Marrying ACD/Labs technologies to eScience Projects at the Royal Society of Chemistry

Antony Williams
ACD/Labs User Meeting
June 2013
RSC eScience

• Royal Society of Chemistry is a member society (>47,000), Publisher and Innovator in eScience
• Host of many online databases and services
  – ChemSpider, SyntheticPages, SpectraSchool,…
• Participant in multiple grant-based projects
  – National Chemical Database Service
  – Open PHACTS
  – PharmaSea
Multiple ACD/Labs Tools in use...

- Structure “checking” routines for data
- Nomenclature generation and conversion
- Physicochemical prediction algorithms
- Web-based spectral display widget
- “Interactive Lab” web-based prediction tools

- But first an intro to ChemSpider...
ChemSpider

• 28 million chemicals with associated data...
I want to know about “Vincristine”

Vincristine

ChemSpider ID: 5758
Molecular Formula: C_{46}H_{56}N_{4}O_{10}
Monoisotopic mass: 824.399644 Da
- Systematic name
(2α,2′β,3α,4α,5β,19β)-22-oxovincaleukoblastine
- SMILES and InChI

Wikibox
Embed
Deprecate
Watch this record
Manage data slice
I want to know about “Vincristine”

Vincristine

ChemSpider ID: 5758

Molecular Formula: C_{46}H_{56}N_{4}O_{10}
Monoisotopic mass: 824.399644 Da

- Systematic name
  (2α,2′β,3α,4α,5β,19β)-22-oxovincacaleukoblastine

- SMILES and InChIs

WikiBox
Embed
Deprecate
Watch this record
Manage data slice
Vincristine: Identifiers and Properties

**Names and Identifiers**

- Validated by Experts, Validated by Users, Non-Validated,
- (2’β)-22-Oxovincaleukoblastine
- 200-318-1 [EINECS/ELINCS]
- 22-Oxovincauleukoblastine
- 57-22-7 [RN]
- vincaleukoblastine, 22-oxo-
- vincaleukoblastine, 22-oxo-(2’β)-
- vincaleukoblastine, 22-oxo-(3β,4’β)-

**Properties**

- **Experimental Physchem Properties**
  - Melting Point: 218 - 220 C

- **Miscellaneous**
  - Appearance: solid
  - Stability: Stable, but may be heat sensitive. Incompatible with strongoxidizing agents.
  - Toxicity: IVN-RAT LD50 1300 mg kg-1, IPR-MUS LD50 5.2 mg kg-1
  - Safety: Safety glasses, gloves, good ventilation.
**Predicted Properties**

<table>
<thead>
<tr>
<th>Properties</th>
<th>ACD/Labs</th>
<th>Predicted - ACD/Labs</th>
<th>Predicted - ChemAxon</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACD/LogP</td>
<td>5.749</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACD/LogD (pH 5.5)</td>
<td>3.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACD/BCF (pH 5.5)</td>
<td>34.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACD/KOC (pH 5.5)</td>
<td>80.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#H bond acceptors</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#Freely Rotating Bonds</td>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Index of Refraction</td>
<td>1.677</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Molar Volume</td>
<td>586.859 cm³</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surface Tension</td>
<td>74 dyne/cm</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Predicted data is generated using the ACD/Labs’ ACD/PhysChem Suite, for more information see their [website](http://www.rsc.org).

- # of Rule of 5 Violations: 3
- ACD/LogD (pH 7.4): 5.20
- ACD/BCF (pH 7.4): 3920.85
- ACD/KOC (pH 7.4): 9090.30
- #H bond donors: 3
- Polar Surface Area: 171.17 Å²
- Molar Refractivity: 221.088 cm³
- Polarizability: 87.646 10⁻²⁴ cm³
- Density: 1.406 g/cm³
Vincristine: Vendors and Sources

Linked by **Structure**

### Chemical Vendors

<table>
<thead>
<tr>
<th>Data Source</th>
<th>External ID(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pharmenten</td>
<td>S-PTN3183</td>
</tr>
<tr>
<td>AvaChem Scientific</td>
<td>1011, 1011B</td>
</tr>
<tr>
<td>AOKChem</td>
<td>aokchem11421</td>
</tr>
</tbody>
</table>

