SPRESIweb 2.1, a Selective Chemical Synthesis and Reaction Database

Dana L. Roth*

Millikan Library 1-32, California Institute of Technology, Pasadena, California 91125

Received July 3, 2005

Introduction. InfoChem's SPRESIweb 2.1, launched in January 2005, offers both structure, chemical reaction, property data, and text searching via the Internet. SPRESIweb's integrated structure and reaction databases are derived from articles indexed from over 1350 journals and conference publications, in addition to chemical patents, published between 1974 and 2002. SPRESIweb provides access to 4.5 million compounds and 3.6 million reactions from 565 000 references including 156 000 patents. In addition, SPRESIweb offers over 28 million chemical, physical, and biological property data values. As an aside, the Chemical Abstracts Service has enhanced its REG File records with InfoChem's physical property data and its CASReact File with InfoChem's chemical reaction data.

The database was originally a joint effort of the All-Union Institute of Scientific and Technical Information of the Academy of Sciences of the USSR (VINIT) and the German Zentrale Informationsverarbeitung Chemie in Berlin (ZIC). InfoChem currently has contracts with the VINITI Institute to continue excerpting data from the current chemical literature for addition to SPRESIweb. It plans to update SPRESIweb during 2005 with the data from 2003 and 2004.

As an indication of the journal article coverage, the percentage of articles indexed for Web of Science between 1982 and 2001, that were also indexed for SPRESIweb, is given for some major organic/organometallic synthesis journals:

Synthesis (4690/6689=70%),

Liebig's Annalen der Chemie (2146/3769=57%),

Tetrahedron Letters (22 156/41 208=54%),

Chemische Berichte (2696/5506=49%),

SynLett (1578/3537=45%),

Journal of Organometallic Chemistry (4167/16 872=25%),

Organometallics (3224/13 413=24%),

Inorganic Chemistry (3,374/20,987=16%)

A similar analysis compares the percentage of articles indexed for Web of Science between 1982 and 2001, that were also indexed for SPRESIweb, for several well-known synthetic organic chemists:

A. G. Myers (46/60=77%),

D. A. Evans (81/126=64%),

J. L. Wood (20/47=43%),

K. B. Sharpless (82/215=38%),

H. C. Brown (166/459=36%),

E. J. Corey (149/488=31%),

* Corresponding author phone: (626)395-6423; fax: (626)792-7540; e-mail: dzrlib@library.caltech.edu.

These results are consistent with SPRESIweb's selective focus on indexing articles describing multiple synthetic procedures for organic/organometallic compounds.

SPRESIweb offers reference links both to publisher's servers via DOI and to the major patent databases (Espacenet, U.S. Patent & Trademark Office, and MicroPatent). SPRESIweb also offers links to several document delivery services (FIZ AutoDoc, CISTI, Subito, and TIBORDER), to ACD/Labs (for subscribers) for predicted properties, and to MetaXchem—a database of commercial sources—which can be searched with structures/substructures or CASRN. SPRESIweb also features an innovative prototype of 'name reactions' search in addition to common search types such as "substructure search", "isomer search", "parent search", and "flex match".

Search. The search/display pages are framed with four quick links (Home, New Query, Help, and Logout) at the top and six module menus on the left: Search, STS (Synthesis Tree Search), Query, Hitlist, Help, and Others. The Search module consists of three parts: Molecules, Reactions, and References.

1. Clinking 'Molecules' displays the New Molecule Query screen as shown in Figure 1.

The two pull-down menus provide a Boolean (and, or, not) combination of any two of the following: Basic Index, Name, Trivial Name, Spresi Regno, Melting Point, Boiling Point, Sublimation Point, Decomposition Point, Transition Temperature, Dissociation Constant, Density, Optical Rotation, Refractive Index, Molecular Formula, Molecular Weight, Element Count, Keywords, Author, Year, Patent Owner, Patent Number, Patent Country.

Search terms can be further qualified (equals, less than, greater than, between, contains, begins with, ends with, etc.) and can also be combined with a structure/substructure search.

'Structure' searching requires double clicking on the 'pencil' icon to open a Structure Window. ISIS/Draw, Java Applet, and ChemDraw are supported. Detailed instructions for advanced structure/reaction searching are available and are linked from the search page.

