

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 Although many of the old databases are no longer active, a number of them continue to run and even prosper.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."

The screenshot shows the ZINC database website interface. At the top, there is a navigation bar with links for UCSF, About UCSF, Search UCSF, and UCSF Medical Center. The main header features the ZINC logo and the text "A free database for virtual screening" and "is not commercial". Below the header, there are navigation buttons for Home, Subsets, HELP!, and Mailing Lists. The main content area is divided into several sections:

- Structure:** This section includes a "Get SMILES" button, a "from editor" dropdown, and an "Update Editor" button. Below these is a text input field containing the SMILES string: CCCCCCCCCCCC(C)C4CC3CCCC(C2CCCC2)C1=CC=CC=C1C=C4. To the right of the input field is a chemical structure editor toolbar with various icons for drawing and editing. Below the toolbar is a vertical list of element symbols (C, N, O, S, P, F, Cl, Br, I, X) and a large empty box for drawing the structure. Below the structure box are buttons for "SMILES", "Beautify", and "Clear".
- Similarity Value:** This section contains a list of radio buttons for selecting a similarity threshold: 99% similar, 90% similar (selected), 80% similar, 70% similar, 60% similar, 50% similar, and Substructure Search. There is also a "Use Java Editor" link.
- Properties:** This section includes a "Subset:" dropdown menu set to "Select Preset". Below this are several rows of input fields for filtering properties: Net Charge, xLogP, Rotatable Bonds, #H-Donors, #H-Acceptors, and Polar Desolvation.
- Run Query:** A button labeled "Run Query" is located at the top right of the main content area.
- Molecule widget by:** A logo for "molinspiration cheminformatics" is displayed.
- Search Engine by:** A logo for "Powered by ChemAxon" is displayed.

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

Frameset Page
http://www.orgsyn.org/orgsyn/default.asp?dbname=orgsyn&dataaction=search&metadata_directive=blind_gui&f Google

Organic Syntheses
Print Search Methods History Queries Hit Lists Help
New Query Preferences

1 of 147 Procedures Found <<< Rec # >>> Return to List Printer Friendly

Organic Syntheses, Vol. 82, p. 188 (2005); Coll. Vol. 11, p.1 (2009).

CATALYTIC REDUCTION OF AMIDES TO AMINES WITH HYDROSILANES USING A TRIRUTHENIUM CARBONYL CLUSTER AS THE CATALYST

A. C1=CC=C2C=CC=CC12 + $\text{Ru}_3(\text{CO})_{12}$ $\xrightarrow[\text{36 h}]{\text{heptane reflux}}$ C1=CC=C2C=CC=CC12Ru3(CO)7 (1)

B. CN(C)C(=O)Cc1ccccc1 (2) $\xrightarrow[\text{rt, 75 min}]{\text{PhMe}_2\text{SiH (2.2 eq. to 2)}}$ CN(C)Cc1ccccc1 (3)
Reaction conditions: 1 (1 mol% to 2) in THP, rt, 30 min.

Submitted by Yukihiko Motoyama, Chikara Itonaga, Toshiki Ishida, Mikihiro Takasaki, and Hideo Nagashima¹.
Checked by Joshua G. Pierce and Peter Wipf.

1. Procedure

Caution! Since reaction intermediates are potentially unstable to air and moisture, preparation of the ruthenium cluster should be carried out under an inert gas atmosphere.

A. ($\mu_3, \eta^3: \eta^3: \eta^5$ -Acenaphthylene) $\text{Ru}_3(\text{CO})_7$ (1). A two-necked, 200-mL round-bottomed flask is equipped with a magnetic stir bar, a reflux condenser, and a stopcock. The top of the condenser is fitted with a three-way stopcock, of which one way is connected to an argon flow line and the other is connected to a vacuum line. The apparatus is flame-dried while under vacuum, and allowed to cool to room temperature under an argon purge. The stopcock is removed, and the flask is charged with $\text{Ru}_3(\text{CO})_{12}$ (Note 1) (639 mg, 1 mmol), acenaphthylene (Note 2) (182 mg, 1.2 mmol), and heptane (Note 3) (75 mL) under an argon stream. The stopcock is again placed on the flask, and the mixture is heated under reflux. At the initial stage, the solution contains a yellow supernatant and an undissolved orange solid of $\text{Ru}_3(\text{CO})_{12}$. The color of the solution gradually turns to red and all of the solid materials are dissolved after 30 min (Note 4). After 36 h, a dark red solution and some red precipitates have formed. The mixture is cooled to room temperature under an argon atmosphere. From this stage, subsequent work can be performed in the air. The precipitates are collected on a glass filter, washed with *n*-hexane (30 mL), and dissolved in dichloromethane (20 mL)

