refining Answer Sets by Document Type /DT 46

3.9 Removing Patent Documents from Answer Sets and the TRANSFER Command 47

3.10 Refining Answer Sets using CAS Roles 48

3.11 Available Roles 49

4. CHEMICAL NAME SEARCHING 51

4.1 Searching for specific compounds using Chemical Names (/CN) 51

5. MOLECULAR FORMULA SEARCHING 53

5.1 Searching for specific compounds using Molecular Formula (/MF) 53

5.3 Shortcuts for locating the compound of interest 55

Using D SCAN to locate the required compound 55

Using Chemical Name Fragments to locate the required compound. 55

6. REGISTRY NUMBER SEARCHES 56

6.1 Searching Compounds Using Registry Numbers 56

6.2 Searching Compositions and Reactions Using Registry Numbers 57

7. SEARCHING FOR SYNTHETIC PREPARATIONS 58

8. SEARCHING FOR ALLOYS 60

9. SEQUENCE SEARCHING TECHNIQUES 61

9.1 Background 61

9.2 Protein Sequence Searching in the Registry file 62

Exact Sequence Search of Proteins/SQEP 63

Exact Family Sequence Search of Proteins/SQEFP 65

Subsequence Search of Proteins /SQSP 66

Subsequence Family Search of Proteins /SQSFP 71

9.3 Display Options for Subsequence Searches in File Registry 72

9.4 Display Options for Subsequence Searches in File CAPLUS 74

9.5 Use of the Registry Number Locator Field in Subsequence Searching 75

9.6 Searching Nucleic Acid Sequences 77

Searching Exact Nucleic Acid Sequences 77

Subsequence Search of Nucleic Acids 79

10. DGENE 80

10.1 Background 80

10.2 Searching Sequence Data with the GETSEQ RUN Package 81

10.3 Types of Sequence Searches in DGENE 82

10.4 Variability Symbols in DGENE 83

10.5 Specifying Gaps in Subsequence Queries /SQSP, /SQSFP, and /SQSN 84

10.6 Similarity Searching of Biosequences with the BLAST RUN and GETSIM RUN Packages 85

10.7 Display Options in DGENE 88

11. CAS REGISTRY BLAST 90

11.1 Background 90

11.2 Operation of the Search 90

11.3 Format of Search Output 91

12. WEB BASED “NEW” STN 92

13. STRUCTURE DRAWING 93

13.1 Background 93

13.2 Defining Bond Characteristics 95

13.3 Normalised and Exact Bonds 96

13.4 Setting Node Characteristics 98

13.5 Defining Hydrogen Attachments 99

13.6 Defining Non-Hydrogen Attachments 100

14. STRUCTURE SEARCHING IN THE REGISTRY FILE 101

14.1 Types of Structure Searches in Registry 101

14.2 Batch Searches 106

14.3 Saving Answer Sets 107

14.4 Extend Option in Structure Searching 108

14.5 Retrieving information upon Substances that do not have a corresponding reference in CAPLUS 120

15. RETRIEVING DOCUMENTS CITED IN FERs 121

15.1 Background 121

15.2 Retrieving Chemical Abstracts 122

15.3 Retrieving Beilstein Abstracts 123

15.4 Registry Numbers 124

15.5 Chemcats Accession Numbers 125

16. EXPORTING THE SEARCH STRATEGY FOR THE SIS 126

16.1 SIS for Keyword or Sequence searches 127

16.2 SIS for Structure Searches 128

17. DOCUMENT ADMINISTRATION 131

|  |
| --- |
| DATABASES AND BACKGROUND |

1.1 Background

This document is intended to provide an overview of the use of STN for searching within this office. A general and basic outline of search terms is given, as well as technology-specific search protocols. It is intended to be a living document, and as examiners develop new commands or approaches they should be incorporated. This is not a comprehensive guide to the search tools that are available on STN, but rather an overview of the tools most commonly used in this office. Information has been sourced from STN user documentation or office material.

STN (The Scientific and Technical Information Network) provides databases in many areas including chemistry, medicine, biotechnology, engineering, mathematics and physics. The service in Australia is provided by the American Chemical Society (ACS), and the local representative is Vicki O’Neill (e-mail vicki@igroupnet.com). Some user documentation is available through the STN website at [**http://www.cas.org**](http://www.cas.org/)

Use of STN in this office has been limited to the chemical/biotechnology areas despite having electrical and mechanical databases. STN is widely considered to be the best source of information on chemistry since the coverage is comprehensive and designed to index, access and retrieve chemical information. Accordingly, the majority of the material presented here pertains to the pure chemical area. However, any suggestions, comments or additional techniques (particularly in non-chemical areas) would be appreciated and should be submitted to the Search Technical Team for inclusion.

It has been assumed that the reader has some knowledge of using STN Express and how to navigate STN files, so this material has not been included in the present notes. However a brief description of some files and the types of search fields has been included in order to provide some background to the search protocols used. Please note that search terms in this document may be given in upper or lower case, but databases in STN are not case-sensitive.

[Return to top](#TOC)

1.2 STN Databases

STN databases are sourced from a number of suppliers, and accordingly each file differs in its content, the way in which it is indexed, and the type of matter that can be accessed. For example, the Registry file is made up of references to chemical compounds, but contains no abstract or descriptive material. Molecular formulae, chemical name segments and other information relating to the chemical compound can be searched, but keywords relating to any use, study or synthesis are not indexed in this file. In contrast this information is available through the CA file, which contains abstracts of the original article and other information.

Accordingly searchers need to consider the following points:

* what is the invention (generic compounds, sequences, concepts etc.)
* what should be searched (e.g. only examples if claims too broad etc.)
* what is the most cost-effective way of searching the invention (e.g. limit search to patents if the invention is most likely to be in patent literature etc.)
* which file(s) are most appropriate to search
* what does each file cover, what are its limitations
* what is the best or most cost effective way to search the file(s).

In considering these points, knowledge of the database is important, as is a good understanding of technology-specific search databases (e.g. GenomeQuest, Registry). Searchers should familiarise themselves with any file they use in STN by consulting the Database Summary Sheets or other available user documentation. This minimises the risk of searching the wrong area or using incorrect or incomplete search terms. Database summary sheets are available at [**http://www.cas.org/products/stn/dbss**](http://www.cas.org/products/stn/dbss)

The files most commonly used in chemical, biotechnology and pharmaceutical searches are the Registry file, File CAPLUS, BIOSIS, and WPIDS. Medline is also widely used since it provides equivalent abstracts at a lower cost compared with other STN files. File FSTA is used for food technology searches. A brief description of these files follows. File AGRICLOA is used for plants and agricultural chemicals.

[Return to top](#TOC)1.3 The REGISTRY File

The Registry File is a chemical structure and dictionary database containing unique substance records that are produced as new substances are identified by the Chemical Abstracts Service CAS Registry System. The Registry File contains records for all the substances cited in the CAS Registry System. These include substances cited in CAPLUS and CA files, and special registrations, for example, registrations for regulatory lists.

All substance records contain a unique CAS Registry Number and index name. Substance records may also have synonyms, molecular formulae, alloy composition tables, classes for polymers, nucleic acid and protein sequences, ring analysis data, and structure diagrams. Each of these fields may be searched. Nucleic acid sequences from GenBank are also included.

Also displayable in the Registry file are the total number of records citing a particular substance in CAPLUS and CA, and the total number of records in the CAPLUS File for non-specific derivatives. Left truncation is available in the Chemical Name Segment /CNS and Notes /NTE fields.

A sample Registry entry is shown on the next page. The RN entry corresponds to the Registry Number for the compound. The CA Index names follow IUPAC nomenclature for the compound. In this instance a change in nomenclature occurred in the 9th Collective Index 9CI, and the previous CAS Index Name 8CI is also given in the record. The record also contains a number of common or Trade names, and a search of any of these would retrieve the record. Molecular formula MF field and structural formula are indexed and can be searched. The entry provides a list of files containing this compound (LC field) and the number of references to the compound in particular files.

[Return to top](#TOC)

### A Sample Registry File Display

=> d ide 57-88-5

ANSWER 1 REGISTRY COPYRIGHT 2015 ACS on STN

RN 57-88-5 REGISTRY

ED Entered STN: 16 Nov 1984

CN Cholest-5-en-3-ol (3β)- (CA INDEX NAME)

OTHER CA INDEX NAMES:

CN Cholesterol (8CI)

OTHER NAMES:

CN (-)-Cholesterol

CN Δ5-Cholesten-3β-ol

CN 3β-Hydroxycholest-5-ene

CN 5:6-Cholesten-3β-ol

CN Cholest-5-en-3β-ol

CN Cholesterin

CN Cholesteryl alcohol

CN Dythol

CN Lidinit

CN Lidinite

CN NSC 8798

CN Provitamin D

CN SyntheChol

FS STEREOSEARCH

DR 80356-14-5, 80356-33-8, 209124-38-9, 218965-24-3, 262418-13-3,

378185-03-6, 676322-57-9, 732297-95-9, 793670-51-6, 849593-11-9,

856708-55-9

MF C27 H46 O

CI COM

SR CA

LC STN Files: ADISNEWS, ANABSTR, BIOSIS, BIOTECHNO, CA, CABA, CAPLUS,

CASREACT, CBNB, CHEMCATS, CHEMLIST, CIN, CSNB, DDFU, DRUGU, IPA,

MEDLINE, MSDS-OHS, NAPRALERT, PIRA, REAXYSFILE\*, RTECS\*, TOXCENTER,

USPAT2, USPATFULL, USPATOLD, VETU

(\*File contains numerically searchable property data)

Other Sources: DSL\*\*, EINECS\*\*, TSCA\*\*

(\*\*Enter CHEMLIST File for up-to-date regulatory information)

Absolute stereochemistry.



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

223298 REFERENCES IN FILE CA (1907 TO DATE)

12035 REFERENCES TO NON-SPECIFIC DERIVATIVES IN FILE CA

226674 REFERENCES IN FILE CAPLUS (1907 TO DATE)

Displays in Registry are generally of limited use since they do not give any bibliographic details, i.e. citation or publication date. (Note: the date that the registry number first entered STN is provided). However the display can be used to locate files in which compounds are indexed. In the example shown below, a search of a Registry Number in File CA has obtained zero answers. The Registry Number is then searched and displayed in the Registry File. The display indicates that the compound is available only in the Chemcats file.

=> File CA

=> S 198757-90-3/RN

L1……..0 198757-90-3/RN

=> File Reg

=> e 198757-90-3

E# FILE FREQUENCY TERM

-- ---- --------- ----

E1 REGISTRY 1 198757-88-9/RN

E2 REGISTRY 1 198757-89-0/RN

E3 REGISTRY 1 --> 198757-90-3/RN

E4 REGISTRY 1 198757-91-4/RN

E5 REGISTRY 1 198757-92-5/RN

E6 REGISTRY 1 198757-93-6/RN

E7 REGISTRY 1 198757-94-7/RN

E8 REGISTRY 1 198757-95-8/RN

E9 REGISTRY 1 198757-96-9/RN

E10 REGISTRY 1 198757-97-0/RN

E11 REGISTRY 1 198757-98-1/RN

E12 REGISTRY 1 198757-99-2/RN

=> s e3

L2 1 198757-90-3/RN

=> d l2

L2 ANSWER 1 OF 1 REGISTRY COPYRIGHT 2000 ACS

RN \*\*\*198757-90-3\*\*\* REGISTRY

CN L-Tyrosine, L-.alpha.-aspartyl-L-valyl-L-seryl-L-threonyl-L-prolyl-L-

prolyl-L-threonyl-L-valyl-L-leucyl-L-prolyl-L-.alpha.-aspartyl-L-

asparaginyl-L-phenylalanyl-L-prolyl-L-arginyl- 9CI CA INDEX NAME

FS PROTEIN SEQUENCE; STEREOSEARCH

MF C83 H124 N20 O26

SR CAS Registry Services

LC STN Files: CHEMCATS

Absolute stereochemistry.

=> fil chemcats

## 1.4 The CAPLUS File

The CAPLUS File is a bibliographic database available from CAS Chemical Abstracts Service covering international journals, patents, patent families, technical reports, books, conference proceedings, and dissertations from all areas of chemistry, biochemistry, chemical engineering, and related sciences from 1907 to the present. Electronic- only journals and Web preprints are also covered.

The records contain bibliographic information, in-depth substance and subject indexing, including CAS Registry Numbers RN, and abstracts, which are concise summaries of the major findings reported in the scientific literature.

CAPLUS is a more recent and more comprehensive file than the CA File. CAPLUS contains all the records included in the CA File plus records for recent publications that have not yet been fully indexed. In addition, CAPLUS contains all articles from more than 1,500 key chemical journals since October 1994, including records for document types not covered in the CA File such as letters to the editor or news announcements. CAPLUS is updated daily with new bibliographic records and weekly with indexing information.

After publication of an article or patent, ACS generates abstracts that are published as Chemical Abstracts by the Chemical Abstracting Service CAS. All new chemical compounds are assigned Registry Numbers, and these are recorded in the Registry database (exceptions to this rule are discussed in later sections). Other information such as Indexing Terms and Supplementary Terms is also indexed.

The Basic Index /BI is the general default index during searching i.e. if you do not specify an index this one will be consulted. It comprises single words from the titles /TI, abstracts /AB, supplementary terms /ST and index terms /IT. An example of an abstract showing these fields is given below. Other fields may also be searched using the appropriate qualifier (for example /AU for author and so on).

The Indexing Term /IT field is controlled text, which means that all material in this field is assigned by the CAS abstracter. Controlled text includes Index Headings, CAS Registry Numbers, roles and other descriptive text.

Index Headings are obtained from the CAS Index Guides (copies are available in Sections A2 and C2). In the example below the Indexing Terms use the Index Heading "Oxidation". Such Index Headings are followed in a single Index Term field by some descriptive text using controlled text such as standard abbreviations or roles that relates the Index Heading to the matter described in the article. Index Headings may also be selectively searched using the qualifier "Controlled Term" /CT.

CAS Registry Numbers also appear as controlled Index Terms. Like the Index Headings they are accompanied by descriptive information and roles. Registry Numbers are only listed for substances that are of some significance to the original work, such as reactants and products. Thus, unless solvents and the like play some important role, they are not indexed.

The title /TI, abstract /AB and supplementary term /ST fields are free text. Abstracts and titles are supplied by the authors, and supplementary terms are added by the CAS indexer to supplement the information in the title. When searching free text you are relying on different authors using the same terms to describe the same concepts—this is rarely the case. Using the basic index can also lead to a number of additional answers, where a key-word may be present in the abstract, but have little relevance to the new work described in the document.

It therefore appears that the use of indexing or control terms as keyword searching provides an efficient method for searching the Chemical Abstract files. In particular, the terms are used consistently, and the culling of less relevant documents may be achieved through the use of the right proximity operators (see the section on Basic Commands). However, despite the potential for improved efficiency, some care must be exercised when using this type of approach. Control Terms are not consistent between each Cumulative Index, so that a search of the most recent Control Terms may not cover previous material. In order to search efficiently in this manner, previous Index Guides must be consulted to ensure that all synonyms are searched. In most cases the Basic Index will be consulted.

### A sample Chemical Abstract display showing different fields

AN 101:89898 LCA

TI Synthesis of `.alpha.-hydroxycarbonyl compounds acyloins: direct oxidation of enolates using 2-sulfonyloxaziridines

AU Davis, Franklin A.; Vishwakarma, Lal C.; Billmers, Joanne G.; Finn, John

CS Dep. Chem., Drexel Univ., Philadelphia, PA, 19104, USA

SO J. Org. Chem. 1984, 4917, 3241-3

CODEN: JOCEAH; ISSN: 0022-3263

DT Journal

LA English

AN - Accession Number CA reference number

TI - Title of the original citation that has been abstracted

AU - Author names

CS - Corporate Source the name and location of the organization

of the first author

SO - Source publication information

DT - Document type

LA - Language of the original document

AB Direct oxidn. of ketone and ester enolates with oxaziridine I

affords `.alpha.-hydroxycarbonyl compds. in higher yield, with fewer side reactions, and with superior stereoselectivity than similaroxidns. using Mo peroxide-pyridine-Me2N3PO or O2. Competitive addn. of enolates to sulfonimide PhSO2:CHPh is unimportant.

ST stereochem oxidn enolate sulfonyloxaziridine; oxaziridine sulfonyl oxidn enolate; acyloin; hydroxy carbonyl compd

IT Oxidation

of enolates with phenylsulfonylphenyloxaziridine

IT 63160-13-4

oxidn. by, of enolates

[Return to top](#TOC)

## **1.5 CA Lexicon**

The CA Lexicon is an online search tool for the CA indexing terms for concepts, chemical classes and taxonomic vocabulary. The thesaurus is available for records from 1967 to present. Further information on how to use the CA Lexicon is provided at <https://www.cas.org/express/help/v8/express/searchaids/ca_lexicon.htm>

[Return to top](#TOC)

## 1.6 BIOSIS

BIOSIS is a bibliographic database covering worldwide literature on all biological and biomedical topics. Records contain bibliographic data, indexing information, and abstracts for most references. For records prior to 1993, indexing includes Biosystematic Codes, Concept Codes, Miscellaneous Descriptors, and CAS Registry Numbers and corresponding chemical names. Records from 1993 to the present contain additional indexing terms such as Major Concepts, Super Taxa, Organism Names, and Organism Superterms. A sample BIOSIS record is shown below. Fields are searchable using the appropriate qualifier, for example /TI for title. Each Index Term entry is further described with supplementary terms that relate the material in the article to the heading. BIOSIS has an on-line thesaurus from which Index Terms, Control Terms etc. can be obtained. Registry Numbers are also indexed in BIOSIS.

### A sample BIOSIS display

ACCESSION NUMBER AN: 1998:370298 BIOSIS

DOCUMENT NUMBER DN: PREV199800370298

TITLE TI: Association between elevated plasma total homocysteine and increased common carotid artery wall thickness.

AUTHORS AU: Voutilainen, Sari; Alfthan, Georg; Nyyssonen, Kristiina; Salonen, Riita; Salonen, Jukka T. 1

CORPORATE SOURCE CS: 1 Res. Inst. Public Health, Univ. Kuopio, PO Box 1627, FIN-70211 Kuopio Finland

SOURCE SO: Annals of Medicine, June, 1998 Vol. 30, No. 3, pp. 300-306. ISSN: 0785-3890.

DOCUMENT TYPE DT: Article

LANGUAGE LA: English

ABSTRACT AB:Homocysteine is increasingly recognized as a risk factor for atherothrombotic arterial diseases. We investigated the relation…

CONCEPT CODE CC: Public Health: Epidemiology - Miscellaneous \*37056

Pathology, General and Miscellaneous - Diagnostic \*12504 Pathology, General and Miscellaneous - Therapy \*12512 Cardiovascular System - General; Methods \*14501

BIOSYSTEMATIC CODE BC: Hominidae 86215

INDEX TERMS IT: Major Concepts

Cardiovascular Medicine Human Medicine, Medical Sciences; Clinical Chemistry Allied Medical Sciences; Epidemiology Population Studies

INDEX TERMS: Diseases

artherosclerosis: vascular disease; atherothrombotic arterial disease: vascular disease

INDEX TERMS: Chemicals & Biochemicals

homocysteine: plasma, risk factor; vitamin C; vitamin E

INDEX TERMS: Methods & Equipment

B-mode ultrasonography: diagnostic method

INDEX TERMS: Miscellaneous Descriptors

carotid artery intima-media wall thickness

GEOGRAPHICAL TERMS: Finland Europe, Palearctic region

ORGANISM: Super Taxa Hominidae:

Primates, Mammalia, Vertebrata, Chordata, Animalia

ORGANISM: Organism Name human Hominidae:

aged, female, male, middle age, patient

ORGANISM: Organism Superterms

Animals; Chordates; Humans; Mammals; Primates; Vertebrates

REGISTRY NUMBER: 454-29-5Q HOMOCYSTEINE

6027-13-0Q HOMOCYSTEINE

1406-18-4 VITAMIN E

[Return to top](#TOC)

## 1.7 MEDLINE

MEDLINE is a bibliographic database produced by the U.S. National Library of Medicine NLM which covers worldwide biomedical literature. Over 99% of MEDLINE's citations are references to journal articles. Approximately 50% of the citations contain abstracts. Records added before 1975 do not have abstracts. CAS Registry Numbers are also present in the file. About 75% of the citations abstracted in MEDLINE are published in English.

OLDMEDLINE, i.e. data from 1960 through to 1965 from the Cumulated Index Medicus CIM, has been added to the file. MEDLINE contains thesauri in the Chemical Name /CN, current Controlled Term /CT, and MeSH Tree Number /MN fields. The thesauri do not apply to terms in the OLDMEDLINE file segment.

MEDLINE is also available on the web (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed>). Most searches of biological inventions and pharmaceuticals include an initial web-based search of possible keywords in order to refine search strategies. MEDLINE is also available through EPOQUE, and abstracts can be displayed in the EPOQUE Viewer. The MEDLINE thesauri are not searchable in EPOQUE.

### A sample MEDLINE display

AN 2000096682

MEDLINEDN 20096682

TI Trk1 and Trk2 define the major K+ transport system in fission yeast.

AU Calero F; Gomez N; Arino J; Ramos J

CS Departamento de Microbiologia, Escuela Tecnica Superior de Ingenieros Agronomos y Montes, 14080 Cordoba, Barcelona, Spain.

SO JOURNAL OF BACTERIOLOGY, 2000 Jan 182 2 394-9. Journal code: HH3. ISSN: 0021-9193.

CY United States

DT Journal; Article; JOURNAL ARTICLE

LA EnglishFS Priority Journals

OS GENBANK-P47946; GENBANK-Q10065EM 200004EW 20000402

AB The trk1+ gene has been proposed as a component of the K+ influx system in the fission yeast...

CT Check Tags: Support, Non-U.S. Gov't Amino Acid Sequence Biological Transport, Active \*Carrier Proteins: PH, physiology Cell Membrane: ME, metabolism \*Fungal Proteins: PH, physiology Hydrogen-Ion Concentration \*Membrane Proteins: PH, physiology Molecular Sequence Data \*Potassium: PH, physiology Rubidium: ME, metabolism \*Schizosaccharomyces: PH, physiology Sodium: ME, metabolism

RN 136956-54-2 TRK1 protein; 7440-09-7 Potassium; 7440-17-7 Rubidium; 7440-23-5 Sodium

CN 0 Carrier Proteins; 0 Fungal Proteins; 0 Membrane Proteins; 0 TRK2 protein

[Return to top](#TOC)

## 1.8 Access to Medline Thesauri

As stated above, the Chemical Name /CN, Controlled Term /CT, and MeSH Tree Number /MN fields all have thesauri that are available online through STN. The thesauri give the scope of the term as defined by Medline, in addition to terms that are broader BT or narrower NT in scope than the one being searched. The search query can be modified in view of the broader and narrower terms.

The online thesaurus is accessed using the command "expand". It is important to expand potential keywords in order to check their spelling and format in Medline before using the thesaurus. This is because the thesaurus is set up in a specific format, which often contains commas, and in a hierarchical fashion e.g. Drugs, Antiplatelet; Inhibitors, Platelet. Using "platelet inhibitor" instead of "inhibitors, platelet" to search in the thesaurus would fail. The command "Expand <keyword> +ALL/CT" would give the online thesaurus of the <keyword>.