A 'Basic Index' search will accept 'Name', 'Trivial Name', or 'Molecular Formula'. The default is 'contains', so that the search term is automatically truncated (both left and right). For example 'thiocyan' retrieves 'AETHYL-ISOTHIOCYANAT'. Please note that chemical names are IUPAC (in German) and that only 162K trivial names (some in English) are indexed for the 4.5M compounds.

The [...] box (on the right margin) is a quick link to a dictionary listing (if available) for the various menu terms.

10.1021/ci050274b CCC: $30.25
Published on Web 08/31/2005

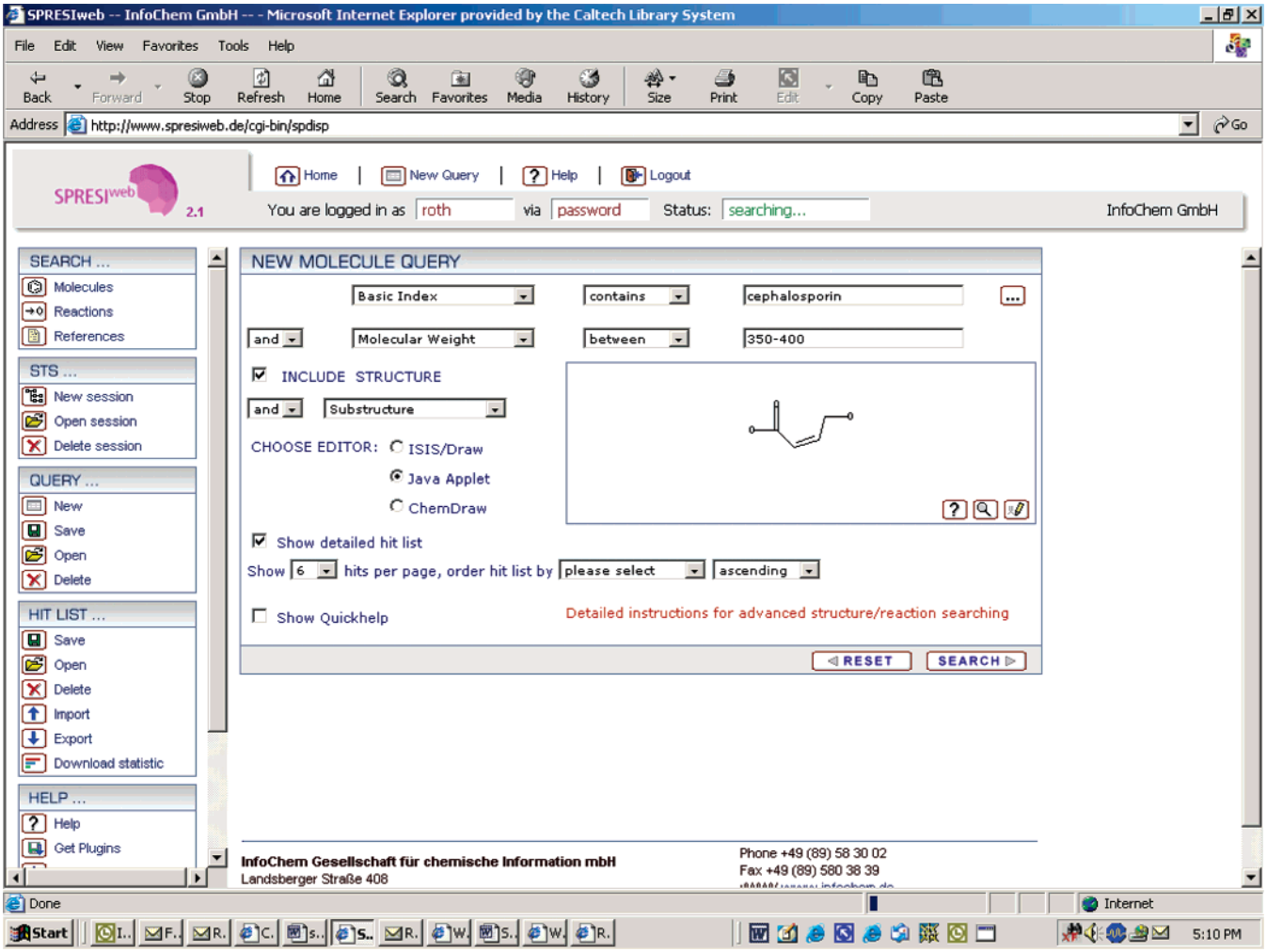

Figure 1.

Clicking on a term in the dictionary rewrites the term in the search box.

Sample Search Results. Searching for 'Benzocaine' in the Basic Index retrieves the record display as shown in Figure 2.

4-Amino-benzoesaeureaethylester. The link to 'More Names/Synonyms' lists nearly 60 synonyms (English, German, French, and the CAS RN).

The links to Journal articles, Patents, and Other references provide brief (au, ti, so) lists which are linked to full records that include 'all molecules' and all 'reactions' as well as links to document suppliers, patent databases, etc.

The link to 'Show all keywords' displays a listing of article/patent titles that are multiply assigned, as appropriate, to the following subject categories:

Analysis: (14 categories)—determination, purity, use, etc.

Colloid: colloid, aerosols

Mechanical/acoustic property: (2 categories)—diffusion, rigidity

Mixture: molecule, discussed

Occurrence: (6 categories)—atmosphere, organism, technical product, etc.

Pharmacology: pharmacologic agent (13 categories)—anesthetic, toxic, etc.

Physical property: melting point, pure

Physiology: organ (4 categories)—eye, kidney, etc.

Reactant: (7 categories)—addition reaction, complex formation, etc.

Reaction: (49 categories)—addition, charge-transfer, hydrolysis, etc.

Special properties: ecological issues

Spectrum: (4 categories)—UV, NMR, IR, MS

Synthesis: (3 categories)-isolation, preparation, etc.

Thermodynamics: (5 categories)—solubility, sorption, etc.

Usage: (5 categories)—cosmetics, fine chemical, drug, etc.

The link to 'Show all physical properties' provides a quick listing of physical property data with titles of literature references (with journal title and year) that are in turn linked to the full article record.

The link to 'Show all detail data' (only first 100 references) provides a listing with full bibliographic detail and document delivery options. In addition, a link to 'Search this article in SPRESI' provides links to all molecules indexed from the article.

The link to 'Use ACD/I-Lab for Predictions' links to the ACD Web site providing additional (calculated) data for ACD subscribers. These data may also be provided in SciFinder/SFS.

The link to 'Search in MetaXchem' transfers the structure to CHEMIE.DE which offers both text and structure/substructure searching with links to supplier catalogs.

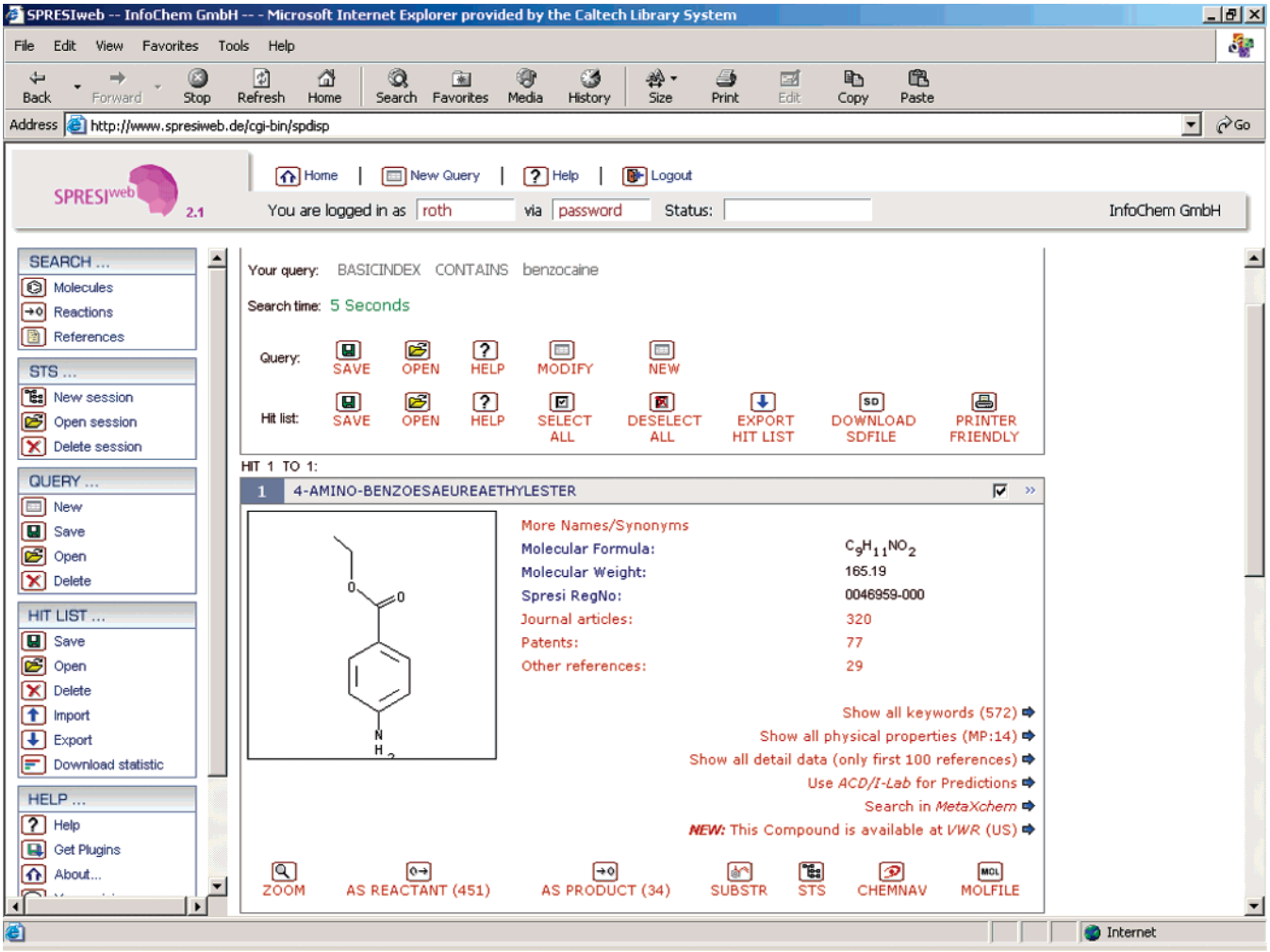

Figure 2.

The link to 'NEW: This Compound is available at VWR (US)' displays the compound record in the VWR International catalog (which includes listings from J. T. Baker, Alfa Aesar, Lancaster Synthesis, and TCI America).

Additional options offered across the bottom of the compound display are as follows:

A 'Zoom' link which magnifies the structural diagram.

'As Reactant' (451 reactions) and 'As Product' (34 reactions) links provide graphical reaction details, with links to literature references, and a search link to 'similar' reactions.

The 'SUBSTR' link runs a substructure search quickly retrieving (in the Benzocaine example) over 7000 compounds.

2. Clicking Reactions displays a 'New Reactions Query' screen that provides pull down menus, for example, to allow searching with a previously retrieved Reaction Number, or SPRESI Regno that can be combined with yield, catalysts/solvents, etc. The results screen offers a link to SIMILAR (reactions). The Reaction query menu also includes an innovative prototype of 'name reactions' search in addition to "substructure search", "isomer search", "parent search", and "flex match" that are available in the 'New Molecule Query'.

3. Clicking References displays a 'New Reference Query' screen that provides pull down menus, for searching, for example, with author names, patent numbers, journal abbreviations, etc. Appropriate dictionary links are provided in the right margin.

STS (Synthesis Tree Search). STS is similar to the Molecule Search options—'As Reactant' or 'As Product'—but with the additional option of using 'yield', 'reaction conditions', and 'Catalysts, solvents' as additional search terms. In a product search, STS displays reactant and product structures, reaction yields with links to reaction details, and a search link to 'similar' reactions. Clicking on a structural diagram, within the STS display, provides a menu offering links to synthesis reactions and MetaXchem and molecule details (i.e. the full SPRESI record). This feature allows each molecule in the Search Tree to be further expanded to quickly find its synthesis reactions, commercial availability, physical properties, and literature references.

Comparison Searches. SPRESIweb, MDL Crossfire Commander (Beilstein), and SciFinder Scholar. 1. 4-Aminobenzoic Acid Propyl Ester. For comparison purposes, a 'Molecule' search for the propyl analogue of Benzocaine (4-Aminobenzoesaeure, propyl ester − SPRESI Regno: 0087331-000) was conducted in SpresiWeb which retrieved 11 references (10 journal articles and 1 patent reference).

A search in Beilstein Crossfire [4-Amino-benzoic acid propyl ester] retrieved 59 references, of which 39 were dated 1972−2002, but only 6 were in common with SPRESI.

A search in SciFinder Scholar (CAS RN 94-12-2) [Benzoic acid, 4-amino-, propyl ester (9CI)] retrieved 193 references,

of which 112 were dated 1972−2002, but only 7 were in common with SPRESI.

The 11 references in SPRESI are listed below, with B or S indicating that the article and compound was indexed in Beilstein or SciFinder Scholar and B* or S* indicating that the article was indexed but the compound was not. The number of compounds, from the article or patent, indexed in each database is given along with the purpose of the research.

1. *Neftekhimija* [*Pet. Chem.*], 1998, 38 (4), 277−281 (Sp6,-B22,S15) − hydrogenation

2. *Zh. Org. Khim*. [*Russ. J. Org. Chem*.], 1997, 33 (11), 1696−1698 (Sp14,B21,S*11) − hydrogenation

3. U.S. Patent 5457128 (1993) (Sp9,S*4) − skin disorders

4. *J. Am. Ceram. Soc.*, 1991, 74 (9), 2189−2196 (Sp11,-S15) − dispersion ... sintering of Al2O3

5. *Indian J. Chem. B*, 1991, 30(5), 494−498 (Sp44,-B48,S*31) − anthelmintics

6. *J. Pharm. Pharmacol*., 1991, 43 (Suppl.), 92 (Sp1) − mechanism of dissolution

7. *Indian J. Chem. B*, 1982, 21(8), 775−777 (Sp36,B34,-S32) − antituberculars

8. J. *Indian Chem. Soc.*, 1980, 57(4), 447−448 (Sp40,-B40,S30) − antimalarials

9. *J. Indian Chem. Soc*., 1976, 53(2), 172−173 (Sp24,-B19,S24) − bactericides and fungicides

10. *J. Indian Chem. Soc.*, 1976, 53(10), 1061−1063 (Sp28,B*16,S20) − antibacterials

11. *Curr. Sci. (India*), 1976, 45(2), 53−54 (Sp25,B*12,-S24) − antimicrobials

Given CAS' policy of only indexing the main point of an article and 'new' information, it is not surprising that, in general, fewer compounds in each article are indexed in SciFinder/Scholar.

The surprising result is that SPRESI may index more compounds/article than Beilstein.

This comparative analysis also highlighted 'transliteration' differences, as the first and third authors of ref 2 were identified as Woronin, M. W. and Kljuew M. W. in SPRESIweb but as Voronin, M. V. and Klyuev, M. V. in both SFS/CA and Beilstein as well as Web of Science. In addition, SPRESIweb shows Neftekhimija, SciFinder Scholar shows Neftekhimiya, and World Cat shows Neftekhimiia.

2. Pentalongin. For comparison purposes, an 'STS' search for the preparation of Pentalongin (1H-Naphtho<2,3-C>Pyran-5,10-dion. SPRESI Regno: 1347417-300) was conducted in SpresiWeb which retrieved 4 references.

A search in Beilstein Crossfire (1H-benzo[*g*]isochromene-5,10-dione. Beilstein RN: 4312410) retrieved 3 references, of which 2 were dated 1972−2002, and 1 was in common with SPRESI.

A search in SciFinder Scholar (1H-Naphtho[2,3-c]pyran-5,10-dione (9CI). CAS RN: 106261-83-0) retrieved 11 references, of which 8 were dated 1972−2002, and 4 were in common with SPRESI.

The 4 references in SPRESI are listed below, with B or S indicating that the article and compound were indexed in Beilstein or SciFinder Scholar and B* or S* indicating that the article was indexed but the preparation of the compound was not indexed.

Chem. Pharm. Bull., 1986, 34(4), 1505−1517 (SP,B,S)

J. Med. Chem., 1987, 30(11), 2005−2008 (SP,B*,S)

J. Org. Chem., 1999, 64(4), 1173−1179 (SP,B*,S)

Synthesis, 1999, (11), 1881−1883 (SP,B*,S)

In conclusion, I found InfoChem's SPRESIweb 2.1 to be an excellent database. While it is obviously selective, it will very likely produce unique results. The search interface is very intuitive and easy to learn. It could be a very reasonable alternative for introducing structure searching in smaller institutions.

Abundant links to help files are provided, and search results are conveniently linked to both additional data and search functionalities. InfoChem's work with both John Wiley & Sons and Thieme-Verlag, in the design and development of the Internet versions of e-EROS and Science of Synthesis as well as the recent announcement of Dialogue's partnering with InfoChem for structure searching, is a testament to both the quality and novel features of their search programs.

Additional information is available at http://www.infochem.de/eng/index.htm.

CI050274B