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.*



ACS
Chemistry for Life®

Scopus¹⁰

PubMed.gov



REAXYS

I U P A C



THE CAMBRIDGE
CRYSTALLOGRAPHIC
DATA CENTRE

ELSEVIER

 **WILEY**
Publishers Since 1807

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 REGISTRYSM

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 REGISTRYSM

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 $(8 \times 1 + 1 \times 2 + 2 \times 3 + 3 \times 4 + 7 \times 5 + 7 \times 6) \bmod 10 = 105 \bmod 10 = 5$.

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

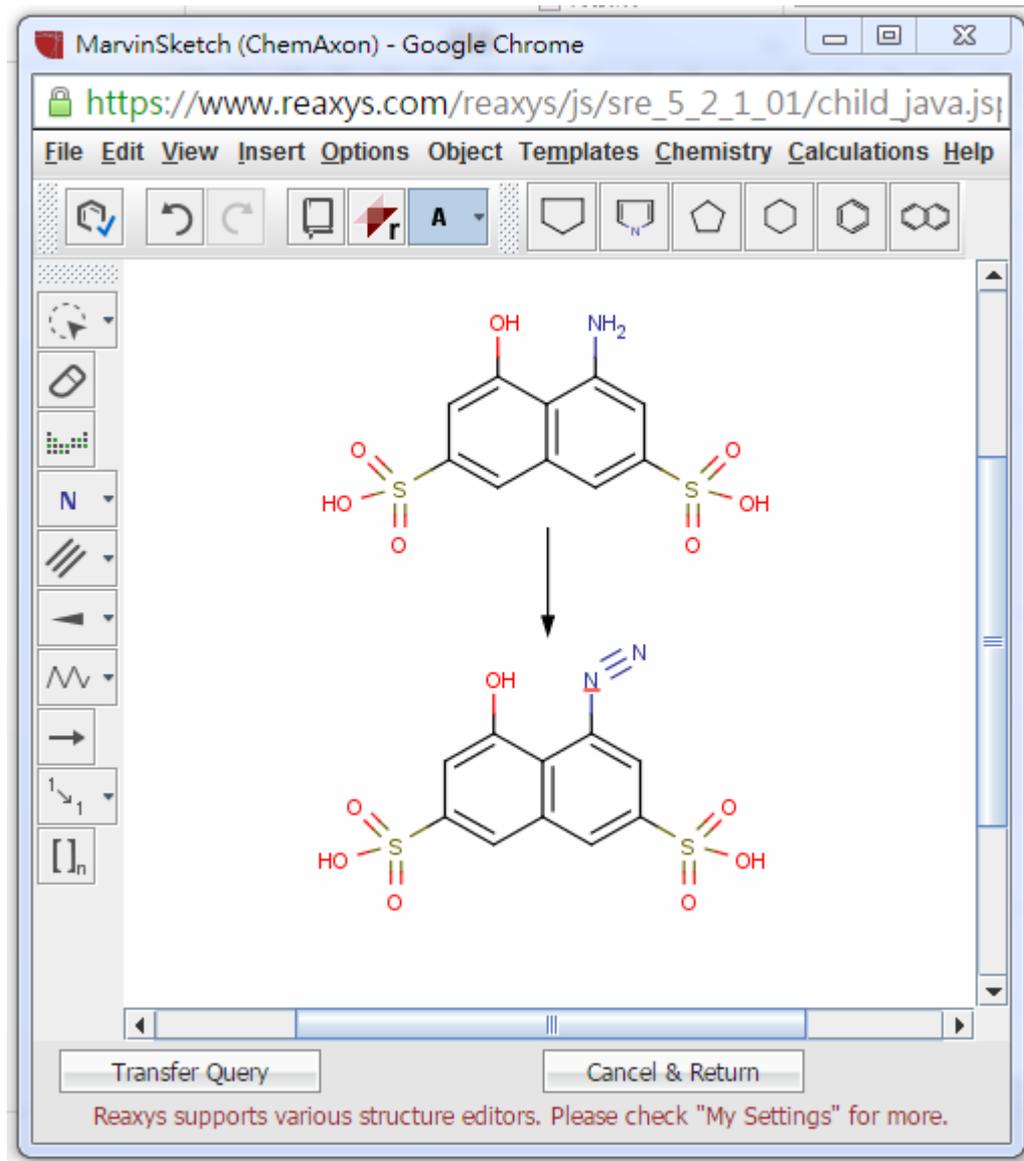
Nodes – linkers

Atoms – bonds

Parallel processing

Reaxys
Reaction
Search

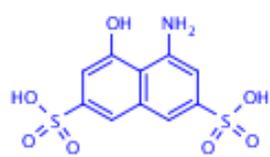
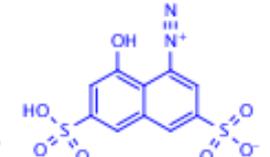
Search and Location of References



Reaxys Reaction Search

Search and Location of References

1 reactions out of 2 substances and 1 citations

Reactions	Substances (Grid)	Substances (Report)	Citations
<p>Limit to Exclude Export Print Zoom in Zoom out Hide Sort by <input type="text" value="Reaxys-Ranking"/></p>			
Yield	Conditions	References	
   Synthesize Find similar	   Synthesize Find similar	Rx-ID: 339170 Find similar reactions	
	Diazotization;	Cassella and Co. Patent: DE67062 ; Fortschr. Teerfarbenfabr. Verw. Industriezweige, vol. 3, p. 466 Full Text Show Details	

