

Bringing data together



The EPSRC National Chemical Database Service...
...and other eScience initiatives at the RSC

Richard Kidd
@rkiddr

The EPSRC National Chemical Database Service...

cds.rsc.org

Available databases

Click boxes for further information and login.

ACD/I-Lab2 	CSD 
DETERM  Gesellschaft für Chemische Technik und Biotechnologie e.V.	SPRESIweb 
Available Chemicals Directory 	ICSD 
Chemicalize 	ARChem 

In partnership with the EPSRC

EPSRC
Engineering and Physical Sciences
Research Council

ChemSpider  The free chemical database	MOS, NPU, CCR 
--	---

Mid-range facilities provide expertise and access to equipment in:

- Aberration-corrected scanning transmission electron microscopy (AC-STEM)
- Beamline at ESRF (XMaS)
- Chemicals database
- Computational chemistry software
- Crystallography
- Dark fibre infrastructure
- Electron paramagnetic resonance spectroscopy (EPR) also known as electron spin resonance (ESR)
- Engineering instruments
- Free electron laser
- Mass spectrometry
- Solid-state nuclear magnetic resonance (NMR)
- X-ray photoelectron spectroscopy (XPS)
- III-V technologies

History

- EPSRC mid-range facility
- CDS run from STFC Daresbury
- Tender in 2012 for 2013-2017

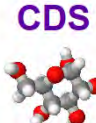
STFC Chemical Database Service

Access Data
Service Overview
Registration
Databases
Training Material
Downloads
Archives

Latest CDS News below
also follow our [Twitter](#) or [RSS](#)

Access to CrystalWorks 'demo' (no searching) continues via the **Access Data** link above. For full search facilities first use the **'Login first..'** option (form to right). This form only works for users within the STFC network domain [more..](#)

Click top menu **Archives** header for older News and other items.



CDS

Daresbury Laboratory

Use menu above for further information

Login first before selecting **Access Data**

Use your CDS username


Username: [Forgot](#)

Password: [Password?](#)

The Chemical Database Service at Daresbury **has now ceased providing general user access**. An EPSRC funded national service is now operated by the Royal Society of Chemistry. For further details of this service go to the [CDS/RSC website](#)

Please note, however, that [access](#) to various components of the original **CDS at Daresbury** is still possible from **within the STFC domain**.


CDS Homepage



Science & Technology
Facilities Council

Send queries and comments to cdsbb@stfc.ac.uk - phone 01925 603162

The STFC CDS is operated at the Daresbury Laboratory by the [SCD](#)



modified May 20, 2013

Vision for the Service

- An aggregate collection of best-in-class databases
- An environment for data validation, curation and model building
- A platform for open collaboration
- An environment for micropublishing
- A unique opportunity

Progress in 2013

- Set up and launch for 2 January
- Addition of databases and services
 - WebCSD, DETHERM, ArChem, Chemicalize
- Matching and exceeding remote access
- Access: Shibboleth (coming very soon)
- Advisory Board in place
- 900+ registered CSD users

Disambiguation...

- CDS (and NCDS, and EPSRC CDS/NCDS)
- CSD from CCDC
- ICSD
- I-Lab from ACD/Labs
- ACD from Accelrys

(thanks to DETHERM, SPRESIweb)