An example of using the CT95 thesaurus for the inhibitors of platelet aggregation is given on the next page. For the meaning of field codes, refer to the Medline STN database summary sheet <http://www.cas.org/File%20Library/Training/STN/DBSS/medline.pdf>. The scope of any MeSH medical subject heading can be obtained online via Pubmed at <http://www.ncbi.nlm.nih.gov/sites/entrez?db=mesh>. This site gives broad and narrow terms. In the below example the ALL/CT operator is used to show both narrower and broader terms, but in order to limit the expanded terms to either of these fields, the operator NT/CT or BT/CT is used.

[Return to top](#TOC)

=> Expand PLATELET AGGREGATION INHIBITORS+ALL/CT95

E1 0 BT3 D Chemicals and Drugs/CT95

E2 0 BT2 Hematologic, Gastric, Renal Agents Non-MeSH/CT95

E3 0 BT1 Hematologic Agents Non MeSH/CT9

E4 5149 --> Platelet Aggregation Inhibitors/CT95  
 MN D19.461.780.  
 DC an INDEX MEDICUS major descriptor

NOTE Drugs or agents which antagonize or impair any mechanism leading to blood platelet aggregation, whether during the phases of activation and shape change or following the dense-granule release reaction and stimulation of the prostaglandin-thromboxane system.

INDX DF: PLATELET INHIB

AQ AD AE AG AN BL CF CH CL CS CT DU EC HI IM IP ME PD

PK PO RE SD ST TO TU UR  
PNTE Blood Platelets 66-87  
PNTE Platelet Adhesiveness 72-87  
PNTE Platelet Aggregation 76-87  
HNTE 88

E5 0 UF Agents, Antiplatelet/CT95  
E6 0 UF Aggregation Inhibitors, Platelet/CT95  
E7 0 UF Antagonists, Blood Platelet/CT95

E8 0 UF Antagonists, Platelet/CT95  
E9 0 UF Antiaggregants, Blood Platelet/CT95  
El0 0 UF Antiaggregants, Platelet/CT95

Ell 0 UF Antiplatelet Agents/CT95  
El2 0 UF Antiplatelet Drugs/CT95  
El3 0 UF Blood Platelet Aggregation Inhibitors/CT95  
El4 0 UF Blood Platelet Antagonists/CT95  
El5 0 UF Blood Platelet Antiaggregants/CT95  
El6 0 UF Drugs, Antiplatelet/CT95  
El7 0 UF Inhibitors, Platelet/CT95  
El8 0 UF Inhibitors, Platelet Aggregation/CT95  
El9 0 UF PLATELET AGGREGATION INHIB/CT95

E20 0 UF Platelet Antagonists/CT95  
E21 0 UF Platelet Antiaggregants/CT95  
E22 0 UF Platelet Inhibitors/CT95  
E23 19540 NT1 spirin/CT95  
E24 5208 NT1 Dipyridamole/CT95  
E25 8845 NT1 Epoprostenol/CT95  
E26 506 NT1 Iloprost/CT95  
E27 1555 NT1 Ketanserin/CT95  
E28 617 NT1 Ticlopidine/CT95

\*\*\*\*\*\*\*\*\* END \*\*\*\*\*\*\*\*\*

[Return to top](#TOC)

1.9 FSTA

FSTA Food Science and Technology Abstracts is a bibliographic database providing worldwide coverage of scientific and technological aspects of the processing and manufacturing of human food products. Citations are in English. The records contain bibliographic information, indexing, and, in most cases (about 95%), an abstract. Information is obtained from journals, books, conference proceedings, reports, patents, pamphlets, legislation and dissertations from 1969 to present. An on-line thesaurus is also available.

### A sample FSTA display

DISPLAY ALL

AN 199802:B0231 FSTA FS FSTA

TI Simultaneous effects of immobilization and substrate protection on the thermodynamics of glucose isomerase activity and inactivation.

AU Converti, A.; Borghi, M. del

CS Fac. of Eng., Inst. of Chem. & Process Eng., Genoa Univ., I-16145 Genoa, Italy

SO Enzyme and Microbial Technology, 1997 21 7 511-517, 23 ref.

ISSN: 0141-0229.

DT Journal

LA English

AB The thermodynamic properties of the glucose-fructose isomerization forward and back reactions, as catalysed by a commercial immobilized glucose isomerase preparation were studied….

CC B Biotechnology

CT FRUCTOSE; GLUCOSE; IMMOBILIZED ENZYMES; ISOMERASES; ISOMERIZATION; STREPTOMYCES; XYLOSE ISOMERASES

TN Sweetzyme T

[Return to top](#TOC)

## 1.10 WPIDS

WPIDS Derwent World Patents Index Subscriber file provides information on patent publications from 41 patent issuing authorities around the world. The records contain bibliographic data, Derwent-assigned titles, abstracts, general indexing, and in-depth chemical and electrical indexing. Additionally, electrical and engineering drawings may be present in records dating back to 1988, and chemical structure drawings may be present in records dating back to 1992. Derwent Chemistry Resource DCR offers structure searching and various other substance identification and indexing fields. DCR search results are linked to bibliographic records. An online thesaurus is available in the Manual Code /MC, Polymer Indexing Enhanced /PLE, and Title Terms /TT fields. There is also a thesaurus-like feature in the DERWENT Compound Number /DCN and DERWENT Registry Number DRN fields.

WPIDS is equivalent to the WPI file accessible through EPOQUE. WPIDS is not part of the IP Australia fixed pricing scheme so any searches in this database should be kept to a minimum (i.e. use EPOQUE whenever possible). If searching WPIDS in STN do not view WPIDS abstracts from STN, instead, accession numbers AN are displayed and then transferred to EPOQUE using STN Extractor (described in the [EPOQUE User Reference Guide](http://docstore.aipo.gov.au/intranet/docstore/technical_communications/Patent_Examiners_Manual/WebHelpFullVersion/patent_searching/epoque/4.12.9.4_Viewing_STN_Search_Results_in_EPOQUE.htm)).

[Return to top](#TOC)

## 1.11 AGRICOLA

AGRICOLA AGRICultural OnLine Access is a bibliographic database of citations to the agricultural literature created by the National Agricultural Library and its cooperators. Production of these records in electronic form began in 1970, but the database covers materials in all formats, including printed works from the 15th century.

Records contain bibliographic information, geographic terms, controlled terms, supplementary terms that include GenBank Numbers, chemical names and CAS Registry Numbers. Abstracts are available for some records. An online thesaurus is available for either the CABA Controlled Term /CT field or the Geographic Term /GT field of the record. The online thesaurus is accessible using the expand command as described in Medline *supra*.

[Return to top](#TOC)

### A sample record from AGRICOLA

Accession Number: 96:51515 AGRICOLA

Document Number: IND20528444

Title: Detection of controlled atmosphere changes in CO2-flushed sealed enclosures for pest and quality management of bagged milled rice.

Authors: Hodges, R.J.; Surendro.

Corporate Source: Natural Resources Institute, Chatham Maritime, Chatham, Kent, UK.

Availability: DNAL 421 J829

Source: Journal of stored products research, Jan 1996. Vol. 32, No. 1. p. 97-104

Publisher: Exeter : Pergamon Press.

CODEN: JSTPAR; ISSN: 0022-474X

Note: Includes references

Pub. Country: England; United Kingdom

Document Type: Article

File Segment: Non-U. S. Imprint other than FAO

Language; English

Abstract:

Long-term pest and quality control for milled rice, by sealing bag stacks into large plastic enclosures flushed with carbon dioxide CO2, is a relatively new technique used by the Indonesian National Logistics Agency BULOG. To improve understanding of this technique, which has potential for more widespread use, an evaluation of gas content was undertaken in a survey of 165 sealed stacks for over 400 days and a limited study made of associated rice quality change yellowing. The CO2 content declined with time in all sealed stacks. In the majority of stacks, oxygen O2 content also declined so that they became anoxic < 2% O2 by 160 days after gassing. In a few cases, this condition was not reached until about 300 days after gassing. Even when stacks failed to conform to current minimum standards of air tightness so that the CO2 content 10 days after gassing fell below the prescribed 50% to as low as 35%, these stacks later became anoxic and no insect-related problems arose. The quality of well or poorly milled rice, judged by the extent of yellowing, was slightly different in samples stored for 9 months, under high CO2,or for the same period, under low O2 conditions, and both were superior to rice stored in a normal atmosphere. It is concluded that in enclosures where the CO2 content has dropped below the BULOG official minimum for long-term storage of 10%, if oxygen levels have fallen to less than 3%, then good quality preservation can still be expected and there is no need for regassing with CO2. A stack classification system is proposed, which can be used to prioritise the discharge of sealed stacks and decide upon the requirement for regassing with CO

Classification: F851 Protection of Stored Plant Products, Insects and Other Arthropods; Ql14 Food Storage, Field Crop Products; Q504 Food Composition, Field Crop Products

Controlled Term CABA: Airtight storage; anoxia; carbon dioxide; change; controlled atmospheres; food quality; fumigation; oxygen; pest control; rice

Supplementary Term: Rice yellowing; sealed stack storage

Geographic Term CABA Indonesia

CAS Registry No: 124-38-9 CARBON DIOXIDE

7782-44-7 OXYGEN

|  |
| --- |
| BASIC COMMANDS IN STN |

2.1 The EXPAND Command

The EXPAND command is useful for choosing search terms, to help you decide how to truncate a term, and to examine what types of search terms are found in the index. When searching for chemical names /cn and molecular formulae /mf, it is essential that you use the EXPAND command to confirm that a particular term exists. The "EXPAND" command is also used in Medline to access the online thesaurus for the Controlled Term /CT. See the previous discussion on the Medline database.

Enter EXPAND or E, at the arrow prompt within a file, followed by the term of interest. If a term is not followed by a field code, the system defaults to the Basic Index of that file. EXPAND shows an alphanumeric listing of the terms that appear around the term entered. Each term is assigned an E-number. The system also shows how many records in the file contain each term. The term entered is always assigned E3. If the term is not in the file, the listing shows 0 postings at the E3 level. To display more than 12 terms, type E or MORE at the next arrow prompt.

=> E PHOTOGRAPHY

E1 6073 PHOTOGRAPHS/BI

E2 1 PHOTOGRAPHSS/BI

E3 23585 --> PHOTOGRAPHY/BI

E4 1 PHOTOGRAPHYCALLY/BI

E5 4 PHOTOGRAPHYING/BI

E6 5 PHOTOGRAPHYS/BI

E7 1 PHOTOGRAPHYU/BI

E8 10 PHOTOGRAPIC/BI

E9 1 PHOTOGRAPICALLY/BI

E10 1 PHOTOGRAPING/BI

E11 1 PHOTOGRAPJIC/BI

E12 1 PHOTOGRAPLIC/BI

[Return to top](#TOC)To view terms in another index, follow the term by a slash and a field code. For example, to view terms in the Title index, follow a term by the field code /TI.

=> E PHOTOGRAPHY/TI

E1 1 PHOTOGRAPHIUC/TI

E2 372 PHOTOGRAPHS/TI

E3 1448 --> PHOTOGRAPHY/TI

E4 1 PHOTOGRAPHYING/TI

E5 6 PHOTOGRAPIC/TI

E6 1 PHOTOGRAPJIC/TI

E7 1 PHOTOGRAPOHIC/TI

E8 1 PHOTOGRAPPHIC/TI

E9 1 PHOTOGRAPYHIC/TI

E10 2 PHOTOGRAVITATIONAL/TI…

To search a term from an EXPAND list, enter in the SEARCH command the E-number corresponding to the search term. You may search one or more terms selected from an EXPAND list, e.g. S E4-E6.

[Return to top](#TOC)In the Basic Index of some files you may also use the command EXPAND LEFT or E LEFT to view terms with any number of characters proceeding the term that you enter. EXPAND LEFT is useful if you want to find out what terms the system searches when you use left truncation.

=> FILE CA

=> E LEFT ASSAY/BI

E1 1 ASSAULT/BI

E2 12 CASSAVA/BI

E3 541 --> ASSAY/BI

E4 1 BIASSAY/BI

E5 85 BIOASSAY/BI

E6 3 ELECTROIMMUNOASSAY/BI

E7 59 IMMUNOASSAY/BI

E8 21 INTERASSAY/BI

E9 4 INTRAASSAY/BI

E10 4 MICROASSAY/BI

E11 3 MICROBIOASSAY/BI

E12 1 PREASSAY/BI

=> E

E13 32 RADIOASSAY/BI

E14 2 RADIOIMMUNASSAY/BI

E15 167 RADIOIMMUNOASSAY/BI

E16 1 RADIORECEPTORASSAY/BI

E17 1 RADIOUMMUNOOASSAY/BI

E18 1 ASSAYABLE/BI

E19 2 BIOASSAYABLE/BI

E20 111 ASSAYED/BI

E21 3 BIOASSAYED/BI

E22 13 ASSAYING/BI

E23 1 RADIOIMMUNOASSAYO/BI

E24 146 ASSAYS/BI

=> S ?ASSAY?

L1 981 ?ASSAY?

[Return to top](#TOC)2.2 Truncation

To obtain variations in word ending, that is, to search on the root of a word, place a question mark ? at the end of the root. The question mark means any number of characters, including zero. For example, a search on ACID? retrieves ACID, ACIDS, ACIDIC, ACIDIFICATION etc.

The hash sign # allows up to one additional character following the word stem. The hash sign is especially useful for searching singular or plural terms. For example, ACID# retrieves terms such as ACID and ACIDS.

The exclamation mark ! is used as a substitute for any single character in the middle of a word, to allow for alternative spellings or meanings. For example, SULF!NYL retrieves SULFINYL and SULFONYL.

In addition, when searching the Basic Index of the CAPLUS file, you can use the question mark at the beginning of a search term. For example, a search on ?ACID?, will retrieve not only ACID, ACIDS and ACIDIFICATION, but also terms such as DIACID or ANTACID.

Note that some files allow both left and right hand truncation when searching. Information on whether left and/or right truncation is permitted in any field may be obtained from the Database Summary Sheet for that file or by using the command HELP SFIELDS at the command prompt within the database of interest.

It should also be noted that for most instances examiners will have the Plurals ON setting enabled (for further information on settings go to **3.2**). When using LEFT truncation, the plurals setting does not work, consequently it will be necessary to use an appropriate RIGHT truncation on the search query.

For example:

L1 323 S (ANIONIC (W) POLYMER) (S) PARTICLE

L2 173 S (ANIONIC (W) POLYMER) (S) ?PARTICLE

L3 403 S (ANIONIC (W) POLYMER) (S) ?PARTICLE?

[Return to top](#TOC)2.3 Boolean Operators

The Boolean search operators are used to specify logical relationship between the terms being searched. They are used in the SEARCH (S) command to combine terms.

Combining a set of terms with the OR operator specifies that at least one term be present in a record in order to retrieve that answer. For example, the search for AIR OR WATER retrieves all records containing one or both of these terms. Using the OR operator usually increases the number of answers that are retrieved.

=> S AIR OR WATER

16289 AIR

42576 WATER

L1 56154 AIR OR WATER

Combining two or more terms with the AND operator specifies that all of the terms occur in the same record. For example, if you search for AIR AND WATER, you retrieve only those records containing both of these words. Use of the AND operator usually decreases the number of answers that are retrieved.

S AIR AND WATER

16289 AIR

42576 WATER

L2 2711 AIR AND WATER

A search expression such as AIR NOT WATER specifies that a record contain the term AIR but not the term WATER. The NOT operator should be used cautiously. A relevant document may contain both the terms AIR and WATER, but the command NOT WATER would eliminate it from the answer set.

=> S AIR NOT WATER

16289 AIR

42576 WATER

L3 13578 AIR NOT WATER

Parentheses should be used where different operators are used (the system has a hierarchy, but in general it is easier to separate terms out using parentheses). For example, a search on acid rain using the term snow as an equivalent, provides two answer sets according to the query used:

S ACID AND RAIN OR SNOW

Results in the answer sets (a) ACID AND RAIN

And (b) SNOW

Whereas;

S ACID AND (RAIN OR SNOW)

Results in the answer sets (a) ACID AND RAIN

And (b) ACID AND SNOW

[Return to top](#TOC)2.4 Proximity Operators

In addition to Boolean operators, proximity operators may be used to limit search sets. Proximity operators are useful in making the search more precise and restricting the number of answers. Proximity operators are used to specify the position of search terms relative to each other. The following proximity operators are available on STN:

(W) terms are adjacent in the specified order

(A) terms are adjacent in either order

(S) terms occur within the same sentence or subfield in any order

(P) terms occur in the same paragraph in any order

(L) terms occur within the same field or information unit

In general a query of two search terms without any proximity operator will obtain abstracts where the terms appear in any field. For example one of the search terms may appear in the abstract, whilst the other is in the Index Term field. Proximity operators can be used to ensure that both terms are in the same field or in a particular order, thereby limiting the answer set to more relevant results.

The use of Control Terms /CT and Indexing Terms /IT with proximity operators is potentially very useful for obtaining highly relevant citations since the indexing by ACS is more consistent than that used by authors. However if using this approach care must be exercised as indexing policies do change over time. A check of older General Subject Indices can confirm this.

[Return to top](#TOC)

### The (W) operator

The (W) operator is used to specify that terms have to be adjacent to each other, in the specified order. The (W) operator also retrieves answers that contain the words separated by punctuation, for example hyphens. This is illustrated by the following example, in which citations concerning high-performance liquid chromatography are sought. This query retrieves answers where the search terms are adjacent to each other or separated by a hyphen. The query would also retrieve answers where the terms were separated by commas, bridged sentences etc., so a more precise search query could be required e.g. High (w) performance (w) liquid (w) chromatography.

=> S HIGH (W) PERFORMANCE

1517809 HIGH

183153 PERFORMANCE

L1 55747 HIGH (W) PERFORMANCE

=> D L1 HIT 1

L1 ANSWER 1 OF 55747 CA COPYRIGHT 1994 ACS

IT Chromatography, column and liquid

\*\*\*high\*\*\* - \*\*\*performance\*\*\* , mobile phase in, for detg.

long chain alkylphenol compns.

Note that (W) may be used with the same terms.

=> S CARBON(W)CARBON

488675 CARBON

488675 CARBON

L2 7126 CARBON(W)CARBON

=> D HIT L2 1

L2 ANSWER 1 OF 7126 CA COPYRIGHT 1994 ACS

AB A miniaturised electrochem. detector cell is described specifically

for microbore liq. chromatog. column internal diams. of 1 mm or

less. The new cell is, in effect, an end-fitting for the microbore

column. The column joins the injection valve and the detector with

no extra column dead vol. A centrosym. radial flow thin-layer

geometry is used with a stainless steel auxiliary electrode and a

ref. electrode without a liq. junction salt bridge. A variety of

working electrodes including glassy \*\*\*carbon\*\*\* , \*\*\*carbon\*\*\*

composites, gold, and platinum can be used, depending on the

application. The working electrodes are disks of 1-6 mm diam.

Improved performance for flow rates of .ltoreq. 200 .mu.L/min is

achieved.

**Implied Proximity (W) in phrases**

When a phrase is entered in the Basic Index of the CA file, the system automatically inserts the (W) operator.

=> S GENETIC ENGINEERING

220764 GENETIC

33899 ENGINEERING

L3 4259 GENETIC ENGINEERING

GENETIC(W)ENGINEERING

You can also separate words in a phrase by using blanks or punctuation.

=> S NUCLEON-NUCLEON INTERACTION

25146 NUCLEON

25146 NUCLEON

404631 INTERACTION

L4 688 NUCLEON-NUCLEON INTERACTION

NUCLEON(W)NUCLEON(W)INTERACTION

The use of the modified proximity operator nW is discussed below.

[Return to top](#TOC)

### The (A) operator

The (A) operator retrieves answers in which the terms are adjacent, but may be in either order. Like the (W) operator the terms may be separated by punctuation.

=> S CHROMATOG?(A)GAS

406097 CHROMATOG?

658176 GAS

L5 110537 CHROMATOG?(A)GAS

Answers:

TI \*\*\*Gas\*\*\* - \*\*\*chromatography\*\*\* method for analysis of

alkylphenols

IT \*\*\*Chromatography\*\*\* , \*\*\*gas\*\*\*

for sepn. and detn. of chlorobenzenes and PCBs and chlorophenols

using silica gel fractionation and electron capture detection

### The (S), (P) and (L) operators

The (S) operator specifies that the search terms must occur within the same subfield in any order, i.e. same title, same supplementary term phrase, same index entry, or the same abstract sentence.

The (P) operator is broader than the (S) operator. The (P) operator specifies that the search terms must occur within the same paragraph in any order.

The (L) operator is broader than the (P) and (S) operators. The (L) operator in the Basic Index of the CA file specifies that terms must occur in the same title, same supplementary term field, same index entry, or the same abstract.

=> S GAS(S)CHROMATOG?(S)PHENOL#

658176 GAS

406097 CHROMATOG?

119326 PHENOL#

L6 1596 GAS(S)CHROMATOG?(S)PHENOL#

Answers:

TI \*\*\*Gas\*\*\* \*\*\*chromatographic\*\*\* analysis for \*\*\*phenol\*\*\*

in the products of vapor-phase hydrolysis of chlorobenzene

ST \*\*\*gas\*\*\* \*\*\*chromatog\*\*\* detn \*\*\*phenol\*\*\* ;

chlorophenol vapor phase hydrolysis analysis phenol

IT 1319-77-3, Cresol

as internal std. in \*\*\*phenol\*\*\* detn. by \*\*\*gas\*\*\*

\*\*\*chromatog\*\*\* .

AB A method has been developed to detect petroleum products and

volatile \*\*\*phenols\*\*\* that included extn., evapn. of

extractant, and \*\*\*gas\*\*\* - \*\*\*chromatog\*\*\* . detection of the

analyzed components in the conc. The conditions for selective extn.

of petroleum products and phenols from water samples when present

together have been established. The method is recommended for

sanitary and toxicol. anal. of the raw and treated sewage.

[Return to top](#TOC)

The (L) operator is particularly useful in searching the specific concepts as indexed in the index term /IT field of CA or CAPLUS files. Chemical Abstracts index the disclosure from an article into the index term /IT field of the CA abstract. The structure of the index term field has two components: (1) a subject heading also called control term /CT, and (2) text modifying phrase. The latter is in parentheses. A list of the Subject heading CT of CA or CAPLUS can be found in Chemical Abstracts - Index Guides.

Thus, searching Chewing Gum/IT and magnesium/IT can retrieve:

(a) Record A

IT Chewing Gum magnesium chloride contg., for dental calculus prevention

(b) Record B

IT Chewing gum bactericidal milt proteins for

IT 9011-16-9d magnesium complexes

(c) Record C

IT Chewing gum magnesium polycarboxylate contg., for dental calculus prevention

Searching Chewing gum/IT (L) Magnesium/IT would only retrieve Records A and C, as these are the only answers in which the two terms are in the same indexing term field.

Searching with the (L) operator can be very specific, but there is always the danger of being too specific in the search strategy and missing important citations. The result of searching with the "(L)" operator should be compared with that from the "and" operator wherein the former has been excluded using the "NOT" operator. Using such a procedure the online searcher can then assess if any important citation would have been lost by using the specific "(L)" operator.