### Data Sources

<table>
<thead>
<tr>
<th>Data Source</th>
<th>External ID(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChemBank</td>
<td>NCI60_026703</td>
</tr>
<tr>
<td>DiscoveryGate</td>
<td>5978</td>
</tr>
<tr>
<td>LeadScope</td>
<td>LS-228</td>
</tr>
<tr>
<td>NIAID</td>
<td>002674</td>
</tr>
<tr>
<td>DrugBank</td>
<td>5978, APRD00495</td>
</tr>
<tr>
<td>Collaborative Drug Discovery</td>
<td>16374</td>
</tr>
</tbody>
</table>
Vincristine: Patents

- **Stable, injectable solutions of vincristine salts**
  This invention provides a stable, injectable aqueous solution of vincristine salts, suitable for intravenous injection for the treatment of neoplastic diseases, ...

- **Produit contenant de la vincristine**
  Procédé de fabrication d’un produit pharmaceutique pour le traitement du cancer. Ledit produit pharmaceutique comprenant ...

- **Synthesis of vinblastine and vincristine type compounds**
  A new process for the stereospecific synthesis of alkaloids of the vinblastine and vincristine type including the synthesis of vinblastine and vincristine as well ...

- **Vincristine immunoassay**
  Novel conjugates and immunogens derived from vincristine and antibodies generated by these immunogens are useful in immunoassays for the quantification ...

- **Synthesis of vinblastine and vincristine type compounds**
  US Pat. EP0292463A2 - Apr 13, 1988 - The University Of Vermont And State Agricultural College
  A new process for the stereospecific synthesis of vinblastine and vincristine type compounds.

- **Synthesis of vinblastine and vincristine type compounds**
  US Pat. EP0215658A1 - Feb 21, 1986 - The University Of Vermont And State Agricultural College
  A new process for the stereospecific synthesis of alkaloids of the vinblastine and vincristine type including the synthesis of vinblastine and vincristine as well ...
Stable, injectable solutions of vincristine salts
EP 0298192 B1

DESCRIPTION

[0001] This invention provides a stable, injectable aqueous solution of vincristine salts, suitable for intravenous injection for the treatment of neoplastic diseases, particularly leukemias, in humans.

[0002] Certain vinca alkaloids, being dimeric indole-dihydroindole compounds, have been used for some time in chemotherapy as oncolytic drugs, particularly for the treatment of leukemias. Among those so-called vinca dimers, there might be mentioned especially vincristine, vinblastine and vindesine (an amide derivative of vinblastine). The treatment consists of intravenous administration of pharmaceutically acceptable salts of the...
Vincristine: Articles
Linked by Name

Please click here to see full list of found articles.

- North KN. Therapeutics for Childhood Neurofibromatosis Type 1 and Type 2. *Curr Treat Options Neurol*, 2011 Aug 18

Validated assay for the simultaneous quantification of total vincristine and actinomycin-D concentrations in human EDTA plasma and of vincristine concentrations in human plasma ultrafiltrate by high-performance liquid chromatography coupled with tandem mass spectrometry.

Validation of an electrospray ionization LC/MS/MS method for quantitative analysis of vincristine in human plasma samples.

An automated method for the bioanalysis of vincristine suitable for therapeutic drug monitoring and pharmacokinetic studies in young children.

Application of dried blood spots combined with high-performance liquid chromatography coupled with electrospray ionisation tandem mass spectrometry for simultaneous quantification of vincristine and actinomycin-D.

Solid-phase extraction of vinblastine and vincristine from plasma and urine: variable drug recoveries due to non-reproducible column packings.
From the Mass Spectrometry Bulletin Database
For the latest techniques and applications in the analytical sciences.