Building a repository

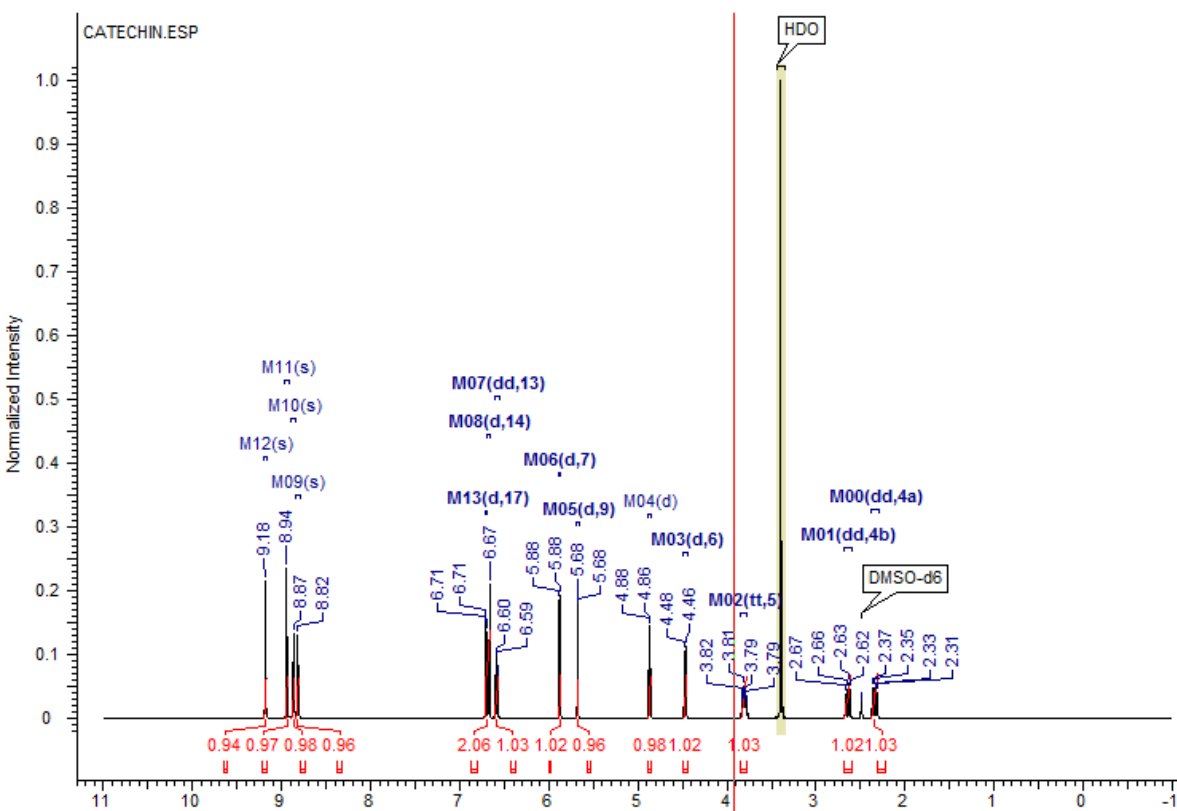
- Depositions of data
- DOI'd data objects for citation purposes
- Other national services to feed the repository
- Integration APIs to blogging tools for chemistry and Electronic Lab Notebooks
- Linking mandates to value and credit

Building the data repository

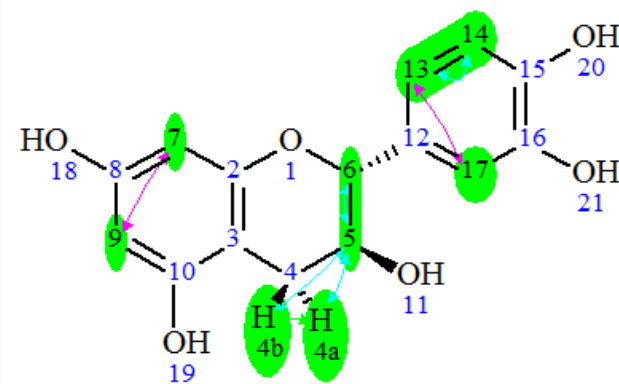
Chemistry data of interest

- Chemical compounds
- Chemical reactions
- Analytical data
- Alphanumeric text – logP, pKa, solubility
- Graphs and plots
- Figures

This is what we really want...



Assignment Quality by Shift Difference (ppm)



We have chosen for now....

Focus on supporting transformed and structured data rather than raw data or dumb files

- Chemicals
- Chemical reactions
- Measured properties
- Analytical data in useful form

What we will deliver...

- Simple interfaces for uploading of data
- Embeddable widgets and programming interfaces to utilise in in-house systems, ELNs
- Automated harvesting approaches – sweeping directories for data
- Data validation where possible
- Embargos

What we will deliver...