Show results per page

1 reactions out of 2 substances and 1 citations

Reaxys Reaction Search

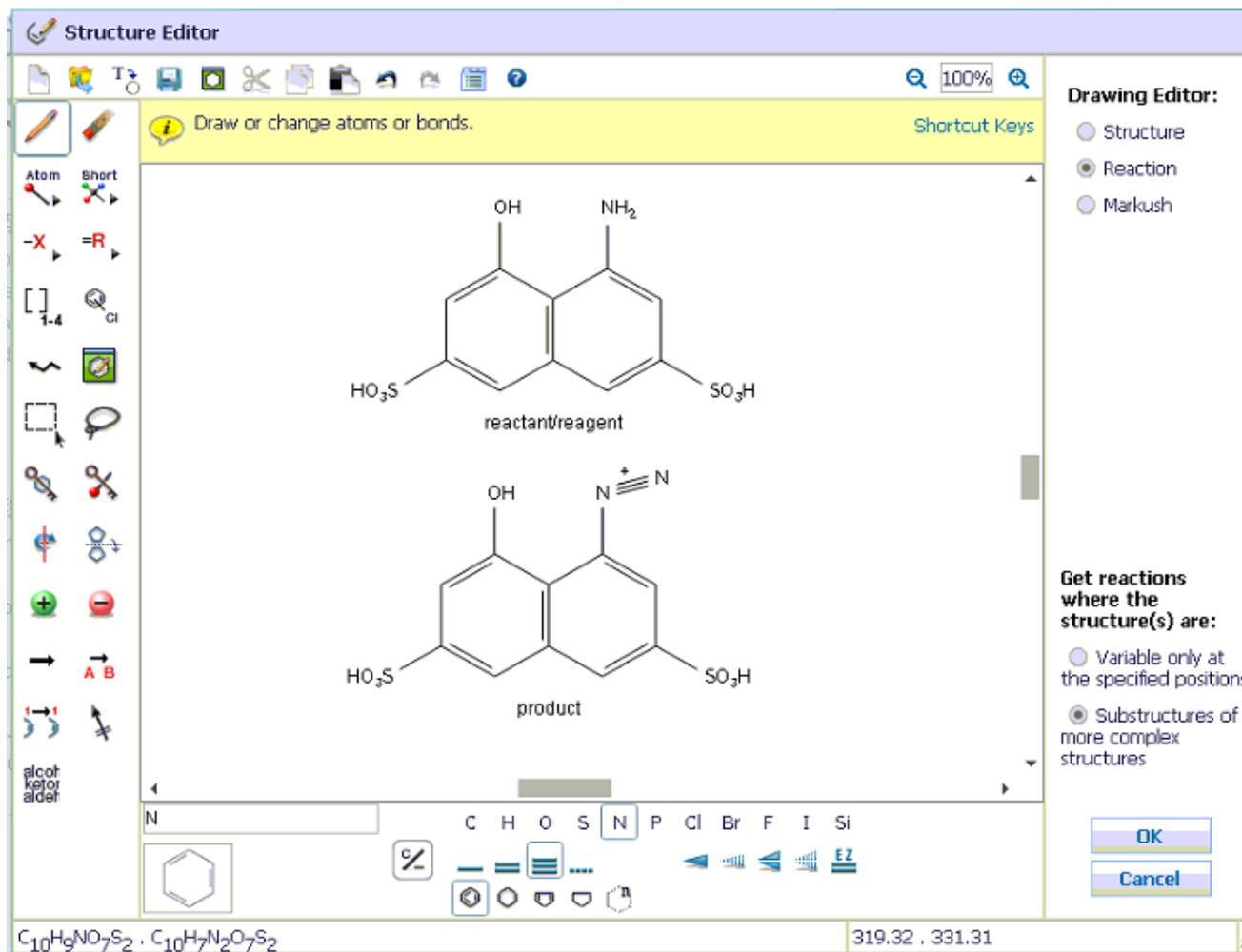
Search and Location of References

1 substances out of 6 reactions and 5 citations

Reactions	Substances (Grid)	Substances (Report)	Citations
Limit to <input type="checkbox"/> Exclude <input type="checkbox"/> Export Print Zoom in Zoom out Hide Sort by No of References <input type="text"/>			
Structure	Structure/Compound Data		N° of preparations All Preps All Reactions
<input type="checkbox"/> 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 ₆ N ₂ O ₇ S ₂ Linear Structure Formula: C ₁₀ H ₆ N ₂ O ₇ S 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 <input type="text" value="9"/> results per page			1 substances out of 6 reactions and 5 citations