Determination of vincristine in infant plasma by liquid chromatography-atmospheric pressure chemical ionization-mass spectrometry

DOI: 10.1016/j.jpba.2005.11.039
Spectra

**Type:** HNMR  
**Associated Hyperlink:** http://rainier.chem.plu.edu/nutsform.html  
**Comments:** These data are obtained from the Pacific Lutheran University FTNMR FID Archive  
**Approved:** No  
**Submitted by:** ChemSpiderman

---

**Type:** CNMR  
**Associated Hyperlink:** http://rainier.chem.plu.edu/nutsform.html  
**Comments:** These data are obtained from the Pacific Lutheran University FTNMR FID Archive  
**Approved:** No  
**Submitted by:** ChemSpiderman

---

**Type:** Electron Impact  
**Associated Hyperlink:** http://webbook.nist.gov/cgi/cbook.cgi?ID=C57885&Units=SI&Mass=200#Mass-Spec  
**Comments:** Provided with permission, March 12 2009. Owner NIST Mass Spectrometry Data Center Collection (C) 2007 copyright by the U.S. Secretary of Commerce on behalf of the United States of America. All rights reserved. Origin T.IIDA NIHON UNIVERSITY, KORIYAMA, FUKUSHIMA-KEN, JAPAN; NIST MS number 67286  
**Approved:** No  
**Submitted by:** ChemSpiderman
Spectra

Cholesterol $^{13}C(1) \text{ CDCI}_3$

Type: CNMR
Approved: Yes
Where do data come from?

- ChemSpider users deposit data
- Some contributions from NIST
- Chemical vendors are starting to provide data. Synthonix are one of our major contributors (www.synthonix.com)
Crowdsourced “Annotations”

- Users can add
  - Compounds
  - Descriptions/Syntheses/Commentaries
  - Links to articles via DOIs
  - Add spectral data
  - Add Crystallographic Information Files
  - Add photos
  - Add MP3 files
  - Add Videos
Crowdsourced Curation

- Crowd-sourced curation: identify/tag errors, edit names, synonyms, identify records to deprecate

<table>
<thead>
<tr>
<th>ID</th>
<th>Index</th>
<th>Date</th>
<th>Priority</th>
<th>New</th>
<th>Description</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>28412</td>
<td></td>
<td>5/1/2008 1:02:27 PM</td>
<td>Extreme</td>
<td>New</td>
<td>This structure should be deprecated. The structure is abysmal</td>
<td><a href="mailto:antony.williams@chemspider.com">antony.williams@chemspider.com</a></td>
</tr>
<tr>
<td>16352795</td>
<td></td>
<td>4/29/2008 3:42:04 PM</td>
<td>Normal</td>
<td>New</td>
<td>Dipodium salt or anthracene-9,10-diol ???</td>
<td><a href="mailto:kolshorn@uni-mainz.de">kolshorn@uni-mainz.de</a></td>
</tr>
<tr>
<td>5558</td>
<td></td>
<td>4/24/2008 2:52:04 AM</td>
<td>High</td>
<td>New</td>
<td>Look for 2-Butenedioic acid (22)-homopolymer (RN 26099-09-2). But I get 50-37-3. It does not seem alright. Searching with key word of 26099-0902, no results can be found. But in ChemIDplus, it does exist.</td>
<td>alexguisina.com</td>
</tr>
<tr>
<td>16815</td>
<td></td>
<td>4/9/2008 4:31:48 AM</td>
<td>Normal</td>
<td>New</td>
<td>Missing stereo bonds ?</td>
<td><a href="mailto:heinz.kolshorn@uni-mainz.de">heinz.kolshorn@uni-mainz.de</a></td>
</tr>
</tbody>
</table>
Spectral Uploading

• Locate the structure of interest and deposit spectrum
Spectral Uploading

- Various types of NMR spectra supported

Spectrum Type:
- 2D NMR 1H-13C Direct correlation
- 2D NMR 1H-13C Long-range correlation
- 2D NMR 1H-1H COSY
- 2D NMR 1H-1H Long range correlation
- 2D NMR 1H-1H NOESY/ROESY
- APC+ Mass Spectrum
- APC- Mass Spectrum
- APPI+ Mass Spectrum
- APPI- Mass Spectrum
- Chemical Ionization +ve
- Chemical Ionization -ve
- CNMR
  - Electron Impact
  - ESI+ Mass Spectrum
  - ESI- Mass Spectrum
  - FNMR
  - HNMR
  - Infrared
  - MALDI+ Mass Spectrum
  - MALDI- Mass Spectrum
  - Near Infrared
  - PNMR
  - Raman
  - UV-Vis
<table>
<thead>
<tr>
<th>ID</th>
<th>Structure</th>
<th>Type</th>
<th>File</th>
<th>Submitted</th>
<th>Open</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>7630</td>
<td><img src="2012-03-20-Acetic_anhydride_CDCl3_APT.png" alt="CNMR spectrum" /></td>
<td>CNMR</td>
<td>2012-03-20-Acetic_anhydride_CDCl3_APT.png</td>
<td>24/03/2012 06:18:47</td>
<td>true</td>
<td>13C (APT) NMR spectrum of acetic anhydride in CDCl3</td>
</tr>
<tr>
<td>7630</td>
<td><img src="2012-03-20-Acetic_anhydride_CDCl3_1H.png" alt="HNMR spectrum" /></td>
<td>HNMR</td>
<td>2012-03-20-Acetic_anhydride_CDCl3_1H.png</td>
<td>24/03/2012 04:32:38</td>
<td>true</td>
<td>1H NMR spectrum of acetic anhydride in CDCl3</td>
</tr>
</tbody>
</table>
Multiple Spectra for One Structure
ChemSpider ID 24528095 H1 NMR
ChemSpider ID 24528095 HSQC
ChemSpider ID 24528095 HMBC
Available Spectra