- Micropublishing platform for submission of
 - Protocols and procedures
 - Reactions
 - Safety and hazard data
- Template-based submissions of procedures

Work in Progress

- Micropublishing platform for ChemSpider SyntheticPages (CSSP) already exists
- Work in progress
 - ChemSpider Reactions in development
 - Integration of ELN templates to CSSP – full protocol captured in ELN and procedure mapped to CSSP for deposition
 - Word-based templates to match CSSP depositions

Micropublishing Syntheses



About | Search | Browse | Most Accessed | Recent Submissions | Help |

Text search | Structure search

Search by Reaction name, Author name, or Compound

Search Reactions



ChemSpider SyntheticPages is a free database of practical synthetic procedures, provided by the community for the community.

[Learn more....](#)

ChemSpider SyntheticPages

Hydrogenation of Ethyl 3-(1-pyrenyl)acrylate; Ethyl 3-(1-pyrenyl)propanoate

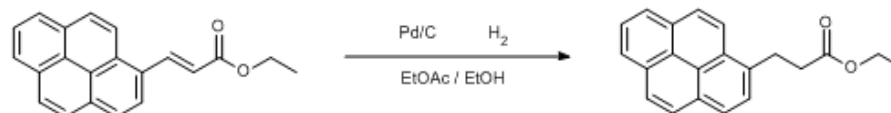
SyntheticPage 510

DOI: [10.1039/SP510](https://doi.org/10.1039/SP510)

Submitted Sep 29, 2011, published Oct 06, 2011

Anish Mistry (a.mistry@warwick.ac.uk)

A contribution from Fox Group, Warwick University



Chemicals Used

Ethyl 3-(1-pyrenyl)acrylate (1 equiv, prepared)

10% Pd/C (2 mol%, Alfa Aeser)

Ethanol

Ethyl Acetate

Hydrogen gas

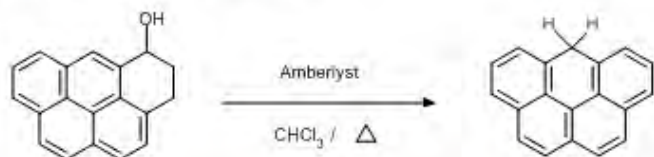
Procedure

Ethyl 3-(1-pyrenyl)acrylate (3.02 g, 10.1 mmol) was firstly dissolved in ethyl acetate (70 ml). Ethanol (70 ml) and the 10% Pd/C (2 mol%) were then added to the mixture and the air evacuated out the system and replaced with hydrogen. The reaction was left to stir under a hydrogen balloon at room temperature for 29 hours. The reaction mixture was then filtered through a pad of celite with ethyl acetate and the solvent removed under *vacuo* to yield a dark yellow solid (3.02 g, 99%).

Dehydration of 3,4-dihydro-5H-Benzo[cd]pyren-5-ol

Anish Mistry

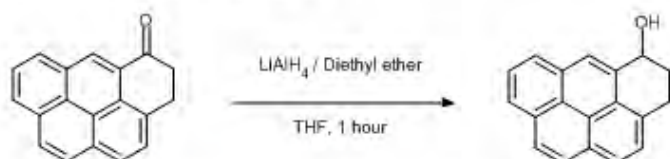
Published: May 31 2012



Reduction of 3,4-dihydro-5H-benzo[cd]pyren-5-one.

