Database Speaks

Ling-Kang Liu (劉陵崗)

Institute of Chemistry, Academia Sinica
Nangang, Taipei 115, Taiwan
Email: liuu@chem.sinica.edu.tw

OUTLINES

-- Personal experiences

Publication types
Secondary publication
From primary to secondary
Databases
Types of chemical databases
CAS registry
Reaxys
Inverted structures of database
Location of exact reference

Publication Types

Full-length papers
Rapid communications
Short communications
Letters to the editor
Case reports
Technical or Laboratory notes
Methods
etc.

Overlapping publications

Duplicate submission
Duplicate publication
Acceptable secondary publication
Manuscripts based on the same database

Acceptable Secondary Publication

Secondary publication of material published in other journals or online may be justifiable and beneficial, especially when intended to disseminate important information to the widest possible audience (e.g., guidelines produced by government agencies and professional organizations in the same or a different language).

Approval from the editors of both journals; secondary publication for a different group of readers; abbreviated version; secondary version citing the primary reference.

From Primary to Secondary

Preparation of an article
Clear communication and impact factor
Structure of a scientific paper
Peer review
Copy right, etc.
Footnotes
etc.

Society publisher *vs* professional publisher
Abstracts *vs Concentrates*
etc.

60+ Free Chemistry Databases

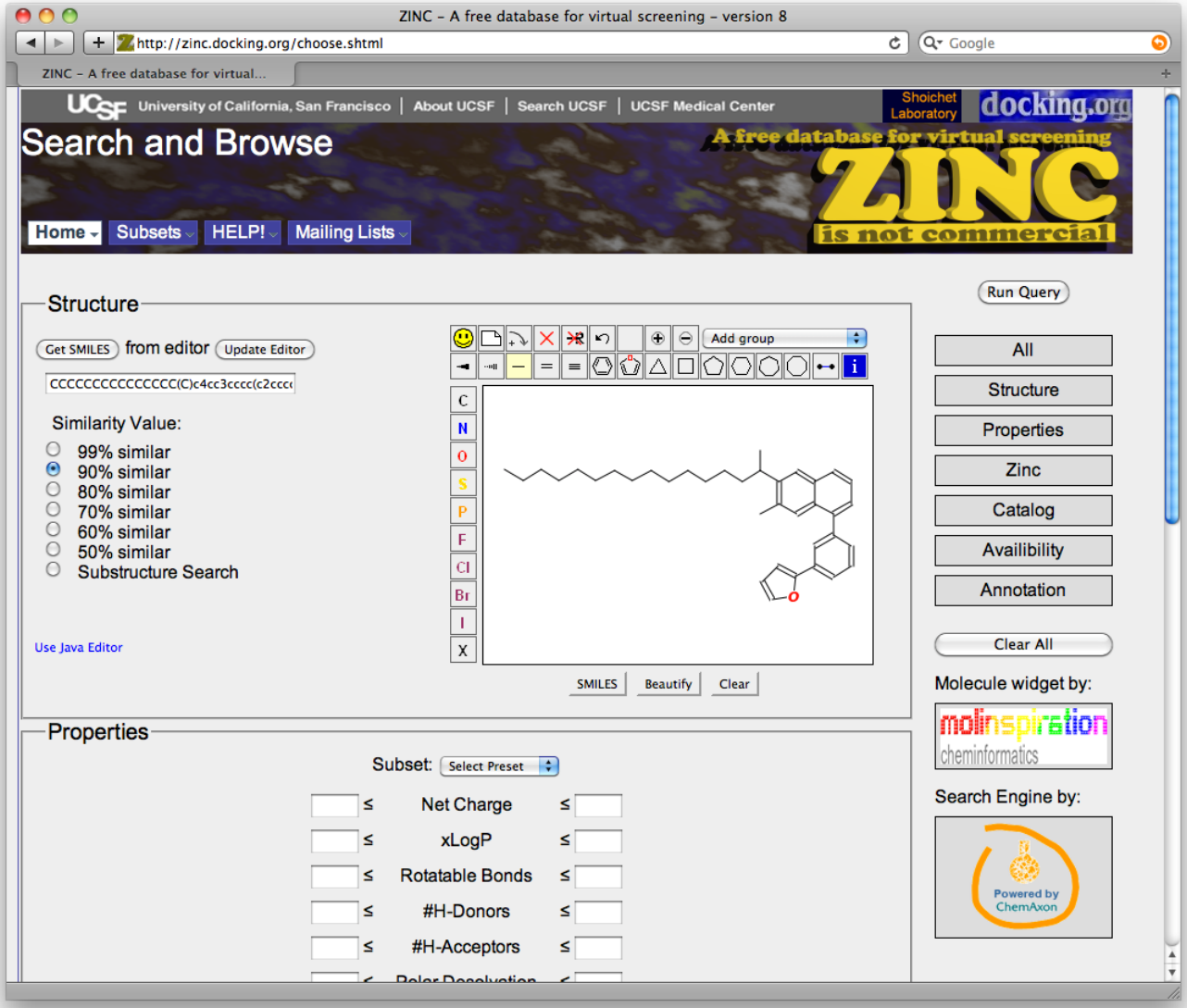
http://depth-first.com/articles/2011/10/12/sixty-four-free-chemistry-databases/

The open Web offers a rich collection of diverse chemical data.