http://www.chemspider.com/spectra.aspx

<table>
<thead>
<tr>
<th>ID</th>
<th>Structure</th>
<th>Type</th>
<th>File</th>
<th>Submitted</th>
<th>Open Date</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>24528095</td>
<td><img src="image1.png" alt="Structure" /></td>
<td>2D NMR 1H-13C</td>
<td>mz2-514-hmnc.JPG</td>
<td>14/03/2010 16:59:37</td>
<td>False</td>
<td>Section hmnc.</td>
</tr>
<tr>
<td>24528095</td>
<td><img src="image2.png" alt="Structure" /></td>
<td>2D NMR 1H-13C</td>
<td>mz2-514-hsqc.JPG</td>
<td>14/03/2010 16:55:39</td>
<td>False</td>
<td>Section HSQC.</td>
</tr>
<tr>
<td>24528095</td>
<td><img src="image3.png" alt="Structure" /></td>
<td>2D NMR 1H-1H COSY</td>
<td>mz2-514-cosy.JPG</td>
<td>14/03/2010 16:53:48</td>
<td>False</td>
<td>Section cosy.</td>
</tr>
</tbody>
</table>
Number of Spectra

- 9746 spectra against 6890 compounds

- IR 5389
- HNMR 1679
- CNMR 1207
- UV-Vis 183
- El 90
- 2D1H13CD 68
- Raman 51
- NIR 32
- 2D1H1HCOSY 21
- 2D1H13CLR 10
- Cl+ve 8
- PNMR 7
Some usage statistics

• ca. 200 visitors at any one time, ~30,000 visits per day
• Mar 4-Apr 3, 2013
  – Visits = 731,656
  – Unique Visitors = 527,008
• Independent servers to support other projects

• Does not include web service calls
ChemSpider as a Foundation

• ChemSpider is a foundation for projects:
  – >500 data sources aggregated and mapped
  – Continually curated and updated with new data
  – Normalized data around a **structure** centric data model
  – Providing an API allows integration to support other internal projects
  – Providing API access outside RSC extends the reach
Micropublishing Syntheses

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
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 Aesar)
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 baloon 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%).
Olympicene

'Olympic rings' molecule olympicene striking image

By Jason Palmer
Science and technology reporter, BBC News

The technique showcases details well under a billionth of a metre in size
Dehydration of 3,4-dihydro-5H-Benzo[cd]pyren-5-ol
Anish Mistry

**Published:** May 31 2012

\[
\text{OH} \xrightarrow{\text{Amberlyst}} \quad \text{H} \quad \text{H}
\]

\[
\text{CHCl}_3 / \Delta 
\]

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

**Published:** Mar 12 2012

\[
\text{LiAlH}_4 / \text{Diethyl ether} 
\]

\[
\text{THF, 1 hour} 
\]

Chlorination of a carboxylic acid
Anish Mistry

**Published:** Nov 27 2011

\[
\text{SOCI}_2, \text{DMF} 
\]

\[
\text{CH}_2\text{Cl}_2/N_2 
\]

Hydrolysis of Ethyl 3-(1-pyrenyl)propanoate
Anish Mistry

**Published:** Oct 11 2011

\[
\text{EtOH} / \text{H}_2\text{O} \xrightarrow{\Delta} 
\]