Anish Mistry

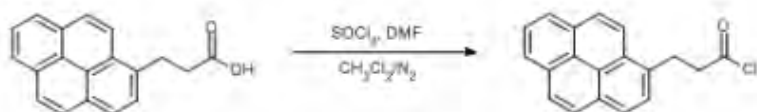
Published: Mar 12 2012



Chlorination of a carboxylic acid

Anish Mistry

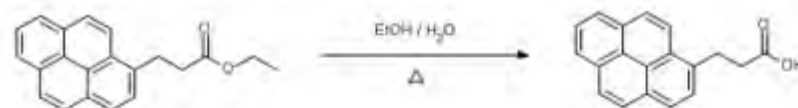
Published: Nov 27 2011



Hydrolysis of Ethyl 3-(1-pyrenyl)propanoate

Anish Mistry

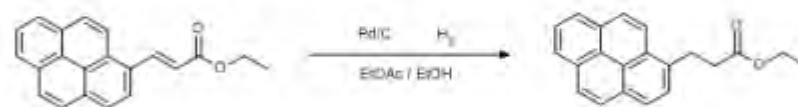
Published: Oct 11 2011



Hydrogenation of Ethyl 3-(1-pyrenyl)acrylate

Anish Mistry

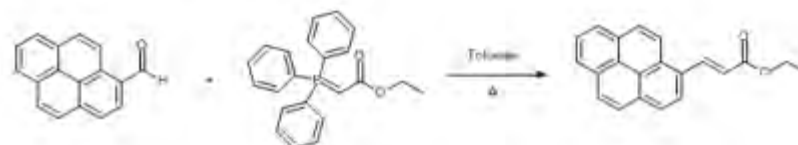
Published: Oct 06 2011



Wittig Reaction

Anish Mistry

Published: Sep 14 2011



Requirements

- Build templates for CSSP/Reactions deposition
- Supporting other EPSRC initiatives, storing successful and failed reactions



Dial-a-Molecule
An EPSRC Grand Challenge Network

Model building with community data

- Community data as a basis of model building
 - Consume newly deposited data
 - Will include us adding in data from publications
 - Build predictive algorithms for the community
 - Regular updating, version stamped prediction
 - Will be using Accelrys Pipeline Pilot
- Provides testing and validation for data

Data validation and curation required

Encouraging participation with
Rewards and **RECOGNITION**



Manual Curation



Rewards & Recognition

- Integrated commenting, curating and validation across all RSC platforms
- All integrated to a central RSC profile and feeding altmetrics



Future recognition in altmetrics?



Total synthesis of all (–)-agelastatin alkaloids†

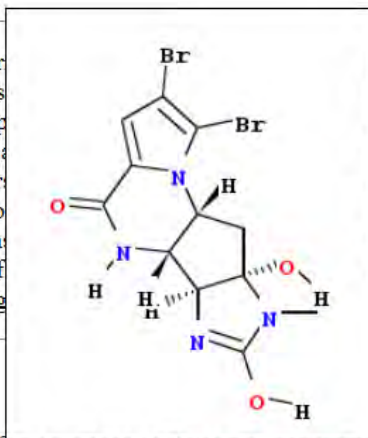
Mohammad Movassaghi *, Dustin S. Siegel and Sunkyu Han

Massachusetts Institute of Technology, Department of Chemistry, 77 Massachusetts Avenue 18-292, Cambridge, MA 02139-4307, USA. E-mail: movassag@mit.edu

Received 2nd July 2010, Accepted 20th July 2010

First published on the web 16th August 2010

The pyrrole-imidazole family of marine alkaloids, derived from a diverse array of structurally complex natural products that possess a tetracyclic molecular framework incorporating a pyrrole ring, provide a hypothesis for the formation of the unique and complex intrinsic chemistry of plausible biosynthetic precursors of all known agelastatin alkaloids including the first to be synthesized. The gram-scale chemical synthesis of agelastatin A was intricate, involving the cyclopentane C-ring and required the development of an annulation reaction and a carbohydroxylative trapping



precursors, constitutes a diverse array of members of this family with unique connectivities. We have now exploited the unique intrinsic chemistry of the pyrrole-imidazole family to provide the first total syntheses of agelastatin A and F. Our strategy involves the synthesis of the tetracyclic imidazolone-forming