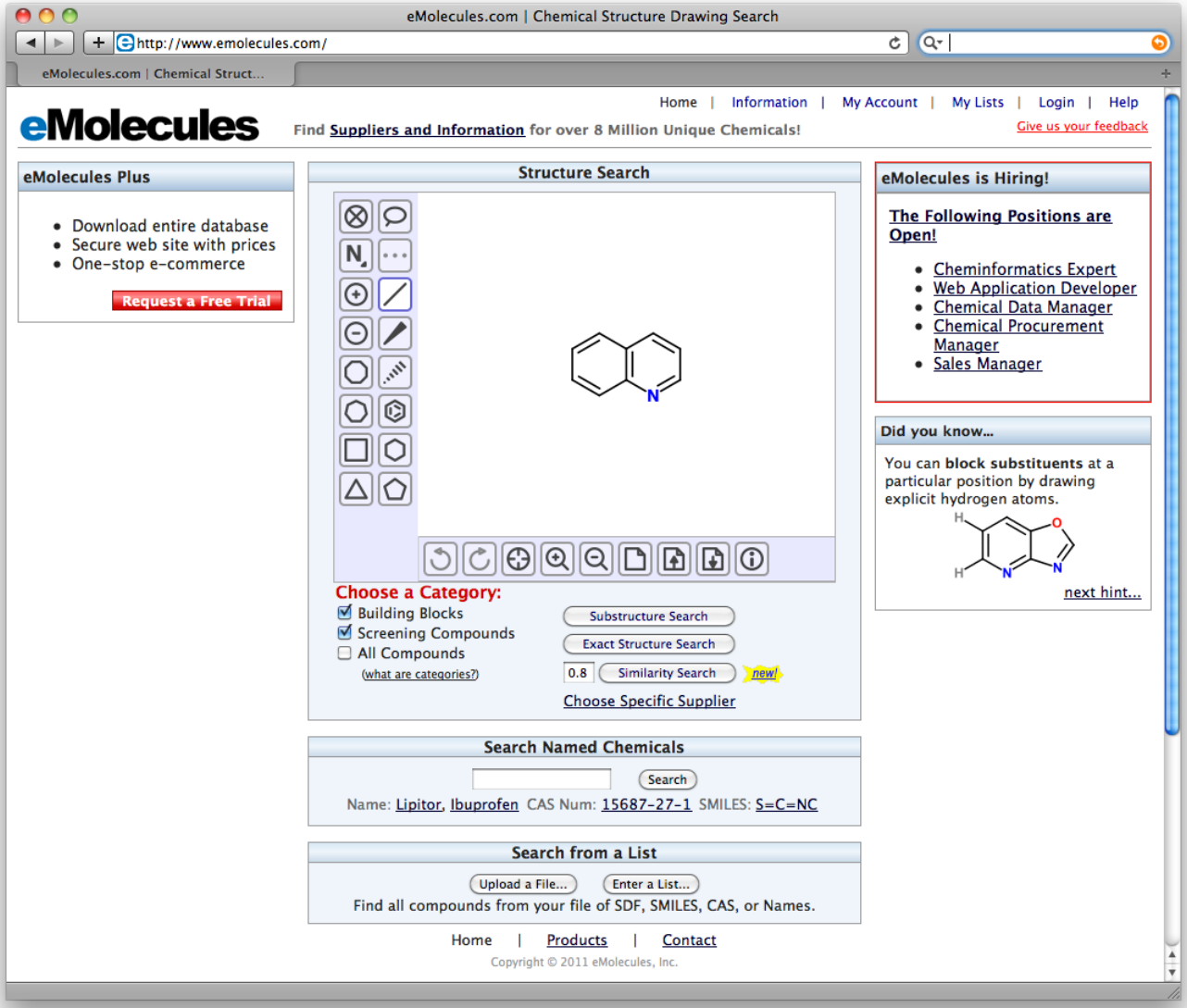
It's of course likely that more services will be created and/or retired in the coming years. Although many of the old databases are no longer active, a number of them continue to run and even prosper.