Hydrogenation of Ethyl 3-(1-pyrenyl)acrylate
Anish Mistry

**Published:** Oct 06 2011

\[
\text{Pd/C} / \text{H}_2 
\]

\[
\text{EtOAc} / \text{EtOH} 
\]

Wittig Reaction
Anish Mistry

**Published:** Sep 14 2011

\[
\text{Toluene} \xrightarrow{\Delta} 
\]
Dehydration of 3,4-dihydro-5H-Benzo[cd]pyren-5-ol; 6H-Benzo[cd]pyrene

Synthetic
Page 542
DOI: 10.1039/SP542
Anish Mistry (a.mistry@warwick.ac.uk)
A contribution from Fox Group, Warwick University

Chemicals Used

3,4-dihydro-5H-Benzo[cd]pyren-5-ol (prepared)
Amberlyst 15 (Sigma-Aldrich)
Chloroform

Procedure

3,4-dihydro-5H-Benzo[cd]pyren-5-ol (0.1 g, 0.39 mmol) was dissolved in chloroform (30 ml) and Amberlyst 15 (0.1 g) added under a dinitrogen atmosphere. The reaction was heated to 30°C and left overnight under the inert atmosphere. The solution was then filtered to separate the Amberlyst and washed with chloroform. The combined solvents were removed under vacuum using a Rotary evaporator. The crude product was column chromatographed under a dinitrogen atmosphere eluting with 1:1 chloroform:petroleum ether 40-60°C. A white solid was obtained using this method (50 mg, 54%).
Web Services
Example: Spectral Data

Spectra

The following operations are supported. For a formal definition, please review the Service Description.

- **GetAllSpectraInfo**
  Returns information for all Open Access spectra in database

- **GetCompoundSpectraInfo**
  Returns information about spectra associated particular compound identified by cmp_id parameter

- **GetSpectraInfoArray**
  Returns information about the particular spectrum identified by spc_id parameter

- **GetSpectrumInfo**
  Returns information about the particular spectrum identified by spc_id parameter
How to play
Each round you'll be presented with a new spectrum. You have to select the molecule that matches the spectrum. For each molecule you identify correctly you'll get one point. The game continues until you get one wrong. Enter your name, choose a group (optional), select the type of spectrum you want and click play.
Spectral Game

Correct. Your Current Score: 1

Click on the molecule that corresponds to the spectrum. Hit F11 to enlarge browser window.

No spectrum? Spectrum reversed? FLAG IT!
No spectrum? Something wrong with the spectrum? Comment. (opens in new window)
Increasing Complexity

![Graph showing increasing complexity with chemical structures](image-url)
SpectralGame in the hand
Welcome | Acknowledgements | What’s new

Welcome to the SpectraSchool website. Here you will find a range of resources to help in your understanding of the principles and practice of spectroscopic and spectrometric methods. Take the opportunity to look at and compare real spectra by zooming and overlaying or watch video clips that show you practical aspects of the techniques.

Getting started

Select learning level:  ○ Basic ○ Advanced

Navigate the site using the buttons in the main toolbar above. The four main sections are as follows:

• Click home to return to this page.
• Click Spectra to start plotting $^1$H and $^{13}$C NMR, IR, Mass and UV/Vis spectra for a range of compounds. You can also use this section to plot your own spectra.
• Click Video to view short videos explaining the techniques.
• Click SIAS to visit the Spectroscopy in a suitcase section, described below.

Click ? at the top of the window for more information at any time.

Spectroscopy in a suitcase

Spectroscopy in a Suitcase (SIAS) is a new initiative that brings spectroscopy to schools and colleges. UV/Visible, FTIR and Mass Spectrometers are now available with trained chemistry postgraduate students and tested activities to allow students to get hands on experience of these techniques in their own classrooms. For more details about the project, please click the SIAS tab above.
Recently Added—THANKS ACD/Labs!

