

# The Royal Society of Chemistry: Research Data and Community Initiatives

Serin Dabb  
Executive Editor, Data

[dabbs@rsc.org](mailto:dabbs@rsc.org)  
@SerinDabb



- Literature alerting services
- Recently acquired *MarinLit*
- **The Merck Index\* *Online***

Open  
Access

Open  
Data

Chemical Sciences  
Article Repository BETA



Home FAQs

Search

Enter your search term e.g. Oxygen

Search

Browse

Articles available: 0 0 1 0 3 8

Select Subject

Select Publisher

Select Article Type

Browse

Clear

Deposit Your Work

Deposit your work and enhance the visibility and impact of your research. It's simple and straight forward, find out more here...

Deposit

Welcome to the Chemical Sciences Article Repository

**CDS** National Chemical  
Database Service

**ChemSpider**  
Search and share chemistry



# Today

1. National Chemical Database Service
  2. ChemSpider
  3. Data Repository (in progress)
  4. Data mining of Royal Society of Chemistry publications (in progress)
- 

# National Chemical Database Service

- An online collection of scientific resources and databases
- Hosted by the RSC,  
- funded by the EPSRC

Free for all UK  
academia

<http://cds.rsc.org>

@cds\_rsc

The screenshot shows the homepage of the CDS National Chemical Database Service. At the top, it features the CDS logo and the text 'National Chemical Database Service'. To the right, it states 'In partnership with EPSRC' and the Royal Society of Chemistry logo. A navigation bar includes links for Home, About CDS, Learning Resources, External Resources, Help, and Contact Us. A paragraph below the navigation bar explains the service's purpose: 'The National Chemical Database Service offers access to a suite of commercial databases and resources, with additional development to create a chemistry data repository to take place. All UK academic institutions are eligible for access to the Service, access will initially be authenticated on institutional IP address - a username and password can be obtained if this is not possible.'

The main content area is a grid of database tiles, each with a logo, name, description, and a 'Further information' link:

- ACD/I-Lab**: Physchem and NMR prediction and database (ACD/Labs Inc.).
- CSD**: Organic and organometallic crystal structures (CCDC).
- DETERM**: Database of thermophysical data for pure substances and mixtures.
- ICSD**: >160,000 inorganic and related crystal structures (FIZ Karlsruhe GmbH).
- Available Chemicals Directory**: Provides supplier information for building block molecules.
- ARChem**: Retrosynthetic tool for chemical analysis of target organic molecules.
- Chemicalize**: Physicochemical property prediction tools with Lipinski-like filters.
- ChemSpider**: An online database of molecules from >400 datasources (RSC).
- SPRESIweb**: Online chemical structure and reaction database (InfoChem GmbH).
- CrystalWorks**: Crystallographic data from the CSD, ICSD and CrystMet (STFC Daresbury).
- Introductory NCDS Video**: A brief introductory video to the National Chemical Database Service.

At the bottom right, it states 'In partnership with the EPSRC' and shows the EPSRC logo with the text 'Engineering and Physical Sciences Research Council'. A 'Watch the video here' link is also present.

ACD/I-Lab



Physchem and NMR prediction and database (ACD/Labs Inc.).

[Further information](#)

*Predicts physicochemical properties, NMR spectra and chemical shifts*  
*Also assesses prediction reliability and includes searchable content databases.*

ACD/I-Lab    Module view    History view    Help    ▼ Options

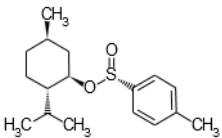
Modules < Predicted Values - <sup>1</sup>H NMR Spectrum (v12.1.0.33936)

Phys Chem

NMR

- C NMR Predictor
- H NMR Predictor
- F 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

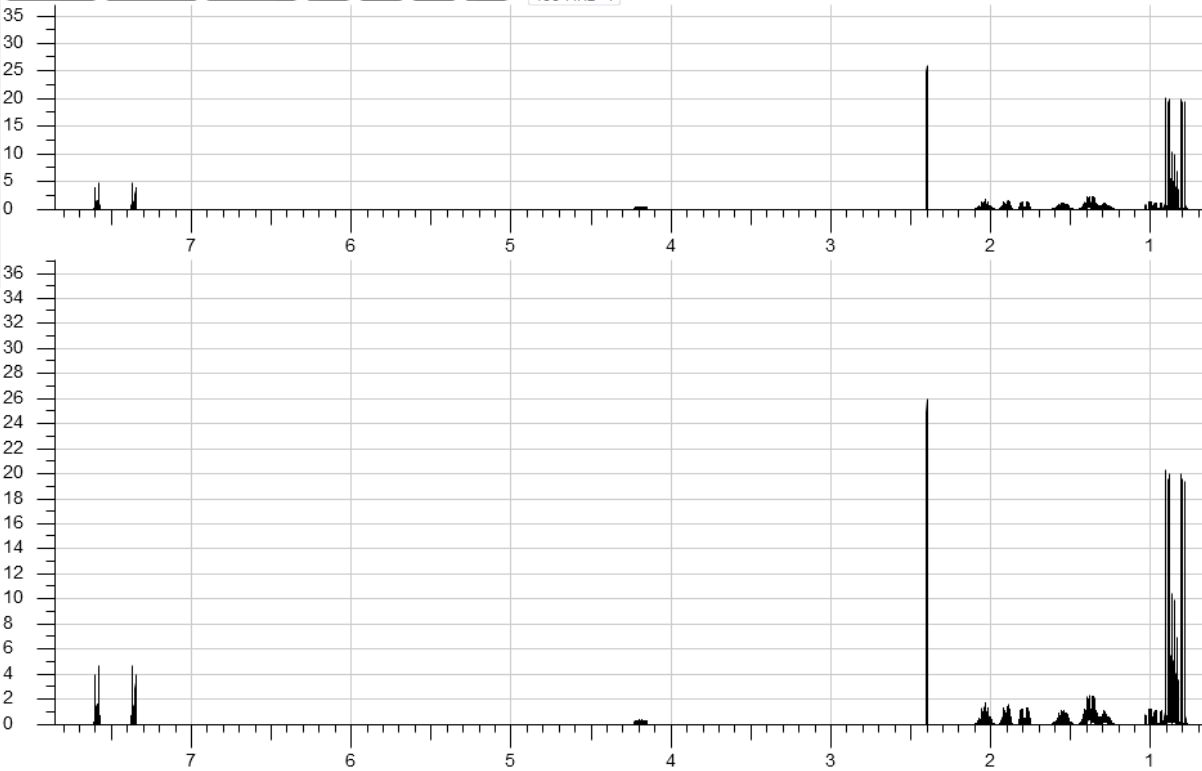
Naming



Shifts		
Atom	Shift	Error
3	0.80	936.62
3	0.80	936.62
3	0.80	936.62
9	0.86	0.73
14	0.86	0.27
14	0.86	0.27
14	0.86	0.27
2	0.90	901.72
2	0.90	901.72

**Coupling constants**

SHIFTS    N<sup>+</sup>    ppm    Hz    Frequency: 400 MHz



Download JCAMP file.    [Download Report](#)



## Learning Resources

### Workshop material

#### W1: Crystallography and Physical Property Prediction (PDF)

This workshop explores crystallography and predicting NMR spectra using:

- [ACD/I-Lab](#)
- [Cambridge Structural Database \(CSD\)](#)
- [Inorganic Crystal Structure Database \(ICSD\)](#)

Download the PDF of the workshop material [here](#).

#### W2: Properties and Reactions of Organic Systems (PDF)

This workshop investigates the properties and reactions of organic systems with:

- [Chemicalize](#)
- [ChemSpider](#)
- [SPRESIweb](#)

Download the PDF of the workshop material [here](#).

### Factsheets

[ACD/I-Lab \(PDF\)](#)

[ARChem \(PDF\)](#)

[Available Chemicals Directory \(PDF\)](#)

[Cambridge Structural Database \(CSD\) \(PDF\)](#)

[Chemicalize \(PDF\)](#)

[DETERM \(PDF\)](#)

[Inorganic Crystal Structure Database \(ICSD\) \(PDF\)](#)

[SPRESIweb \(PDF\)](#)

### Videos

#### ACD/I-Lab

[Webinar](#) outlining [ACD/I-Lab](#), an online structure-based prediction engine and database of physicochemical properties and spectral information.