Alternatively, if answer sets are not too large they can be displayed for free using the D SCAN command. This provides a display of the title, supplementary term and indexing term fields. An assessment of whether other keywords, the use of linking search terms using proximity operators and the like should be used can be made with some confidence that important results are not being omitted. This can be time consuming, but does allow the most relevant search terms and proximity operators to be identified.

[Return to top](#TOC)2.5 Further Qualification of Proximity Operators nA etc.

In addition to the above, some proximity operators can be used with the NOT command, e.g. NOTL, NOTP, NOTS, NOTA, NOTW. These specify that the terms are not linked, not in the same paragraph and so on. Again care must be exercised in their use. In general they could be used to limit answer sets after a DSCAN has indicated that certain non-relevant answers could be excluded in this way.

The proximity operators (P), (S), (A) and (W) can also be further qualified using numbers. The proximity (W) or (A) operators are in fact a subset of the (nW) and (nA) proximity operators. By specifying the number of words in "n" the searcher can search the keywords with n or less words separating the specified key words used in the search strategy.

Thus if instead of using "CHROMATOG?(A)GAS" the search query

=> S CHROMATOG?(3A)GAS

was used, the search would retrieve records with the phrase:

1 gas chromatog?;

2 gas xxx chromatog?

3 gas xxx xxx chromatog?

4 gas xxx xxx xxx chromatog?

xxx stands for a word

Note: answer set 3 clearly includes the phrase "gas and adsorption chromatography".

Whilst using the "(nA)" proximity operator in an online search strategy broadens the search when compared with using "(A)", there are situations where such broadening may be beneficial or even justified.

[Return to top](#TOC)2.6 Display Options

Some of the more commonly used formats are provided below. Display formats vary between files, and Database Summary Sheets should be consulted for a comprehensive list of available formats.

### Display Scan

This is a free format that can be used in a search to confirm that queries are correct, or to find additional search terms in order to narrow the search. In File CAPLUS a SCAN gives the fields IC, ICA, ICI, NCL, CC, TI, ST and IT. In File Registry the SCAN gives the fields IN, SQL, MF, CI, STR and COMP. For a complete explanation of what these fields refer to see the relevant Database Summary Sheets. In each case the answers are given in random order and without answer numbers. Individual fields noted above can also be scanned. For example in File CAPLUS the command D SCAN TI gives only the title field.

In Medline and the Derwent files, such as DGENE and WPIDS, the free scan term is D TRIAL.

### Display Abs (D Abs) And Display Bib Abs (D Bib Abs)

These are used in the CAPLUS and Medline files. D Abs is generally used where the bibliographic details are already known and all that is required is the abstract (for example where the Chemical Abstract is cited in a European Search Report). D BIB ABS is used for original searches where the bibliographic details are not known and are required for ordering the citation or for citing the abstract.

For display options used in substructure and subsequence searching, refer to the relevant sections of this manual.

[Return to top](#TOC)

|  |
| --- |
| KEYWORD SEARCHING |

3.1 Background

The first stage of any keyword search is to consider synonyms that may have previously been used to describe the matter being searched. In the chemistry field some guidance is given by the CA Index Guides, and in particular the General Subject Index. For example, the General Subject Heading for dendrimers is "dendritic polymers". Thus a search for dendrimer? gives 734 results, whereas a search for dendrimer? or dendritic polymer# is more comprehensive and gives 848 answers.

The Index Guides also give hierarchies of subject headings in order of increasing specificity, as well as cross references to other areas that may be of relevance. If using the CA Index Guides, a comprehensive search of synonyms can only be carried out if previous Index Guides have been consulted.

The CA Lexicon is an online search tool for the CA indexing terms for concepts, chemical classes and taxonomic vocabulary. The thesaurus is available for records from 1967 to present. Further information on how to use the CA Lexicon is provided at <https://www.cas.org/express/help/v8/express/searchaids/ca_lexicon.htm>.

For searching Biotech and pharmaceutical inventions the Medical Subject Headings- Annotated Alphabetic List is a useful source of synonyms. Some files also have an on-line thesaurus (e.g. Medline – see <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=mesh>)

3.2 Searching for Abbreviations, Alternate Spelling and Plurals

It is good practice to set the system to search for abbreviations, alternate spellings (e.g. color and colour) and for plurals. This is done using the commands:

=> SET ABB ON

(OR => SET ABB ON PERM to set the command on permanently)

=> SET SPELLINGS ON

(OR => SET SPELLINGS ON PERM to set the command on permanently)

=> SET PLU ON

(OR =>SET PLU ON PERM to set the command on permanently)

The system then searches for abbreviated equivalents, alternate spellings and plurals of the query terms without incurring additional search costs. This is particularly useful for searching terms such as "preparation" where the term would be given in the abstract as "preparation", but in the Indexing Terms as truncated terms such as "prep." and "prepn.". The alternative to using abbreviation is to search by truncated terms, for example "prep?", but this leads to additional answers beginning with prep (preponderance, preposition etc.).

[Return to top](#TOC)

3.3 Keyword Searching

Keywords may be searched in one or more files by entering the files and searching terms using the SEARCH command. The search term can be expanded to confirm that the term is used in that field. The example below shows a search of the term mango in CAPLUS. The term MANGO is expanded to confirm that it is present in the index. Note that no field is defined in the query, so the system defaults to the Basic Index /BI. The expanded result indicates that the additional terms MANGOE and MANGOES contain results that may be of relevance. Accordingly the term MANGO## is searched in order to include these results. The approach of expanding the search term should be used for all terms that are not given in the Index Guides.

=> FILE CAPLUS

=> E MANGO

E1 1 MAGNOSITE/BI

E2 1 MANGNOUS/BI

E3 1620 MANGO/BI

E4 1 MANGOBINDPUR/BI

…

E10 1 MANGOE

E11 213 MANGOES

…

=> S MANGO##

L1 1724 MANGO##

To search multiple files the file names are given in the FILE command without commas, for example:

=> FIL CAPLUS MEDLINE

{both the CAPLUS and MEDLINE files are entered}

Note that search terms and on-line costs are charged for each file entered and searched.

Once the keyword search has been performed, it is a useful practice to D SCAN the results. This gives an indication of the types of results that the query has retrieved, and a scan of the Indexing Terms can provide additional terms to either widen or narrow the search. Other fields, such as the titles, can also be scanned by entering the appropriate field qualifier in the display query, e.g. D SCAN TI. The system displays the first answer then asks how many more answers are required. The remaining answers can then be scanned by entering the required number. Only fields that are generally given by D SCAN are available to be displayed in this way.

[Return to top](#TOC)

3.4 Multifile Searching Using INDEX

Multifile searching can also be carried out using STNINDEX. Index allows the search strategy to be tested and can be used with multiple files or clusters. Single files cannot be tested with Index, but in such cases a second 'dummy' file may be entered. When searching in the STNINDEX environment there is no cost per search term, even if there generally is when searching in the single file. Note that any file-specific indexing policies should be taken into account when multifile searching. The following example shows a multifile search using index on chiral glyceryl derivatives.

=> index

ENTER FILE OR CLUSTER NAMES NONE:CAPLUS wpids

INDEX 'CAPLUS, WPIDS' ENTERED AT 19:14:07 ON 17 MAY 2000

2 FILES IN THE FILE LIST IN STNINDEX

Enter SET DETAIL ON to see search term postings or to view

search error messages that display as 0\* with SET DETAIL OFF.

=> s glycer? and chiral?

FILE 'CAPLUS'

185433 GLYCER?

65735 CHIRAL?

748 GLYCER? AND CHIRAL?

FILE 'WPIDS'

37432 GLYCER?

4622 CHIRAL?

38 GLYCER? AND CHIRAL?

L1 QUE GLYCER? AND CHIRAL?

=> fil wpids

=> s l1

37432 GLYCER?

4622 CHIRAL?

L2 38 GLYCER? AND CHIRAL?

=> d an 1-38

L4 ANSWER 1 OF 38 WPIDS COPYRIGHT 2000 DERWENT INFORMATION LTD

AN 1999-600520 [51] WPIDS

In this case the CAPLUS file and the WPIDS file were chosen. The initial answer set for CAPLUS was considered too large, but that found for WPIDS was considered reasonable. Once an answer set is determined in Index, the query is searched by entering the relevant file and searching the L-answer set obtained in Index.

[Return to top](#TOC)

3.5 Removing Duplicates in Multifile Searches

Duplicate answers can also be removed when searching multiple files. The initial step is to SET DETAIL OFF and SET MSTEPS ON. The DETAIL command provides information on the postings, any error messages and the query interpretation abbreviations, plurals, all terms searched when truncation is used etc.. Turning this feature off gives a less cluttered transcript file. The MSTEPS ON command is essential when multifile searching. When the search query is run in multiple files, each of the files is assigned an L-number. This is used for removing duplicates, as seen below.

=> index

ENTER FILE OR CLUSTER NAMES NONE:medline CAPLUS wpids biosis

INDEX 'MEDLINE, CAPLUS, WPIDS, BIOSIS' ENTERED AT 23:55:31 ON 29 MAY 2000

4 FILES IN THE FILE LIST IN STNINDEX

=> set detail off

SET COMMAND COMPLETED

=> set msteps on

(OR => SET MSTEPS ON PERM to set the command on permanently)

SET COMMAND COMPLETED

The next step is to enter the search terms.

=> s antibod? or immunoglob?

606297 FILE MEDLINE

325251 FILE CAPLUS

35899 FILE WPIDS

533045 FILE BIOSIS

L1 QUE ANTIBOD? OR IMMUNOGLOB?

…..

=> s l1 and l2

120 FILE MEDLINE

18 FILE CAPLUS

40 FILE WPIDS

52 FILE BIOSIS

L3 QUE L1 AND L2

Once the search query has been refined in Index, the search is then repeated in the abstract files. If MSTEPS has been set ON, then the files may be simultaneously searched. In this example, only the non-patent literature files are being searched in this way. The desired L-number, in this case L3, is searched and the answer set obtained contains an individual L-answer set for each of the files, as well as an answer set comprising the total number of answers found (this includes duplicate answers).

=> fil medline caplus biosis

COST IN U.S. DOLLARS SINCE FILE TOTAL

ENTRY SESSION

FULL ESTIMATED COST 3.42 3.69

FILE 'MEDLINE' ENTERED AT 23:59:01 ON 29 MAY 2000

FILE 'CA' ENTERED AT 23:59:01 ON 29 MAY 2000

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

PLEASE SEE "HELP USAGETERMS" FOR DETAILS.

COPYRIGHT C 2000 AMERICAN CHEMICAL SOCIETY ACS

FILE 'BIOSIS' ENTERED AT 23:59:01 ON 29 MAY 2000

COPYRIGHT C 2000 BIOSISR

=> s l3

L5 120 FILE MEDLINE

L6 18 FILE CAPLUS

L7 52 FILE BIOSIS

TOTAL FOR ALL FILES

L8 190 L3

Duplicate answers are then removed.

=> set duporder file

(OR => SET DUPORDER FILE PERM to set the command on permanently)

SET COMMAND COMPLETED

=> duplicate remove

ENTER L# LIST OR END:l5 l6 l7

DUPLICATE PREFERENCE IS 'MEDLINE, CAPLUS, BIOSIS'

KEEP DUPLICATES FROM MORE THAN ONE FILE? Y/N:n

PROCESSING COMPLETED FOR L5

PROCESSING COMPLETED FOR L6

PROCESSING COMPLETED FOR L7

L9 146 DUPLICATE REMOVE L5 L6 L7 44 DUPLICATES REMOVED

ANSWERS '1-120' FROM FILE MEDLINE

ANSWERS '121-131' FROM FILE CAPLUS

ANSWERS '132-146' FROM FILE BIOSIS

=> d bib abs 1-146

If the searcher did not put in the set duplicate order command prior to the duplicate command, STN would just duplicate, continue with the removal and produce the same result in L9 but with the records arranged in a chronological order, and not arranged according to the source Medline, Biosis, CA. In other words, record 1 may be from Medline, record 2 from CAPLUS record 3 from Biosis, record 4 from Biosis, and so on.

[Return to top](#TOC)3.6 Refining answer sets by publication year /PY

Publication year of the citation is searched in the /PY index. Publication year is specified as a 4-digit year, e.g. 1990. You may search for specific years or ranges of years, but it is sometimes more efficient to use the RANGE command (see the next section). The following examples illustrate the use of year ranges:

To find only documents published in a given year, *e.g*. 1994, add 1994/PY or PY=1994 to the search query.

S L1 and PY=1994

To find documents published in a range of years, e.g. 1990-1993, add 1990-1993/PY to the search query.

S L1 and 1991-1993/PY

To find documents published after a certain year, for example, after but not including 1992, add PY>1992 to the search query.

S L1 and PY>1992

To find documents published in or after a certain year, i.e. 1993 and beyond, add PY>=1993 to your search query.

S L1 and PY>=1993

To find documents published earlier than a particular year, for example earlier than 1978, and not including 1978, search PY<1978.

S L1 and PY<1978

[Return to top](#TOC)To restrict search results to a specific publication year or a range, the L-number for the search is combined with the appropriate search term in the /PY index, as shown below.

=> S CYSTIC FIBROSIS

3981 CYSTIC

7110 FIBROSIS

L1 2740 CYSTIC FIBROSIS

CYSTIC(W)FIBROSIS

=> S L1 AND PY>=1993

623245 PY>=1993

L2 462 L1 AND PY>=1993

[Return to top](#TOC)

3.7 Using the RANGE command to search for Publication Year

Using RANGE parameters allows the search to be limited to a specified portion of a file. To limit the search to a particular range, enter the range limits in the SEARCH command. For example:

=> S CARDIAC RAN=(1995,1997) (searches the range to 1995-1997)

=> S CARDIAC RAN=(,1997) (searches the range up to 1997)

=> S CARDIAC RAN =(1997,) (searches the range after 1997)

If doing series of searches in a given range, the range limits can be set in the SET RANGE option and subsequent searches are then executed in that specified file portion until the range is changed or another file entered. Enter HELP SET RANGE and HELP SEARCH RANGE for details of using RANGE parameters. For example, to set the range in a series of searches:

=> SET RANGE=(1995,1997) (sets the range to 1995-1997)

=> SET RANGE=(,1997) (sets the range up to 1997)

=> SET RANGE =(1997,) (sets the range after 1997)

RANGE parameters are file-dependent. Enter HELP RANGE at an arrow prompt (=>) in the file to see what ranges may be specified for the file. In most files, searches can be restricted to Accession Numbers or to entry date years, i.e. years in which the records were entered into the database. In the CAPLUS file, the range can also be limited to particular Cumulative Indexes.

When using years to define the range, it is important to note that these are the dates that records entered the file, not publication dates. If you want to restrict a search to publication years, add the Publication Year (/PY) to your search strategy rather than including RANGE parameters. However, the RANGE command may have some advantages over the /PY command. For example, in the WPIDS file, if the /PY command is used the system searches each answer for the publication date (in some cases a very slow process). By searching in a given range, the answer set is given more quickly.

[Return to top](#TOC)

3.8 Refining Answer Sets by Document Type /DT

The results of a search can be restricted to certain kinds of publications by searching document types in the /DT field. The following document types may be searched in the /DT index:

B/DT or BOOK/DT

P/DT or PATENT/DT

C/DT or CONFERENCE/DT

D/DT or DISSERTATION/DT

J/DT or JOURNAL/DT

T/DT or TECHNICAL REPORT/DT

GR/DT or GENERAL REVIEW/DT or REVIEW/DT

The following example shows a search for reviews on cystic fibrosis published in or after 1993.

=> S CYSTIC FIBROSIS

3981 CYSTIC

7110 FIBROSIS

L1 2740 CYSTIC FIBROSIS

CYSTIC(W)FIBROSIS

=> S L1 AND PY>=1993 AND REVIEW/DT

623245 PY>=1993

973977 REVIEW/DT

L4 65 L1 AND PY>=1993 AND REVIEW/DT

=> D

L4 ANSWER 1 OF 65 CA COPYRIGHT 1994 ACS

AN 121:32067 CA

TI \*\*\*Cystic\*\*\* \*\*\*fibrosis\*\*\*

AU Aoki, Takahisa

CS Cancer Inst., Japanese Found. Cancer Res., Tokyo, 170, Japan

SO Jikken Igaku \*\*\*1994\*\*\* , 126, 790-1

CODEN: JIIGEF; ISSN: 0288-5514

DT Journal; \*\*\*General Review\*\*\*

LA Japanese

[Return to top](#TOC)

3.9 Removing Patent Documents from Answer Sets and the TRANSFER Command

In the pure chemical area, this is a particularly useful way in which to reduce the cost of searches. For example, patent documents may be removed from answer sets in the CAPLUS file by searching for patent documents using the qualifier p/dt. The patent documents thus obtained can be transferred to the WPIDS file and then transferred to the EPOQUE Viewer using STN Extractor. The non-patent literature results can then be further manipulated with additional keywords, publication dates and the like. The following example shows a search of this type. The answer set L6 is searched for patent documents using the qualifier P/DT. The resulting answer set is then transferred into the WPIDS file using the TRANSFER command in conjunction with the display fields PN APPS. This finds Derwent abstracts having the same priority number or application number. The accession numbers are then displayed free of charge. These answers are transferred to the EPOQUE Viewer using STN Extractor.

=> s l6 and p/dt

2552546 P/DT

L7 223 L6 AND P/DT

=> fil wpids

COST IN U.S. DOLLARS SINCE FILE TOTAL

ENTRY SESSION

FULL ESTIMATED COST 3.50 4.19

FILE 'WPIDS' ENTERED AT 17:06:12 ON 28 FEB 2000

COPYRIGHT C 2000 DERWENT INFORMATION LTD

=> transfer

ENTER L# L7 OR ?:l7

ENTER ANSWER NUMBERS, RANGES 1-, OR ?:1-

ENTER DISPLAY FIELDS TI OR ?:pn apps

SELECT IS APPROXIMATELY 49% COMPLETE

L8 TRANSFER L7 1- PN APPS : 1822 TERMS

SEARCH OF L8 IS APPROXIMATELY 34% COMPLETE

SEARCH OF L8 IS APPROXIMATELY 61% COMPLETE

SEARCH OF L8 IS APPROXIMATELY 86% COMPLETE

L9 252 L8

=> d l9 an 1-252

L9 ANSWER 1 OF 252 WPIDS COPYRIGHT 2000 DERWENT INFORMATION LTD

AN 1999-610737 [52] WPIDS

.….

[Return to top](#TOC)3.10 Refining Answer Sets using CAS Roles

CAS Roles are CAS Indexing Terms consisting of 3- or 4-letter codes which describe the new or novel information reported about a substance or class of compounds. Roles are assigned to every indexed substance and to controlled index terms for classes of compounds. They are only available in the CA, HCA, CAPLUS and HCAPLUS files, and may be searched and displayed in these files.

An example of a search using roles to find synthetic preparations of fullerenes is shown below. The Chemical Name Segment /CNS fullerene is searched in the Registry file to obtain an answer set L2 comprising compounds containing this name fragment. Then Control Term “fullerenes” obtained from the General Subject Index is searched in the CAPLUS file to supplement the answer set from Registry. The terms are searched with the qualifier /SPN to limit the answer set to answers describing the synthetic preparation of fullerenes.

=>FILE REGISTRY

=>S FULLERENE/CNS

L2 4370 FULLERENE/CNs

=>FILE CAPLUS

=> S L2/SPN OR FULLERENES/SPN

6515 L2

1056276 SPN/RL

791 L2/SPN

L2 L SPN/RL

2289 FULLERENES/CT

1056276 SPN/RL

306 FULLERENES/SPN

FULLERENES/CT L SPN/RL

L3 928 L2/SPN OR FULLERENES/SPN

=>D TI HIT 101

L3 ANSWER 101 OF 928 CAPLUS COPYRIGHT 1996 ACS

TI Process for the synthesis of fullerenes

IT \*\*\*Fullerenes\*\*\*

RL: \*\*\*SPN Synthetic preparation\*\*\* ; PREP Preparation

prepn. of fullerene contg. soots from hydrocarbons and

halocarbons in plasma reactor

IT \*\*\*99685-96-8P\*\*\* , C60 Fullerene \*\*\*115383-22-7P\*\*\* , Fullerene

[Return to top](#TOC)3.11 Available Roles

For a list of the roles available in CAS see <http://www.cas.org/File%20Library/Training/STN/User%20Docs/casroles.pdf>

Roles can be combined with CAS Registry Numbers to limit answer sets to answers in which the compounds have the desired use or property. This approach can be taken for searching reactions, as in the following example, wherein two specific compounds having the Registry Numbers shown are alternative starting materials in a reaction process. The Registry Numbers are located in the Registry file, then searched in the CAPLUS file with the role Reactant /RCT.

=> s 152044-54-7/rn

L1 1 152044-54-7/RN

…

=> s 152044-53-6/rn

L2 1 152044-53-6/RN

=> s l1 or l2

L4 2 L1 OR L2

=> fil ca

=> s l4/rct

141 L4

2261051 RCT/RL

L5 19 L4/RCT

L4 L RCT/RL

=> d all

Pre-1994 abstracts are not retrieved using specific roles. A rough way of retrieving pre-1994 abstracts using a super role is to search the year range PY<1994 together with the role. However, some abstracts may be missed using this approach as indexing of specific roles began late in 1994. A more accurate way is to use the proximity operator NOTL. In post-1994 abstracts, both specific and super roles are given in individual Index Term fields, as exemplified below.

IT 9028-13-1P Homoserine dehydrogenase

RL: PRP properties; PUR purification and recovery; PREP preparation

In order to limit answers to those in which only the super roles are given, the role PREP must not be present in an Index Term field with a specific role. The NOTL proximity operator limits the answer set to those in which the PREP role is not in the same information as a specific role. The specific roles associated with the PREP super role are BMF, BPN, BYP, IMF, PUR, PNU and SPN (see the list of roles above), and the search statement used to achieve this is given as:

S L1/PREP NOTL (BMF OR BPN OR BYP OR IMF OR PUR OR SPN OR PNU)/RL

[Return to top](#TOC)

|  |
| --- |
| CHEMICAL NAME SEARCHING |

4.1 Searching for specific compounds using Chemical Names (/CN)

This approach to searching requires correct IUPAC nomenclature for the compound of interest. If the structure is complicated it is preferable to search the compound by molecular formula. The first step is to EXPAND the name in the REGISTRY file to ensure that the compound is in the file or at least that the nomenclature is correct. The E-number corresponding to the compound is then searched. The EXPAND command is free, and saves any unnecessary search costs if the compound is not in the file.

=> FILE REGISTRY

=> E ALLYL CHLORIDE

E1 1 ALLYL CHLORIDE ETHER/CN

E2 1 ALLYL CHITOSAN/CN

E3 1 🡪 ALLYL CHLORIDE/CN

E4 1 ALLYL CHLORIDE OLIGOMER/CN

E5 1 ALLYL CHLORIDE OLIGOMERS/CN

E6 1 ALLYL CHLORIDE POLYMER/CN

E7 …..

=> S E3

L1 1 ALLYL CHLORIDE/CN

Generally only one result will be obtained for the chemical name. If more than one result is obtained then the query should be checked. If the qualifier /CN is omitted the system defaults to the basic index. A large number of answers are obtained since the basic index will search name fragments instead of complete names. It is also possible for there to be more than one record or Registry number with the same name. If the query is checked and considered correct then the search should be continued.