Examples follow…

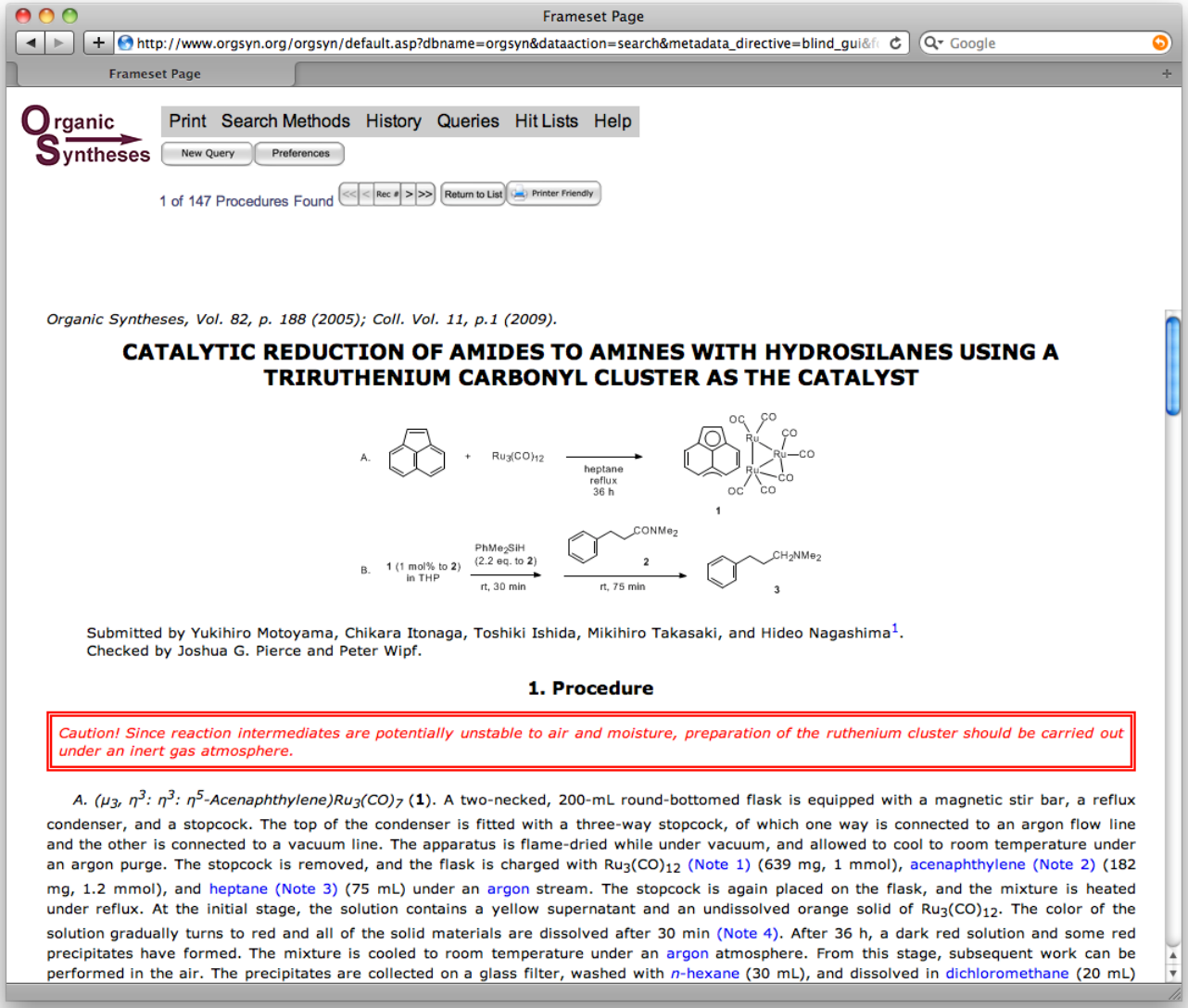
ZINC "... a free database of commercially-available compounds for virtual screening. ZINC contains over 13 million purchasable compounds in ready-to-dock, 3D formats."



eMolecules discovers sources of chemical data by searching the Internet, and receives submissions from data providers such as chemical suppliers and academic research institutions.



Organic Syntheses "... Procedure is written in detail as compared to typical procedures in journals, and reaction and characterization data has been checked for reproducibility."



Chemical Databases

A **chemical database** is a database specifically designed to store chemical information. This information is about chemical and crystal structures, spectra, reactions and syntheses, thermophysical data, *etc*.

Types of Chemical Databases

Chemical structures are traditionally represented using bonds between atoms and drawn as 2D structural formulae.

While these are ideal visual representations for the chemist, they are unsuitable for computational search and storage.

Small molecules (also called ligands in drug design applications), are usually represented using lists of atoms and their connections. Large molecules such as proteins are however more compactly represented using the sequences of their amino acid building blocks.

!!! Terabytes of physical memory !!!

Types of Chemical Databases

Chemical Literature databases correlate structures or other chemical information to relevant references such as academic papers or patents. This type of database includes STN, Scifinder and Reaxys.

Crystallographic databases: noted examples are Protein Data Bank and Cambridge Structural Database.

NMR spectra databases often include FTIR and mass data.

Reactions databases store products, intermediates and temporarily created molecules, and also reaction mechanisms.

Thermophysical databases …

Chemical Database

CAS REGISTRY[SM]

More than 95 million small molecules information, including the structures, names, predicted and experimental properties, tags, and spectra

literature from 1957

up to 10 digits, divided by hyphens into 3 parts

15,000 new substances added /day

numbers assigned in sequential order to unique, new substances identified by CAS scientists for inclusion

66,068,339 sequences

CAS Number

CAS REGISTRY℠

CAS Number, up to 10 digits long with format xxxxxxx-yy-z, is assigned to a compound as CAS registers a new compound.

CAS Number [64-17-5] refers to ethanol, also to ethyl alcohol, ethyl hydrate, absolute alcohol, grain alcohol, hydroxyethane.

D-glucose is also called dextrose and has CAS number [50-99-7]. L-glucose is the mirror image of D-glucose and has a CAS Number of [921-60-8].

Water has a CAS Number [7732-18-5]; the control 5 from (8×1 + 1×2 + 2×3 + 3×4 + 7×5 + 7×6) mod 10 = 105 mod 10 = 5.

Reaxys Databases

Reaxys and Reaxys Medicinal Chemistry combine comprehensive databases of chemistry literature and data with a powerful interface. They return relevant extracted data and citations in optimal formats. Validate promising leads, investigate chemical synthesis possibilities, and understand relationships between chemicals and bioactivity data.

