

# Marvin 5.4 – A new generation of structure indexing at Elsevier

---

Dr. Michael Maier, Dr. Heike Nau, Elsevier



# Agenda

- Elsevier: Reaxys database
- Compound classes
- Structure requirements
- Marvin 5.4
  - *Decision process*
  - *Development of functionality*
  - *Integration into our software*
- Experience in use

# What is Reaxys?

Reaxys is the web-based workflow solution for research chemists.

## • Chemistry

- ~200 years of organic, organometallic and inorganic chemistry
- Experimentally validated structural, chemical, and physical data from Journal articles and patents
- Based on CrossFire Beilstein, CrossFire Gmelin and Patent Chemistry Database

## • Intuitive

- Easy to use, appealing web-interface, available anytime and anywhere without limits and without installation needs
- Allowing chemists to query and filter by property data, design the optimum synthesis route, and view multi-step reactions.
- Created from scientist for scientist

## • Workflow Solution

- Supporting scientists in life science and chemistry in industry and academia in their daily work
- Providing decision support in finding relevant and validated chemical information



J. Am. Chem. Soc.

Heterocycles

Applied Spectroscopy

J. Of Mass Spectrometry

Biomolecular Spectroscopy

Biochemistry

Biochemistry

Chemical Compounds

Inorganic

Acta

Angew.

Chemm. Cor

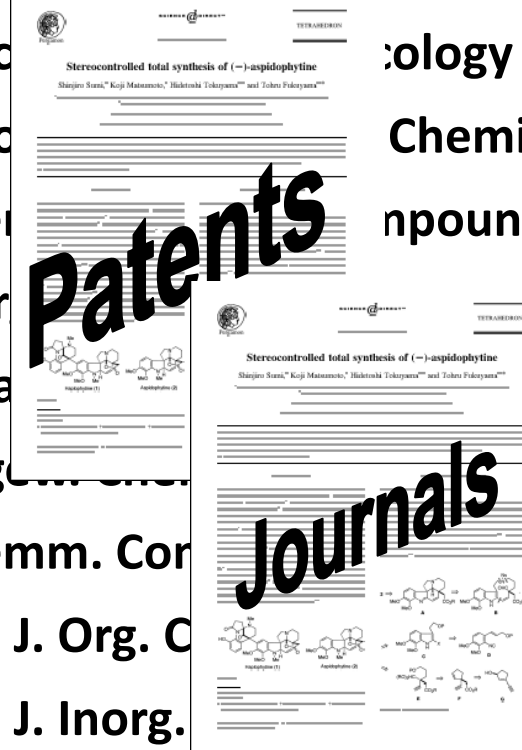
Eur. J. Org. C

Eur. J. Inorg.

Drug Discovery Today

J. Mol. Catal.

J. Solid State Chem.



# Reaxys extracts valuable chemical structures, properties and reaction details from documents...

**reaxys**

Query Results Synthesis Plans History My Alerts My Settings Help Logout

Query: 78 reactions → 35 reactions (filtered by Journal Title)

Filter by: Yield, Record Type, Reagent/Catalyst, Solvent, Reaction Type, No. of Steps, Document Type, Authors, Patent Assignee, Journal Title

35 reactions out of 15 citations

Reactions Citations

Limit to Selection Output Sort by: Reaxys-Ranking Hide Details

Yield Conditions References

Structure: CC(C)C(C)C(=O)N1C=CC=C1 (1-(2,2-diethylbutanoyl)indole)

35% Chromat. With  $\text{SnCl}_4 \cdot 2\text{H}_2\text{O}$ ;  $\text{RuCl}_3 \cdot 3\text{H}_2\text{O} / \text{PPH}_3$  in dioxane;  $\text{H}_2\text{O}$ ;  $T = 180^\circ\text{C}$ ; 20 h;

Authors: Cho, Chan Sik; Kim, Jin Hwang; Kim, Tae-Jeong; Shim, Sang Chul

Tetrahedron, 2001, vol. 57, # 16, p. 3321-3330

Structure: CC(C)C(C)C(=O)N1C=CC=C1

Chemical Name: 1-(2,2-diethylbutanoyl)indole

N° of preparations: 1 prep out of 19 reactions.