The following conversions are required when writing out names with special characters:

(a) Ignore italics.

(b) Enter superscript or subscript characters on the same line as regular characters.

(c) Spell out Greek letters with a period (.) before and after the name of the letter.

(d) Replace square brackets [] with parentheses .

(e) Put quotes (") around names that contain apostrophes or primes (').

(f) Put quotes around names that contain parentheses when using the SEARCH command (this is not necessary when using the EXPAND command).

For example:

Tricyclo[3.3.2.02,8]deca-3,6-diene is written as Tricyclo(3.3.2.02,8)deca-3,6-diene

α-Methylbenzoin is written as .alpha.-methylbenzoin

N,N'-dimethyl-p-phenylenediamine is written as "N,N'-dimethyl-p-phenylenediamine"

If zero answers are obtained from the EXPAND then the name should be checked and re-entered if necessary.

This approach is fraught with difficulties. Some of the more common errors are:

(a) the name may need to be inverted, as in 2-amino-3-vinylpyridine.

=> E 2-AMINO-3-VINYLPYRIDINE/CN

E1 1 2-AMINO-3-SULFONAMIDOPYRIDINE/CN

E2 1 2-AMINO-3-TRIFLUOROMETHYLPYRIDINE/CN

E3 0🡪 2-AMINO-3-VINYLPYRIDINE/CN

.

.

=> E PYRIDINE, 2-AMINO-3-VINYL-/CN

E1 1 PYRIDINE, 2-AMINO-3-NITRO-6-PROPYL-/CN

E2 1 PYRIDINE, 2-AMINO-3-NITRO-6-PROPYL, PICRATE/CN

E3 1🡪 PYRIDINE, 2-AMINO-3-VINYL-/CN

(b) typographical errors and errors in punctuation. If zero answers are obtained and the nomenclature is correct it is most likely due to a typographical error. The query should be checked for incorrect numbers and to ensure that the conversion of special characters has been carried out correctly. Punctuation is a particularly trying problem. Insertion or omission of spaces, commas and the like can result in zero answers. For example "methyl paraben" may have no postings if the convention is to name it as "methylparaben".

(c) portions of the name have been entered in the wrong order. The correct order generally follows IUPAC rules.

(d) an incomplete name has been entered. For example a search of dichloroacetone may obtain zero postings since the position of the chloro atoms has not been specified (1,1- or 1,3-).

As noted above the searching of compounds by chemical name alone is particularly difficult, and as a result is generally avoided. Within the pure chemical area compounds are generally initially searched by their molecular formula in order to obtain the correct chemical name, and then a specific search of the chemical name carried out. This is described in the section on molecular formula searching.

[Return to top](#TOC)

|  |
| --- |
| MOLECULAR FORMULA SEARCHING |

5.1 Searching for specific compounds using Molecular Formula (/MF)

This approach to searching for specific compounds is particularly useful when the IUPAC name is not known, or where a chemical name is not found using the command EXPAND name/CN.

The first step is to calculate the molecular formula and arrange the elements in Hill order. For carbon-containing compounds, the Hill order means that carbons are listed first, followed by hydrogens, and then all other elements are listed in alphabetical order. For compounds that do not contain carbon, simply arrange the elements in alphabetical order and indicate the number of each element. EXPAND on the molecular formula followed by /MF to indicate that you want to look in the molecular formula index. EXPAND is useful to both verify the search term, find the number of compounds with that search term, and to see related molecular formulae. In this example, EXPAND on the molecular formula shows that there are many substances with this molecular formula. Consequently, further refining of the answer needs to be done.

=> FILE REGISTRY

=> E C14H18N4O3/MF

E1 1 C14H18N4O2TL/MF

E2 1 C14H18N4O2ZR/MF

E3 248 --> C14H18N4O3/MF

E4 1 C14H18N4O3.C2H4ONC15H24O/MF

E5 1 C14H18N4O3.CH2OX/MF

E6 1 C14H18N4O3.1/2CL4PT.H/MF

E7 1 C14H18N4O3.1/2H2O4S/MF

E8 1 C14H18N4O3.2C2H6O3S/MF

E9 1 C14H18N4O3.2C7H3IN2O3/MF

E10 1 C14H18N4O3.2C7H6O2/MF

E11 3 C14H18N4O3.2CLH/MF

E12 3 C14H18N4O3.BRH/M

=> S E3

L1 248 C14H18N4O3/MF

=> D SCAN

L 248 ANSWERS CA COPYRIGHT 1994 ACS

[Return to top](#TOC)5.2 Multicomponent Substances

Multicomponent substances such as salts, addition compounds, mixtures, polymers and solvates have a complete molecular formula in the MF index that is made up of two or more complete formulae. These are given as dot-disconnected formulae, and the following rules are followed:

1. In the case of acid salts, the metal ion is removed and replaced with hydrogen i.e. to give the parent acid. The formula is then searched as the dot-disconnected combination of the parent acid and the metal, e.g. sodium acetate CH3CO2Na is searched as CH3CO2H.Na/MF in the Registry file.
2. In the case of primary and secondary amine salts, one of the hydrogens is removed from the amine and added to the anion, e.g. trimethyl ammonium chloride HN(CH3)3Cl is searched as C3H9N.ClH/MF. Note that the second component does not contain a carbon atom, so the atoms are given in alphabetical order rather than hydrogen before chlorine.
3. If the compound is a tertiary amine salt i.e. the nitrogen does not have an hydrogen atom attached, then the formula of the ammonium component is dot-disconnected with the anion atom, e.g. CH3-(CH2)13-N(CH3)3+Br- is searched as C17H38N.Br/MF.
4. Mixtures are given as each of the individual components separated by a dot following the order:

(i) If there is a carbon-containing and a non-carbon-containing component, then the carbon containing component is first;

(ii) If there are no carbon-containing components, then the order is alphabetical using the first element of each component. This arrangement is also used for the indexing of alloys;

(iii) If all components contain carbon, then the component with the highest number of carbons is first with the remaining components arranged in descending order;

(iv) If two or more components have the same number of carbons, then the order is determined by the number of hydrogen atoms in each component, with the highest number being first;

(v) If two or more components have the same number of carbons and hydrogens, then the order is alphabetical using the remaining elements.

E.g. C2H4F2.C2ClF5.CHClF2. The first component has a higher number of hydrogens than the second, which in turn has a higher number of carbons than the third component.

1. Fractions can be used in salts and the like:

e.g. Ba.1/10Fe2O13

1. Polymers are entered as for multicomponent substances, with the entry in parentheses followed by x:

e.g. (C8H8.C2H4)X.

[Return to top](#TOC)5.3 Shortcuts for locating the compound of interest

### Using D SCAN to locate the required compound

Once the molecular formula has been expanded and searched, the answer set may be scanned for free using the D SCAN command. This is used where there are not too many answers found with the desired molecular formula.

The chemical structure and the IUPAC name are given by the D SCAN. A simple and quick method of narrowing the answer set to the specific compound is to copy and paste the name from the scan, then expand and search the name as for a chemical name search. Note that any unusual symbols must be treated as described for searching chemical names.

### Using Chemical Name Fragments to locate the required compound.

If too many answers are obtained from a molecular formula search and it is inconvenient to D SCAN the answer set to locate the required compound, then the molecular formula search results may be combined with chemical name fragments. The basic or default index is used to add the name fragments (the molecular formula index includes only whole names, but the basic index includes name fragments). Thus, if the term butyl was added as a molecular formula term, zero hits would be obtained. However, by adding the term as a name fragment from the basic index the answer set is reduced to compounds including the name fragment, as shown below.

=> S L1 AND BUTYL

248 C14H18N4O3/MF

526169 BUTYL

L2 28 C14H18N4O3/MF AND BUTYL

=> S L2 AND AMINO AND METHYL

1927517 AMINO

6665678 METHYL

L3 10 L2 AND AMINO AND METHYL

[Return to top](#TOC)

|  |
| --- |
| REGISTRY NUMBER SEARCHES |

6.1 Searching Compounds Using Registry Numbers

Registry number searches can be used where the invention relates to a known compound, for example new uses for known compounds, new formulations, new crystal structures polymorphs, and the like. In general, the Registry Number can be obtained from standard references, such as the Merck Index or Dictionary of Organic Compounds. The EXPAND command is used to confirm that the desired Registry Number is in the Registry file, then searched to give a Registry answer set. This answer set is then searched in the CAPLUS file (other bibliographic files that index CAS Registry Numbers can also be searched to retrieve any citations referring to that compound).

=> e 208783-24-8/rn

E# FILE FREQUENCY TERM

-- ---- --------- ----

E1 REGISTRY 1 208783-22-6/RN

E2 REGISTRY 1 208783-23-7/RN

E3 REGISTRY 1 --> 208783-24-8/RN

E4 REGISTRY 1 208783-25-9/RN

E5 REGISTRY 1 208783-26-0/RN

E6 REGISTRY 1 208783-27-1/RN

E7 REGISTRY 1 208783-28-2/RN

E8 REGISTRY 1 208783-29-3/RN

E9 REGISTRY 1 208783-30-6/RN

E10 REGISTRY 1 208783-31-7/RN

E11 REGISTRY 1 208783-32-8/RN

E12 REGISTRY 1 208783-33-9/RN

=> s e3

L1 1 208783-24-8/RN

=> fil ca

=> s l1

L2 2 L1

=> d bib abs 1-2

The following should be noted:

1. if the invention relates to a specific stereoisomer, then the Registry Number entered must be that of the required stereoisomer. Registry Numbers are assigned to each individual stereoisomer. For example, L-alanine is assigned the Registry Number 56-41-7, whilst DL-alanine is assigned 302-72-7;
2. similarly, different salts, hydrates and the like are assigned different Registry Numbers. The Registry Number for the specific salt or solvate must be used.

6.2 Searching Compositions and Reactions Using Registry Numbers

Registry numbers can also be used to search reactions, pharmaceutical preparations and the like. For example, if a pharmaceutical or agricultural preparation contains several known components for which Registry numbers can be readily obtained, a search for abstracts containing each of these Registry numbers will retrieve the composition.

Similarly, reactions can be searched using the Registry numbers of the reactants. The following example shows a search of this type. The claims in question define a process wherein fluoroalcohols are prepared by reacting tetrafluoroethylene or hexafluoropropylene with methanol in the presence of a free radical source. The Registry numbers of methanol L14, tetrafluoroethylene L15 and hexafluoropropylene L16 are first located and searched. Tetrafluoroethylene L15 and hexafluoropropylene L16 are alternative reactants and a search of L15 OR L16 gives the answer set L17. A search of this answer set with methanol L14 gives an answer set containing methanol together with tetrafluorethylene or hexafluoropropylene. The answer set in such searches can be further limited by adding the registry numbers of other reactants or products, or by the addition of other keywords.

=> FIL CAPLUS

=> s 67-56-1/rn

L14 81591 67-56-1/RN

=> s 116-14-3/rn

L15 2350 116-14-3/RN

=> s 116-15-4/rn

L16 1621 116-15-4/RN

=> s l15 or l16

L17 3543 L15 OR L16

=> s l17 and l14

L18 69 L17 AND L14

[Return to top](#TOC)

|  |
| --- |
| SEARCHING FOR SYNTHETIC PREPARATIONS |

To find references on the synthetic preparation of a specific substance, the CAS Registry Number is searched using the qualifier /P in the CAPLUS file.

=> FILE CAPLUS

=> S 84-74-2P

L1 138 84-74-2P

If a search has already been carried out in the Registry File, the L-answer set from the Registry file is searched in File CAPLUS with the qualifier /P. This approach is also applicable to any other type of search in the Registry file. For example a substructure search gives an answer set that comprises Registry Numbers. Therefore a similar protocol of searching the answer set in File CAPLUS with the qualifier /P will limit results to preparations.

=> FILE REG

=> S 84-74-2/RN

L1 1 84-74-2/RN

=> FILE CAPLUS

=> S L1/P

L2 138 84-74-2P

The following example shows a search for the CAS Registry Number of dibutyl phthalate followed by a search for references dealing with preparation of that substance. CAS Roles may also be used to search for methods of preparation, and have the advantage of further characterising the preparation as Biosynthetic manufacture, Biosynthetic Preparation, Byproduct, Industrial Manufacture, Purification or Recovery, Preparation Unclassified, and Synthetic Preparation.

=> FILE REGISTRY

=> E DIBUTYL PHTHALATE/CN

E1 1 DIBUTYL PHOSPHORODIAMIDATE/CN

E2 1 DIBUTYL PHOSPHORODITHIOATE/CN

E3 1 --> DIBUTYL PHTHALATE/CN

E4 1 DIBUTYL PHTHALATE-ETHYLENE GLYCOL-TEREPHTHALIC ACID POLYMER/CN

E5 1 DIBUTYL PHTHALATE-VINYL ACETATE-VINYL CHLORIDE COPOLYMER/CN

…

=> S E3

L1 1 "DIBUTYL PHTHALATE"/CN

=> FILE CAPLUS

=> S L1/P

L2 138 L1/P

=> D SCAN

L2 138 ANSWERS CA COPYRIGHT 1994 ACS

CC 22-8 Physical Organic Chemistry

TI Thermal analysis of .beta.-substituted phthalic acid diesters

ST phthalate dialkyl thermal stability; pyrolysis dialkyl phthalate

kinetics

IT Kinetics of thermal decomposition

of dialkyl phthalates

IT \*\*\*84-74-2P\*\*\* 2553-24-4P 4131-84-4P 83415-90-1P

83646-71-3P

prepn. and thermal decompn. of, kinetics and mechanism of

HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? 1:0

[Return to top](#TOC)

|  |
| --- |
| SEARCHING FOR ALLOYS |

In the Registry file the constituent elements of an alloy are arranged in order of decreasing percentage composition in the chemical name field, but in alphabetical order in the molecular formula field. This should be taken into account when searching for alloys by chemical name or formulae.

Note also that CAS Registry Numbers were not used for all alloys in the 8CI. Thus the 8CI Index Guide should be consulted for alloy indexing practices during this period.

An example of a search of an alloy containing aluminium, zinc and copper is shown below.

=> S al/els and zn/els and cu/els and ays/ci

200919 AL/ELS

107220 ZN/ELS

284744 CU/ELS

538458 AYS/CI

L2 11108 AL/ELS AND ZN/ELS AND CU/ELS AND AYS/CI

Alloys are retrieved by searching the element symbols in the Element Symbol index /ELS. The AYS in the Class Identifier then limits the answer set to alloys. This search retrieves alloys containing the specified elements but other elements may also be present.

In STN alloys can also be retrieved by searching for the element symbols in the Material Composition index /MAC.

=> S al and zn and cu/mac

In this case the ays/ci term is not required since the MAC index contains only alloys. However this approach does not retrieve manually registered alloys containing the metals of interest or alloy compositions containing an Al component such as Al2O3.

[Return to top](#TOC)

|  |
| --- |
| SEQUENCE SEARCHING TECHNIQUES |

9.1 Background

Sequences are generally searched using GenomeQuest (<http://www.genomequestlive.com>), a subscription website which contains databases of patent and non-patent sequences. Examiners should contact Section A2 or C2 for assistance with sequence searching.

It is also possible to search sequences using the Registry file and DGENE in STN. DGENE only contains sequences from patent literature and is therefore used only rarely at IP Australia. Alternatively the Registry file may be searched, but this option does not provide alignment displays and has the further disadvantage that it is restricted to sequence lengths of less than 250 single letter residues due to system limits. Homology also cannot be searched in the Registry file. However the Registry file is useful for searching smaller exact sequences or subsequences, and also allows for the searching of sequences containing gaps and alternate residues at certain positions. GenomeQuest should normally be considered the primary search tool for searching sequences.

[Return to top](#TOC)9.2 Protein Sequence Searching in the Registry file

The Registry file contains protein information from research articles, patent claims and patent examples. Sequences of four or more amino acids are searchable by 1-letter or 3-letter amino acid codes. Di- and tripeptides are also registered, but may only be searched by name or structure, not by sequence representation. Entries are made for the first occurrence of a complete sequence or any subsequent major structure updates. Partial sequences without internal gaps claimed in patent claims or related to the novelty of a claim have been indexed from 1988. Some other partial sequences are also indexed. Protein sequences deduced from the corresponding nucleic acid sequences in GENBANK are also indexed up to 2005:

“From that point, CAS (1) limited the registration of GenBank sequences to those reported in the journal literature or patents and (2) stopped adding to CAS REGISTRY any sequences from patents that contained more than approximately 4,000 sequences. In 2007, CAS stopped adding to CAS REGISTRY GenBank sequences in journal articles that referenced more than 1,000 GenBank accession numbers.”

Four options are available for searching: Exact Sequence, Exact Family Sequence, Subsequence and Subsequence Family search of proteins. Each of these is discussed in turn below.

[Return to top](#TOC)

### Exact Sequence Search of Proteins/SQEP

This retrieves sequences that exactly match the search query. The search query must be exactly defined. Common amino acids and specific amino acids may be included in the query. The minimum query is four amino acids. There is no upper limit to the length of the protein, but there is a system limit of 256 characters in the query including all punctuation, spaces and the /sqep field code. This allows a single-letter sequence of up to 250 amino acids to be searched.

Amino acids may be entered as either the one- or three-letter codes, but three letter codes must be enclosed in single quotes with individual codes separated by hyphens. Uncommon amino acids must be entered as three-letter codes. This applies to all sequence queries. The three-letter codes are available in the STN manual "Registry File:Biosequence Searching Manual" and in the Quick Reference Guide at:

<http://www.cas.org/File%20Library/Training/STN/User%20Docs/protseq.pdf>.

Note that the three-letter code Und refers to an undefined amino acid. This cannot be used to search for uncommon amino acids of known structure in a peptide, only those in which the original article did not define the amino acid at this position.

Eg => s 'thr-ala-leu-lys-arg'/sqep

The EXPAND command may be used to verify that the query sequence is contained in the Registry file. Note that more than one answer may be shown for a particular sequence since the entry includes modifications to the peptide, such as tritylation and the like. The EXPAND command gives only the first 200 amino acids in a protein. Thus, any E-answer obtained in this way should not be used for searching peptides of greater than 200 amino acids as it will result in zero answers (the exact sequence of 200 amino acids will be searched rather than the desired longer sequence). However the subsequence /sqsp command could be used with the E-answer obtained in this way.

Eg => e talkr/sqep 5

(the 5 indicates that only five answers will be shown in the EXPAND result)

A combination of one- and three-letter codes may be used, but the three letter codes must be in single quotes and hyphenated as above. The following example shows a query of this type. Time is saved by using three-letter codes only for the uncommon amino acids.

Eg => s cpf'hcy-sta'lf/sqep

Multiple sequences may be searched in the same query, and are separated by the term "|" (a shorthand form of the term OR). The search is then carried out as separate searches, and charged accordingly.

Eg => s mpdn|talkr/sqep

(Searches for the sequences mpdn and talkr)

[Return to top](#TOC)

### Exact Family Sequence Search of Proteins/SQEFP

This search retrieves answers that exactly match the query and in which family-equivalent substitution of the query amino acids occurs. Equivalents of uncommon and undefined amino acids are not searched, but may be included in the query. The EXPAND command is not available in the SQEFP field.

The family equivalents are based on a conservative substitution of amino acids having similar chemical properties, as shown below. As can be seen from the table, the family groupings are very specific.

|  |  |
| --- | --- |
| Family Class Name | Class Members 3 Letter Code |
| Neutral-Weakly Hydrophobic  A, G, P, S, T  Hydrophilic-Acid Amine  N, D, Q, E  Hydrophilic-Basic  R, H, K  Hydrophobic  I, M, L, V  Hydrophobic-Aromatic  F, W, Y  Crosslinking  C | Ala, Gly, Pro, Ser, Thr  Asn, Asp, Gln, Glu  Arg, His, Lys  Ile, Met, Leu, Val  Phe, Trp, Tyr  Cys |

[Return to top](#TOC)

### Subsequence Search of Proteins /SQSP

This search retrieves exact answers plus sequences in which the query sequence is embedded in a larger sequence, or the subsequence matches exactly an entire sequence in length and sequence backbone. The minimum query size for /SQSP is one amino acid. The maximum number of characters in the query statement is 256.

Multiple sequences may be searched by separating the individual sequences with the character "|". In contrast to the corresponding exact search, only a single subsequence search cost is incurred.

E.g. => s mpdn|talkr/sqsp

This query will give any sequences containing the sequences mpdn and talkr. Thus if a large number of exact sequences need to be searched, an option is to search them as a list of sequences separated by the character "|", and then limit the answer set using sequence length.

E.g. => s mpdn|talkr|tgieq|aefi|tiky|typi/sqsp

L1 ….

=> s l1 and sql<6

Variability symbols are also available in the /SQSP field. These allow the following:

1. Search the beginning or end of a sequence or limit the query to proteins where the query sequence is at the beginning or end. The symbol "^" is used to indicate that the sequence is at the beginning or end of a protein.

E.g. => file Registry

=> s ^afpam/sqsp

This results in proteins having the sequence afpam at the beginning of the sequence field.

E.g. => file Registry

=> s afpam^/sqsp

This results in proteins having the sequence afpam at the end of the sequence field.

[Return to top](#TOC)

1. Require two or more alternate residues at one or more positions. Alternate residues at a particular position are indicated using square brackets.

E.g. => file Registry

=> s lgp[vl]q/sqsp

This query results in sequences containing lgpvq or lgplq or in the exact sequences lgpvq or lgplq.

1. Exclude one or more residues at one or more positions. Particular amino acids are excluded from a position by specifying the excluded amino acids preceded by a tilde ~ or a dash - and enclosed in brackets.

E.g. => s ptgk[-H]ea/sqsp

This query results in proteins containing the query sequence wherein the 5-position is other than histidine, as well as exact sequences of this formula.

To exclude more than one residue at a position the excluded amino acids are contained in the same square brackets, and are not separated by any spaces.

E.g. => s ptgk[-HN]ea/sqsp

This query results in proteins as in the previous example, but in which histidine and asparagine are excluded from the fifth position of the query sequence.

1. Repeat sequence expressions. A number of repeat operators are available to specify that the preceding sequence expression is to be repeated an exact number of times or a range of times. The unit to be repeated is enclosed in round brackets, followed by the repeat operator. A summary of the repeat operators is given on the next page.

[Return to top](#TOC)

|  |  |  |  |
| --- | --- | --- | --- |
| **Repeat Operator** | **Function** | **Search Examples**  **each takes the form s query/sqsp** | **Search Retrievals** |
| {m} | Repeat m times | (FL){2} | FLFL |
| {m,u} or {m-u} | Repeat m to u times | GG(FL){1,2} | GGFL or GGFLFL |
| ? or {0,1} or {0-1} | Repeat zero or one time | FLRRIRP?K  or  FLRRIRP{0,1}K  or  FLRRIRP{0-1}K | FLRRIK or FLRRIRPK |
| \* or {0,} or {0-} | Repeat zero or more times | K(WD)\*N  or  K(WD){0,}N  or  K(WD){0-}N | KN, KWDN, KWDWDN etc. |
| + or {1,} or {1-} | Repeat one or more times | K(WD)+N  or  K(WD){1,}N  or  K(WD){1-}N | KWDN, KWDWDN etc. |

In addition to the above, other repeat units such as, for example, {3,} and {3-}, can be used to indicate that a unit must be repeated at least three times. If a single residue is to be repeated then the round brackets need not be used. For example, TA{2-}T would result in answers TAAT, TAAAT, TAAAAT etc. Repeat operators may be used with a specific residue, a sequence of residues, alternate residues, E-numbers, L-numbers or saved names for queries.