Introduction

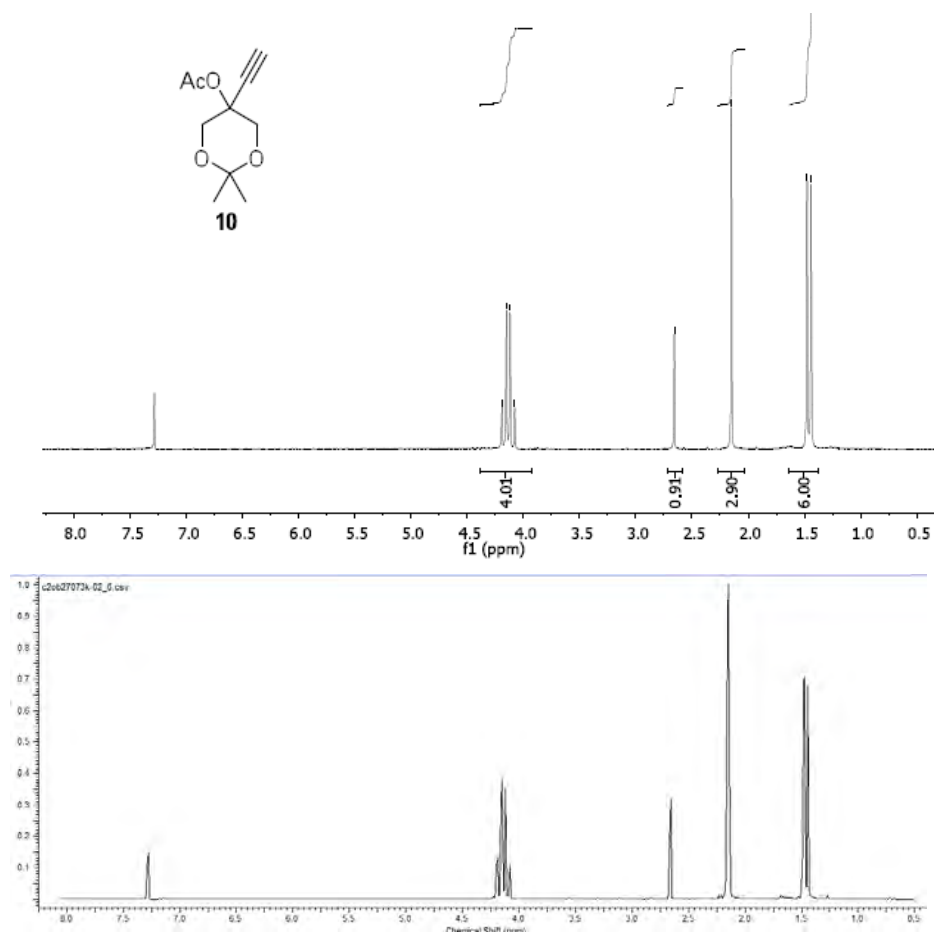
The agelastatin alkaloids constitute an intriguing subclass of the pyrrole-imidazole family of marine alkaloids that are likely derived from linear biogenetic precursors such as clathrocin (7),¹ hymenidin (8),² and oroidin (9, Fig. 1).^{3,4} (–)-Agelastatins A (1) and B (2) were first isolated from the Coral Sea sponge *Agelas dendromorpha* by Pietra *et al.* in 1993 who successfully identified and chemically studied their unique

DERA : data enable the RSC Archive

How much **data** is in the archive, in the publications and in the supplementary info?

- How many compounds?
- How many syntheses?
- How many characterisation measurements?
 - Property data
 - Spectral data
 - Graphs and charts to be used for modelling

Top: Figure; Bottom: Converted



What if we could capture it all?



Aims in summary

- Provide best-in-class data resources to (initially) UK academia
- Provide domain data repository to support models, grand challenges, reuse and analysis
- Support micropublications
- Support curation
- Supporting chemical data interoperability

The Merck Index* Online

RSC Advancing the Chemical Sciences | Sign in | Publishing | ChemSpider

THE MERCK INDEX Online

Home | Search | Structure Search | Named Reactions | Reference Tables | Help

Home | Search | Structure Search | Named Reactions | Reference Tables | Help

Home

Quick search

Enter a Name, CAS Registry Number, Molecular Formula or Molecular Weight
Enclose Molecular Formula in Brackets (e.g. [C3H6O])
Molecular Weight can be input as a single value, or a range (e.g. 168.23 or 77-78)