• Storage and display of ASSIGNED spectra
Access ChemSpider

• APIs
  – Programmatic access used by Mobile Apps, Funded Consortia projects, many Academic groups

• Widgets
  – UI components for embedding in other websites

• Data
  – Data access, downloads, reuse, licensing
Flexible ChemSpider API
Flexible ChemSpider API

**Xanax**

ChemSpider ID: 2034
Molecular Formula: C₁₇H₁₁ClN₄
Average mass: 308.764893 Da
Monoisotopic mass: 308.082886 Da
Systematic name: 8-Chloro-1-methyl-6-phenyl-4H-[1,2,4]triazolo[4,3-a][1,4]benzodiazepine
Search this compound on [Google Scholar](https://scholar.google.com)

---

**Synthesis of Essential Drugs**
by Ruben Vardanyan, Victor Hruby
2000 - 034 pages
books.google.com

**The Mood Cure: The 4-Step Program to Take Charge of Your Emotions-Today**
by Julia Ross
2004 - 400 pages
books.google.com

**Medication Madness: A Psychiatrist Exposes the Dangers of Mood-Altering ...**
by Peter R. Breggin, MD
2009 - 400 pages
books.google.com
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 the agelas sponge growers, constitutes a diverse array of structurally complex natural products that possess a tetracyclic molecular framework incorporating pyrrole, imidazole, pyridine and pyrazidine substructures. We provide a hypothesis for the formation of the unique tetracyclic core in this family by exploiting the unusual chemistry of plausible biosynthetic precursors and present preliminary total syntheses of all known agelastatin alkaloids including the first total synthesis of agelastatin A. Our gram-scale chemical synthesis of agelastatin A was initiated by a base-mediated ring closure of the cyclohexane C-ring and required the development of a new route for the synthesis of the 3-(4-fluorophenyl)pyrrole moiety.

Introduction
The agelastatin alkaloids constitute an intriguing subset of the marine pyrrole-imidazole family of marine alkaloids that are likely derived from linear biogenic precursors such as clathrodin (7), hymenidin (8) and eonidin (9, Fig. 1).1-4 (-)-Agelastatins A (1) and B (2) were first isolated from the Coral Sea sponge Agelas dendronotus by Pietra et al. in 1993 who successfully identified and chemically studied their unique structure.
It is so difficult to navigate...

Chemical genetics reveals a complex functional ground state of neural stem cells

Phedias Diamandis¹,⁴, Jan Wildenhein⁵, Ian D Clarke¹,², Adrian G Sacher¹,², Jeremy Graham¹,², David S Bellows⁶, Erick K M Ling¹,²,⁵, Ryan J Ward¹,²,⁵, Leanne G Jamieson¹,²,⁵, Mike Tyers³,⁴ & Peter B Dirks¹,²,⁵,⁶


Figure 2: Identification of potent NPC-specific compounds. (a-f) Dose-response curves and chemical structures of controls: cycloheximide (a), etoposide (b) and carboplatin (d), and of selected newly identified compounds: dihydrocapsaicin (d), apomorphine (e) and PAPP (f). Each plot shows the fitted sigmoidal logistic curve to MTT proliferation assay readings of both astrocytes (- - -) and neurosphere cultures (- - -). Values represent the mean and

IP?
What's the structure?
Are they in our file?
What's similar?
Pharmacology data?
What's the target?
Known Pathways?
Competitors?
Working On Now?
Connections to disease?
Expressed in right cell type?
• 3-year Innovative Medicines Initiative project

• Integrating chemistry and biology data using semantic web technologies

• Open source code, open data and open standards

• Academics, Pharma companies, Publishers
ChemSpider Contributions

• The host of the chemistry services
  – Supplier of “standardized” chemical data files
  – Chemistry searching (structure, substructure etc)
  – Curator and data quality checking

• Presently rolling out the Open PHACTS chemical registration system
• FP7 Initiative. PharmaSea: increasing value and flow in the marine biodiscovery pipeline (2012-2017)

• Improve the quality, volume and value of active agents discovered in the marine environment and increase the speed at which they can be delivered
• Dereplication via ChemSpider
• Hosting of natural products datasets
• Integrated storage of analytical data (ACD/Labs)
• Analytical data algorithms & integration
  – Mass spec searching – predicted fragmentation
  – NMR feature searching – NMR prediction
  – Computer-assisted structure elucidation
• Integration to ACD/Structure Elucidator
UK Chemical Database Service

National Chemical Database Service
The RSC is now hosting the EPSRC National Chemical Database Service.

What is the RSC's vision for the Service?
We intend to build the Service for the future - to develop a chemistry data repository for UK academia, and to build tools, models and services on this data store to increase the value and impact of researchers' funded work. We will continue to develop this data store through the lifetime of the contract period and look forward to working with the community to make this a world-leading exemplar of the value of research data availability.

The Service also offers access to a suite of commercial databases and services. Building on the databases most popular with the user community we will deliver new data and services and optimize the offering based on user feedback.

What services are available?
The currently available database services are the Cambridge Structural Database, the Inorganic Crystal Structure Database, ACDI/Lab2, SPRESiWeb and Accelrys' Available Chemicals Directory. The Service will also include integrated access to the RSC's award winning ChemSpider database and other selected RSC current awareness databases. As 'live' dates for other services become clear, they will appear here.

Frequently asked questions
Click the links to read the answers:
• Why isn't the Service continuing to offer the same database services as before?
• What about the other databases currently provided? What is different to the Daresbury CDS?
Ilab Integration – NMR DB Searching

**Modules**
- PhytoChem
- ADME
- Toxicity
- NMR
  - C NMR Predictor
  - H NMR Predictor
  - N NMR Predictor
  - P NMR Predictor
- C NMR DB
- C/H NMR DB
- H NMR DB
- F NMR DB
- N NMR DB
- P NMR DB
- Naming

**C NMR Query setup**
- **Structure**
  - Similar Structure
  - Substructure
  - Exact Structure
  - None

**Common**
- Molecular Formula
- Molecular Weight
- Shifts/Tolerance
- Coupling constants
- Compound Name

**Reference**
- Source
- Year
- Volume
- Number
- Page

**Search results of CNMR DB**
Select a compound from the list on the left.
Each time you select a compound to view its database entry a charge will be applied.

1. C_{27}H_{46}O
   (3b)-cholest-5-en-3-ol
Ilab Integration – NMR Prediction
National Chemistry Data Repository

• Imagine all chemistry related data from all academic projects in the UK in ONE system
• Security model for the data to be embargoed, private or public (available to the entire world!)
• Provide tools for easy data upload, review, automated validation – chemicals, reactions, spectral data, alphanumerical data
• Use the data for algorithm training...
In Discussions At Present

- Develop the world's largest online spectroscopy database of integrated data
- Does ACD/Labs have tools to help?
  - Automated depositions – Silent Automation
  - Processing and validation – Spectrus
  - Databasing – Spectrus DB
  - Web-based integration into ChemSpider
Where else can we get RICH data?

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 derives from diverse array of structurally complex natural products that possess a tetracyclic molecular framework incorporating an imidazole ring. We provide a hypothesis for the formation of the unique and intrinsic chemistry of plausible biosynthetic precursors of agelastatin alkaloids including the first total synthesis of all known agelastatin alkaloids. The gram-scale synthesis of agelastatin A was initiated by the annulation reaction of diazadiazinone and N-alkylation of cyclopentane C-ring and required the development of a carboxydiisopropylidene trapping reagent.

Introduction

The agelastatin alkaloids constitute an intriguing subset of the diverse pyrrole-imidazole family of marine alkaloids that are likely derived from linear biogenic precursors such as clathrodiol (7), hymenidin (8), and oroidin (9, Fig. 1). (−)-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 properties.
DERA : Data Enable the RSC Archive

• How much **data** is in the archive, in the publications and in the supplementary info?
  – How many compounds for ChemSpider?
  – How many syntheses for ChemSpider reactions?
  – How many characterization measurements?
    • Property Data
    • Spectral Data
    • Graphs and charts to be used for modeling?
What if we could capture it all?
The Future of Data

• In Publications
  – Interactive plots, spectra, buy that compound, predict that property
  – Validation of data going INTO publications – NMR prediction, CASE validation, PhysProp comparisons

• From the lab
  – How much data NEVER gets published and is still useful? Failed Reactions? More Open Data...
Acknowledgements

• RSC eScience Team
• ACD/Labs – Pranas Japartas and Karim Kassam
• GGA – Indigo Toolkit and Bingo Cartridge
• The community of depositors
• The Open Source Community
Thank you

Email: williamsa@rsc.org
Twitter: @ChemConnector
Personal Blog: www.chemconnector.com
SLIDES: www.slideshare.net/AntonyWilliams