Search and Location of References

Scifinder Reaction Search



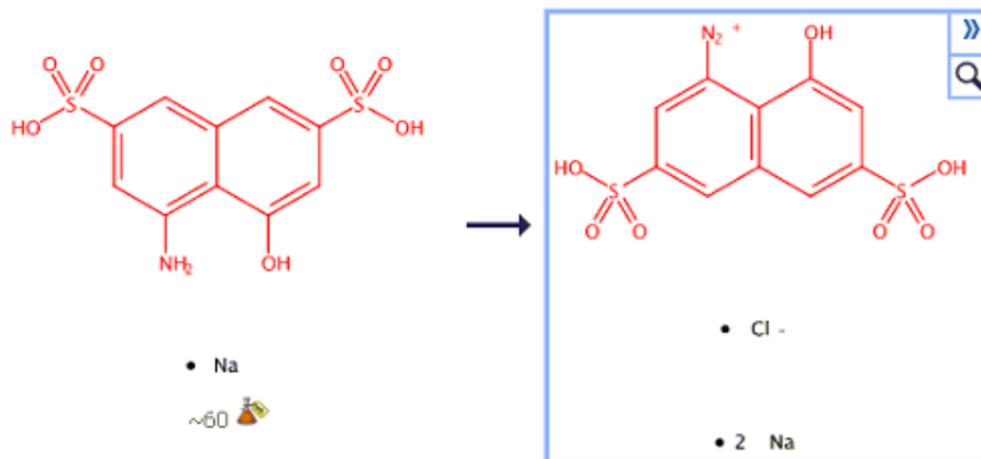
Search and Location of References

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:NaNO₂, R:HCl, S:H₂O, 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. (2007a). *Acta Cryst. E63*, i170.
- Liu, T., Wang, Z.-W., Wang, Y.-X. & Xie, Z.-P. (2007b). *Acta Cryst. E63*, m1887–m1888.
- Liu, T., Wang, Z.-W., Wang, Y.-X. & Xie, Z.-P. (2007c). *Acta Cryst. E63*, m2020–m2021.
- Liu, T., Wang, Z.-W., Wang, Y.-X. & Xie, Z.-P. (2007d). *Acta Cryst. E63*, m2027–m2028.
- Liu, T., Wang, Z.-W., Wang, Y.-X. & Xie, Z.-P. (2007e). *Acta Cryst. E63*, m2080–m2081.
- Liu, T., Wang, Z.-W., Wang, Y.-X. & Xie, Z.-P. (2007f). *Acta Cryst. E63*, m2196–m2197.
- Liu, T., Wang, Z.-W., Wang, Y.-X. & Xie, Z.-P. (2007g). *Acta Cryst. E63*, m2198–m2199.
- Liu, T. & Xie, Z.-P. (2007a). *Acta Cryst. E63*, m1820.
- Liu, T. & Xie, Z.-P. (2007b). *Acta Cryst. E63*, m1908.
- Liu, T. & Zhu, J. Y. (2007a). *Acta Cryst. E63*, m2506–m2507.
- Liu, T. & Zhu, J.-Y. (2007b). *Acta Cryst. E63*, m2592–m2593.
- Liu, T. & Zhu, J. Y. (2007c). *Acta Cryst. E63*, m2659–m2660.
- Liu, T. & Zhu, J. Y. (2007d). *Acta Cryst. E63*, m2775–m2776.
- Liu, T. & Zhu, J. Y. (2007e). *Acta Cryst. E63*, m2809.
- Liu, T. & Zhu, J. Y. (2007f). *Acta Cryst. E63*, m2912.
- Liu, T. & Zhu, J.-Y. (2007g). *Acta Cryst. E63*, m2977–m2978.
- Liu, T. & Zhu, J. Y. (2007h). *Acta Cryst. E63*, m3108.
- Liu, T. & Zhu, J. Y. (2007i). *Acta Cryst. E63*, m3144.
- Liu, T. & Zhu, J.-Y. (2007j). *Acta Cryst. E63*, o3829.
- Liu, T. & Zhu, J. Y. (2007k). *Acta Cryst. E63*, o3830.
- Liu, T. & Zhu, J.-Y. (2007l). *Acta Cryst. E63*, o3860.
- Liu, T. & Zhu, J. Y. (2007m). *Acta Cryst. E63*, o4112.
- Liu, T. & Zhu, J. Y. (2007n). *Acta Cryst. E63*, o4267.
- Liu, T. & Zhu, J. Y. (2007o). *Acta Cryst. E63*, o4441.
- Liu, T. & Zhu, J. Y. (2007p). *Acta Cryst. E63*, o4527.
- Liu, T. & Zhu, J. Y. (2007q). *Acta Cryst. E63*, o4574.
- Liu, T. & Zhu, J.-Y. (2007r). *Acta Cryst. E63*, o4660.
- Liu, T. & Zhu, J. Y. (2007s). *Acta Cryst. E63*, o4874.
- Liu, T. & Zhu, J. Y. (2008). *Acta Cryst. E64*, m28.