[Return to top](#TOC)

1. Specify gaps and gap ranges. A number of expressions may be used to specify gaps in a sequence expression.

|  |  |  |
| --- | --- | --- |
| **Symbol** | **Function** | **Query Example and what the query retrieves** |
| .  may be used more than once, as in "…" for three residues | A gap of one residue | SY.RGP/SQSP  SY followed by one residue followed by RGP |
| .{m} or [m.] | A gap of m residues | SY.{2}RGP/SQSP  or  SY[2.]RGP/SQSP  SY followed by any two residues followed by RGP |
| .{m,u} or .{m-u} | A gap of m to u residues | SY.{2,10}RGP/SQSP  Or  SY.{2-10}RGP/SQSP  SY followed by any 2 to 10 residues followed by RGP |
| .? or : or .{0,1} or .{0-1} | A gap of zero or one residue | SY.?RGP/SQSP  or  SY:RGP/SQSP  or  SY.{0,1}RGP/SQSP  or  SY.{0-1}RGP/SQSP  SY followed by zero or one residue followed by RGP |
| .\* or .{0,} or .{0-} | A gap of zero or more residues | SY.\*RGP/SQSP  or  SY.{0,}RGP/SQSP  or  SY.{0-}RGP/SQSP  SY followed by zero or more residues followed by RGP |
| .+ or .{1,} or .{1-} | A gap of one or more residues | SY.+RGP/SQSP  or  SY.{1,}RGP/SQSP  or  SY.{1-}RGP/SQSP  SY then any number of residues then RGP |

A special concatenation symbol & may also be used to create subsequence or exact sequence queries by joining together L-numbers, E-numbers or saved names for sequence queries. Boolean operators AND, NOT and OR can also be used to limit or exclude certain answers. The link operator L can be used to specify that certain sequence fragments must be present in the same protein. Conversely the command NOTL can be used to specify that two sequences are not in the same protein. The proximity operator NOTL and the Boolean operator give different answers, as shown below. As seen from these results, the choice of operator can greatly affect the relevancy of the answers obtained.

|  |  |
| --- | --- |
| **Query** | **Results obtained** |
| S EAY/SQSP AND IVE/SQSP | The exact sequence EAY Or Sequences containing EAY And The exact sequence IVE or Sequences containing IVE *where the two sequences appear in the same abstract*  And  Sequences containing both EAY and IVE |
| S EAY/SQSP (L) IVE/SQSP | Sequences containing both EAY and IVE |
| S EAY/SQSP NOT IVE/SQSP | The exact sequence EAY  And  Sequences containing EAY but not IVE |
| S EAY NOTL IVE/SQSP | The exact sequence EAY  And  Sequences containing EAY but not IVE  And  The exact sequence IVE  And  Sequences containing IVE but not EAY |

[Return to top](#TOC)

### Subsequence Family Search of Proteins /SQSFP

This retrieves exact sequences, subsequences, and answers in which family-equivalent substitution of the query amino acids occurs. The types of family equivalents searched are as for the Exact Family Search of Proteins. Variability gaps etc., alternate (|) and concatenation (&) symbols discussed in the Subsequence Search of Proteins may be used for SQSFP queries.

[Return to top](#TOC)

9.3 Display Options for Subsequence Searches in File Registry

Whilst percentage homology cannot be searched in the Registry file, it is possible to display the hit sequences and positions corresponding to the query sequence. This can simplify the culling of results. The following example shows how the command D SEQ can be used in this way. A search of the subsequence HGWSYGGYL/SQSP gives several answers in the Registry file. The display format D SEQ is then used to display the answer set. As shown below, the query sequence is underlined in the displayed sequence, and the positions at which the hits occur are also given.

=> d seq l1

L1 ANSWER 1 OF 3 REGISTRY COPYRIGHT 2000 ACS

SEQ 1 MESRTQMNQS NQQQQQQRNR LPSSPSPPLS …….

901 QLRMHMLAAQ GYCVICIDSR GSRHRGKRFE SHIRGRMGQV…

951 SLSDQLGYID MDRVAIHGWS YGGYLSLMGL VQYPKIFKVA IAGAPVTNWE

===== ======

1001 YYDTGYTERY MDMPQNNEAG YSAGSVLEYV NSFPEEDKRL LLIHGLIDEN

1051 VHFCHTSRLI SALNKANKPY EVHLFPEERH SLRNLESNKN YETKLLSFLQ

1101 NL

HITS AT: 967-975

This option is probably of limited use if too many answers are retrieved. Furthermore, the result will still need to be searched in the bibliographic file to obtain abstract material and bibliographic details.

An alternative display format is D SQIDE. This option has the same limitations as D SEQ above. An example of a sequence displayed using SQIDE is shown below. (Note that in this case the sequence appears in a patent document and the location of the sequence has been specified).

=> d sqide l6

L6 ANSWER 1 OF 16 REGISTRY COPYRIGHT 2007 ACS on STN

RN 913217-84-2 REGISTRY

CN L-Aspartic acid, L-lysyl-L-alanylglycyl-L-phenylalanyl-L-phenylalanyl-L-

lysyl-L-arginyl-L-glutaminyl-L-tyrosyl-L-lysyl-L-seryl-L-isoleucyl-L-

leucyl-L-glutaminyl-L-a-glutamyl-L-a-glutamyl-L-asparaginyl-L-

arginyl-L-arginyl-L-a-aspartyl-L-seryl-L-tryptophyl-L-seryl-L-

tyrosyl-L-isoleucyl-L-asparaginyl-L-seryl-L-lysyl-L-seryl-L-asparaginyl-L-

a-aspartyl- (9CI) (CA INDEX NAME)

OTHER NAMES:

CN 25: PN: WO2006112738 SEQID: 25 claimed sequence

FS PROTEIN SEQUENCE

SQL 32

PATENT ANNOTATIONS (PNTE):

Sequence |Patent

Source |Reference

=========+============

Not Given|WO2006112738

|claimed

|SEQID 25

SEQ 1 KAGFFKRQYK SILQEENRRD SWSYINSKSN DD

= ========

HITS AT: 20-28

\*\*RELATED SEQUENCES AVAILABLE WITH SEQLINK\*\*

MF Unspecified

CI MAN

SR CA

LC STN Files: CA, CAPLUS

DT.CA CAPLUS document type: Patent

RL.P Roles from patents: BIOL (Biological study); PRP (Properties); USES

(Uses)

1 REFERENCES IN FILE CA (1907 TO DATE)

1 REFERENCES IN FILE CAPLUS (1907 TO DATE)

9.4 Display Options for Subsequence Searches in File CAPLUS

D HITSEQ is used to display the results of subsequence searches in CAPLUS. HITSEQ gives the Registry number, role, chemical name and sequence diagram of the “hit” sequences, i.e. those retrieved by the subsequence query. This display format is useful for non-patent literature citations, as the abstracts can be quickly culled, and there may be no need to order the full text article once the exact sequence disclosed in the article is known.

9.5 Use of the Registry Number Locator Field in Subsequence Searching

Once a subsequence search has been conducted in Registry, the resulting answer set is usually searched in CAPLUS in order to obtain relevant abstracts. However in some cases the answer set may contain sequences that occur in databases other than CAPLUS, for example GenBank. The Registry Number Locator field LC may be used to identify such sequences.

In the example below, the command s L9 and CAPLUS/LC is used to determine the number of sequences located in file CAPLUS (in this case 29). This answer set L10 is then searched in file CAPLUS as per usual. The remaining 6 sequences that do not occur in CAPLUS (answer set L11) are displayed using the D SQIDE command.

=> s dswsy|yinsk/sqsp

L8 905 DSWSY|YINSK/SQSP

=> s l8 and sql<100

9705668 SQL<100

L9 35 L8 AND SQL<100

=> s l9 and caplus/lc

53572683 CAPLUS/LC

L10 29 L9 AND CAPLUS/LC

=> s l9 not l10

L11 6 L9 NOT L10

=> d sqide

L11 ANSWER 1 OF 6 REGISTRY COPYRIGHT 2007 ACS on STN

RN 489699-15-2 REGISTRY

CN GenBank AAC12739 (9CI) (CA INDEX NAME)

OTHER NAMES:

CN GenBank AAC12739 (Translated from: GenBank AF054873)

FS PROTEIN SEQUENCE

SQL **89**

SEQ 1 ISVSTEDTPP QEYDATKVRE YLRQVEEYQI SLILQLCKVA LVPEVLSQIN

51 AMNAGILEDW QLGFVPTPEN AVHDMYRYIN SKATKCPDA

=== ==

HITS AT: 78-82

MF Unspecified

CI MAN

SR GenBank

=> fil caplus

FILE COVERS 1907 - 8 Mar 2007 VOL 146 ISS 12

FILE LAST UPDATED: 8 Mar 2007 (20070308/ED)

New CAS Information Use Policies, enter HELP USAGETERMS for details.

This file contains CAS Registry Numbers for easy and accurate

substance identification.

=> s l10

L12 19 L10

=> s l12 and p/dt

4704092 P/DT

L13 10 L12 AND P/DT

=> s l12 not l13

L14 9 L12 NOT L13

=> d bib abs hitseq

L14 ANSWER 1 OF 9 CA COPYRIGHT 2007 ACS on STN

AN 143:320049 CA Full-text

TI Genome sequence of the chlorinated compound-respiring bacterium

Dehalococcoides species strain CBDB1

AU Kube, Michael; Beck, Alfred; Zinder, Stephen H.; Kuhl, Heiner; Reinhardt,

Richard; Adrian, Lorenz

CS Max-Planck-Institut fuer Molekulare Genetik, Berlin-Dahlem, 14195, Germany

SO Nature Biotechnology (2005), 23(10), 1269-1273

CODEN: NABIF9; ISSN: 1087-0156

PB Nature Publishing Group

DT Journal

LA English

AB Dehalococcoides species are strictly anaerobic bacteria, which catabolize

many of the most toxic and persistent chlorinated aroms. and aliphatics by

reductive dechlorination and are used for in situ bioremediation of

contaminated sites. Sequencing of the complete 1,395,502 base pair genome

of Dehalococcoides strain CBDB1 revealed the presence of 32 reductive dehalogenase-

homologous (rdh) genes, possibly conferring on the bacteria an

immense dehalogenating potential. Most rdh genes were associated with

genes encoding transcription regulators such as two-component regulatory

systems or transcription regulators of the MarR-type. Four new paralog

groups of rdh-associated genes without known function were detected.

Comparison with the recently sequenced genome of Dehalococcoides

ethenogenes strain 195 reveals a high degree of gene context conservation

(synteny) but exceptionally high plasticity in all regions containing rdh

genes, suggesting that these regions are under intense evolutionary

pressure. The complete genome sequence is deposited in GenBank/EMBL/DDBJ

under accession number AJ965256.

IT **865394-87-2**

RL: BSU (Biological study, unclassified); PRP (Properties); BIOL

(Biological study)

(amino acid sequence; genome sequence of the chlorinated

compound-respiring bacterium Dehalococcoides species strain CBDB1)

RN 865394-87-2 CA

CN Protein (Dehalococcoides strain CBDB1 87-amino acid) (9CI) (CA INDEX

NAME)

SEQ 1 MQSSLIGKIE KAKRYAQEKD RIRFNQLSVQ FRGENDIHTT EYINSKWHCT

51 CNFFAAWGMC SHTMAMEKVL ENMLPDEALS NVFAELP

[Return to top](#TOC)9.6 Searching Nucleic Acid Sequences

### Searching Exact Nucleic Acid Sequences

Exact sequences of nucleotides can be searched in the Registry and DGENE files using the /SQEN field. The SEARCH command is used in Registry and the RUN GETSEQ command in DGENE. DGENE is discussed in greater detail elsewhere. The following codes may be used in exact nucleic acid sequence searches:

|  |  |
| --- | --- |
| **Code** | **Name or Definition** |
| A | adenine |
| C | cytosine |
| G | guanine |
| T | thymine in DNA |
| U | uracil in RNA |
| N | unknown only in DGENE |

An example of searching for an exact nucleic acid sequence is shown below. The initial step of the exact search is to EXPAND the query sequence to confirm whether or not the sequence is in the Registry file (this saves the cost of searching for a term that is not in the file). If zero answers are obtained the query should be checked for typographical errors. Once the term has been found in the Registry file it may be searched. In the example below this is shown as a search of the sequence, but if the EXPAND command has been used the E-answer may be searched (generally E3 in the answer list), but can only be used if the nucleic acid is less than 200 residues.

=> FILE REGISTRY

=> E CGCCCCTGCGTTACCCTCCCCGCCG/SQEN

……

=> S CGCCCCTGCGTTACCCTCCCCGCCG/SQEN

1 CGCCCCTGCGTTACCCTCCCCGCCG/SQEN

5259 SQL=25

L1 1CGCCCCTGCGTTACCCTCCCCGCCG/SQEN

CGCCCCTGCGTTACCCTCCCCGCCG/SQEN AND

SQL=25 => FILE CAPLUS

=> FILE CAPLUS

=> S L1

L2 1 L1

=> D 1 BIB AB HITRN

Ambiguity codes are found in answers from GenBank, and these have been added to the Registry file. In SQEN searches all codes, including the ambiguity codes other than Z, may be searched. However it should be noted that the ambiguity codes match only the specified ambiguity codes in answers. For example Y matches Y but not B or T.

|  |  |  |
| --- | --- | --- |
| **Ambiguity Codes** | **Name or Definition** | **Complement** |
| Y | pyrimidine | R |
| R | purine | Y |
| M | amino | K |
| K | keto | M |
| S | strong interaction 3 H bonds | S |
| W | weak interaction 2 H bonds | W |
| B | not-A | V |
| V | not-T, not-U | B |
| D | not-C | H |
| H | not-G | D |
| N | Unknown nucleotide | N |
| X | Uncommon nucleic acid | X |
| Z | Non-specific nucleic acid matches Y, R, M, K, S, W, B, V, D, H, N, X | Z |

In nucleic acid searches the system allows a search of the complementary strand to either be included or excluded. If included, the system matches the query sequence as well as its complement. For example, a query containing ACG will search both ACG and its complement TGC. Alternatively the complementary strand can be excluded using the command SIN singlestr and, for example "S L1 SIN", or responding "Y" at the prompt "EXCLUDE SEARCH OF COMPLEMENTARY STRAND Y?N".

[Return to top](#TOC)

### Subsequence Search of Nucleic Acids

A subsequence search of nucleic acids /SQSN retrieves exact answers plus sequences in which the query sequence is embedded. Variability symbols are allowed. In subsequence searches the symbols for the specific nucleotides A, C, G, T and U match only on those specific residues in sequences in the database. The ambiguity codes are translated in the following ways:

|  |  |
| --- | --- |
| **Ambiguity Code** | **/SQSN Matches on** |
| B | Any code and its complement except A or X |
| D | Any code and its complement except C or X |
| H | Any code and its complement except G or X |
| K | Any code and its complement except A, C, M or X |
| M | Any code and its complement except G, T, U, K or X |
| N | Any code |
| R | Any code and its complement except C, T, U, X or Y |
| S | Any code and its complement except A, T, U, W or X |
| U | T or U and their complements |
| V | Any code and its complement except T, U or X |
| W | Any code and its complement except G, C, S or X |
| Y | Any code and its complement except A, G, R or X |
| X | X |
| Z | Any code and its complement except G, C, A, T or U |

Thus if the query was GGGNNNAAT/SQSN, an answer set comprising the sequences GGG followed by any three nucleotides followed by AAT would be obtained (N translates as any code). In order to retrieve the specific answer GGGNNNAAT, the ambiguity code must be enclosed in square brackets, i.e. GGG[N][N][N]AAT/SQSN.

[Return to top](#TOC)

|  |
| --- |
| DGENE |

10.1 Background

DGENE Derwent Geneseq is a unique database of nucleotide and amino acid sequences that have been extracted from the original patent disclosure, i.e. basic patents from 40 issuing authorities. The file covers 1981 to present. All nucleotide sequences of 10 or more bases, all amino acid sequences of 4 or more residues and probes and primers of any length are indexed.

Records contain a Derwent enhanced title, an English abstract that has been written especially for this database by Derwent, patent information, detailed indexing, and the standard features table and sequence data. Both text and sequence data are fully searchable and displayable. DWPI family information can be directly displayed within DGENE. For file crossover to the Derwent World Patents Index files WPIDS, the DWPI accession number is available in all DGENE records.

For sequence searching, DGENE contains three packages which can be started with the RUN command: GETSEQ for nucleic acid sequence and protein sequence searching, and GETSIM and BLAST for homology searches of amino acid and nucleotide sequences. The commands are much the same as those used in Registry searches, and are summarised in the following tables.

NOTE: The DGENE database only covers sequences derived from patent documents. DGENE is rarely used in sequence searching at IP Australia.

[Return to top](#TOC)10.2 Searching Sequence Data with the GETSEQ RUN Package

**Sequence Search Terms**

|  |  |
| --- | --- |
| **Terms** | **Examples** |
| Single letter codes for common amino acids | QUE LAGLL/SQSP |
| Three-letter codes for common amino acids. Enclose codes or strings of codes in single quotes. Use dashes to separate codes in strings. | QUE 'THR-SER-GLY-MET THR'/SQSFP  RUN GETSEQ 'CYS-ASN-THR-ALA'/SQSP |
| Single letter codes for nucleic acids | QUE ATGAAN/SQEN  RUN GETSEQ ATGAAN/SQSN |

Single letter and three letter codes for the common amino acids are as for those noted above under the Registry file searching of sequences. Uncommon amino acids are represented in the sequence either by a related parent amino acid or by an X or XXX.

[Return to top](#TOC)10.3 Types of Sequence Searches in DGENE

Sequence data for nucleic acid and protein sequences are displayed in the SEQ field with 1-letter codes and the SEQ3 field with 3-letter codes for proteins only.

|  |  |  |  |
| --- | --- | --- | --- |
| **Type** | **Definition** | **Code** | **Examples** |
| Sequence Exact, Protein | Search for sequences that match the query. The query must be completely defined. | /SQEP | QUE AFFFF/SQEP  QUE 'ALA-PHE-PHE-PHE-PHE'/SQEP |
| Sequence Exact Family, Protein | Search for sequences that match the query and those in which family-equivalent substitution of the query amino acids occurs. | /SQEFP | RUN GETSEQ YGGFL/SQEFP  QUE 'TYR-GLY-GLY-PHE-LEU'/SQEFP |
| Subsequence, Protein | Search for exact answers plus sequences in which the query sequence is embedded. | /SQSP | RUN GETSEQ LAGLL/SQSP  QUE GLPGY/SQSP |
| Subsequence, Family, Protein | Search for exact sequences, subsequences, and answers in which family-equivalent substitution of the query amino acids occurs. | /SQSFP | RUN GETSEQ ATCXAW/SQSFP  QUE 'LEU-ALA-GLY-LEU-LEU'/SQSFP |
| Sequence Exact, Nucleic Acid | Search for sequences  that match the query.  Ambiguity codes for nucleic acids are allowed. | /SQEN | QUE ATGAAN/SQEN |
| Subsequence, Nucleic Acid | Search for exact answers, plus sequences in which the query sequence is embedded. Ambiguity codes for nucleic acid symbols and variability symbols are allowed. | /SQSN | RUN GETSEQ ATGAAN/SQSN |

## 10.4 Variability Symbols in DGENE

|  |  |  |
| --- | --- | --- |
| **Symbol** | **Function** | **Examples** |
| [ ] | to specify alternate residues | RUN GETSEQ LGP[VL]/SQSP  QUE LGP['VAL''LEU''LYS']/SQSP |
| [-] | to exclude a specific residue or alternate residues | QUE LGP[-H]/SQSP  QUE LGP[-'HIS']/SQSP  QUE LGP[-HL]/SQSP |
| {m} | to repeat the preceding sequence or sequence query or L-number m times | RUN GETSEQ L4{2}/SQSP  RUN GETSEQ TATAAA{2}/SQSN |
| {m,u} or  {m-u} | to repeat the preceding sequence or sequence query or L-number m to u times | RUN GETSEQ CTG{1,3}/SQSN  RUN GETSEQ GGFL{1,2}/SQSP  RUN GETSEQ L3{1,3}/SQSP |
| ? or {0,1} or {0-1} | to repeat the preceding sequence or sequence query or L-number zero or one time | RUN GETSEQ FLRRIRP?K/SQSP  RUN GETSEQ L1{0-1}NN/SQSP  RUN GETSEQ TCGA{0,1}C/SQSN  RUN GETSEQ L1{0,1}NN/SQSP |
| \* or {0,} or  {0-} | to repeat the preceding sequence or sequence query or L-number zero or more times | RUN GETSEQ KLKWD\*N/SQSP  RUN GETSEQ L1{0-}NN/SQSP  RUN GETSEQ CCTG{0,}TA/SQSN |
| + or {1,} or  {1-} | to repeat the preceding sequence or sequence query or L-number one or more times | RUN GETSEQ KLKDLE+/SQSP  RUN GETSEQ KLKDLE{1,}/SQSP  RUN GETSEQ L2{1-}/SQSP  RUN GETSEQ CACG{1,}TT/SQSN |
| ^ | To specify that the sequence fragment is at the beginning or end of the sequence field | QUE ^MCGIL/SQSP  {Query fragment at beginning}  QUE VCDS^/SQSFP  {Query fragment at end} |
| | | To specify alternate sequence expressions or L-numbers | RUN GETSEQ ACD|KLM  {equivalent to ACD or KLM}  RUN GETSEQ CAGAA|TTCT/SQSN  {searches for CAGAA and TTCT} |
| & | to join expressions or L-numbers together | RUN GETSEQ L2&L5{1,3}/SQSP  RUN GETSEQ L1&L3/SQSFP |

[Return to top](#TOC)

10.5 Specifying Gaps in Subsequence Queries /SQSP, /SQSFP, and /SQSN

|  |  |  |
| --- | --- | --- |
| **Symbol** | **Function** | **Examples** |
| . | a gap of one residue | QUE SY.RPG/SQSP  {a gap of one residue}  RUN GETSEQ SY..RPG/SQSP  {a gap of two residues}  QUE AAG...TGC/SQSN  {a gap of three residues} |
| .{m} or [m.] | a gap of m residues | QUE SY.{2}RPG/SQSP  QUE SY[2.]RPG/SQSP |
| .{m,u} or.{m-u} | a gap of m to u residues | RUN GETSEQ GF.{2-10}LSS/SQSP  RUN GETSEQ GFF.{2,10}LSS/SQSP  QUE AAG.{2,5}TGC/SQSN |
| : or .? or .{0,1}  or .{0-1} | a gap of zero or one residues | RUN GETSEQ AGA:SRI/SQSFP  RUN GETSEQ AGA.?SRI/SQSFP  QUE AGA.{0,1}SRI/SQSFP  QUE AGA.{0-1}SRI/SQSFP |
| .\* or .{0,} or .{0-} | a gap of zero or more residue | RUN GETSEQ HLC.\*TYG/SQSP  QUE HLC.{0,}TYG/SQSP  RUN GETSEQ HLC.{0-}TYG/SQSP  QUE AAGGAGATG.\*GCAA/SQSN |
| .+ or .{1,} or.{1-} | a gap of one or more residues | RUN GETSEQ SY.+TH/SQSFP  QUE SY.{1-}TH/SQSFP  QUE SY.{1,}TH/SQSP  RUN GETSEQ TCCTG.+GTG/SQSN |

[Return to top](#TOC)10.6 Similarity Searching of Biosequences with the BLAST RUN and GETSIM RUN Packages

Homology searching is available for protein and nucleotide sequences using the RUN BLAST and RUN GETSIM commands. RUN BLAST is based on the BLAST algorithm whilst RUN GETSIM is based on the FASTA algorithm. The modules retrieve sequences that include the exact or a similar sequence query and assign a similarity score. Sequences can be subjected to a homology search in various ways. A query may be prepared with the QUERY command and saved beforehand, it may be entered directly on the RUN GETSIM/BLAST command line, or it may be uploaded from an ASCII file.