Available Data: Identification Spectra (3)

N° of ref.: 2

Boiling Point:

Structure/Compound Data

Reaxys Registry Number: 8318639

CAS Registry Number: 242805-86-3

Chemical Name: 1-(2,2-diethylbutanoyl)indole

Type of Substance: heterocyclic

Molecular Formula:  $\text{C}_{16}\text{H}_{23}\text{NO}$

Linear Structure Formula:  $\text{C}_{16}\text{H}_{23}\text{NO}$

Molecular Weight: 243.349

InChI Key: CBOJNJZOMZAK-UHFFFAOYSA-N

Physical Data, Spectroscopic Data, Bioactivity, Natural Product, Document Type, Authors, Patent Assignee, Journal Title, Publication Year

1H NMR 300 MHz

1H NMR Spectroscopy (2)

Description	Nucleus	Solvents	Frequency	Reference
Spectrum	1H	CDCl3	300MHz	Fukuda, Tsutomu; Maeda, Ryoichi; Iwao, Masatomo Tetrahedron, 1999, vol. 55, # 30, p. 9151-9162 Title/Abstract Full Text View citing articles

TETRAHEDRON

1-(1999) 9151-9162

TETRAHEDRON

TETRAHEDRON

Directed C-7 Alkylation of 1-(2,2-Diethylbutanoyl)indoles

Tsutomu Fukuda, Ryoichi Maeda, and Masatomo Iwao

Department of Applied Chemistry, Faculty of Engineering, Nagasaki University, 1-8-1 Sakamoto-2, Nagasaki 852-8581, Japan

Received 16 April 1999; accepted 26 May 1999

Abstract: Introduction of a nucleophile-derived C-7 alkylation... (text continues)

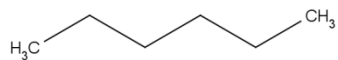
Some biologically significant natural products and synthetic drugs comprise 7-substituted indole moieties as a key structural unit. For the synthesis of 7-substituted indoles, de novo ring construction of the indole nucleus from appropriately substituted benzene precursors has been utilized as a major approach. This includes Longmire-Bassett's sequence,<sup>1</sup> Isakogawa's<sup>2</sup> method,<sup>3</sup> and Kambe-Sakamoto's<sup>4</sup> procedures. Another ring construction route, benzene precursors is also useful, especially for the synthesis of complex natural products such as tryptamine.<sup>5</sup> Compared to these ring construction methods, Tsuchida<sup>6</sup> and Iwao<sup>7</sup> utilized 1-protected indoles (2,3-dihydroindoles) as stable equivalents for 7-selective alkylation and bromination, respectively. Reported general benzene-lithium exchange route to the benzimidazole ring-substituted indole including 7-substituted ones.<sup>8</sup> This approach, however, requires the synthesis of benzimidazole from efficient bromides. Nucleophilic alkylation<sup>9</sup> and lithiation<sup>10</sup> systems using  $\text{Li}^+$ -indole- $\text{Cr}(\text{CO})_5$  complexes have also been reported. Some of these procedures are lengthy and tedious for the synthesis of rather simple 7-substituted indole derivatives, but directed indole ring construction is more straightforward.

It will be shown that the lithiation of 7-substituted indoles occurs at C-7 preferentially.<sup>14</sup> Although a quantity due to much higher thermodynamic stability of C-2 proton compared to C-7 situation. This is if the inducting moiety in the directing group were fixed toward H-7. In this respect, however, results by complex indole lithiation using  $\text{C}(\text{CF}_3)_3$  (Scheme 1).

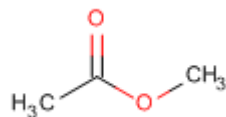
Scheme 1

1-(1999) 9151-9162

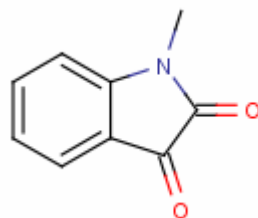
# Reaxys covers a broad variety of compound classes



**n-hexane**

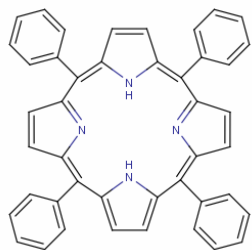


**methylformiate**

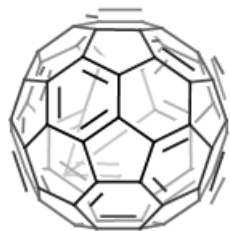


**indolin-2,3-dione**

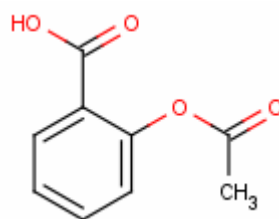
- Simple molecules
- Bulk chemicals
- Heterocycles
- Tetrapyrroles
- Fullerenes
- Drugs
- Pigments
- Agrochemicals
- Natural products, steroids
- ...



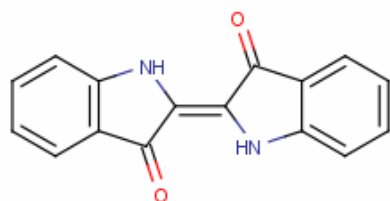
**porphyrin**



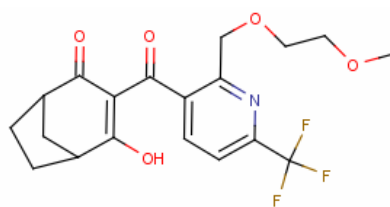
**C60 fullerene**



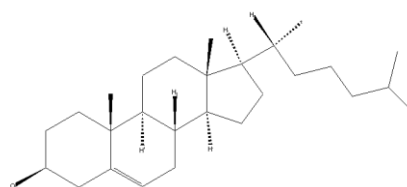
**aspirin**



**indigo**



**bicyclopyron**

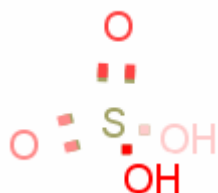


**cholesterol**

# Reaxys covers a broad variety of compound classes ... also in the Inorganic area



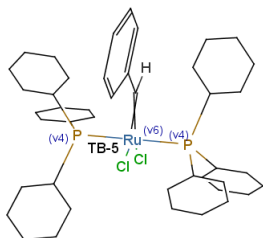
hydrogen



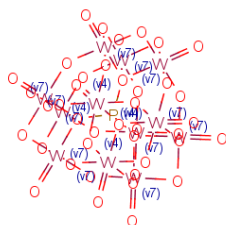
sulfuric acid



hexacyanoferrate



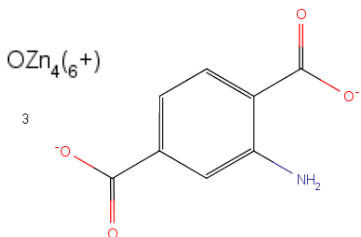
Grubbs catalyst



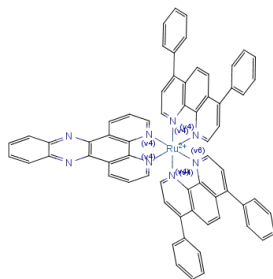
dodecatungstate



cis-platin



Zn-aminobenzenedicarboxylate



Ru-phenanthroline

YBa<sub>2</sub>Cu<sub>3</sub>O<sub>7</sub>

- Simple molecules
- Bulk chemicals
- Coordination compds
- Catalysts
- Clusters
- Drugs
- MOFs
- Biological compounds biosenor
- Solid-state compounds
- ...

# New structure editor: requirements

## General requirements

- Intuitive & easy to use
- Support of atoms with charge, as isotope, radical, unusual hydrogen-count
- Support of all bond types
- Expand of abbreviated groups
- Structure templates (standard templates & own templates)
- Stereochemistry functions
- Functionalities like rotation, mirror
- Support of ...
- Integ ... software

**state-of-the-art of most structure editors – in 2D**

## Special requirements

- Drawing and visualization in **3D** (with xyz coordinates)
- Stereochemistry functions beyond tetrahedron

**missing elements**

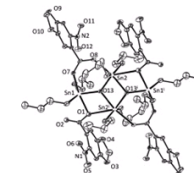
## MarvinSketch

- flexible and configurable to our needs

## ChemAxon

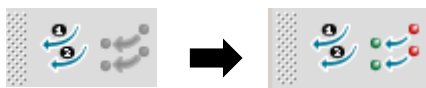
- offered the possibility to improve MarvinSketch with new features

# Improvement of MarvinSketch

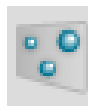


Development of new features and enhancement of existing features:

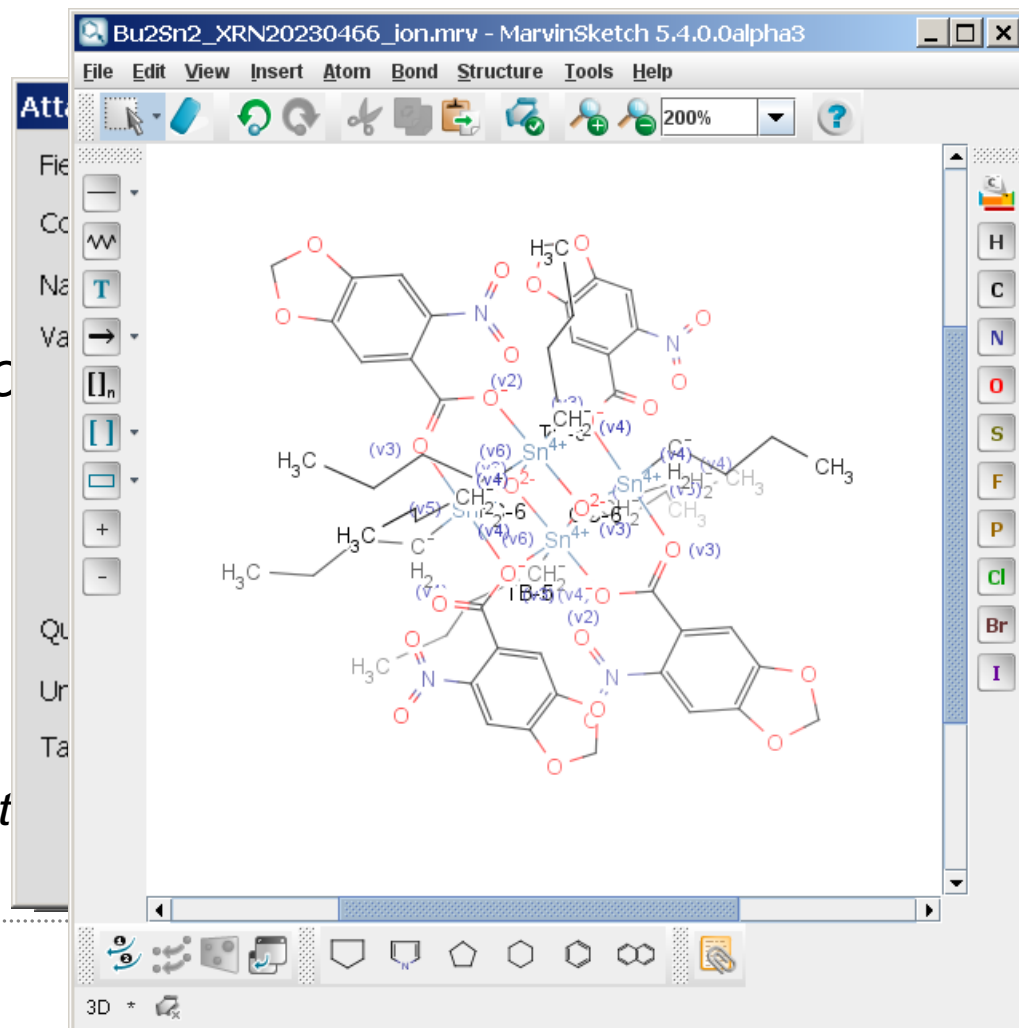
- 3D geometry and conformation generation
- 3D editing
- 3D visualization with fog effect
- *Example: Organotin complex*  
 $\{[(CH_2O_2C_6H_2(o-NO_2)COO)SnBu_2]_2C$ 
  - *merge two 3D polyhedra*



- *align 3D structures*



- *Advanced stereochemistry function dialog*



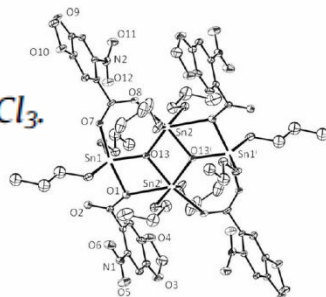
# The Reaxys View ...

## In vitro biological studies and structural elucidation of organotin(IV) derivatives of 6-nitropiperonylic acid: Crystal structure of $\{[(\text{CH}_2\text{O}_2\text{C}_6\text{H}_2(\text{o-NO}_2)\text{COO})\text{SnBu}_2]_2\text{O}\}_2$

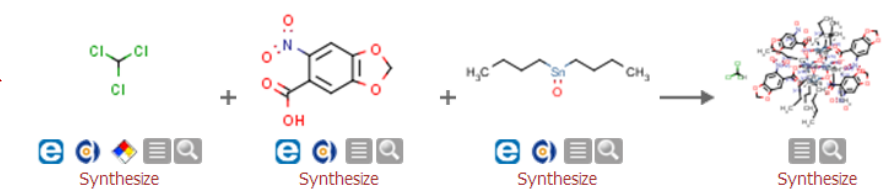


Compound (4)  $[(n\text{-Bu}_2\text{SnL})_2\text{O}]_2 \cdot \text{CHCl}_3$ .

L = 6-nitropiperonylate



Structure	Chemical Name	N° of preparations All Preps   All Reactions	Available Data
	$[(\text{Bu}_2\text{Sn}(6\text{-nitropiperonylate}))_2\text{O}]_2 \cdot 2\text{CHCl}_3$	1 prep out of 1 reactions.	Physical Data (9) Spectra (4) Bioactivity/ECotox (1)
Yield	Conditions	References	



65% in toluene  
Bu<sub>2</sub>SnO, 6-nitropiperonylic acid refluxed in toluene for 6 h with removal of H<sub>2</sub>O; cooled to room temp.; evapn.; recrystd. from CHCl<sub>3</sub>/hexane; elem. anal.;

Hanif, Muhammad; Hussain Bhatti, Moazzam H.; Ahme al. Polyhedron, 2010, vol. 29.

▲ Melting Point (1)

Melting Point	Solvent	Comment	Reference
196 - 198 °C	hexane CHCl <sub>3</sub>	from Gmelin	Hanif, Muhammad; Hussain Bhatti, Moazzam H.; Ahme al. Polyhedron, 2010, vol. 29. Title/Abstract Full Text

Nucleus	Coupling Nuclei	Solvents
<sup>1</sup> H	<sup>1</sup> H <sup>119</sup> Sn	CDCl <sub>3</sub>

2.3.3.1.4. Compound (4)  $[(n\text{-Bu}_2\text{SnL})_2\text{O}]_2 \cdot \text{CHCl}_3$ . Prepared by method B by adding ligand acid (1 g, 4.74 mmol) and *n*-Bu<sub>2</sub>SnO (0.59 g, 2.37 mmol) in 2:1 ratio, respectively.

Physical data; Yield 65%, m.p. 196–198 °C, CHN Anal. Calc.: C, 38.77; H, 4.4; N, 2.74. Found: C, 38.02; H, 4.22; N, 2.62%.

<sup>1</sup>H NMR data (CDCl<sub>3</sub>, ppm, <sup>1</sup>J(<sup>1</sup>H, <sup>1</sup>H), <sup>2</sup>J(<sup>119</sup>Sn, <sup>1</sup>H) in Hz); 7.3 (s, 4H, aromatic protons), 6.91 (s, 4H, aromatic protons), 6.15 (s, 8H, O–CH<sub>2</sub>–O), {1.82 (t, 16H, (8), [73]), 1.51–1.38 (m, 32H), 0.91 (t, 24H, (7.4), SnCH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>)}

### 2.3.3. Synthesis of organotin derivatives

2.3.3.1. General procedure. Two different methods have been employed for the synthesis of the organotin derivatives of the 6-nitropiperonylic acid. In method A the organotin chloride was refluxed with sodium salt of the acid in dry toluene for 5–6 h in 1:2 (diorganotin dichloride) or 1:1 (triorganotin chloride) molar ratio. After reflux the insoluble material were filtered off and solvent was evaporated under reduce pressure. The resultant solid masses were recrystallized from chloroform and *n*-hexane mixture. In method B appropriate R<sub>2</sub>SnO and ligand were refluxed for 6 h in 1:2 molar ratios in dry toluene (100 mL) using Dean-stark apparatus for azeotropic removal of water formed during the condensation reaction. The reaction mixture was then cooled to room temperature and the solvent was rotary evaporated. The solid product obtained was recrystallized from a mixture of chloroform and *n*-hexane.



# Integration of Marvin into Elsevier software

- Used as .Net application
- Integrated as component (COM dll)
- Integration into our application faced some difficulties to synchronize the rendering of a structure with the information in the MOL file  
solved by our developers

# Experience in Use

- Used in production for about 8 months
- without problems
- Launch of the new structure editor MarvinSketch was smoothly
- with nearly no training of users
- high user acceptance

# Résumé

- Reviewing the cooperation with ChemAxon regarding the enhancement of MarvinSketch, we can give a positive summary.
- The development was completed within budget and time.
- The final product has met our expectations

# Thank you very much for your attention

- Questions?