#### Chemicalize

This [webinar](#) introduces [Chemicalize](#), a public web resource developed by ChemAxon which uses ChemAxon's Name to Structure parsing to identify chemical structures on webpages and other text. Structure based predictions and

Learning materials:  
Videos  
Factsheets  
Worked examples

# National Chemical Database Service


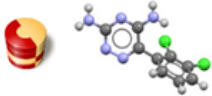







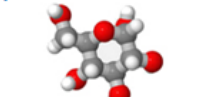


**CDS National Chemical Database Service**

In partnership with **EPSRC**

ROYAL SOCIETY OF CHEMISTRY

Home About CDS Learning Resources External Resources Help Contact Us

The National Chemical Database Service offers access to a suite of commercial databases and resources, with additional development to create a chemistry data repository to take place. All UK academic institutions are eligible for access to the Service, access will initially be authenticated on institutional IP address - a username and password can be obtained if this is not possible.

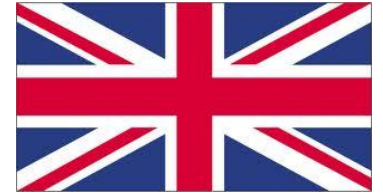
<p><b>ACD/I-Lab</b></p>  <p>Physchem and NMR prediction and database (ACD/Labs Inc.).</p> <p><a href="#">Further information</a></p>	<p><b>CSD</b></p>  <p>Organic and organometallic crystal structures (CCDC).</p> <p><a href="#">Further information</a></p>	<p><b>DETERM</b></p>  <p>Database of thermophysical data for pure substances and mixtures.</p> <p><a href="#">Further information</a></p>	<p><b>ICSD</b></p>  <p>&gt;180,000 inorganic and related crystal structures (FIZ Karlsruhe GmbH).</p> <p><a href="#">Further information</a></p>
<p><b>Available Chemicals Directory</b></p>  <p>Provides supplier information for building block molecules.</p> <p><a href="#">Further information</a></p>	<p><b>ARChem</b></p>  <p>Retrosynthetic tool for chemical analysis of target organic molecules.</p> <p><a href="#">Further information</a></p>	<p><b>Chemicalize</b></p>  <p>Physicochemical property prediction tools with Lipinski-like filters.</p> <p><a href="#">Further information</a></p>	<p><b>ChemSpider</b></p>  <p>An online database of molecules from &gt;400 datasources (RSC).</p> <p><a href="#">Further information</a></p>
<p><b>SPRESIweb</b></p>  <p>Online chemical structure and reaction database (InfoChem GmbH).</p> <p><a href="#">Further information</a></p>	<p><b>CrystalWorks</b></p>  <p>Crystallographic data from the CSD, ICSD and CrystMet (STFC Daresbury).</p> <p><a href="#">Further information</a></p>	<p><b>Introductory NCDS Video</b></p>  <p>A brief introductory video to the National Chemical Database Service.</p> <p><a href="#">Watch the video here</a></p>	<p>In partnership with the EPSRC</p>  <p>Engineering and Physical Sciences Research Council</p>

.....AND

A data repository for the UK academic community



# New data policies



*EPSRC-funded research data is a public good produced in the public interest and should be made freely and openly available with as few restrictions as possible in a timely and responsible manner.*

*Institutional and project specific data management policies and plans should be in accordance with relevant standards and community best practice and should exist for all data. Data with acknowledged long term value should be preserved and remain accessible and useable for future research*

EPSRC Policy Framework on Research Data



**EPSRC**

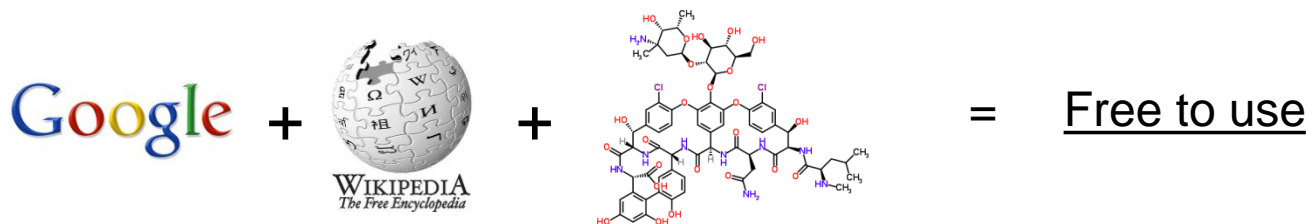
Engineering and Physical Sciences  
Research Council





# ChemSpider

Search and share chemistry



- 28 million structures
- Over 400 different data sources, including:
  - PubChem, ChEBI
  - GSK Malarial compounds
  - Chemical Suppliers Catalogues
  - Patents
  - RSC journals
- Physical properties, spectra, safety information and much more
- Access it anywhere there's internet

[www.chemspider.com/](http://www.chemspider.com/)

# ChemSpider

Search and share chemistry



[About](#) | [More Searches](#) | [Web APIs](#) | [Help](#)

eg. Pyridine

Search

**Simple search** | [Structure search](#) | [Advanced search](#)

eg. Aspirin

Systematic names

1,2-dihydroxybenzene

Synonyms

AIBN

Trade names  
Aspirin

Registry numbers  
7732-18-5

SMILES

O=C(OCC)  
C

InChI

InChI=1/CH4/h1H4

Search



[Advertise](#) | [Sponsor](#)

Advertisements



## What is ChemSpider?

ChemSpider is a free chemical structure database providing fast text and structure search access to over 30 million structures from hundreds of data sources.

Watch [our introduction video](#).

## Search by chemical names

- Systematic names
- Synonyms
- Trade names
- Database identifiers

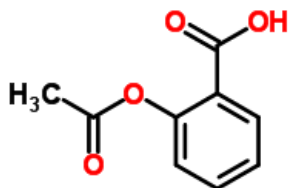
## Search by chemical structure

- Create structure-based queries
- Draw structures in the web page
- Use structure files from your computer

## Find important data

- Literature references
- Physical properties
- Interactive spectra
- Chemical suppliers





Cell 2D 3D Save Zoom

## Aspirin

ChemSpider ID: **2157**

Molecular Formula:  $C_9H_8O_4$

Average mass: 180.157394 Da

Monoisotopic mass: 180.042252 Da

- ▼ Systematic name  
2-Acetoxybenzoic acid
- ▶ SMILES and InChIs
- ▶ Cite this record

### Featured data source



The Merck  
*Index Online*  
has more data on  
this compound

Want to comment  
on this record?

[Leave Feedback](#)

▶ **Names and Identifiers**

▶ ChemSpider Searches

▶ Properties

▶ Spectra

▶ CIFs

▶ Articles

▶ Chemical Vendors

▶ Data Sources

▶ Wikipedia Article(s)

▶ Patents

▶ RSC Databases

## ▼ Spectra



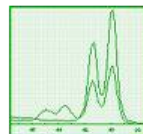
Add

- Type: Infrared

Comments: These data were collected from standards supplied by legitimate manufacturers or synthesized in forensic laboratories under controlled conditions. Supplied by John Meyers, Member of the ChemSpider Advisory Group.

Approved: Yes

Submitted by: [Antony Williams](#)



- 

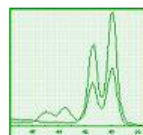
Type: HNMR

Associated Hyperlink: <http://wwwchem.uwimona.edu.jm/spectra/JSpecView>

Comments: HNMR spectrum of Aspirin

Approved: No

Submitted by: [Antony Williams](#)



[OPEN DATA](#)

Download

- 

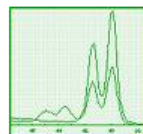
Type: CNMR

Associated Hyperlink: <http://wwwchem.uwimona.edu.jm/spectra/JSpecView>

Comments: CNMR spectrum of aspirin

Approved: No

Submitted by: [Antony Williams](#)



[OPEN DATA](#)

Download

- 



Alt:



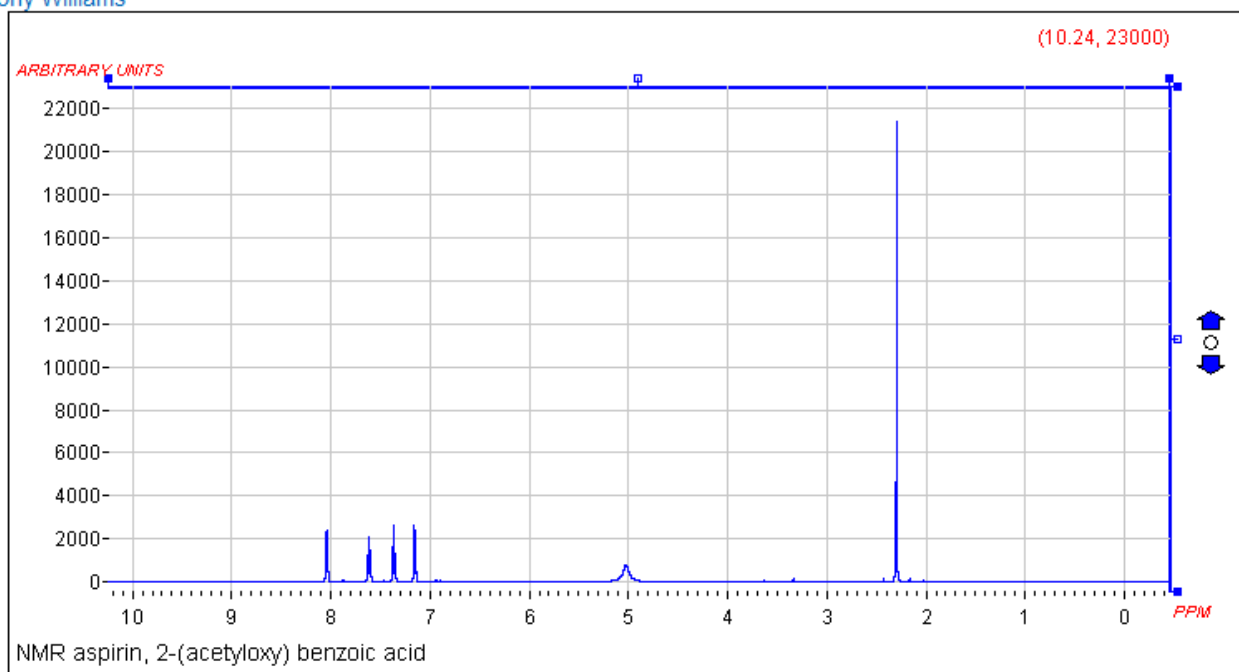
Type: HNMR

Associated Hyperlink: <http://www.chem.uwimona.edu.jm/spectra/JSpecView>

Comments: HNMR spectrum of Aspirin

Approved: No

Submitted by: [Antony Williams](#)



[OPEN DATA](#)

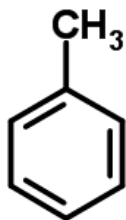

[Download](#)



# Adding and curating data

## *Adds quality and quantity...*

- Users can comment for others to action
- Registered users can:
  - Curate names
  - Add links, spectra, and multimedia resources
- Depositors can add new compounds
- Curators and Master Curators assess and approve additions

Search term: **toluene** (Found by approved synonym)  [2D](#) [3D](#) [Save](#) [Zoom](#)

## toluene

ChemSpider ID: **1108**Molecular Formula: C<sub>7</sub>H<sub>8</sub>

Average mass: 92.138397 Da

Monoisotopic mass: 92.062599 Da

▼ Systematic name

Toluene

[▶ SMILES and InChIs](#)[▶ Cite this record](#)[Wikibox](#)[Embed](#)[Watch this record](#)[Manage data slice](#)

### Featured data source



The Merck  
Index *Online*  
has more data on  
this compound

Want to comment  
on this record?

[Leave Feedback](#)

### Add data to this record

Identifier	Description
<a href="#">Image</a>	<a href="#">Spectrum</a>
<a href="#">CIF</a>	<a href="#">Data source</a>
<a href="#">Publication</a>	<a href="#">DOI</a>
<a href="#">PubMed ID</a>	<a href="#">URL</a>



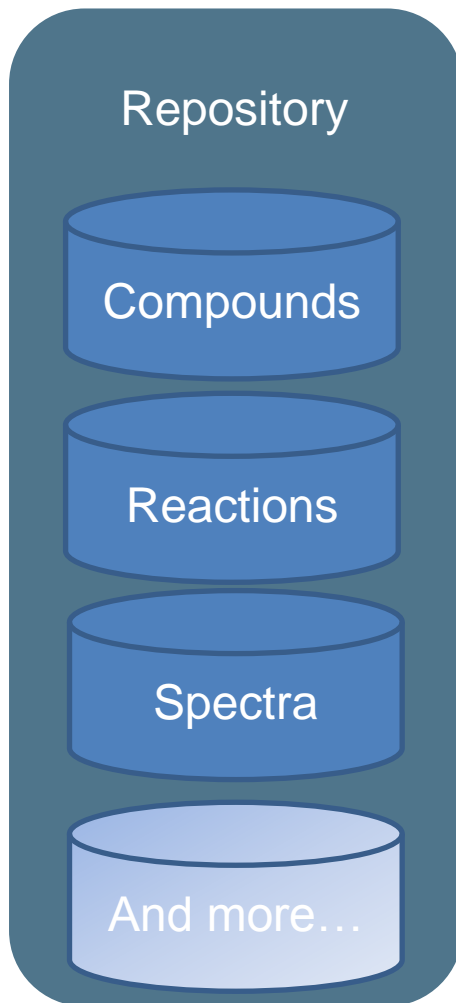
# Chemical Sciences Data Repository

**CDS** National Chemical Database Service

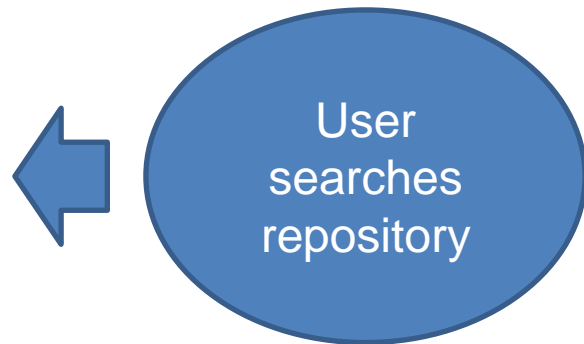
Researcher collects data  
(ELNs, drop box, PC,  
documents)



Validation of  
data



*private*  
*embargoed*  
*public*





# Data Repository: Compounds

## Data Repository Compounds



Simple search Structure search

*eg. Aspirin*

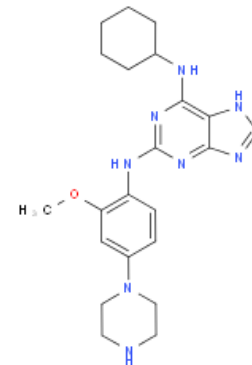
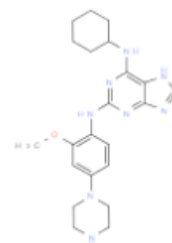
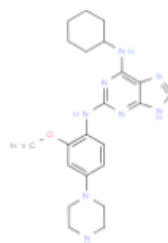
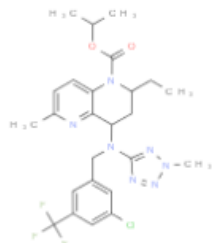
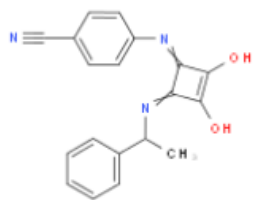
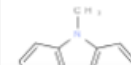
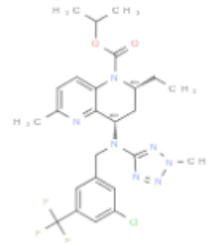
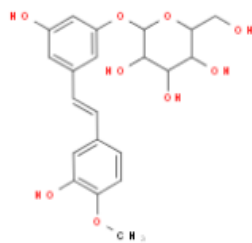
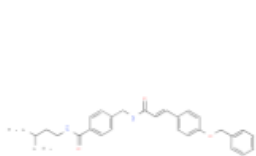
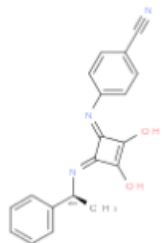
Systematic names	Synonyms	Trade names	Registry numbers	SMILES	InChI
1,2-dihydroxybenzene	AIBN	Aspirin	7732-18-5	<chem>O=C(OCC)C</chem>	InChI=1/CH4/h1H4

Search

1929149 compounds loaded!

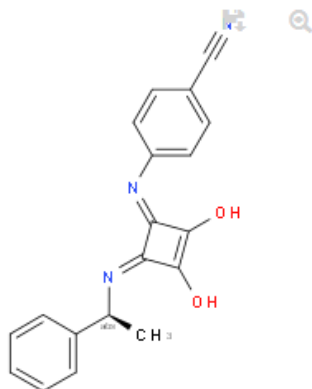
# Data Repository: Compounds

1929149 record(s) found



# Data Repository: Compounds

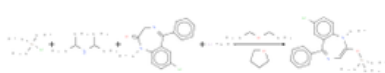
## Compound Record



<b>ID</b>	1
<b>Virtual</b>	No
<b>Molecular Formula</b>	C <sub>19</sub> H <sub>15</sub> N <sub>3</sub> O <sub>2</sub>
<b>Monoisotopic Mass</b>	317.116425 Da
<b>Molecular Weight</b>	317.341309 Da
<b>SMILES</b>	<chem>C[C@H](/N=C1/C(=N/C2C=CC(=CC=2)C#N)/C(O)=C1O)C1C=CC=CC=1</chem> <a href="#">🔗</a>
<b>Non Std. InChI</b>	InChI=1/C19H15N3O2/c1-12(14-5-3-2-4-6-14)21-16-17(-19(24)18(16)23)22-15-9-7-13(11-20)8-10-15/h2-10,12,23-24H,1H-3/b21-16-,22-17-/t12-/m0/s1 <a href="#">🔗</a>
<b>Non Std. InChIKey</b>	IQZAVINXCHHEGZ-CCTGVCNWWA-N <a href="#">🔗</a>
<b>Std. InChI</b>	InChI=1S/C19H15N3O2/c1-12(14-5-3-2-4-6-14)21-16-17(-19(24)18(16)23)22-15-9-7-13(11-20)8-10-15/h2-10,12,23-24H,1-H3/b21-16-,22-17-/t12-/m0/s1 <a href="#">🔗</a>
<b>Std. InChIKey</b>	IQZAVINXCHHEGZ-CCTGVCNWSA-N <a href="#">🔗</a>
<b>ChemSpider ID</b>	28513355

# Data Repository: Reactions

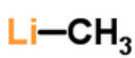
Reaction Record



**ID** 361532  
**RInChI** RInChI=0.02.1S///C6H15N/c1-5(2)7-6(3)4/h5-7H,1-4H3//CH3.Li/h1H3:///X/d-  
**RInChIKey Long** bSA-BEANN---UAOMVDZJSHZZME-UHFFFAOY-N-DVSDBMFJEQPWNO-UHFFFAOY-N-X  
**RInChIKey Short** bSA-BEANN-EANNATPGMB-BULSEIXIGJ-BIQAGPPMEE-NEANN-NEANN-NEANN

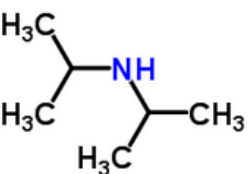
Reaction Components **7**

1




**Li-CH<sub>3</sub>**

2



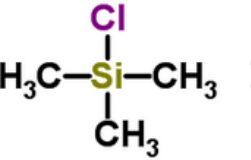
**H<sub>3</sub>C**  
**H<sub>3</sub>C**  
**NH**  
**H<sub>3</sub>C**  
**CH<sub>3</sub>**

3



**H<sub>3</sub>C**  
**Cl**

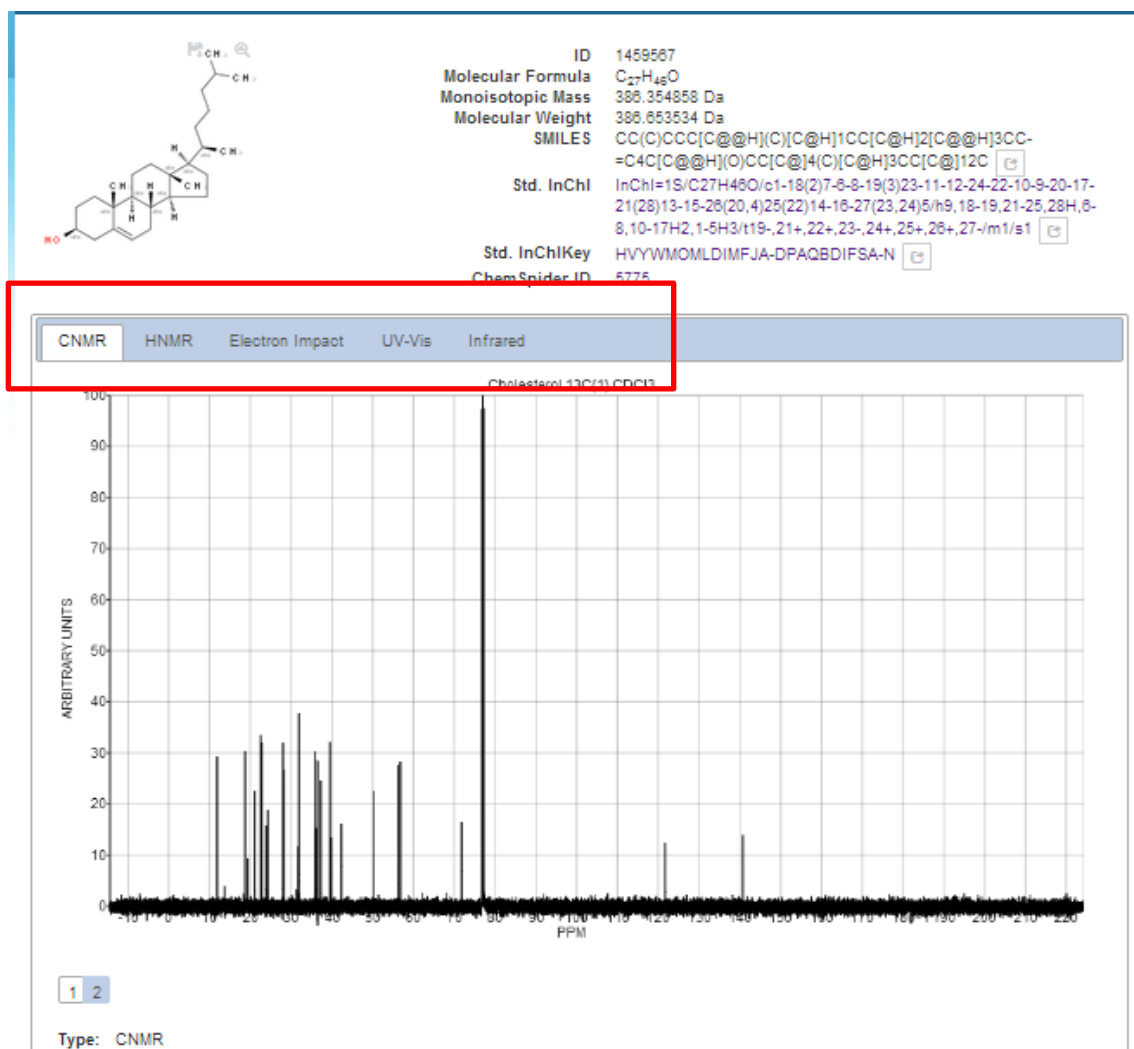
4



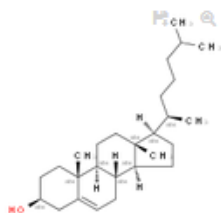
**Cl**  
**H<sub>3</sub>C-Si-CH<sub>3</sub>**  
**CH<sub>3</sub>**

ID: 10915 Type: Reactant ChemSolder ID: 10254338	ID: 8290 Type: Reactant ChemSolder ID: 7624	ID: 9876 Type: Reactant ChemSolder ID: 2908	ID: 8205 Type: Reactant ChemSolder ID: 6157
--	---	---	---

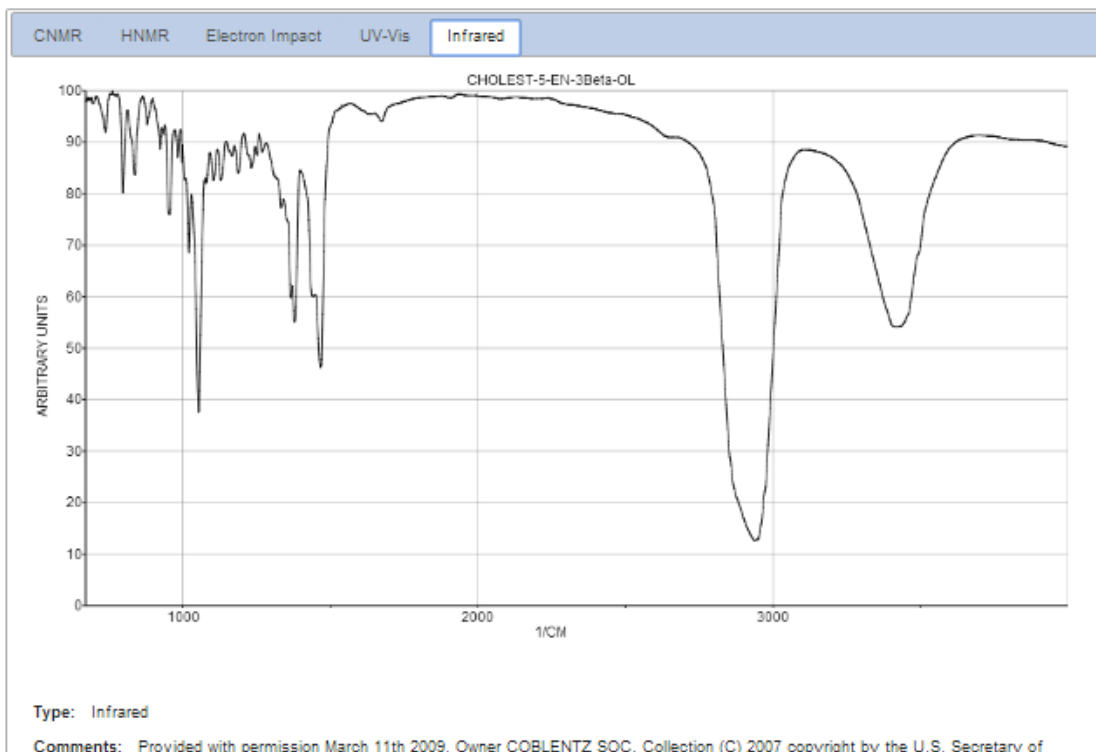
# Data Repository: Spectra



# Data Repository: Spectra



ID 1459567  
Molecular Formula  $C_{27}H_{48}O$   
Monoisotopic Mass 386.354858 Da  
Molecular Weight 386.653534 Da  
SMILES CC(C)CCC[C@H](C)[C@H]1CC[C@H]2[C@@H]3CC=C4C[C@@H](O)CC[C@]4(C)[C@H]3CC[C@]12C  
Std. InChI InChI=1S/C27H46O/c1-18(2)7-8-9-19(3)23-11-12-24-22-10-9-20-17-21(28)13-15-26(20,4)25(22)14-16-27(23,24)5/h9,18-19,21-25,28H,6-8,10-17H2,1-5H3/t19-,21+,22+,23-,24+,25+,28+,27-/m1/s1  
Std. InChIKey HVYWMOMLDMFJA-DPAQBDFISA-N  
ChemSpider ID 5775





# ChemSpider SyntheticPages

## Text

\* mandatory fields

To link to another SyntheticPage, use [[xxx]] where xxx is the number of the SyntheticPage. To link some text in this way use the following [[xxx|text to be linked]], e.g. [[123|Page 123]]

### \* Procedure or Transformation

e.g. "Addition of 2-litioindole to tributyltin chloride" or "alkylation of zirconium tetrachloride with benzyl Grignard" or "Epoxidation of

### \* Compound being made

e.g. "2-indolylstannane", "Zirconium tetrabenzyl" or "Aromatic epoxides".

### \* Chemicals Used (each on a different line)

e.g. TBDMS-Cl, 97% (Sigma-Aldrich)

### \* Procedure

### \* Author's Comments

e.g.

- Particular care should be taken when handling (name of reagent)
- This reaction has been performed with a range of alkenes, in our experience good yields were obtained for electron-rich alkenes, but using certain electron-poor alkenes as substrates (for example, acrylates) we obtained lower yields (40-50%)
- It is important to use freshly distilled tributyltin hydride, we observed that there was only a small fall off in yield (<5%) if we

### Lead Reference

e.g. S. Caddick, K. Aboutayab, K. Jenkins, R. I. West, J. Chem. Soc. Perkin Trans. 1, 1996, 675. DOI: 10.1039/P19960000675

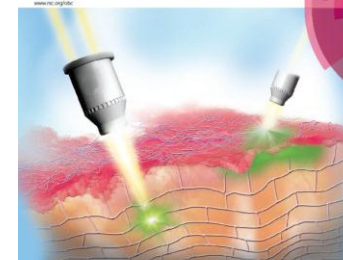


# Unanswered questions

- Level of curation from us
- Number of 'required fields'
- Who can deposit? Who can have access (even if 'private)?
- Standardisation of metadata
- What do people want to search on (properties? spectral features?)
- How does it link to the ESI?
- Should we assign DOIs for data?



# Data mining of historic content



ROYAL SOCIETY OF CHEMISTRY  
 Organic & Biomolecular Chemistry  
 Organic Chemistry of the Royal Society of Chemistry



View Article Online  
 View Journal | View Issue

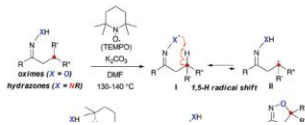
## TEMPO-mediated allylic C–H amination with hydrazones†

Xu Zhu and Shunsuke Chiba\*

Cite this: *Org. Biomol. Chem.*, 2014, **12**, 4567  
 Received 23rd April 2014  
 Accepted 10th May 2014  
 DOI: 10.1039/c4ob00839a  
 www.rsc.org/obc

TEMPO-mediated reactions of alkenyl hydrazones afforded azaheterocycles via  $sp^3$  C–H allylic amination. The transformation is featured by a sequence of remote allylic H-radical shift and allylic homolytic substitution with hydrazone radicals.

Development of methods for oxidative functionalization of  $sp^3$  C–H bonds that provide direct and step-economical approaches to construct functionalized organic structures has



View PDF Version | Previous Article | Next Article

DOI: 10.1039/C4OB00839A (Communication) *Org. Biomol. Chem.*, 2014, **12**, 4567–4570

## TEMPO-mediated allylic C–H amination with hydrazones†

Xu Zhu and Shunsuke Chiba\*

Division of Chemistry and Biological Chemistry, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore. E-mail: [shunsuke@ntu.edu.sg](mailto:shunsuke@ntu.edu.sg); Fax: +65-67911961

Received 23rd April 2014, Accepted 10th May 2014

First published on the web 13th May 2014

TEMPO-mediated reactions of alkenyl hydrazones afforded azaheterocycles via  $sp^3$  C–H allylic amination. The transformation is featured by a sequence of remote allylic H-radical shift and allylic homolytic substitution with hydrazone radicals.

Development of methods for oxidative functionalization of  $sp^3$  C–H bonds that provide direct and step-economical approaches to construct functionalized organic structures has been one of the hottest trends in the area of synthetic organic chemistry. To realize this goal in a regio- and chemo-selective manner, the use of organometallic intermediates as well as metal-carbene and -nitrene species with various transition-metal catalysts has prevailed. On the other hand, free-radical mediated  $sp^3$  C–H bond functionalization by a remote H-radical shift has been recognized for a long time as represented by the Hofmann–Löffler–Freitag reaction,<sup>1</sup> while the inherent violent chemical reactivity of the free radical species often renders these processes difficult to control along with undesired side reactions such as fragmentation and intermolecular H-radical abstraction.

Our group has been interested in the use of stabilized O- and N-radicals derived from oximes and hydrazones, respectively, for remote  $sp^3$  C–H bond oxidation, and recently reported  $sp^3$  C–H oxygenation and amination with oximes and hydrazones mediated by 2,2,6,6-tetramethylpiperidin-1-oxyl (TEMPO) (Scheme 1).<sup>2</sup> In  $\beta$ - $sp^3$  C–H oxygenation with oximes (X =

## Electronic Supplementary Information

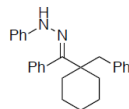
### TEMPO-Mediated Allylic C–H Amination with Hydrazones

Xu Zhu and Shunsuke Chiba\*

Division of Chemistry and Biological Chemistry, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 637371, Singapore.

E-mail: [shunsuke@ntu.edu.sg](mailto:shunsuke@ntu.edu.sg)

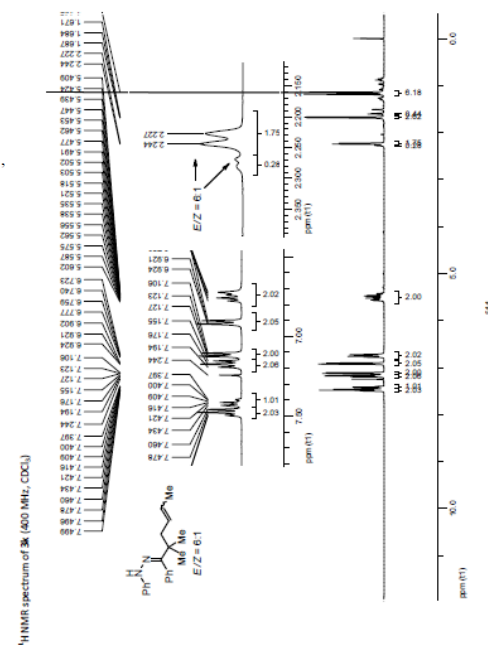
### (Z)-1-((1-Benzylcyclohexyl)(phenyl)methylene)-2-phenylhydrazine (3a):



68% yield as a yellow viscous liquid (single Z-isomer) from (1-benzylcyclohexyl)(phenyl)methanone.<sup>2</sup>

IR (NaCl) 3337, 3053, 2934, 1647, 1601, 1503, 1452, 1308, 1265  $\text{cm}^{-1}$ ;  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ )  $\delta$  1.19–1.35 (1H, m), 1.40–1.49 (4H, m), 1.53–1.56 (3H, m), 1.81–1.86 (2H, m), 2.99 (2H, s), 6.78 (1H, t,  $J = 8.0$  Hz), 6.90–6.92 (5H, m), 7.17–7.26 (7H, m), 7.35–7.43 (3H, m);

$^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ )  $\delta$  22.7, 26.0, 26.9, 34.4, 45.6, 112.5, 119.2, 125.9, 127.8, 128.39, 128.41, 129.0, 129.1, 131.1, 133.7, 138.9, 145.4, 151.5; ESIHRMS: Found:  $m/z$  369.2333. Calcd for  $\text{C}_{26}\text{H}_{29}\text{N}_2$ : (M+H)<sup>+</sup> 369.2331.





# Text Mining

The **N-(β-hydroxyethyl)-N-methyl-N'-(2-trifluoromethyl-1,3,4-thiadiazol-5-yl)urea** prepared in Example 6 , **thionyl chloride** ( 5 ml ) and **benzene** ( 50 ml ) were charged into a glass reaction vessel equipped with a mechanical stirrer , thermometer and reflux condenser .

The reaction mixture was heated at reflux with stirring , for a period of about one-half hour .

After this time the **benzene** and unreacted **thionyl chloride** were stripped from the reaction mixture under reduced pressure to yield the desired product **N-(β-chloroethyl)-N-methyl-N'-(2-trifluoromethyl-1,3,4-thiadiazol-5-yl)urea** as a solid residue

## Results and discussion

### Effect of pH of **tyramine** – **oxalic acid** **eluent** on retention behavior of mono- and divalent **cations**

Since the silanol group on the surface of **silica** gel is a weakly acidic **cation** -exchanger, the **cation** -exchange capacity of a silica gel column is strongly influenced by the pH of the **eluent** . Hence, in order to separate common mono- and divalent **cations** (Li<sup>+</sup>, Na<sup>+</sup>, NH<sub>4</sub><sup>+</sup>, K<sup>+</sup>, Mg<sup>2+</sup> and Ca<sup>2+</sup>) on the Zorbax BP-SIL column in a reasonable time, the effect of the pH of 1 mmol dm<sup>-3</sup> **tyramine** – 0.2 mmol dm<sup>-3</sup> **oxalic acid** as the **eluent** on the retention behavior of these **cations** was investigated. The pH of the **eluent** was adjusted with 1 mol dm<sup>-3</sup> **HNO<sub>3</sub>** .

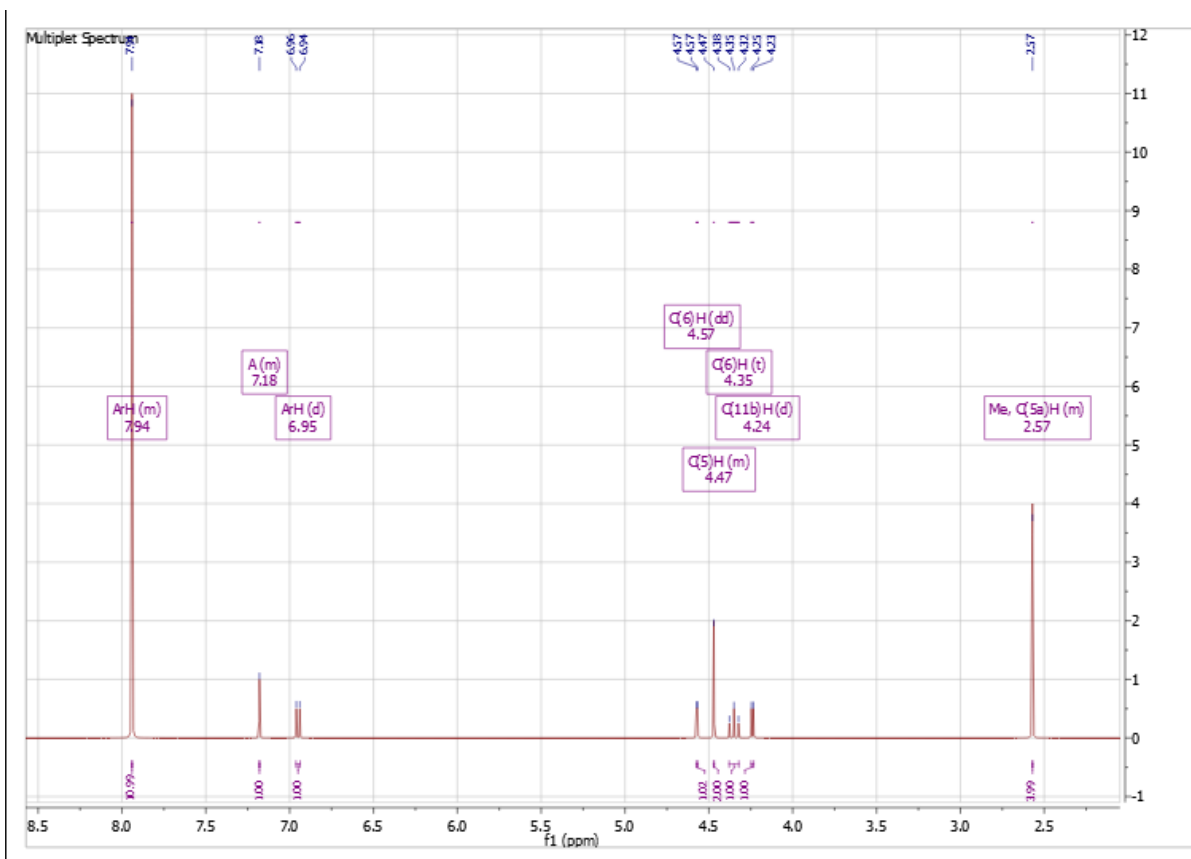
Fig. 1 shows the relationship between the pH of the **eluent** and the retention volumes of the mono- and divalent **cations** . The retention volumes of these **cations** increased with increasing pH of the **eluent** . This is due mainly to an increase in the **cation** -exchange capacity of the Zorbax BP-SIL column by promoting the dissociation of the **silanol** group as a **cation** -exchanger. The retention volumes of the divalent **cations** markedly increased in comparison with those of the monovalent **cations** . This is because the retention volumes of divalent **cations** are strongly influenced by the **cation** -exchange capacity.<sup>9</sup> Complete separation of the divalent **cations** and incomplete separation of the monovalent **cations** were achieved at pH ≥ 4.5. This indicated that **oxalic acid** acted as a complexing agent for the divalent **cations** .<sup>5,6</sup>

Effect of pH of 1 mmol dm<sup>-3</sup> **tyramine** – 0.2 mmol dm<sup>-3</sup> **oxalic acid** as **eluent** on retention volumes of common mono- and divalent **cations** . Column, Zorbax BP-SIL (150 × 4.6 mm id); column temperature, 35 °C; **eluent** , 1 mmol dm<sup>-3</sup> **tyramine** – 0.2 mmol dm<sup>-3</sup> **oxalic acid** . The pH of the **eluent** was adjusted with 1 mol dm<sup>-3</sup> **HNO<sub>3</sub>** . Flow rate, 1 ml min<sup>-1</sup>; detection, indirect UV at 265 nm; injection volume, 20 μl. Sample concentration, 0.2 mmol dm<sup>-3</sup> for monovalent

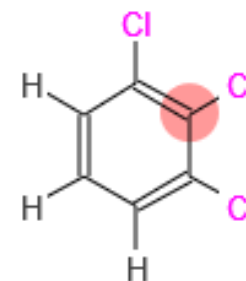
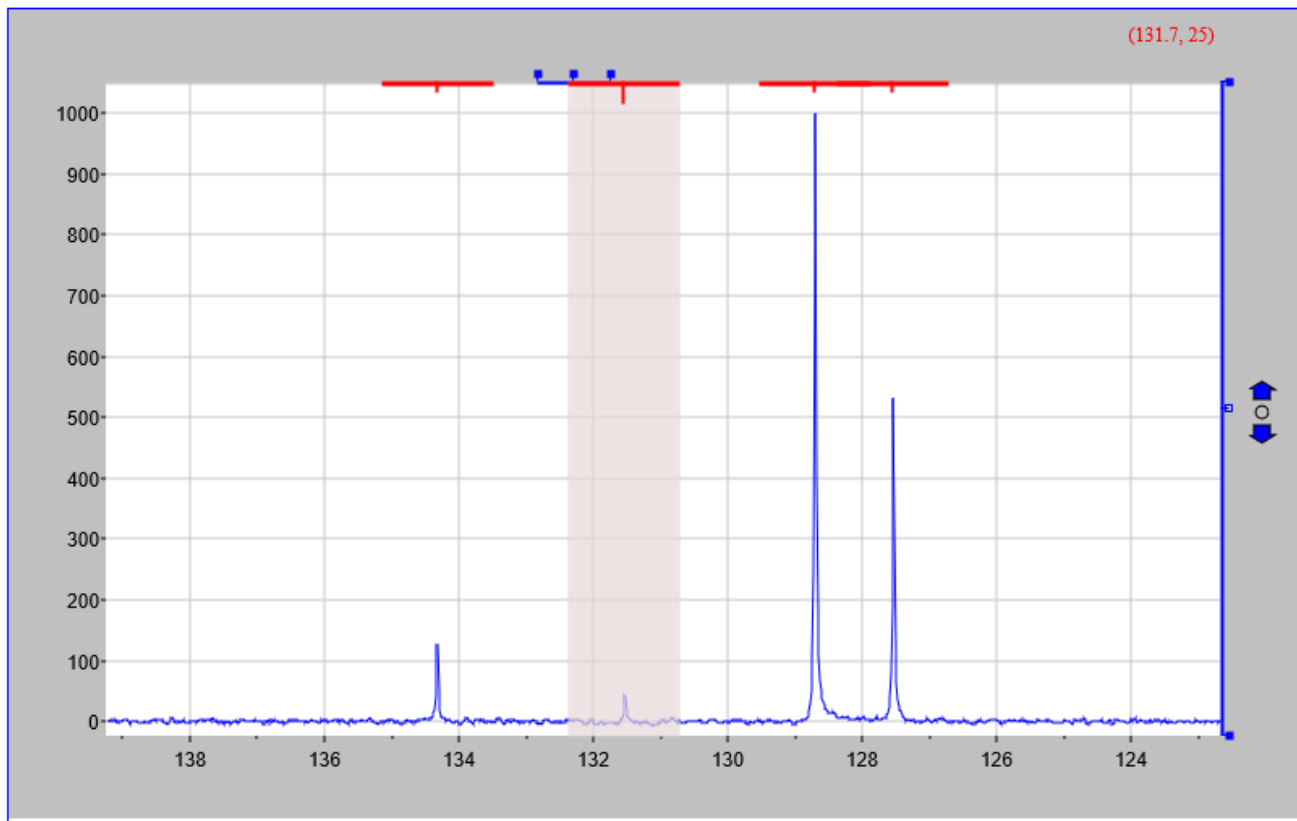


# $^1\text{H}$ NMR ( $\text{CDCl}_3$ , 400 MHz):

$\delta = 2.57$  (m, 4H, Me, C(5a)H), 4.24 (d, 1H,  $J = 4.8$  Hz, C(11b)H), 4.35 (t, 1H,  $J_b = 10.8$  Hz, C(6)H), 4.47 (m, 2H, C(5)H), 4.57 (dd, 1H,  $J = 2.8$  Hz, C(6)H), 6.95 (d, 1H,  $J = 8.4$  Hz, ArH), 7.18–7.94 (m, 11H, ArH)




# Visualisation of spectra



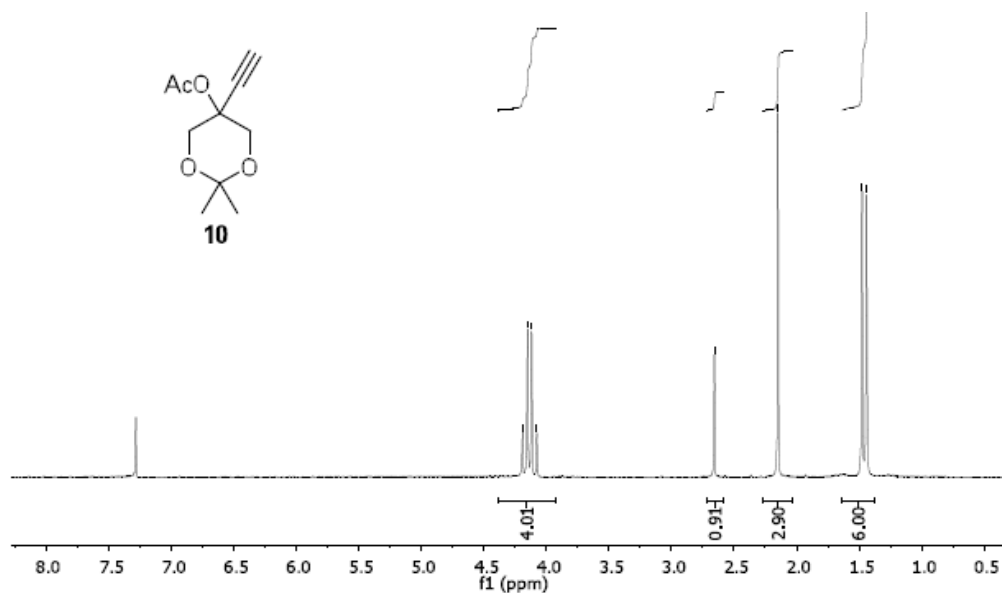
- Javascript viewer with Jmol



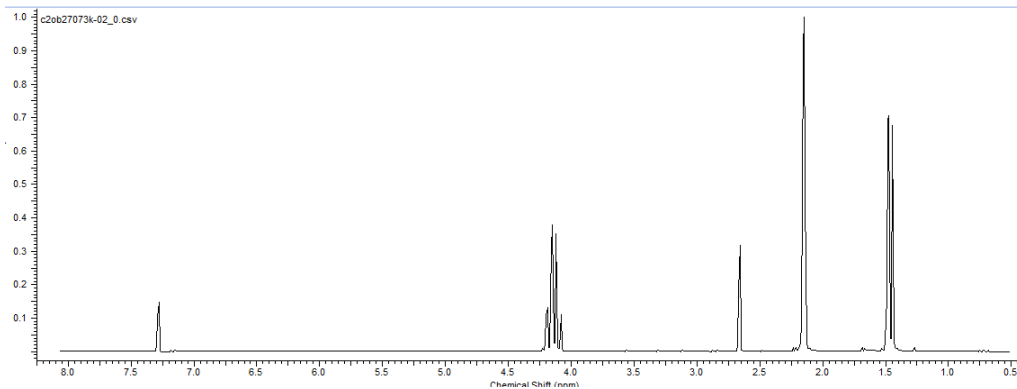
# Figure Spectra into “Real Spectra”?

- We are turning text into structures
  - We are turning text into spectra
  - And we are turning figures into spectra
- 

# Turn "Figures" Into Data

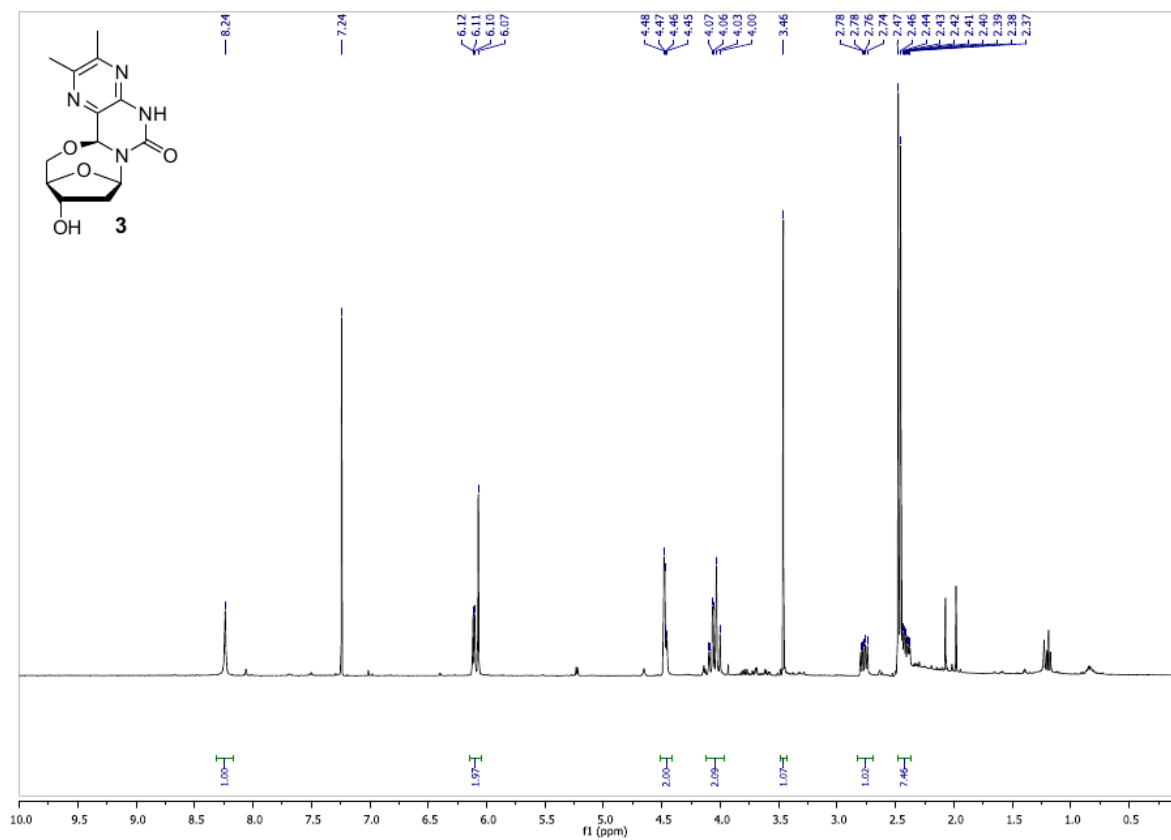


- Figure

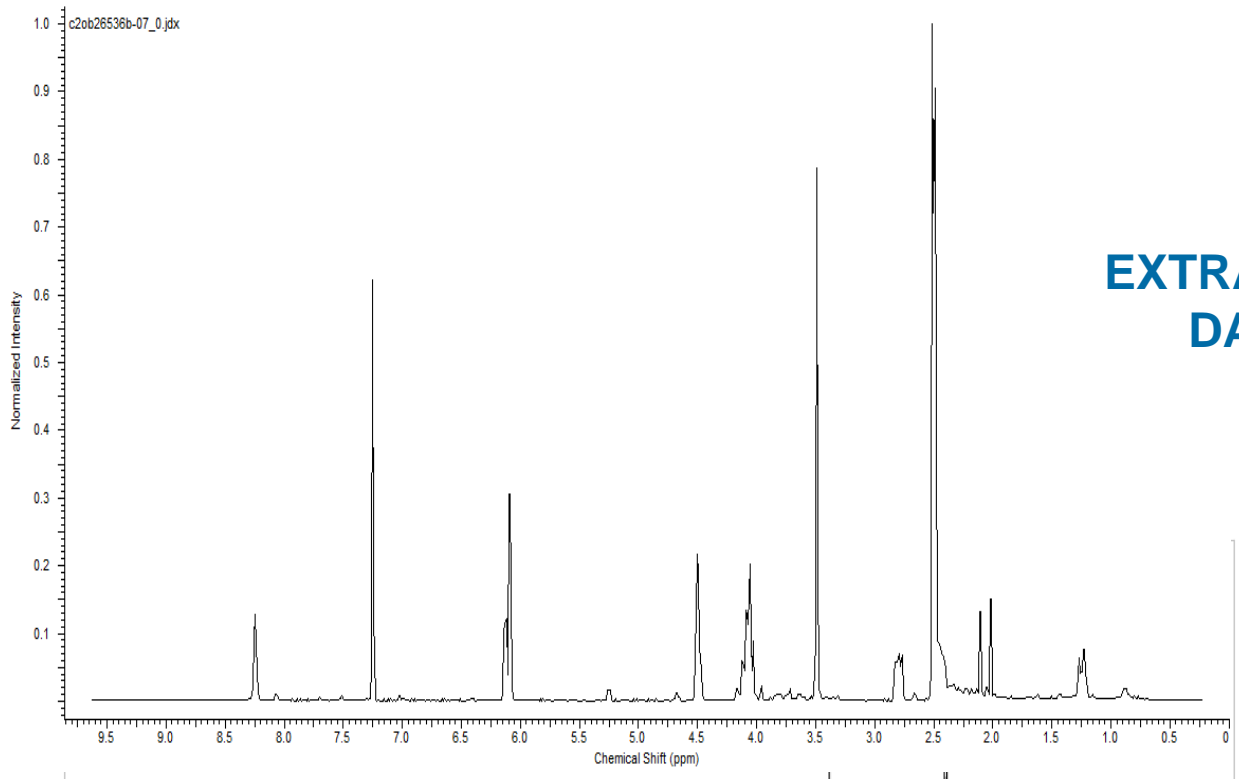


- Extracted Data

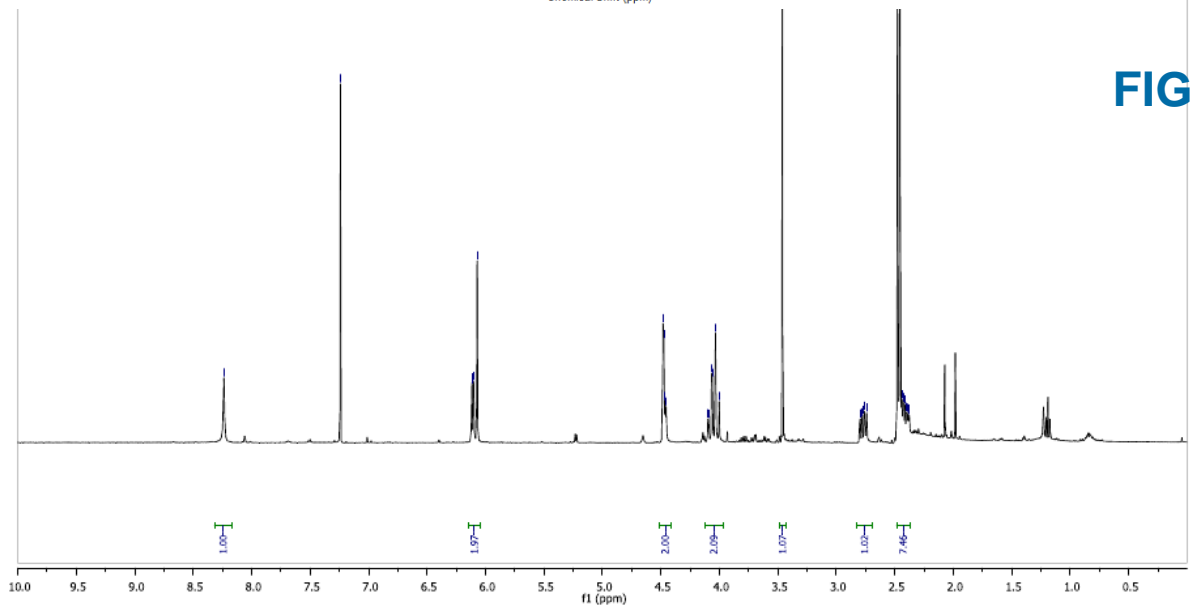
NMR-spectra







**EXTRACTED  
DATA**



**FIGURE**



# Progress: Figures - Spectra

- Validation tests performed with William Brouwer. Good enough to proceed with larger test set
- Ready to run process across larger collection
- Focus on 21<sup>st</sup> century articles only for now

## Example

- Input : 74 supplementary data documents/ 3444 pages
  - 1151 spectra
    - > 80% of peaks extracted to within 1-2 decimal places (ppm)



# Acknowledgments

- Bill Brouwer (Penn State) – Plot2Txt Development
- Carlos Cobas and Santi Dominguez
- Bob Hanson and Bob Lancashire for Jmol/JSpecView Javascript version
- Leah McEwan and Will Dichtel
- ACD/Labs – Provider of spectroscopy tools

Royal Society of Chemistry eScience team:

Antony Williams, Colin Batchelor, Peter Corbett, Ken Karapetyan and Valery Tkachenko

Contact me: [dabbs@rsc.org](mailto:dabbs@rsc.org)





# Questions for you

- **Will you deposit your NMR spectra with ChemSpider?**
  - **Would you like to be involved in future trials of the data repository?**
  - **Feedback on ChemSpider or the National Chemical Database Service?**
  
  - Are there any repositories or data banks members of your community currently use?
  - What incentives would you need to deposit data?
  - What functionality in a repository would you want (for example, searching capabilities)?
  - How is your institution currently handling research data management? What plans are being put in place?
- 