An example of RUN BLAST is given below (RUN GETSIM produces similar results). A diagram is generated that shows the similarity between the retrieved sequences and the query. The x-axis represents the number of answers with a specific degree of similarity represented by the y-axis. The whole answer set or only the most relevant answers of your choice can be kept. The generated L-number contains these answers, but they are sorted by descending accession number. This L-number may be arranged by descending similarity score. Enter SOR SCORE D and the corresponding L-number at an arrow prompt. It is possible to see the alignment between the retrieved sequence and the query sequence with the ALIGN display format. This displays a line between the two sequences, thus giving information about the degree of similarity: two dots represent identical amino acids, one dot indicates amino acids of the same family, and a blank occurs if there is no match. Gaps in the sequence are shown with an underscore.

Note: The "align" and "sorting" commands are part of the RUN GETSIM and RUN BLAST modules, and as such they do not incur further costs. The query file can contain more than one sequence and allows the searcher to continue with the next sequence after the first one has been searched.

[Return to top](#TOC)

=> run blast

MAAQVATTKVPEVRDITRIERIGAHSHIRGLGLDDALEPRQVSQGMVGQLASRRAAGLILEMIKDGQIAGRAVLIAGQPGTGKTAIAMGIAQSLGPDTPFTALAGSEIFSLEMSKTEALSQAFRKAIGVRIKEETEIIEGEVVEIQIDRPATGTGAKVGKLTLKTTEMETIYDLGTKMIESLSKERVQAGDVITIDKATGKISKLGRSFTRARDYDAMGAQTQFVQ/sqp

BLAST Version 2.2

The BLAST software is used herein with permission of the

National Center for Biotechnology Information (NCBI) of

the National Library of Medicine (NLM). See also, Altschul,

Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui

Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),

"Gapped BLAST and PSI-BLAST: a new generation of protein

database search programs." Nucleic Acids Res. 25:3389-3402

BLAST SEARCHING

Database DGENE AA

Posted date: Mar 9, 2007 6:41 PM

Number of letters in database: 504,419,037

Number of sequences in database: 2,838,436

Lambda K H

0.316 0.132 0.354

Gapped

Lambda K H

0.267 0.0410 0.140

Matrix: BLOSUM62

Gap Penalties: Existence: 11, Extension: 1

Number of Hits to DB: 167,877,916

Number of Sequences: 2838436

Number of extensions: 5497665

Number of successful extensions: 24895

Number of sequences better than 10.0: 2107

Number of HSP's better than 10.0 without gapping: 807

Number of HSP's successfully gapped in prelim test: 1300

Number of HSP's that attempted gapping in prelim test: 23548

Number of HSP's gapped (non-prelim): 2553

length of query: 226

length of database: 504,419,037

effective HSP length: 116

effective length of query: 110

effective length of database: 175,160,461

effective search space: 19267650710

effective search space used: 19267650710

T: 11

A: 40

X1: 16 ( 7.3 bits)

X2: 38 (14.6 bits)

X3: 64 (24.7 bits)

S1: 41 (21.6 bits)

S2: 69 (31.2 bits)

2107 ANSWERS FOUND BELOW EXPECTATION VALUE OF 10.0

Similarity

Score

392 |

|

|

|

|

|

|

|

|

|

196 |

|

||

||

||

||

||

||

||||||||||||||||||||||||||||||||||||||||||||||||||

||||||||||||||||||||||||||||||||||||||||||||||||||

Answer Count 420 840 1260 1680 2100

HOW MANY ANSWERS WOULD YOU LIKE TO KEEP ? (ALL) OR ?:all

L1 RUN STATEMENT CREATED

L1 2107 MAAQVATTKVPEVRDITRIERIGAHSHIRGLGLDDALEPRQVSQGMVGQL

ASRRAAGLILEMIKDGQIAGRAVLIAGQPGTGKTAIAMGIAQSLGPDTPF

TALAGSEIFSLEMSKTEALSQAFRKAIGVRIKEETEIIEGEVVEIQIDRP

ATGTGAKVGKLTLKTTEMETIYDLGTKMIESLSKERVQAGDVITIDKATG

KISKLGRSFTRARDYDAMGAQTQFVQ/SQP. -E 10.0

Answer set arranged by accession number; to sort by descending

similarity score, enter at an arrow prompt (=>) "sor score d".

=> sor score d

PROCESSING COMPLETED FOR L1

L2 2107 SOR L1 SCORE D

=> d blastalign

L2 ANSWER 1 OF 2107 DGENE COPYRIGHT 2007 The Thomson Corp on STN

BLASTALIGN

Query = 226 letters

Length = 463

Score = 392 bits (1008), Expect = e-114

Identities = 209/226 (92%), Positives = 209/226 (92%)

Query: 1 MAAQVATTKVPEVRDITRIERIGAHSHIRGLGLDDALEPRQVSQGMVGQLASRRAAGLIL

MAAQVATTKVPEVRDITRIERIGAHSHIRGLGLDDALEPRQVSQGMVGQLASRRAAGLIL

Sbjct: 1 MAAQVATTKVPEVRDITRIERIGAHSHIRGLGLDDALEPRQVSQGMVGQLASRRAAGLIL

Query: 61 EMIKDGQIAGRAVLIAGQPGTGKTAIAMGIAQSLGPDTPFTALAGSEIFSLEMSKTEALS

EMIKDGQIAGRAVLIAGQPGTGKTAIAMGIAQSLGPDTPFTALAGSEIFSLEMSKTEALS

Sbjct: 61 EMIKDGQIAGRAVLIAGQPGTGKTAIAMGIAQSLGPDTPFTALAGSEIFSLEMSKTEALS

Query: 121 QAFRKAIGVRXXXXXXXXXXXXXXXXXDRPATGTGAKVGKLTLKTTEMETIYDLGTKMIE

QAFRKAIGVR DRPATGTGAKVGKLTLKTTEMETIYDLGTKMIE

Sbjct: 121 QAFRKAIGVRIKEETEIIEGEVVEIQIDRPATGTGAKVGKLTLKTTEMETIYDLGTKMIE

Query: 181 SLSKERVQAGDVITIDKATGKISKLGRSFTRARDYDAMGAQTQFVQ 226

SLSKERVQAGDVITIDKATGKISKLGRSFTRARDYDAMGAQTQFVQ

Sbjct: 181 SLSKERVQAGDVITIDKATGKISKLGRSFTRARDYDAMGAQTQFVQ 226

10.7 Display Options in DGENE

An abstract is given for each sequence retrieved by the query. This means that there may be duplication of results when there is more than one sequence in a single citation. This problem may be overcome in the following manner. The results are scanned using D TRIAL, which is a free display format. The display gives the title, DGENE accession number which is different from the DWPI accession number, keyword, description and molecule type fields. Duplicates may be identified using the title. Each of the answers in the D TRIAL is assigned an answer number shown below as answer 1 of 100 and so on, so that the results can then be selectively displayed. By displaying the OTHER SOURCE /OS field, the DWPI accession number can be obtained. Once the DWPI accession number is obtained, the abstract may be viewed in the EPOQUE Viewer by using STN Extractor.

=> d trial l4 1-7

L4 ANSWER 1 OF 100 DGENE COPYRIGHT 2000 DERWENT INFORMATION LTD

XYX 1999P-Y28603 Protein DGENE

TI Steroidal and juvenile hormone receptors and partner proteins, useful for identification of modulators and insecticidal compounds

DESC EcR polypeptide subunit of L. cuprina ecdysone receptor

KW Lucilia cuprina EcR polypeptide …

SQL 757

…..

L4 ANSWER 7 OF 100 DGENE COPYRIGHT 2000 DERWENT INFORMATION LTD

XYX 2000P-Y70357 Protein DGENE

TI Novel fusion protein comprising a protease cleavage site, a ligand binding domain, and a DNA binding domain useful for characterizing proteases, detecting viral infection, and screening for protease inhibitors -

DESC Heterodimeric nuclear receptor encoded by vector pVgRXR

KW Protease cleavage site…

SQL 746

Other display formats are shown below.

=> d ti psl os

L2 ANSWER 1 OF 2107 DGENE COPYRIGHT 2007 The Thomson Corp on STN

TI Diagnosing whether a subject has, or is at risk of developing, a disease

or condition related to Reptin (e.g. myocardial infarction or cancer), by

determining whether the subject has a mutation in a gene encoding the

Reptin protein.

PSL Claim 16; Fig 5

OS 2003-381576 [36]

=> d ti pi psl

L2 ANSWER 1 OF 2107 DGENE COPYRIGHT 2007 The Thomson Corp on STN

TI Diagnosing whether a subject has, or is at risk of developing, a disease

or condition related to Reptin (e.g. myocardial infarction or cancer), by

determining whether the subject has a mutation in a gene encoding the

Reptin protein.

PI WO 2003028537 A2 20030410 61

PSL Claim 16; Fig 5

=> d ti pi seq psl

L2 ANSWER 1 OF 2107 DGENE COPYRIGHT 2007 The Thomson Corp on STN

TI Diagnosing whether a subject has, or is at risk of developing, a disease

or condition related to Reptin (e.g. myocardial infarction or cancer), by

determining whether the subject has a mutation in a gene encoding the

Reptin protein.

PI WO 2003028537 A2 20030410 61

SEQ

1 maaqvattkv pevrditrie rigahshirg lglddalepr qvsqgmvgql

51 asrraaglil emikdgqiag ravliagqpg tgktaiamgi aqslgpdtpf

101 talagseifs lemskteals qafrkaigvr ikeeteiieg evveiqidrp

151 atgtgakvgk ltlkttemet iydlgtkmie slskervqag dvitidkatg

201 kisklgrsft rardydamga qtqfvqcpeg elqkrkevvh tvslheidvi

251 nsrtqgflal fsgdtgeiks evreqinakv sewreegkae iipgvlfide

301 vhmldiecfs flnralesdl spvlimatnr gitrirgtny qsphgipidm

351 ldrlliiatt pyteketrqi lkirceeedv elseeahtvl trigqetslr

401 yaiqlistag lvcrkrrgte vqvedikrvy slfldearss qymkeyqdsf

451 lfnetqtsqm dts

PSL Claim 16; Fig 5

OS – other source (DWPI accession number)

PI – patent information

PSL – patent sequence location

SEQ - sequence

TI - title

Further information on DGENE is available at <http://www.cas.org/ONLINE/DBSS/dgene.pdf>

[Return to top](#TOC)

|  |
| --- |
| CAS REGISTRY BLAST |

11.1 Background

CAS Registry Blast is a CAS proprietary package for similarity searching using Blast2 as the searching algorithm.

11.2 Operation of the Search

\* Database on which the search is done is File Registry.

\* Access to the package is via an internet environment through the STN Express logon. To initiate a search click on the BLAST icon. This brings up a further STN logon screen (uses normal STN userid and password, select ‘Columbus’ from the ‘STN Service Center’ drop down menu), and also a popup for the office firewall authentication login.

\* Can search either polynucleotide or polypeptide sequences.

\* A single search can cover patent, non-patent or both.

\* Sequence input into the "Result Set Manager" module:  
 Upload a sequence by   
 a) direct typing into the dialog box, or  
 b) keying in (cut & paste) the text into the box;  
 c) retrieving a file directly from PC  
 (the latter two are faster and less error prone).

\* There is a limit for the length of the sequence to be searched:

Line limit: Each line containing the search statement cannot exceed 256 characters:   
 NB space is a character.

Other uploading limit: Line limit can be circumvented using the "upload" command: upload a sequence using a manual line break rather the paragraph mark at the end of each line (maximum of 256 characters per line). This extends the limit to 50K.

\* Search rate is fast and comparable to BLAST searches in NCBI.

\* Notwithstanding the actual search time, the process may take longer because getting into the BLAST package can take some time as it is dependent on the traffic/telecommunication at the time of logon.

\* CAS only indexes the sequence disclosed in the document if the indexing person regards the disclosure as noteworthy: The sequence is indexed in the Registry file.

\* CAS indexes primers and probes; to search them, select the option "less than 50 in length".

\* Sensitivity of the search can be increased by changing the parameters from their default values.

11.3 Format of Search Output

\* The search result is a list of sequences with decreasing order of identity or positives in the "Sequence Explore Result" module.

\* Each cited sequence is listed by its registry number (RN) followed by the title of the sequence, the score (identity or positives) and the expect value.

\* Further information on the degree and region of identity/homology between the cited sequence and the one searched can be viewed online by selecting "show all details" or click the box marked "+" on the right hand side of the sequence in the list.

\* The search result gives only local alignment between the cited sequence and the one searched.

\* No bibliographic data on the individual cited sequence is provided with the search result. To retrieve any bibliographic data, search the RN in CAPLUS file using the appropriate button provided under the drop down "tools" menu. This process automatically initiates an online STN session where normal (text) searching can be carried out (i.e. normal command line searching).

\* The search result can be also be printed.

Additional information on how to conduct a CAS Registry BLAST search is available at the following site:

<https://www.cas.org/File%20Library/Training/STN/User%20Docs/blast.pdf>

[Return to top](#TOC)

|  |
| --- |
| WEB BASED “NEW” STN |

“New” STN (<https://www.stn.org/stn/>) has been developed jointly by the Chemical Abstracts Service and FIZ Karlsruhe as a new platform offering increased speed, power, and usability. Currently at version 2.0, “New” STN is being constantly improved and is available to examiners (using existing STN logins) should the features of the existing (STN Express) not be suitable for their particular requirements.

STN Express will continue to function as normal and is still the recommended tool for searching of chemical related applications in IP Australia.

[Return to top](#TOC)

|  |
| --- |
| STRUCTURE DRAWING |

13.1 Background

Structure searching provides a simple route to obtaining information on specific compounds and classes of compounds. Whilst chemical names may be complex and difficult to derive, and many compounds may have the same molecular formula, each compound has a unique chemical structure. Unfortunately, structure searching tends to be a more expensive option than name or molecular formula searching for specific compounds, and is rarely used for such searches.

The best way of learning how to draw and search chemical structures is to practice on actual cases. Close supervision and checking of all structure queries is essential, as mistakes with structure searches can be expensive.

The STN files available for structure searching are REGISTRY, ZREGISTRY, GMELIN, BEILSTEIN, CASREACT, MARPAT, MARPATPREV, SPECINFO, CHEMINFORMRX, CHEMREACT, DJSMONLINE, DJSMDS, and DRUGU along with the LREGISTRY, LBEILSTEIN, LCASREACT, and LMARPAT learning files. The WPIDS/LWPI file may also be searched using Fragmentation Codes generated from a structure query. Generally structure searches in this office are carried out in the Registry File.

Structure searches give the most up-to-date information from Registry File searches. When new compounds are indexed in the Registry file, in addition to the chemical name, molecular formula and other chemical information, a number of structure attributes are recorded. This is the first information that is available for a compound in the Registry file. A structure search is a search of compounds in the Registry file which have particular structure attributes. When a substructure is built in STN Express the structure is broken down into standard structure attributes. When the search query is entered on-line, a number of filters screen the Registry File for compounds having the required structure attributes. The resulting answer set is a list of Registry Numbers corresponding to compounds which match the structure query. This can be displayed in the Registry File, but generally the answer set is searched in File CAPLUS.

This section will provide a basic outline of the tools used to draw structures, the type of structure searches available and the on-line procedures. The general procedure used for structure searching is given in the following flow chart. Structures are built off-line using the Prepare Query file in STN Express. The structure is then saved. After logging on to STN International, the Registry File is entered, and the structure uploaded using the Upload Query Structure icon also available under the Query pulldown menu. The structure query is then searched to give the Registry Number answer set. One or more bibliographic files is then entered, the Registry Number answer set searched, and the answers displayed. After logging off the answer set is printed out.



[Return to top](#TOC)13.2 Defining Bond Characteristics

Bond Characteristics is used to set a bond type and a bond value for one or more bonds in a structure query. Select the bonds to be modified, and then choose Bond Characteristics from the QueryDef menu. The Bond Characteristics dialog box is displayed and includes options to set the Bond Type ring or chain and the Bond Value bond order, e.g. single or double. Shortcut: right-click on a bond, then click the radio button next to the Bond Type and Bond Value that you want. Click the Bond Types button to select a bond default for single or multiple use, and then click OK.

The bond types and bond values include:

**Bond Types**

1. Chain: Search results include structures which only contain the selected bonds in a chain. This is the default type for drawn chain bonds.
2. Ring/Chain: Search results include structures containing the selected bonds either as part of a ring or a chain. Nodes on either side of a bond set to ring/chain are set to ring or chain.
3. Ring: Search results include substances containing the specified bonds in rings. This is the default type for ring bonds. You may change a chain bond to a ring bond, but you cannot alter a bond within a drawn ring. The nodes on either side of the new ring bond are also set to ring.
4. Mixture: Structure Drawing activates this to note that bonds with a mixture of bond types have been selected. This option cannot be selected manually.

**Bond Values**

1. Exact/Normalized: Search results match answers either with the exact bond value drawn or with normalized bonds in the corresponding position. Using these bonds ensures maximum recall without sacrificing precision.
2. Exact: Search results exactly match the bond value drawn. This is the default value for carbon-carbon bonds in chains and most bonds in isolated rings.
3. Normalized: Search results only match answers with corresponding normalized bonds. This is the default for even-numbered rings and ring systems with alternating single and double bonds around the entire ring and for tautomeric bonds.
4. Unspecified: Search results give any bond value to an unspecified bond. This is selected when the chosen bond has a value of Unspecified, but is greyed because it is not user-selectable in this dialog box. Select Bond from the Draw menu to change bonds to Unspecified.
5. Mixture: Structure Drawing activates this choice to remind you that you have selected bonds with a mixture of bond values. You may not select this choice yourself.

After specifying the bond values and types, click OK.

[Return to top](#TOC)13.3 Normalised and Exact Bonds

Care must be taken when defining bond types. The following considerations must be made:

1. Most aromatic systems would be assumed to contain normalised bonds. However, this definition does not directly correlate in STN. Normalised bonds are only assigned to ring systems having alternating double and single bonds and an even number of ring atoms.

Normalised Bonds



No Normalised Bonds



1. Tautomers have the general formula:



1 2 3

wherein:

H may be : H, D, T or a charge

Y may be: C, N, P, S, As, Sb, Se, Te, Cl, Br, I

X or Z may be: N, O, S, Se, Te.

Tautomers are also defined by the system as normalised bonds. Tautomers require two structural characteristics:

(i) either the 1- or 3-position of a three atom chain must have either a hydrogen attachment or a charge; and,

(ii) neither the 1- nor the 3-position may be carbon.

Some examples are given below. Normalised bonds are shown in bold.



The requirement that the 1- or 3-position of the tautomer must be hydrogen means that carboxylic acids and salts thereof are treated as tautomers and contain normalised bonds, whereas esters contain only exact bonds. Similarly, primary and secondary amides contain normalised bonds, whereas tertiary amides do not.

Normalised bonds in bold



Exact bonds



In these structures R is other than hydrogen

Due to the difficulties involved in defining normalised and exact bonds, this should be avoided unless absolutely necessary. The system defaults to exact/normalised for most bonds, and in most cases the definition of exact or normalised bonds generally does not greatly limit the answer set.

[Return to top](#TOC)

13.4 Setting Node Characteristics

Node Characteristics is used to set an attribute for one or more nodes in a structure query. Select the nodes using the Select tool, then select Node Characteristics from the QueryDef menu. The Node Characteristics dialog box is displayed. Shortcut: right-click on a node and select Node Characteristics.

Click the radio button next to the desired attribute. These include:

1. Chain: This setting retrieves search results which include structures with the selected node in a chain. This is the default for nodes drawn in chains. This characteristic may not be applied to a node in a ring.
2. Ring: This setting retrieves search results which have the selected node in a ring, and this is the default for ring nodes. A chain node may be redefined as a ring node, but the system will not allow a ring node to be redefined as a chain. *Changing node characteristics of atoms on both ends of a bond does NOT change the bond type. Therefore both the bond and the atom nodes must be assigned*.
3. Ring/Chain: This setting retrieves search results which have the selected nodes in either a chain or a ring.
4. Mixture: The Structure Drawing software activates this choice to remind you that you have selected nodes that have a mixture of node characteristics. You may not select this choice yourself.

Changing the node characteristics changes all selected nodes regardless of their previous setting.

After defining the node characteristics, click OK.

[Return to top](#TOC)

13.5 Defining Hydrogen Attachments

Hydrogen Attachments is used to specify the number of hydrogen atoms attached to a particular node in a structure query. Select the desired nodes, and then select Hydrogen Attachments from the QueryDef menu. The Hydrogen Attachments dialog box is displayed.

Shortcut: right-click on a node and select Hydrogen Attachments.

Select the radio button next to the required value:

1. Any: This setting retrieves search results which include structures with any number of hydrogen attachments at the selected node including none. This is the default setting for all nodes.
2. Specific: This setting allows the number of hydrogens attached to a selected node to be specified. The exact or minimum number may be specified in the range 0 to 99.
3. Exact: This setting retrieves search results which include only structures with exactly the specified number of hydrogens attached to the selected nodes.
4. Minimum: This setting retrieves search results which include only structures with at least the specified number of hydrogens attached to the selected node. Alternately, the hydrogens can be drawn on the node directly using the Pencil or Chain tool.
5. Mixture: Structure Drawing activates this choice to remind you that you have selected nodes that have a mixture of hydrogen attachments. You may not select this choice yourself.

The value for hydrogen attachments must include the count of hydrogen atoms drawn as attachments to the node. If the value for hydrogen attachments is less than the count of hydrogens attached to a node, a warning comes up during Save. Changing the number of hydrogen attachments changes all selected nodes regardless of their previous setting.

Click OK.

[Return to top](#TOC)

13.6 Defining Non-Hydrogen Attachments

Non-H Attachments is used to specify the degree of substitution for one or more selected nodes in a structure query. Select the required nodes, and then select Non-H Attachments from the QueryDef menu. The Non-Hydrogen Attachments dialog box is displayed.