addenda and errata

Table 1

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

Title	Reference	DOI	Refcode
<i>Tetrakis(pyrazine-κN)bis(thiocyanato-κN)manganese(II)</i> <i>(Dihydroxyglyoxime-κ^2N,N')bis(1,10-phenanthroline-κ^2N,N')copper(II) dinitrate dihydrate</i>	Liu & Xie (2007a) Liu, Wang, Wang & Xie (2007b)	10.1107/S1600536807026852 10.1107/S1600536807028255	EDUMAS EDUVAB
<i>Tetrakis(pyrazine-κN)bis(thiocyanato-κN)zinc(II)</i> <i>Tetrakis(μ-2-pyridyloxyacetato)bis[(1,10-phenanthroline)(2-pyridyloxyacetato)-lanthanum(III)]</i> <i>Polymeric KNOF₂</i> <i>(Dihydroxyglyoxime-κ^2N,N')bis(1,10-phenanthroline-κ^2N,N')cobalt(II) dinitrate dihydrate</i>	Liu & Xie (2007b) Liu, Wang, Wang & Xie (2007c) Liu Wang, Wang & Xie (2007a) Liu, Wang, Wang & Xie (2007d)	10.1107/S1600536807028735 10.1107/S1600536807030917 10.1107/S1600536807027195 10.1107/S1600536807031224	RIGQAA UDUMIQ ICSD 240891 WIHJED
<i>Tetrakis(μ-2-pyridyloxyacetato)bis[(1,10-phenanthroline)(2-pyridyloxyacetato)-praseodymium(III)]</i> <i>Tetrakis [μ-(2-pyridyloxy)acetato-κ^2O:O']bis[(1,10-phenanthroline-κ^2N,N')-(2-pyridyloxy)acetato-κO]neodymium(III)]</i> <i>(Dihydroxyglyoxime-κ^2N,N')bis(1,10-phenanthroline-κ^2N,N')manganese(II) dinitrate dihydrate</i>	Liu, Wang, Wang & Xie (2007e) Liu, Wang, Wang & Xie (2007f) Liu, Wang, Wang & Xie (2007g)	10.1107/S1600536807032679 10.1107/S1600536807035349 10.1107/S1600536807035076	WIHQEK TIGDAP TIGDET
<i>2-Amino-3,5-dinitrobenzoic acid-ammonia (1/1)</i> <i>2-Hydroxy-3,5-dinitrobenzamide monohydrate</i> <i>2-(1-Hydroxy-2-pyridyl)acetamide monohydrate</i> <i>Bis(2,2'-bipyridine-κN,N')bis(thiocyanato-κN)iron(II)</i> <i>catena-Poly[hexakis(μ-2-anilinoacetamide)bis(1,10-phenanthroline)disamarium(III)]</i> <i>3-Hydroxy-2,4,6-trinitropyridine monohydrate</i> <i>catena-Poly[hexakis(μ-2-anilinoacetamide)bis(1,10-phenanthroline)-dipraseodymium(III)]</i> <i>catena-Poly[[tetra-μ-anilinoacetamidato-bis(1,10-phenanthroline)dicerium(III)]-di-μ-anilinoacetamidato]</i>	Liu & Zhu (2007j) Liu & Zhu (2007k) Liu & Zhu (2007l) Liu & Zhu (2007a) Liu & Zhu (2007b) Liu & Zhu (2007m) Liu & Zhu (2007c) Liu & Zhu (2007d)	10.1107/S1600536807040068 10.1107/S1600536807039712 10.1107/S1600536807040652 10.1107/S1600536807043486 10.1107/S1600536807045485 10.1107/S1600536807045230 10.1107/S1600536807047733 10.1107/S1600536807050969	KIKQAX KIKQEB CIKQOD XIFXOA XILNAI PILNOO SILZET GIMZOS
<i>Tetrakis(pyridine-κN)bis(thiocyanato-κN)chromium(II)</i> <i>2-Ammonio-3-carboxy-5-nitrobenzoate monohydrate</i> <i>2-(Benzoylhydrazinocarbonyl)benzoic acid</i> <i>Tetrakis(pyridine-κN)bis(thiocyanato-κN)vanadium(II)</i>	Liu & Zhu (2007e) Liu & Zhu (2007n) Liu & Zhu (2007o) Liu & Zhu (2007f)	10.1107/S1600536807051756 10.1107/S1600536807048477 10.1107/S160053680705204X 10.1107/S1600536807054529	WINFAB GINFEP TINZIA HIPZIQ
<i>catena-Poly[[(nitrato-κO)(1,10-phenanthroline-κ^2N,N')nickel(II)]-μ-acetamido-κ^2O:N]</i>	Liu & Zhu (2007g)	10.1107/S1600536807056504	XIRGIP
<i>catena-Poly[[(nitrato-κO)(1,10-phenanthroline-κ^2N,N')copper(II)]-μ-acetamido-κ^2O:N]</i>	Liu & Zhu (2007h)	10.1107/S1600536807059077	HIQROP
<i>catena-Poly[[(nitrato-κO)(1,10-phenanthroline-κ^2N,N')cobalt(II)]-μ-acetamidato-κ^2O:N]</i>	Liu & Zhu (2007i)	10.1107/S1600536807060631	YIQMER
<i>N-Benzoyl-4-nitronicotinohydrazide</i> <i>N-(3-Nitro-4-pyridylcarbonyl)pyridine-4-carbohydrazide</i>	Liu & Zhu (2007p) Liu & Zhu (2007q)	10.1107/S1600536807053068 10.1107/S1600536807054876	CIPVON KRWVEV

Academic Anticorruption in China's Universities

Haohao Li, Xiaoqing He, Shuai Wang¹

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 Jingtangshan 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 rise³. After the incident, according to a press survey by www.chinanews.com.cn, Zhong and Liu were indeed teachers of Jingtangshan 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 24

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 Jingtangshan 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/S0108767307043930).

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