Access over 16,000 periodicals with 500 million experimental facts

Find structures, properties, synthesis possibilities, reactions and bioactivity data

Get inorganic, organic and organometallic chemistry and bioactivity data through one interface

Search from various intuitive options such as Ask Reaxys and Reaxys Tree

Inverted Structures of Database

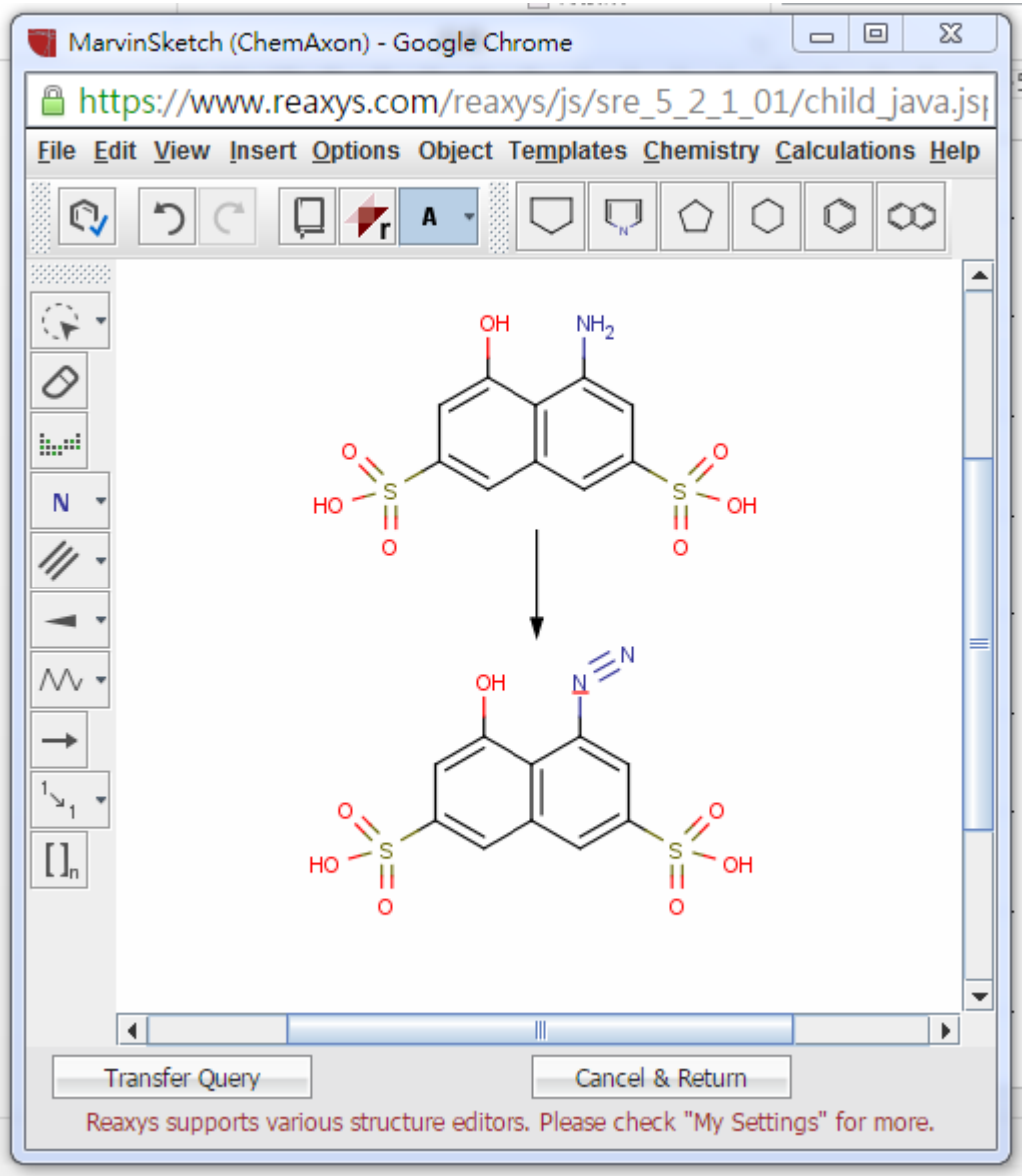
Sequential processing

Sorting

Nodes – linkers
Atoms – bonds
Parallel processing

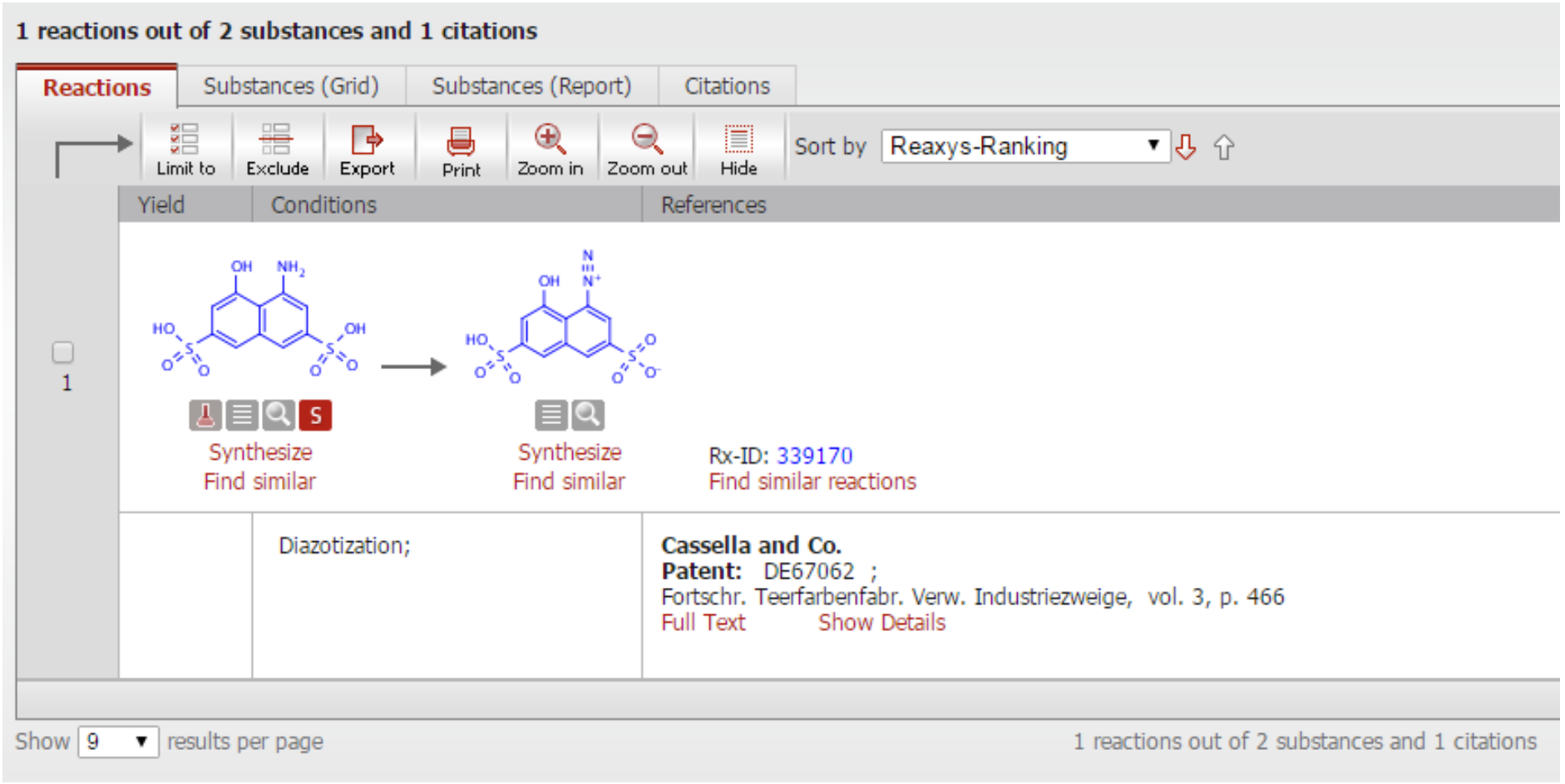
Reaxys Reaction Search

Search and Location of References



Reaxys Reaction Search

Search and Location of References



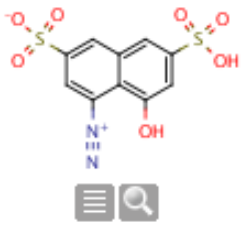
Reaxys Reaction Search

Search and Location of References

1 substances out of 6 reactions and 5 citations

Reactions | Substances (Grid) | **Substances (Report)** | Citations

Limit to | Exclude | Export | Print | Zoom in | Zoom out | Hide | Sort by: No of References

	Structure	Structure/Compound Data	N° of preparations All Preps | All Reactions
1	 Synthesize | Hide Details Find similar	**Chemical Name:** 8-hydroxy-3,6-disulfo-naphthalene-1-diazonium-betaine **Reaxys Registry Number:** 3722124 **CAS Registry Number:** 34154-01-3 **Type of Substance:** isocyclic **Molecular Formula:** $C_{10}H_6N_2O_7S_2$ **Linear Structure Formula:** $C_{10}H_6N_2O_7S_2$ **Molecular Weight:** 330.299 **InChI Key:** HZEXCYFNUXRNJF-UHFFFAOYSA-N	3 prep out of 6 reactions.

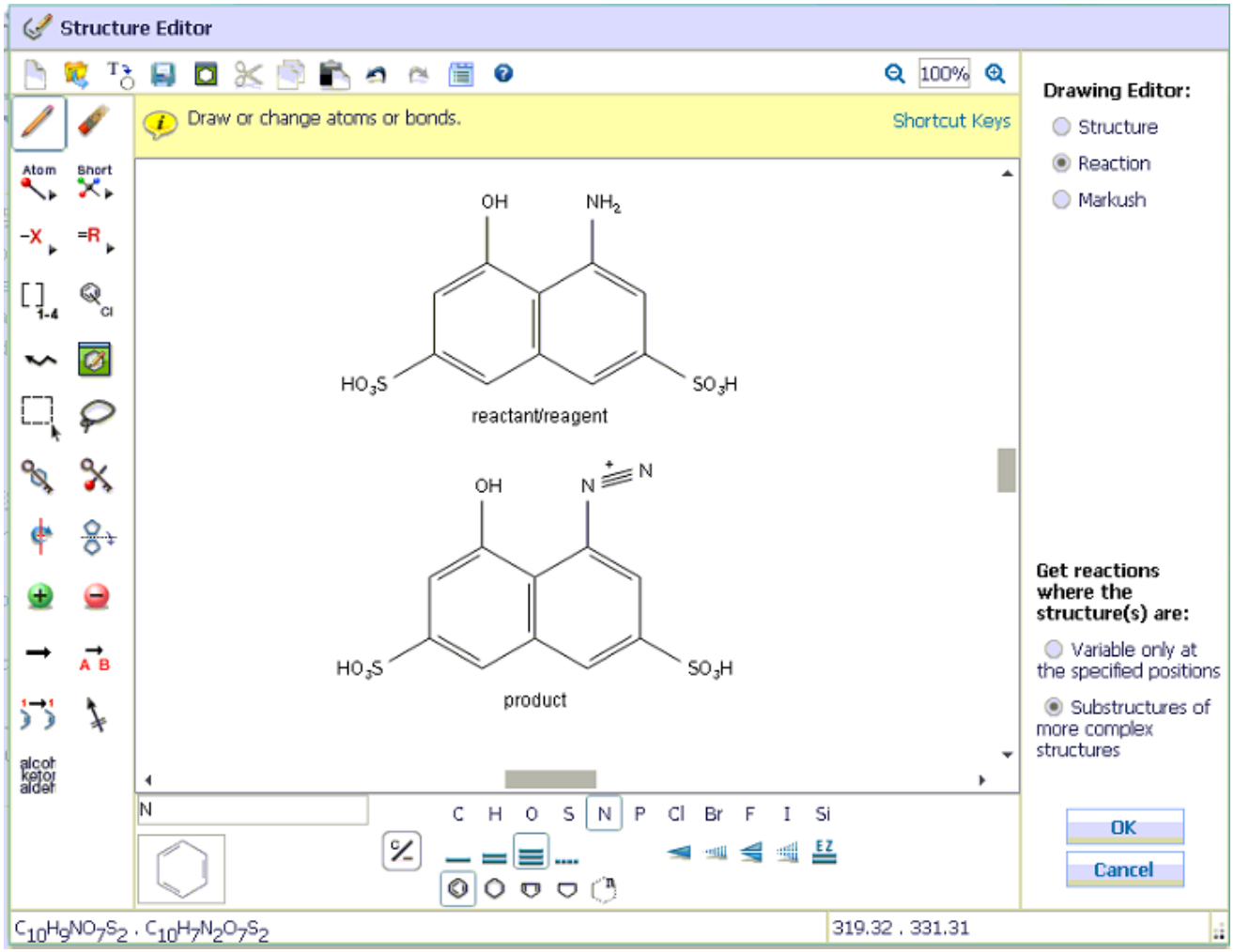
Chemical Names and Synonyms

8-hydroxy-3,6-disulfo-naphthalene-1-diazonium-betaine, 8-Hydroxy-3,6-disulfo-naphthalin-1-diazonium-betain, 8-Diazo-naphthol-(1)-disulfonsaeure-(3.6)

Show 9 results per page — 1 substances out of 6 reactions and 5 citations

Search and Location of References

Scifinder Reaction Search



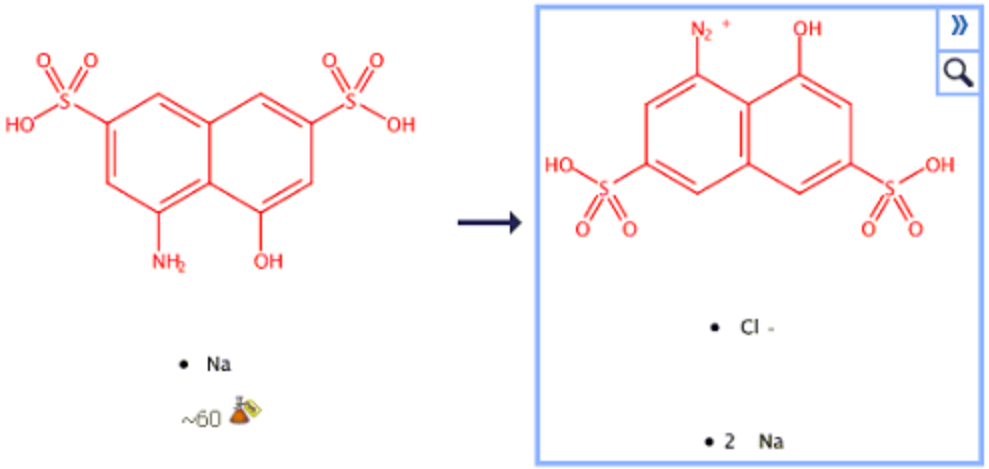
Search and Location of References

Scifinder Reaction Search

1 of 1 Reaction Selected

1. View Reaction Detail Link Similar Reactions

Single Step *Hover over any structure for more options.*



Overview

Steps/Stages

1.1 R:$NaNO_2$, R:HCl, S:H_2O, 0°C; 1 h, 5-8°C

Notes

Reactants: 1, Reagents: 2, Solvents: 1, Steps: 1, Stages: 1,
Most stages in any one step: 1

References

Synthesis of leather dark brown A
Quick View Other Sources
By Song, Nei-li
From Zhongguo Pige, 41(17), 37; 2012

Search and Location of References

Author search
Journal search
Limited by years

References

Liu, T., Wang, Z.-W., Wang, Y.-X. & Xie, Z.-P. (2007*a*). *Acta Cryst.* E**63**, i170.
Liu, T., Wang, Z.-W., Wang, Y.-X. & Xie, Z.-P. (2007*b*). *Acta Cryst.* E**63**, m1887–m1888.
Liu, T., Wang, Z.-W., Wang, Y.-X. & Xie, Z.-P. (2007*c*). *Acta Cryst.* E**63**, m2020–m2021.
Liu, T., Wang, Z.-W., Wang, Y.-X. & Xie, Z.-P. (2007*d*). *Acta Cryst.* E**63**, m2027–m2028.
Liu, T., Wang, Z.-W., Wang, Y.-X. & Xie, Z.-P. (2007*e*). *Acta Cryst.* E**63**, m2080–m2081.
Liu, T., Wang, Z.-W., Wang, Y.-X. & Xie, Z.-P. (2007*f*). *Acta Cryst.* E**63**, m2196–m2197.
Liu, T., Wang, Z.-W., Wang, Y.-X. & Xie, Z.-P. (2007*g*). *Acta Cryst.* E**63**, m2198–m2199.
Liu, T. & Xie, Z.-P. (2007*a*). *Acta Cryst.* E**63**, m1820.
Liu, T. & Xie, Z.-P. (2007*b*). *Acta Cryst.* E**63**, m1908.
Liu, T. & Zhu, J. Y. (2007*a*). *Acta Cryst.* E**63**, m2506–m2507.
Liu, T. & Zhu, J.-Y. (2007*b*). *Acta Cryst.* E**63**, m2592–m2593.
Liu, T. & Zhu, J. Y. (2007*c*). *Acta Cryst.* E**63**, m2659–m2660.
Liu, T. & Zhu, J. Y. (2007*d*). *Acta Cryst.* E**63**, m2775–m2776.
Liu, T. & Zhu, J. Y. (2007*e*). *Acta Cryst.* E**63**, m2809.
Liu, T. & Zhu, J. Y. (2007*f*). *Acta Cryst.* E**63**, m2912.
Liu, T. & Zhu, J.-Y. (2007*g*). *Acta Cryst.* E**63**, m2977–m2978.
Liu, T. & Zhu, J. Y. (2007*h*). *Acta Cryst.* E**63**, m3108.
Liu, T. & Zhu, J. Y. (2007*i*). *Acta Cryst.* E**63**, m3144.
Liu, T. & Zhu, J.-Y. (2007*j*). *Acta Cryst.* E**63**, o3829.
Liu, T. & Zhu, J. Y. (2007*k*). *Acta Cryst.* E**63**, o3830.
Liu, T. & Zhu, J.-Y. (2007*l*). *Acta Cryst.* E**63**, o3860.
Liu, T. & Zhu, J. Y. (2007*m*). *Acta Cryst.* E**63**, o4112.
Liu, T. & Zhu, J. Y. (2007*n*). *Acta Cryst.* E**63**, o4267.
Liu, T. & Zhu, J. Y. (2007*o*). *Acta Cryst.* E**63**, o4441.
Liu, T. & Zhu, J. Y. (2007*p*). *Acta Cryst.* E**63**, o4527.
Liu, T. & Zhu, J. Y. (2007*q*). *Acta Cryst.* E**63**, o4574.
Liu, T. & Zhu, J.-Y. (2007*r*). *Acta Cryst.* E**63**, o4660.
Liu, T. & Zhu, J. Y. (2007*s*). *Acta Cryst.* E**63**, o4874.
Liu, T. & Zhu, J. Y. (2008). *Acta Cryst.* E**64**, *m*28.

addenda and errata

Table 1
Details of articles to be retracted, in order of publication.

Title	Reference	DOI	Refcode
Tetrakis(pyrazine-κN)bis(thiocyanato-κN)manganese(II)	Liu & Xie (2007*a*)	10.1107/S1600536807026852	EDUMAS
(Dihydroxyglyoxime-κ^2N,N')bis(1,10-phenanthroline-κ^2N,N')copper(II) dinitrate dihydrate	Liu, Wang, Wang & Xie (2007*b*)	10.1107/S1600536807028255	EDUVAB
Tetrakis(pyrazine-κN)bis(thiocyanato-κN)zinc(II)	Liu & Xie (2007*b*)	10.1107/S1600536807028735	RIGQAA
Tetrakis(μ-2-pyridyloxyacetato)bis[(1,10-phenanthroline)(2-pyridyloxyacetato)-lanthanum(III)]	Liu, Wang, Wang & Xie (2007*c*)	10.1107/S1600536807030917	UDUMIQ
Polymeric $KNOF_2$	Liu Wang, Wang & Xie (2007*a*)	10.1107/S1600536807027195	ICSD 240891
(Dihydroxyglyoxime-κ^2N,N')bis(1,10-phenanthroline-κ^2N,N')cobalt(II) dinitrate dihydrate	Liu, Wang, Wang & Xie (2007*d*)	10.1107/S1600536807031224	WIHJED
Tetrakis(μ-2-pyridyloxyacetato)bis[(1,10-phenanthroline)(2-pyridyloxyacetato)-praseodymium(III)]	Liu, Wang, Wang & Xie (2007*e*)	10.1107/S1600536807032679	WIHQEK
Tetrakis[μ-(2-pyridyloxy)acetato-$\kappa^2O:O'$]bis[(1,10-phenanthroline-κ^2N,N')-[(2-pyridyloxy)acetato-κO]neodymium(III)]	Liu, Wang, Wang & Xie (2007*f*)	10.1107/S1600536807035349	TIGDAP
(Dihydroxyglyoxime-κ^2N,N')bis(1,10-phenanthroline-κ^2N,N')manganese(II) dinitrate dihydrate	Liu, Wang, Wang & Xie (2007*g*)	10.1107/S1600536807035076	TIGDET
2-Amino-3,5-dinitrobenzoic acid–ammonia (1/1)	Liu & Zhu (2007*j*)	10.1107/S1600536807040068	KIKQAX
2-Hydroxy-3,5-dinitrobenzamide monohydrate	Liu & Zhu (2007*k*)	10.1107/S1600536807039712	KIKQEB
2-(1-Hydroxy-2-pyridyl)acetamide monohydrate	Liu & Zhu (2007*l*)	10.1107/S1600536807040652	CIKQOD
Bis(2,2′-bipyridine-κN,N′)bis(thiocyanato-κN)iron(II)	Liu & Zhu (2007*a*)	10.1107/S1600536807043486	XIFXOA
catena-Poly[hexakis(μ_2-anilinoacetamide)bis(1,10-phenanthroline)disamarium(III)]	Liu & Zhu (2007*b*)	10.1107/S1600536807045485	XILNAI
3-Hydroxy-2,4,6-trinitropyridine monohydrate	Liu & Zhu (2007*m*)	10.1107/S1600536807045230	PILNOO
catena-Poly[hexakis(μ_2-anilinoacetamide)bis(1,10-phenanthroline)-dipraseodymium(III)]	Liu & Zhu (2007*c*)	10.1107/S1600536807047733	SILZET
catena-Poly[[tetra-μ-anilinoacetamidato-bis(1,10-phenanthroline)dicerium(III)]-di-μ-anilinoacetamidato]	Liu & Zhu (2007*d*)	10.1107/S1600536807050969	GIMZOS
Tetrakis(pyridine-κN)bis(thiocyanato-κN)chromium(II)	Liu & Zhu (2007*e*)	10.1107/S1600536807051756	WINFAB
2-Ammonio-3-carboxy-5-nitrobenzoate monohydrate	Liu & Zhu (2007*n*)	10.1107/S1600536807048477	GINFEP
2-(Benzoylhydrazinocarbonyl)benzoic acid	Liu & Zhu (2007*o*)	10.1107/S160053680705204X	TINZIA
Tetrakis(pyridine-κN)bis(thiocyanato-κN)vanadium(II)	Liu & Zhu (2007*f*)	10.1107/S1600536807054529	HIPZIQ
catena-Poly[[(nitrato-κO)(1,10-phenanthroline-κ^2N,N')nickel(II)]-μ-acetamido-$\kappa^2O:N$]	Liu & Zhu (2007*g*)	10.1107/S1600536807056504	XIRGIP
catena-Poly[[(nitrato-κO)(1,10-phenanthroline-κ^2N,N')copper(II)]-μ-acetamido-$\kappa^2O:N$]	Liu & Zhu (2007*h*)	10.1107/S1600536807059077	HIQROP
catena-Poly[[(nitrato-κO)(1,10-phenanthroline-κ^2N,N')cobalt(II)]-μ-acetamidato-$\kappa^2O:N$]	Liu & Zhu (2007*i*)	10.1107/S1600536807060631	YIQMER
N′-Benzoyl-4-nitronicotinohydrazide	Liu & Zhu (2007*p*)	10.1107/S1600536807053068	CIPVON
N′-(3-Nitro-4-pyridylcarbonyl)pyridine-4-carbohydrazide	Liu & Zhu (2007*q*)	10.1107/S1600536807054876	RIRWEV

Academic Anticorruption in China's Universities

Haohao Li, Xiaoqing He, Shuai Wang[1]
University of Shanghai for Science & Technology, 200093 Shanghai, P. R. of China

Case Two: In December 2009, an editorial in *Acta Crystallographica* Section E (Structure Reports) by the scholars William T. A. Harrison, Jim Simpson and Matthias Weil announced that it had withdrawn 41 articles by Dr. Zhong and 29 articles by Professor Liu from China's Jinggangshan University. The reason was that, from 2006 to 2008, the authors of these reports just modified the raw intensity data and published a series of articles with data frauds. The editorial said that the investigation was continuing and the number of papers with problems might continue to rise[3]. After the incident, according to a press survey by www.chinanews.com.cn, Zhong and Liu were indeed teachers of Jinggangshan University, but not doctors or professors, only general lecturers. What's more surprising is that Liu studies engineering, while crystal structure belongs to chemistry. It is inconceivable that Liu could have published as many as 29 papers which did not belong to his research field.

Dr. Haohao Li, Professor of Business School;
Xiaoqing He and Shuai Wang, Postgraduate students Business School

Published online 12 January 2010 | *Nature* **463**, 142-143 (2010) | doi:10.1038/463142a

News

Publish or perish in China

The pressure to rack up publications in high-impact journals could encourage misconduct, some say.

Jane Qiu

The latest in a string of high-profile academic fraud cases in China underscores the problems of an academic-evaluation system that places disproportionate emphasis on publications, critics say. Editors at the UK-based journal *Acta Crystallographica Section E* last month retracted 70 published crystal structures that they allege are fabrications by researchers at Jinggangshan University in Jiangxi province. Further retractions, the editors say, are likely.

Under pressure: one-third of researchers surveyed in China admit to plagiarism, falsification or fabrication of data.

K. Brofsky/Getty

Search and location of references

Journal **impact factor** is the ratio of number of citations in one year to papers published during the 2 preceding years

For Acta Cryst. A, the 2009 calculation was as follows:
Total number of citations in 2009 to items published in 2007 and 2008 (A) = 6091
Total number of papers published in 2007-2008 (B) = 122
Impact factor = A/B = 6091/122 = 49.9

At 49.9, the **impact factor** of Acta Cryst. A is the second highest for a scientific journal for 2009, with only CA: A Cancer Journal for Clinicians scoring higher.

Acta Cryst. A covers diffraction physics and the theory of crystallographic structure determination by diffraction methods using X-rays, neutrons and electrons.

Search and location of references

Thomson Reuters' Journal Citation Report for 2009:
Acta Cryst. A: Foundations of Crystallography I.F. = 49.9

Primary cause is a single feature article by George Sheldrick, A short history of **SHELX**, published in a special issue to celebrate 60 years of journal and IUCr (Acta Cryst. (2008). A64, 112-122; doi:10.1107/S010876730704393O).

The article has provided the crystallographic community with a citable reference when one or more of the programs, amongst the most widely used in structure determination, are employed in the course of a crystal structure study. 43,000 articles published in IUCr journals alone have referenced SHELX programs.

The journal impact factor is normal at 2.0-2.5.

Thank you