The following selections can be made:

1. Any: This setting retrieves search results which include structures with any number of ring or chain non-hydrogen substituents at the selected nodes. This is the default setting for all nodes.
2. Specific: This setting allows you to specify the number and type of non-hydrogen connections attached to a selected node. You may specify either the exact or minimum number in the range of 1 to 16. You may also specify that the bonds to the attached nodes are Chain, Ring, or Ring/Chain.
3. Exact: This setting only retrieves search results which include structures with the exact number of Non-H connections of the specified type chain, ring, or ring/chain attached to the selected nodes.
4. Minimum: This setting only retrieves search results which include structures with at least the specified number for Non-H connections of the specified type chain, ring, or ring/chain attached to the selected node.
5. Maximum: This setting retrieves search results that include structures with no more than the specified number of non-hydrogen connections of the specified type chain, ring, or ring/chain attached to the selected nodes.
6. Mixture: Structure Drawing activates this choice to remind you that you have selected a mixture of non-hydrogen attachments. You may not select this choice yourself.
7. Chain: This setting specifies that the bonds to the attached nodes must be chain.
8. Ring: This setting specifies that the bonds to the attached nodes must be ring.
9. Ring/Chain: This setting specifies that the bonds to the attached nodes may be ring or chain.

Click OK after defining any non-hydrogen attachments.

[Return to top](#TOC)

|  |
| --- |
| STRUCTURE SEARCHING IN THE REGISTRY FILE |

14.1 Types of Structure Searches in Registry

There are three types of structure-based searches in Registry:

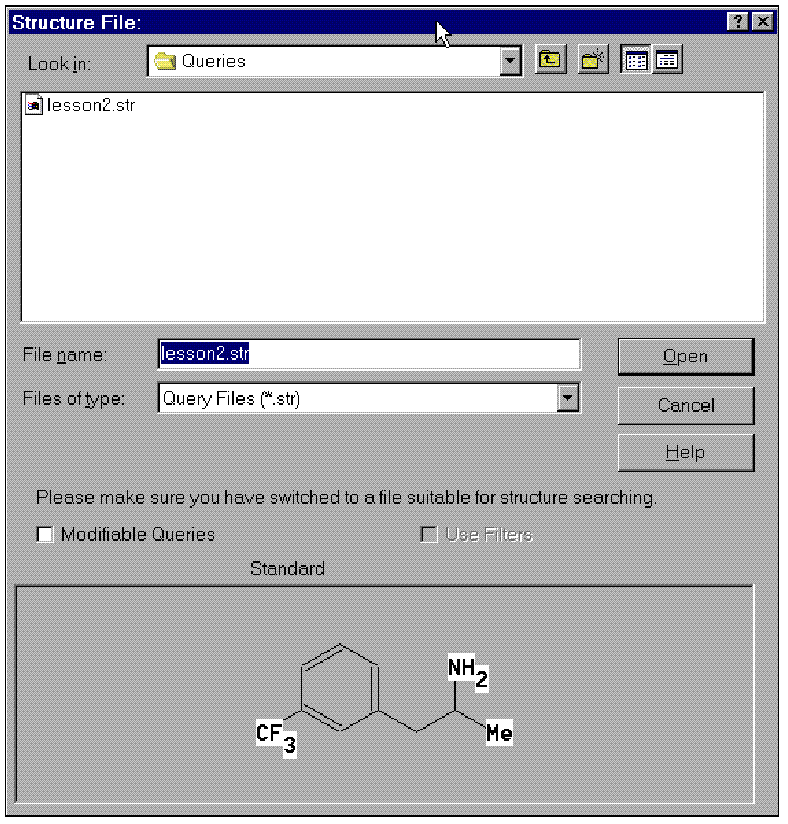
1. Exact search: retrieves compounds with structures which match exactly the unsubstituted query structure.
2. Family search: results include the exact search answers, as well as multicomponent substances such as salts, mixtures, or copolymers in which the structure of one of the components matches the query structure.
3. Substructure search: retrieves answers in which the defined fragment structure is embedded within a larger structure. Only substructure searches retrieve answers in which any type of substituent may occur on the query structure.

The following table summarizes the differences in the types of substances retrieved in each case.

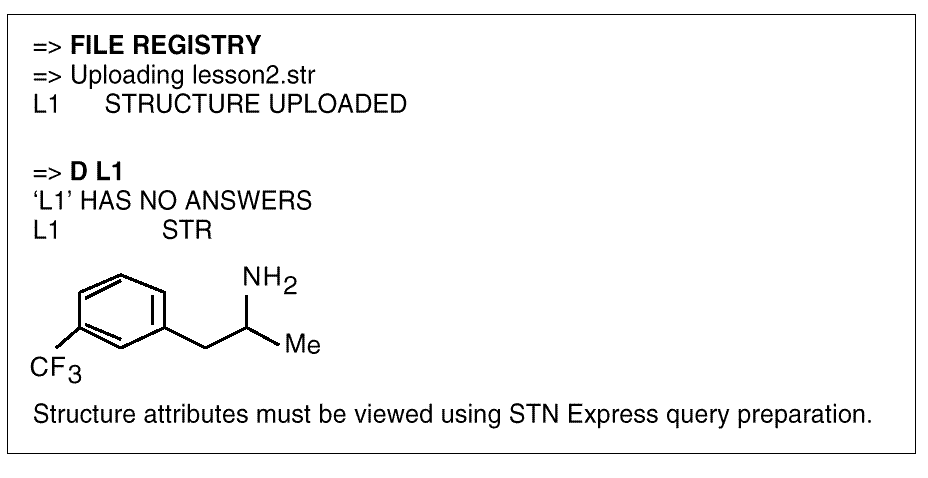
|  |  |  |  |
| --- | --- | --- | --- |
| **Type of substance retrieved** | **EXACT** | **FAMILY** | **SSS** |
| Structure with no additional substitution | X | X | X |
| Radiolabelled forms of the substance with no additional substitution | X | X | X |
| Stereoisomers with no additional substitution | X | X | X |
| Homopolymers with no additional substitution | X | X | X |
| Free radicals with no additional substitution | X | X | X |
| Charged compounds with no additional substitution | X | X | X |
| Salts and mixtures of unsubstituted compounds |  | X | X |
| Molecular addition compounds of unsubstituted compounds |  | X | X |
| Copolymers of unsubstituted compounds |  | X | X |
| Structures with additional substitution |  |  | X |

All three types of structure search generally follow a similar procedure. The Registry file is entered and the structure query uploaded by either clicking on the Upload Query icon  in the toolbar, by using Upload Structure Query through the Query pulldown menu, or by using Ctrl+U.

The following dialog box appears. Choose the required file (the most recently prepared file will appear in the File name box) and click Open.



The structure file will be uploaded, and the query assigned an L-number. Once complete, it is good practice to display the file using the display command D in order to double check that the correct file has been uploaded. Note: large structures will not be displayed.



A structure search can now be carried out using the L-number for the structure. To search structures in Registry, enter the SEARCH command, the L-number for the structure which you want to search and provide information on the type and scope of the search. To conduct an exact structure search, enter EXACT. To conduct a family structure search, enter FAMILY or FAM. To conduct a substructure search, enter SSS (the default). To conduct a sample structure search, enter SAM or SAMPLE. A full search is performed by entering the command FULL or ful.

Always perform a SAMPLE search first. This searches a small portion of the file (about 5%) in order to verify the types of answers the query retrieves and to see whether the search will run to completion in the full file. If neither SAM or FUL are assigned to the search, the system defaults to a sample search.

Structure searches are conducted in two steps: screening followed by iteration. In the screening process candidate answers are retrieved based on some general analysis of the query structural characteristics. The iterative process involves a detailed atom-by-atom and bond-by-bond comparison of the candidate answers with the query structure. Based on sample search results, projections are made for the range of answers and iterations in the full file. If those projections exceed the limits for the iterative process, a message that the full file projection is incomplete if the search is conducted online appears. If a message of incomplete projections is given after running a sample search, then the query must be further limited.

SAMPLE SEARCH INITIATED 23:23:12

SAMPLE SCREEN SEARCH COMPLETED - 1 TO ITERATE

100.0% PROCESSED 1 ITERATIONS 1 ANSWERS

SEARCH TIME: 00.00.01

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

BATCH \*\*COMPLETE\*\*

PROJECTED ITERATIONS: 1 TO 80

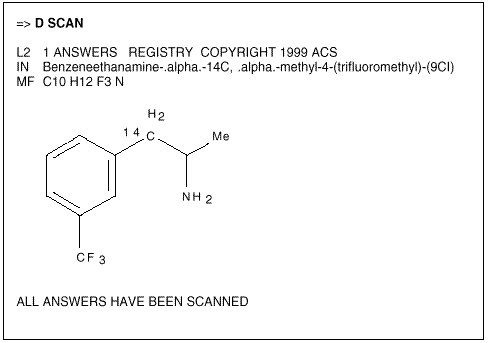
PROJECTED ANSWERS: 1 TO 80

L2 1 SEA SSS SAM L1

The following should be noted:

1. In some cases the sample search is incomplete, in which case the second line of the above transcript reads "Sample Search Incomplete". This is not a concern if the Full File Projections are both complete.
2. If the Full File Projection indicates that the Online search is incomplete, then the Full on-line search must not be carried out. If the Batch is also incomplete then the query must be reconsidered and further limited off-line.
3. If the Online search is incomplete and the Batch is complete, then a Batch search may be carried out. Batch searches are left to run overnight.
4. If both the Online and Batch are complete then the Full search may be carried out.

The sample search creates an L-answer set. In the above example this is a single compound. Display Scan or D Scan is used to browse the answer set at no cost. Answers are displayed in random order, and answer numbers are not displayed. Answers include the CAPLUS index name IN, the molecular formula MF, and the structure diagram.



Once the answers have been checked and the query considered satisfactory, the full search is carried out. In the example below a full substructure search is carried out.

=> s l1 sss ful

FULL SEARCH INITIATED 23:23:35

L3 41 SEA SSS FUL L1 EXTEND

CANDIDATE STRUCTURE SEARCH COMPLETED - 41 TO ITERATE

100.0% PROCESSED 41 ITERATIONS 36 ANSWERS

SEARCH TIME: 00.00.01

L4 36 SEA SSS FUL L1

These answers can be displayed in the Registry file, but no bibliographic or abstract material is given. To retrieve this information the result is then searched in a bibliographic file such as File CAPLUS.

=> fil caplus

=> s l4

L5 222 L4

In this case 36 compounds were found that matched the query in the Registry File. However these corresponded to 222 abstracts in the CAPLUS File. These may be limited in the usual way by keyword, removing patents etc. Family and exact searches are carried out in a similar manner, but with the commands FAM or EXA instead of SSS.

[Return to top](#TOC)

14.2 Batch Searches

Occasionally, searches will be too large to carry out a complete search on-line. In such cases an option is available to load the query on to the STN system, where it is searched at a time when there is less demand on the system (usually overnight).

Batch searches are carried out if:

1. the on-line search is projected to be incomplete; and,
2. the batch search is projected to be complete; and,
3. there is no other way of restricting the search to obtain a complete on-line search.

The command to run the batch search is:

=> BATCH Lx QUERYNAME/B

The Lx is the L-value of the uploaded structure query. The QUERYNAME is simply a name by which the results can be retrieved initials, application number etc.. The system will then ask what type of search is required (SSS, CSS, FAM or EXA) and the scope of the search (full or range).

Once the batch search has been carried out, it is saved onto the system as a stored answer QUERYNAME/A. Stored answers can be displayed using the command:

=> D SAVED/A

The stored answer set can then be used by activating it:

=> ACT QUERYNAME/A

This results in a Registry answer set that is then searched as per the usual Registry answer sets obtained from substructure queries.

[Return to top](#TOC)

14.3 Saving Answer Sets

Answers can be saved on the system for later manipulation using the SAVE command. This can be used for structures, but can also be used for keywords and other queries.

To save a substructure result, the required L-answer set is used (generally the full search answer):

=> SAVE Lx QUERYNAME/A

The answer set is then saved for later use. Note that the registry answer sets can only be saved and used in the Registry file. To display files use the command:

=> D SAVED/A

To activate files:

=> ACT QUERYNAME/A

To delete files:

=> DEL QUERYNAME/A

[Return to top](#TOC)

14.4 Extend Option in Structure Searching

As previously indicated, structure searches are conducted in two steps: screening followed by iteration. The final answer set is that obtained after the iterative process. The extend option allows the results of the initial screening process to be viewed, in addition to the answers retrieved after the iterative process. This is particularly useful when the final answer set gives zero results.

The extend option is applied to structure searches using the command SET EXTEND ON (alternatively use SET EXTEND ON PERM to set the command on permanently) – see the example below.

=> FILE REGISTRY

=> SET EXTEND ON PERM

SET COMMAND COMPLETED

=>

Uploading extend3.str

L1 STRUCTURE UPLOADED

=> D L1

L1 HAS NO ANSWERS

L1 STR



=> S L1

'EXTEND' DOES NOT APPLY TO SAMPLE SEARCHES

SAMPLE SEARCH INITIATED 16:17:18

SAMPLE SCREEN SEARCH COMPLETED - 1 TO ITERATE

100.0% PROCESSED 1 ITERATIONS 0 ANSWERS

SEARCH TIME: 00.00.01

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

BATCH \*\*COMPLETE\*\*

PROJECTED ITERATIONS: 1 TO 80

PROJECTED ANSWERS: 0 TO 0

L2 0 SEA SSS SAM L1

=> S L1 FULL

FULL SEARCH INITIATED 16:17:31

L3 60 SEA SSS FUL L1 EXTEND

CANDIDATE STRUCTURE SEARCH COMPLETED - 60 TO ITERATE

100.0% PROCESSED 60 ITERATIONS 0 ANSWERS

SEARCH TIME: 00.00.01

L4 0 SEA SSS FUL L1

=> D HIS

FILE REGISTRY ENTERED...

SET EXTEND ON PERM

L1 STRUCTURE UPLOADED

L2 0 S L1

L3 60 S L1 FULL EXTEND

L4 0 S L1 FULL

=> D SCAN L3

L3 60 ANSWERS REGISTRY COPYRIGHT 2004 ACS on STN

IN 1H-Pyrrole-3-carbothioic acid, 4,5-dibromo-2-

(trifluoromethyl)-, S-methylester (9CI)

MF C7 H4 Br2 F3 N O S



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

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

L3 60 ANSWERS REGISTRY COPYRIGHT 2004 ACS on STN

IN 3-Pyridinecarbothioic acid, 6-(difluoromethyl)-4-

(2-methylpropyl)-5-[[(methylthio)methylene]amino]-

2-(trifluoromethyl)-, S-methyl ester (9CI)

MF C15 H17 F5 N2 O S2



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

HOW MANY MORE ANSWERS DO YOU WISH TO SCAN? (1):0

=> D L3 1-5

L3 ANSWER 1 OF 60 REGISTRY COPYRIGHT 2004 ACS on STN

RN 264880-65-1 REGISTRY

CN 1H-Pyrazole-4-carbothioic acid, 1-(3-chloro-4-

fluorophenyl)-3-methyl-5-(4-methyl-1,2,3-

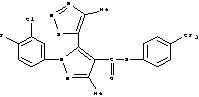
thiadiazol-5-yl)-, S-[4-(trifluoromethyl)phenyl]

ester (9CI) (CA INDEX NAME)

FS 3D CONCORD

MF C21 H13 Cl F4 N4 O S2

SR CAS Client Services



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

Note that the extend answer set may be viewed using D Scan, or transferred to File CAPLUS in the usual way.

Further information on the extend option is available at <http://www.stn-international.com/fileadmin/be_user/STN/pdf/STNotes/note34.pdf>

**Search Techniques**

In any substructure search the following techniques should be considered:

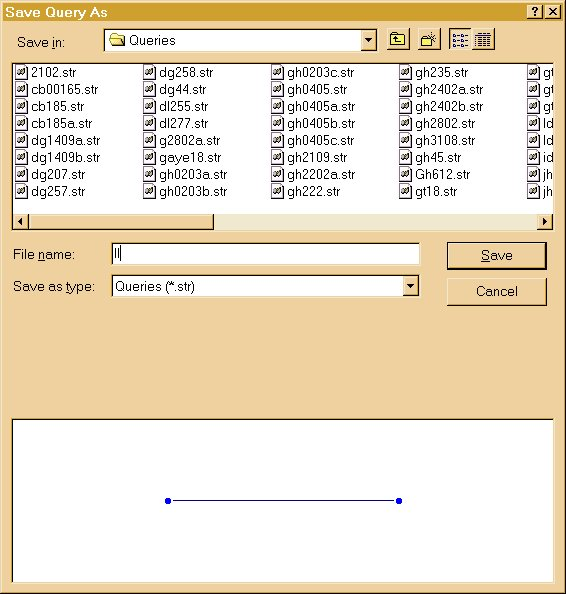
* addition of more functional groups to further limit answer sets;
* using Ring Isolation to specify that rings are either isolated or embedded in condensed ring systems;
* defining specific G-groups to limit answer sets to certain more specific substitution rather than leaving specifying hydrogen or non-hydrogen attachments, defining more functional groups in the structure and so on.

Some more advanced approaches include:

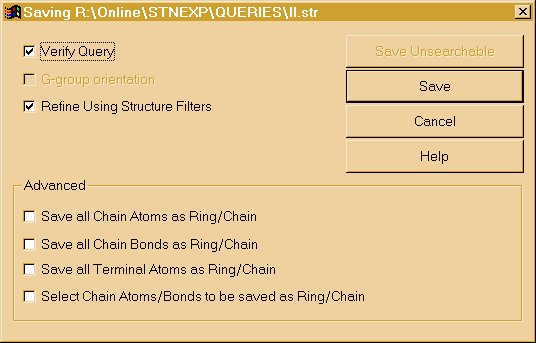
1. Using Manual Filters

Filters are generated by STN Express when a structure is drawn. The Registry database is then screened for compounds that match these filters. While the system automatically generates filters, it is occasionally useful to manually set filters to further restrict answer sets.

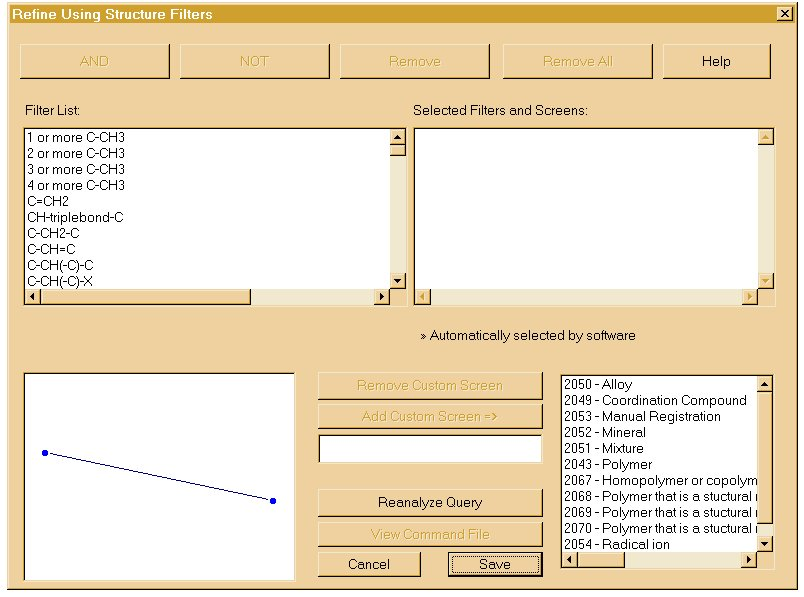
Filters can be added when saving the structure. When the structure is saved the Save Query As dialog box appears. A filename is given to the structure query, and the Save box clicked.



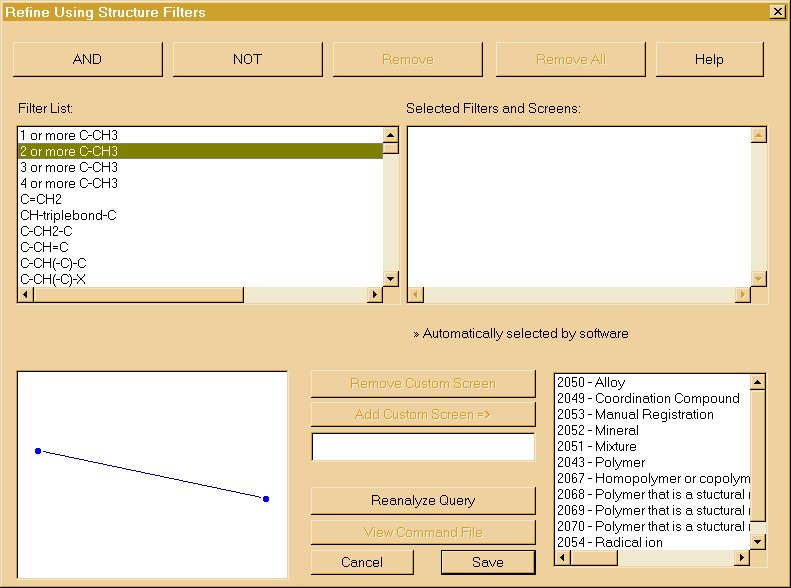
A second dialog box appears Saving R:\Online\STNExpress\Queries\Filename.str. This dialog box contains the options to Verify Query and to Refine Using Structure Filters. If G-groups containing two points of attachment have been defined, then the G-group Orientation box is also active. To add filters the Refine Using Structure Filters must be marked.



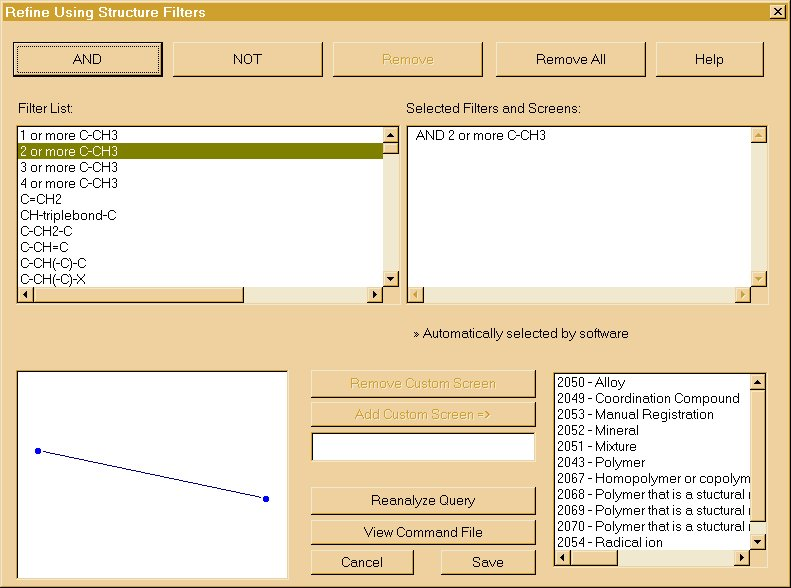
After verifying the query in the usual way, the Refine Using Structure Filters dialog box appears. This contains Filter List and Custom Screen windows, as shown in the diagram below.



To choose filters from the filter list, simply click on the required filter. The icons AND and NOT become active.



Clicking AND specifies that the structure contains the defined filter. Clicking NOT specifies that the structure does not contain the group.



As shown in the diagrams above, some filters define "1 or more" groups, "2 or more" groups and so on. If, for example, exactly one methyl group was required, the filters would be added as "AND 1 or more C-CH3", then "NOT 2 or more C-CH3". This would give only compounds containing a single C-CH3 group.

The Custom Screens window allows the query to be limited to fields such as alloys, mixtures, polymers and the like.

[Return to top](#TOC)

1. Using Separate functional groups in queries

This technique is used for searching compounds in which functional groups are separated by generic links, for example, as in (I) wherein A, B and C are defined as a number of different linking groups.