Other search options

Search
Text Search
Compound Name
CAS Registry Number


Structure search
H
C
N
O
S
P
F
Cl
Br
I

Reference tools

Named Reactions
Search or browse over 500 organic named reactions

Reference Tables
Useful reference matter including: common abbreviation chemical structures, constants and conversions used in the The Merck Index Online

About The Merck Index Online

 For over 120 years The Merck Index has been regarded as the most authoritative and reliable source of information on drugs and biologicals. Now this trusted resource is available online from RSC Publishing

The Merck Index Online offers the same highly authoritative information as the print edition text database. It contains over 11,500 monographs – including historic records not available in the print edition. The Merck Index Online will be updated regularly with accurate information curated by experts.

*The name THE MERCK INDEX is owned by Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Whitehouse Station, N.J., U.S.A., and is licensed to The Royal Society of Chemistry for use in the U.S.A. and Canada.

THE MERCK INDEX Online

Home | Search | Structure Search | Named Reactions | Reference Tables | Help

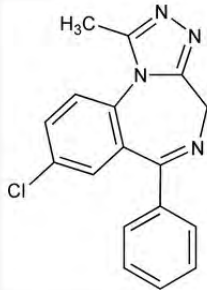
Home > Monograph

Alprazolam

Monograph ID: MONO150000308
Title: Alprazolam
CAS Registry Number: 28981-97-7
CAS Name: 8-Chloro-1-methyl-6-phenyl-4H-[1,2,4]triazolo[4,3-*a*][1,4]benzodiazepine
Additional Names: 8-chloro-1-methyl-6-phenyl-4H-s-triazolo[4,3-*a*][1,4]benzodiazepine
Manufacturer Code: D-65MT, U-31889
Trademark Names: Alplax (Gador), Cassadan (Temmler), Espanon (Orion), Tafil (Pharmacia), Tranquinal (Bago), Trankimazin (Pharmacia), Xanax (Pharmacia & Upjohn)
Molecular Formula: C₁₇H₁₃ClN₄
Molecular Weight: 308.77
Percent Composition: C 66.13%, H 4.24%, Cl 11.48%, N 18.15%
Standard InChI: InChI=1S/C17H13ClN4
/c1-11-20-21-16-10-19-17(12-5-3-2-4-6-12)14-9-13(18)7-8-15(14)22(11)16/h2-9H,10H2,1H3
Standard InChIKey: VREFGVLTWBCJP-UHFFFAOYSA-N

Properties
Crystals from ethyl acetate, mp 228-228.5°C. Freely sol in chloroform; sol in alc; sparingly sol in acetone; slightly sol in ethyl acetate. Insol in water. uv max (ethanol): 222 nm (ε 40250). LD₅₀ in mice, rats (mg/kg): 1020, >2000 orally; 540, 610 i.p. (Nakajima).

References
Benzodiazepine anxiolytic. Prepn: J. B. Hester, DE 2012190; *idem*, US 3987052 (1970, 1976 both to Upjohn); J. B. Hester *et al.*, *Tetrahedron Lett.* 1971, 1609; A. Walsler, G. Zenchoff, *J. Med. Chem.* 20, 1694 (1977) DOI: 10.1021/jm00222a035 PMID: 592339.
Central Nervous System. *Drugs* 1977, 24, 147-153. DOI: 10.1007/BF02701003 PMID: 860873


Save
Image Structure Zoom

RSC Content
Powered by ChemSpider

- ChemSpider Record
- ChemSpider Name Search
- ChemSpider Skeleton Search
- Search RSC Platform

*The name THE MERCK INDEX is owned by Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Whitehouse Station, N.J., U.S.A., and is licensed to The Royal Society of Chemistry for use in the U.S.A. and Canada.

See also...

- Open PHACTS data interoperability
(see Lee's talk)
- PharmaSea (Natural Products to drug leads)
(www.pharma-sea.eu)
- Topic modelling for chemistry
(ask me)

Richard Kidd
@rkiddr
@cds_rsc