(I)

These types of searches present difficulties. It is not possible to search structures in which two G-groups are directly linked, so the above compound cannot be drawn with A and B defined as two G-groups. However, such structures can be searched by drawing the structure as separate structures, and the system retrieves any answer that contains the two groups. For (I) the structure query could be drawn as separate phenyl and pyridyl moieties, which the system searches as the two groups separated by any link. This retrieves compounds of formula (I), but also gives compounds where the groups are directly linked e.g. (II), linked through rings eg. (III) etc.

(II) (III)

Alternatively the structure query could be drawn as separate phenyl, pyridyl and carbonyl groups. This retrieves compounds corresponding to I, but also gives any other compounds containing phenyl and pyridyl rings, for example II and III with further substitution by acyl, e.g. (IV).



(IV)

G-groups can also be drawn as separate fragments, and the system then searches these as alternative structures. Thus for the structure I, in addition to the separate phenyl, pyridyl and carbonyl groups, three separate G-groups could be drawn corresponding to the A, B and C links. The search then requires that all five groups are present in any answer retrieved though not necessarily in the exact arrangement in (I).

In any search of this type it is best to include as much definition as possible in each fragment. This reduces the number of irrelevant answers retrieved.

[Return to top](#TOC)

1. Searching multiple structures in a single query

It is possible to search multiple structures in the one query. This is done by entering each of the structures as a separate G-group in the one structure query. For example the two structures below could be searched in the one query by building the structure query as G-H, in which the two structures are entered as fragments under G-groups.



The point of attachment and other requirements of the G-groups are entered in the usual manner. In the present example, the hydrogen of G-H can be defined as one of the hydrogens attached to the rings. The point of attachment is therefore defined as being to the corresponding ring carbons.

Searches of this type result in a single answer set. Accordingly, such searches are generally only done where answers are easily distinguished (for example when HITSTR is used to display answers, or the searches relate to the same or related cases).

[Return to top](#TOC)

1. Removing peptide answers from answer sets

Occasionally searches retrieve answers that contain polypeptide groups. These are sometimes difficult to remove as no manual filter is available that can be used to exclude polypeptides. However the following procedure can be used to exclude such answers:

(i) search the structure and get the answer set Lx;

(ii) enter the command S Lx NOT PS/FS

This excludes answers that contain protein sequences as a file segment. Unfortunately this does not appear to remove all protein sequences. If the SQL field is expanded in the CAPLUS file, 689 entries are found having a sequence length of 3. When this is searched with the qualifier ps/fs, only 654 results are obtained. This shows that 35 tripeptides have not been indexed in the CAPLUS file under the ps/fs field. Thus, whilst the ps/fs qualifier will remove a number of polypeptides from the structure search answer set, there may still be some peptides retrieved.

=> e 1/sql

\*\*\*\* START OF FIELD \*\*\*\*

E3 3040 --> 1/SQL

E4 827 2/SQL

E5 689 3/SQL

E6 46381 4/SQL

E7 39299 5/SQL

E8 38937 6/SQL

E9 26274 7/SQL

E10 28447 8/SQL

E11 33910 9/SQL

E12 40773 10/SQL

=> s e5 and ps/fs

689 3/SQL

748557 PS/FS

L2 654 3/SQL AND PS/FS

An alternative is to use the command:

S L1 not sql>1

This will exclude answers where a peptide length of 2 or more has been recorded.

[Return to top](#TOC)

**Useful Tips for Conducting Structure Searches**

The following is a list of tips that may be useful when conducting structure searches.

***G groups***

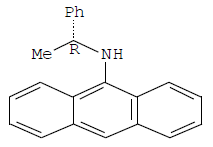
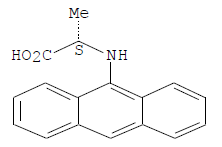
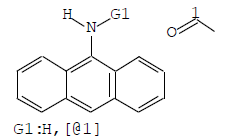
\* A structure query may have up to 20 different G groups and each G group may contain as many as 20 optional substituents. Note however that a query with many G groups may be too large to search.

\* A G group may be drawn as a variable point of attachment.

\* A G group may form part of another G group.

\* DO NOT use hydrogen in G groups! If one atom/node has G-groups that include hydrogen, you will retrieve unexpected answers (but you won’t miss answers you should get)

The search system looks at the definition of the G-group and the atom to which the G-group is attached, and then asks, “Is this a match?”

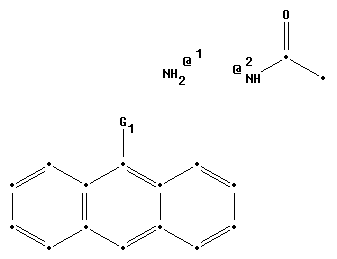
In the example below the G-group includes a hydrogen atom. The G-group is attached to the nitrogen atom of 9-anthraceneamine. The search system looks at the candidate answers, and then asks, “Is there a hydrogen atom attached to this nitrogen atom?” In each of the answers we found (including all of the answers we did not expect to find), the answer is, “Yes, there is a hydrogen atom attached to this nitrogen atom, so this compound is a match.”

Will retrieve (for example)

Unfortunately, the search system did not recognize that the hydrogen atom we found is the same hydrogen atom that explicitly is included on the nitrogen atom of the query structure. The system did not recognize that we want retrieved structures to contain both the explicit hydrogen and the hydrogen in the G-group.

The solution to this problem is simple: Create a G-group that includes the atom to which the explicit (or implicit) hydrogens and the optional hydrogens are attached. Include in this G-group all of the alternatives needed to represent all possible combinations of explicit/implicit hydrogens and optional hydrogens (see over).

\*



***Coordination Compounds and N-oxides***

\* Bonds to metal atoms in coordination compounds should be drawn as single bonds, with the bond type unspecified.

\* N-oxides should be drawn as a double bond between the nitrogen atom and the oxygen atom, e.g.



14.5 Retrieving information upon Substances that do not have a corresponding reference in CAPLUS

The CAS Registry database comprises more than 100,000,000 substances, most of which are referenced by Patents and/or NPL that are indexed in CAPLUS (the primary bibliographic database for chemists). However, there are many substances in the Registry file that do not have an associated CAPLUS reference. In order to cover such a deficiency the examiner must use the /LC (locator) field as shown in the example below (**bolded** text).

L1 STRUCTURE UPLOADED

=> s l1 sss ful

FULL SEARCH INITIATED 00:40:34 FILE 'REGISTRY'

L3 1844 SEA SSS FUL L1 EXTEND

CANDIDATE STRUCTURE SEARCH COMPLETED - 1844 TO ITERATE

100.0% PROCESSED 1844 ITERATIONS 110 ANSWERS

SEARCH TIME: 00.00.01

L4 110 SEA SSS FUL L1

**=> s l4 not caplus/lc**

**96349902 CAPLUS/LC**

**L5 1 L4 NOT CAPLUS/LC**

In some cases the examiner may wish to limit the answer set further by the entry date of the substance into the Registry database. This is done by using the ED (entry date) field using the syntax of YYYYMMDD as shown in the example below (for a Filing date of 5 October 2009).

**=> s l5 and ED<20091005**

|  |
| --- |
| RETRIEVING DOCUMENTS CITED IN FERs |

15.1 Background

International and Foreign Search Reports often have references to Chemical Abstracts, Beilstein Registry Numbers BRN and CAS Registry Numbers CAS RN. These can be retrieved through STN quickly.

[Return to top](#TOC)

## 15.2 Retrieving Chemical Abstracts

Chemical Abstracts are obtained on-line from File CAPLUS, as shown below. The bibliographic details are generally given in the Search Report, so only a display of the abstract is required D Abs. However, if the bibliographic details are required the command D Bib Abs can be used. For a complete record of all indexing information, Registry numbers and supplementary terms use the command D All.

The Chemical Abstract number (document number) is entered in the format AB:CDEFG. The first two numbers refer to the volume number and the numbers following the colon refer to the abstract number. Chemical Abstracts may end with a letter, for example 73:6000i, but this is not entered in the search query. Alternatively, the Chemical Abstracts accession number may be searched using the format ABCD:EFGHIJ wherein the first four numbers denote the year of publication.

=> fil caplus

=> D abs 73:6000/DN

ANSWER 1 CA COPYRIGHT 2000 ACS

AB The geol., mineralogical, and petrographic characteristics of the

Mukhor-Taly, Mukhor-Balyk, and Baryn-Arshan areas of perlite development

are briefly described. Perlite rocks have similar chem. compn. despite

structural-textural differences in volcanic glasses and glasslike rocks

easily distinguishable visually. This was substantiated by microscopic

studies. Perlites and rocks accompanying them are acidic glasses of

liparite-dacite series. The rocks have elevated alky. and a const.

predominance of K over Na. An excess in SiO2 content in rocks promoted

the segregation of the free quartz or cristobalite and tridymate, during

devitrification of primary crystn., and formation of secondary quartz

during recrystn. of primary structures. The supersatn. of rocks with

Al2O3 was related to the widely distributed zeolites and montmorillonite.

Formation of volcanic glasses was related to the activity of the

Babryn-Arshan volcano during Mesozoic time. The possible use of volcanic

rocks, accompanying perlite, in various industries is discussed.

=> log y

[Return to top](#TOC)

15.3 Retrieving Beilstein Abstracts

European Search Reports occasionally cite Beilstein Registry Numbers. These may be obtained from File Reaxysfile, as shown below. The Beilstein Registry number is generally cited in Search Reports in the format “BRN ABCDEF”. The initial step in retrieving the Registry entry is to search the Beilstein Registry number, entered as ABCDEF. The resulting answer set is then displayed. Reaxysfile also provides other search and display fields, but in some cases the display format can be costly (>~$300), so the cost of the display should be checked before carrying out the command.

=> fil reaxysfile

FILE LAST UPDATED: 6 MAR 2000

FILE COVERS 1779 TO 2000.

=> s 392785

L1 1 392785

=> d l1 ide

L1 ANSWER 1 OF 1 REAXYSFILE COPYRIGHT 2013 Elsevier Properties SA. on STN

Accession Number (AN): 392785

Chemical Name (CN): N,N-dimethyl-4-4,5,6,7-tetrahydro-3H-azepin-2-yl-aniline

Autonom Name (AUN): dimethyl-<4-4,5,6,7-tetrahydro-3H-azepin-2-yl-

phenyl>-amine

Lin. Struct. Formula (LSF): C14H20N2

Molec. Formula (MF): C14 H20 N2

Formula Weight (FW): 216.326

Compound Type (CTYPE): heterocyclic

InChi Key: (INCHI): BNRLKMHWJRQOIX-UHFFFAOYSA-N

Alternate InChi Key: (AINCHI): BNRLKMHWJRQOIX-UHFFFAOYAF

Markush Ref. Count (MARKREF): 0

Entry Date (DED): 1988/11/28

Update Date (DUPD): 2008/02/20

….

=> log y

A single Reaxysfile entry does not provide any bibliographic details. If bibliographic details are required they may be obtained by searching the compound in the Chemical Abstracts file once the compound details have been obtained from File Reaxysfile.

[Return to top](#TOC)

15.4 Registry Numbers

CAS Registry Numbers may also be cited in Search Reports, generally as part of a cited Chemical Abstract or other non-patent literature citation and given in the format “CAS RN ABCDE-XY-Z”. These are particularly useful in determining whether or not a citation is of relevance. They are obtained from the Registry file using the display command. The display provides the compound name, any synonyms, molecular formula, files in which the compound is indexed and the number of references in each, and the chemical structure.

=> fil reg

=> d 36587-34-5

ANSWER 1 REGISTRY COPYRIGHT 2000 ACS

RN 36587-34-5 REGISTRY

CN Benzenesulfonic acid, 2,2'-1,2-ethenediylbis[5-[[4-[bis2-

hydroxyethylamino]-6-[[4-diethylaminophenyl]amino]-1,3,5-triazin-2-

yl]amino]-, disodium salt 9CI CA INDEX NAME

MF C48 H62 N14 O10 S2 . 2 Na

LC STN Files: CA, CAPLUS

1 REFERENCES IN FILE CA 1967 TO DATE

1 REFERENCES IN FILE CAPLUS 1967 TO DATE

=> log y

[Return to top](#TOC)

15.5 Chemcats Accession Numbers

Periodically, EPO generated search reports will contain Chemcats accession numbers and/or order numbers. These may be searched using fil chemcats using the syntax ABCDEF/on (for an order number) or ABCDEF/an (for an accession number). The chemcats database is, however, problematic as records older than 2 years are deleted from the database, making older EP ISRs (that cite chemcats) practically redundant.

=> fil chemcats

=> s 123456/an

L1 1 123456/AN

(0000123456/AN)

=> d l1

L1 ANSWER 1 OF 1 CHEMCATS COPYRIGHT 2013 ACS on STN

Accession No. (AN): **0000123456** CHEMCATS

Catalog Name (CO): Alinda Chemical Ltd. Screening Compounds

Publication Date (PD): 1 Jan 2013

Order Number (ON): IBS-L0152614

Chemical Name (CN): 3-Pyrrolidinecarboxamide,

N-(5-chloro-2-methoxyphenyl)-1-(1,1-dimethylethyl)-5-

oxo-

CAS Registry No. (RN): 873095-84-2

Supplementary Term (ST): CHEMICAL LIBRARY

Structure :



=> log y

[Return to top](#TOC)

|  |
| --- |
| EXPORTING THE SEARCH STRATEGY FOR THE SIS |

To obtain a history of an STN search the command D HIS should be used. Depending upon the type of search performed, it may be necessary to use the command D HIS FULL as using D HIS will only retrieve single lines of a search statement.

=> d his

(FILE 'HOME' ENTERED AT 01:48:28 ON 13 JUL 2015)

FILE 'CAPLUS' ENTERED AT 01:49:04 ON 13 JUL 2015

L1 1 WO2011047174/PN

SEL RN

FILE 'REGISTRY' ENTERED AT 01:49:12 ON 13 JUL 2015

L2 23 E1-E23

=> d his full

(FILE 'HOME' ENTERED AT 01:48:28 ON 13 JUL 2015)

FILE 'CAPLUS' ENTERED AT 01:49:04 ON 13 JUL 2015

L1 1 SEA SPE=ON ABB=ON PLU=ON WO2011047174/PN

SEL RN

FILE 'REGISTRY' ENTERED AT 01:49:12 ON 13 JUL 2015

L2 23 SEA SPE=ON ABB=ON PLU=ON (1019771-89-1/BI OR 1019771-90-4/BI

OR 10288-72-9/BI OR 1068-55-9/BI OR 1222543-67-0/BI OR

1222544-85-5/BI OR 1222548-05-1/BI OR 1222548-07-3/BI OR

1245647-80-6/BI OR 1259933-15-7/BI OR 1259933-16-8/BI OR

1286255-36-4/BI OR 1286255-37-5/BI OR 1286255-38-6/BI OR

139-85-5/BI OR 17515-77-4/BI OR 215867-86-0/BI OR 29668-44-8/BI

OR 533-31-3/BI OR 864264-98-2/BI OR 91-56-5/BI OR 912656-46-3/

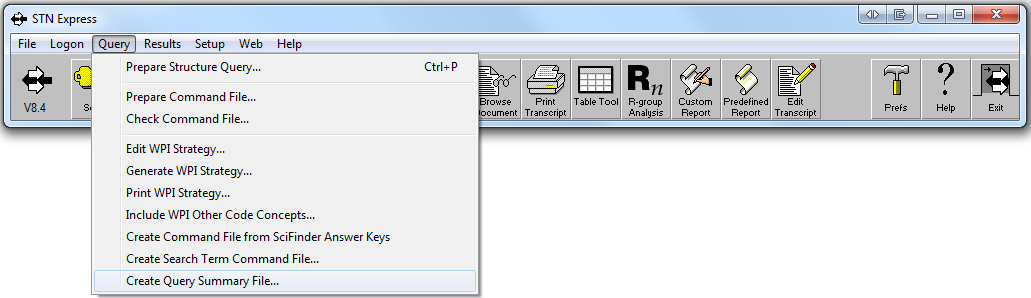
BI OR 912661-88-2/BI)

Alternatively it is good practice to set the system to automatically create a history listing when logging off (using LOG H or LOG Y). This is done using the command:

=> SET LHISTORY ON

(OR => SET LHISTORY ON PERM to set the command on permanently)

Following logging out of the system a query summary may be produced using the following steps:

1. From the “Query” menu of the main toolbar choose “Create Query Summary File”
2. Open the transcript from the request box (it will default to the most recent transcript saved)
3. Immediately a second query will open asking for the location to save the summary file (e.g. the desktop). Ensure RTF is selected as the file type as PDF is the default (using RTF makes it easier to transfer to the information to Infiniti/Intellidox).

## 16.1 SIS for Keyword or Sequence searches

A search summary for a keyword search should look similar to the following:

FILE 'CAPLUS' ENTERED AT 21:23:29 ON 06 JUL 2015

L5 5637 SEA SPE=ON ABB=ON PLU=ON L4

L6 4284 SEA SPE=ON ABB=ON PLU=ON NLRP3 OR NALP3 OR NACHT OR CRYOPYRIN OR (CATERPILLER (W) LIKE (W) RECEPTOR) OR INFLAMMASOME

L7 0 SEA SPE=ON ABB=ON PLU=ON L5 AND L6

L8 3424527 SEA SPE=ON ABB=ON PLU=ON MODULAT? OR INHIBIT?

L9 360969 SEA SPE=ON ABB=ON PLU=ON INTERLEUKIN?

L10 22050 SEA SPE=ON ABB=ON PLU=ON L8 (3A) L9

L11 4 SEA SPE=ON ABB=ON PLU=ON L5 AND L10

L12 47828 SEA SPE=ON ABB=ON PLU=ON HELPER (2A) CELL

L13 1170 SEA SPE=ON ABB=ON PLU=ON L8 (3A) L12

L14 0 SEA SPE=ON ABB=ON PLU=ON L5 AND L13

L15 0 SEA SPE=ON ABB=ON PLU=ON L5 AND L12

L16 13 SEA SPE=ON ABB=ON PLU=ON L5 AND L9

D BIB ABS HITSTR 1-

L17 25621 SEA SPE=ON ABB=ON PLU=ON CARRI? (2A) PROTEIN

L18 3 SEA SPE=ON ABB=ON PLU=ON L5 AND L17

L19 3 SEA SPE=ON ABB=ON PLU=ON L18 NOT L16

D BIB ABS HITSTR 1-

1. Copy the text from the search summary directly into the SIS screen for Infiniti (Intellidox)

## 16.2 SIS for Structure Searches

A structure search summary file should look similar to the following:

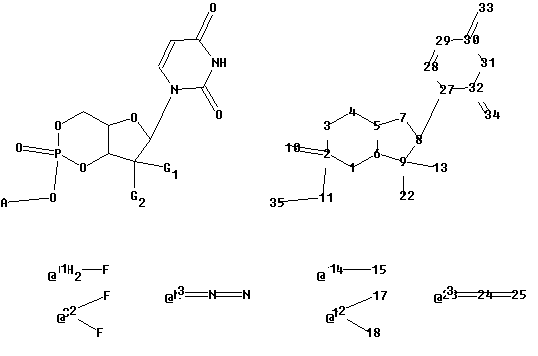
STN Express Query Summary

(FILE 'HOME' ENTERED AT 23:11:19 ON 17 MAR 2014)

FILE 'REGISTRY' ENTERED AT 23:11:23 ON 17 MAR 2014

L1 STRUCTURE UPLOADED

STRUCTURE: R:\Chem\_supersection\C2\Search & Examination\STNEXP\QUERIES\au2009257647 gav.str



chain nodes :

10 11 13 14 15 16 17 18 22 23 24 25 33 34

ring nodes :

1 2 3 4 5 6 7 8 9 27 28 29 30 31 32

ring/chain nodes :

35

chain bonds :

2-10 2-11 8-27 9-13 9-22 11-35 14-15 16-17 16-18 23-24 24-25 30-33 32-34

ring bonds :

1-2 1-6 2-3 3-4 4-5 5-6 5-7 6-9 7-8 8-9 27-28 27-32 28-29 29-30 30-31 31-32

exact/norm bonds :

1-2 1-6 2-3 2-10 2-11 3-4 4-5 5-6 5-7 6-9 7-8 8-9 8-27 9-13 9-22 11-35 23-24 24-25 27-28 27-32 28-29 29-30 30-31 30-33 31-32 32-34

exact bonds :

14-15 16-17 16-18

G1:CH3,CF3,CN,F,[@1],[@2]

G2:X,MeO,NH2,[@3]

Match level :

1:Atom 2:Atom 3:Atom 4:Atom 5:Atom 6:Atom 7:Atom 8:Atom 9:Atom 10:CLASS 11:CLASS 13:CLASS 14:CLASS 15:CLASS 16:CLASS 17:CLASS 18:CLASS 22:CLASS 23:CLASS 24:CLASS 25:CLASS 27:Atom 28:Atom 29:Atom 30:Atom 31:Atom 32:Atom 33:CLASS 34:CLASS 35:CLASS

L2 3 SEA FILE=REGISTRY SSS SAM L1

L3 80 SEA FILE=REGISTRY SSS FUL L1 EXTEND

L4 33 SEA FILE=REGISTRY SSS FUL L1

L5 0 SEA FILE=REGISTRY SPE=ON ABB=ON PLU=ON L4 NOT CAPLUS/LC

FILE 'CAPLUS' ENTERED AT 23:11:59 ON 17 MAR 2014

L6 6 SEA FILE=CAPLUS SPE=ON ABB=ON PLU=ON L4

L7 2 SEA FILE=CAPLUS SPE=ON ABB=ON PLU=ON L4 AND WO2009152095/PN

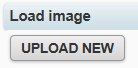
L8 1 SEA FILE=CAPLUS RAN=(,2009) SPE=ON ABB=ON PLU=ON L4

D L6 BIB ABS HITSTR 1-6

1. Copy (or cut) the image of the searched structure from the saved RTF file and paste into a program such as ChemDraw or Paint (ChemDraw is a better option as you can safely resize your image).
2. Save the image as a Bitmap (BMP image) or JPEG (JPG image).
3. Open Infiniti (Intellidox), and ensure that “Will your [...] (including SIS if required) contain images?” is enabled.
4. Copy the text that is before the image into the “Detailed Search Strategy” box on the SIS page.
5. Check the box below this text (in Infiniti) to Insert an Image.



1. Click on “Upload New”.



1. Click on “Browse” and select the saved picture from step 5.



1. In some cases a second Text field will automatically become available, otherwise click on “Insert additional text”.



1. Copy the text that is below the image (see above) into the new text field.
2. Indicate which search statements were viewed (e.g. Results of search statement L6 viewed.)

|  |
| --- |
| DOCUMENT ADMINISTRATION |

**DOCUMENT LOCATION**

**This document is electronically controlled. Its accuracy can only be guaranteed when viewed electronically**

The master copy of this document is held at the following location:

|  |
| --- |
| BRIK File C2007/11139 |

**APPROVALS**

| Name | Signature | Title | Date | Version |
| --- | --- | --- | --- | --- |
| Bob Bartram |  | Supervising Examiner, C4 | 22 March 2007 | V1.0 |
| Jacob Elijah |  | Supervising Examiner, B1 | 31 March 2010 | V2.0 |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |