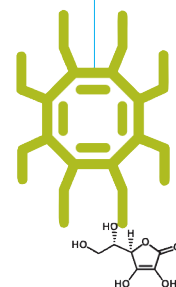


ChemSpider

Building community for chemists



Guide to Database Curation and New Structure Deposition

January 2010

Table of Contents

1.	Introduction	1
	Overview of content of ChemSpider	
	What is Data Curation	
<hr/>		
2.	How to Get Started	2
	Site Registration Logging in & MyChemSpider	
<hr/>		
3.	Posting Comments on ChemSpider	3
<hr/>		
4.	Curating Identifiers on ChemSpider	4-6
<hr/>		
5.	Associating Data Sources with the Record View	7-9
<hr/>		
6.	Depositing New Structures	10-13
	How to upload a structure	
	How to add links to the structure	
<hr/>		
7.	Uploading Spectra to ChemSpider	14

Introduction

1. Introduction

[ChemSpider](#) is a chemistry search engine. It has been built with the intention of aggregating and indexing chemical structures and their associated information into a single searchable repository.

In addition to searching [ChemSpider](#), users can be given the right to deposit their own structures and associated information to the database. Users of [ChemSpider](#) can also be given the right to remove erroneous or add information associated with a chemical compound, not just their own depositions.

[ChemSpider](#) is a rich but, for the present, impure source of information for the chemistry community. As an aggregator of information from multiple sources the challenge is one of curating the data to provide the highest quality information to users.

What is Curation?

For the [ChemSpider](#) database curation is the manual annotation and correction of data. This may include changes to the chemical structures (structure layout for example) addition or removal of identifiers associated with the individual chemical entities, adding links to new data sources, adding new links to associated publications or adding data such as reaction syntheses.

There are two ways to help curate data on [ChemSpider](#), these are:

1. Post comments on a record in order for a Master curator to validate your concern and take appropriate action.
2. As a registered user who has requested curation rights where you participate directly in marking data for master curation or removing erroneous data.

This manual is intended to give guidance in regards to the steps of posting comments on erroneous data, curation of synonyms, adding various types of links to an existing record and how to deposit new structures or analytical data from your personal collections or recent research.

How to get Started

2. How to get Started

In order to curate data, upload structures or add associated information you need to be a registered user.

See <http://www.chemspider.com/controls/Login/RegForm.aspx>

Curators and depositors have additional capabilities relative to registered users. On the sign up form select the roles you wish to have from the options presented.

Your roles:	Role Name	
	Beta tester	<input type="checkbox"/>
	Curator	<input type="checkbox"/>
	Depositor	<input type="checkbox"/>

In order to post comments or upload a structure you must be logged into ChemSpider as a registered user.

Login	Search	Services	Support	Contact Us
ChemSpider - Building Community for Chemists				
LOG IN				
User Name: <input type="text"/>				
Password: <input type="password"/>				
<input type="checkbox"/> Remember me next time.				
<input type="button" value="Log In"/>				
Create new account				
Forgotten your password?				

Posting Comments on ChemSpider

3. Posting Comments on ChemSpider

Anyone can Post Comments regarding erroneous data, this could be an incorrect name or a structure that is incorrectly drawn. Having conducted a search, when you display the Record Table, there are several options to ADD your comments.

The screenshot shows the 'INHERENT PROPERTIES, IDENTIFIERS AND REFERENCES' section of a ChemSpider record. A blue arrow points to the 'Comments' button in the top navigation bar. Another blue arrow points to the 'Comments' button in the 'Quick Links' section. The record details include:

ChemSpider ID:	5775	Quick Links:	Permalink Similar Isomers Wikibox Embed
Empirical Formula:	C ₂₇ H ₄₆ O		
Molecular Weight:	386.6535		
Nominal Mass:	386 Da		
Average Mass:	386.6535 Da		
Monoisotopic Mass:	386.354866 Da		

A chemical structure is also visible on the left side of the record.

If you simply click on the Comments Button you can fill in your comment.

The screenshot shows a dialog box titled 'Please leave your feedback here'. The text inside reads: 'Thank you for your assistance in curating the data on ChemSpider. We welcome your feedback. Please add any comments you have regarding observed errors in any of the associated text, properties or chemical structures.' Below this is a text area containing the comment: 'I believe the registry number for this molecule is incorrect. I recommend checking it.' At the bottom, there is an 'E-mail:' field with the address 'antony.williams@chemspider.com'. A dropdown menu is open, showing options: 'Normal', 'Low', 'Normal', 'High', and 'Extreme'. The 'Normal' option is selected. There are 'Submit Feedback' and 'Cancel' buttons at the bottom right.

You also need to select what you consider to be the appropriate priority for the change and select Submit Feedback.

If you are not logged in then we will need your e-mail address in order to respond but we do respect your privacy and our privacy policy is posted here:

<http://www.chemspider.com/Privacy.aspx>

Curating Identifiers on ChemSpider

4. Curating Identifiers on ChemSpider

As a Curator there are several different ways to add a new identifier or edit existing identifiers.

Adding an Identifier

If a chemical compound has no identifiers, or you simply want to ADD an identifier to a chemical record, then click IDENTIFIER.

[Add DOI](#) [Publication](#) [Description](#) [Link](#) **Identifier** [CIF](#) [Spectrum](#) [Image](#) [Comments](#)
Add: URL

INHERENT PROPERTIES, IDENTIFIERS AND REFERENCES

ChemSpider ID: [Quick Links](#): [Permalink](#) [Similar](#) [Isomers](#) [Wikibox](#) [Embed](#) [Deprecate](#)
Empirical Formula: C₂₇H₄₆O
Molecular Weight: 386.6535
Nominal Mass: 386 Da
Average Mass: 386.6535 Da
Monoisotopic Mass: 386.354866 Da

Systematic Name: (3S,8S,9S,10R,13R,14S,17R)-17-[(1R)-1,5-dimethylhexyl]-10,13-dimethyl-2,3,4,7,8,9,11,12,14,15,16,17-dodecahydro-1H-cyclopenta[a]phenanthren-3-ol
SMILES: O[C@@H]4C/C3=C/C[C@@H]1[C@H](CC[C@]2([C@H]1CC[C@@H]2[C@H](C)CCCC(C)C)[C@@]3(C)CC4
InChI: InChI=1/C27H46O/c1-18(2)7-6-8-19(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+,26+,27-/m1/s1

[Monitor this Article](#)

This will open the Names and Synonyms panel which again has the option to ADD.

NAMES AND SYNONYMS

[Validated by Experts](#), [Validated by Users](#), [Non-Validated](#), [Removed by Users](#), [Redirected by Users](#), [Redirect Approved by Experts](#)

<input type="checkbox"/>	(3S,8S,9S,10R,13R,14S,17R)-17-[(1R)-1,5-dimethylhexyl]-10,13-dimethyl-2,3,4,7,8,9,11,12,14,15,16,17-dodecahydro-1H-cyclopenta[a]phenanthren-3-ol	<input type="button" value="edit"/>	Antony Williams
<input type="checkbox"/>	(+)-Cholesterol	<input type="button" value="edit"/>	
<input type="checkbox"/>	(3b)-cholest-5-en-3-ol	<input type="button" value="edit"/>	
<input type="checkbox"/>	(3beta)-Cholest-5-en-3-ol	<input type="button" value="edit"/>	
<input type="checkbox"/>	(3bêta)-Cholest-5-én-3-ol	<input type="button" value="edit"/>	Antony Williams Antony Williams

The identifier entry box will appear where you can add a Synonym and click an appropriate check box to further define the nature of the identifier.

The checkboxes have the following meanings

- Wiki: This name is on Wikipedia and when linked in this way will open up Wikipedia and display the associated record. Please check the name IS on Wikipedia as typed.
- EINECS/ELINCS numbers are as defined here: <http://en.wikipedia.org/wiki/EINECS>
- Chemical Abstracts registry number as defined here: http://en.wikipedia.org/wiki/CAS_number
- WLN is the Wiswesser Line Notation as defined here: http://en.wikipedia.org/wiki/Wiswesser_Line_Notation
- Beilstein indicates a Beilstein number
- Chemical Name is an invisible field but is useful for internal purposes
- DB ID is a database identifier and is populated into separate areas of the record view. Examples include an NCI number, a PubChem ID etc
- INN is an International Nonproprietary Name as defined here: http://en.wikipedia.org/wiki/International_Nonproprietary_Name
- USAN is the US Adopted Name, the official US Drug Name
- USP, JAN, NF, JP15 and BAN are all other types of names issued by various bodies and will be understood by nomenclature experts but not expanded here for brevity.
- If the name entered is in a foreign language then the pulldown menu can be used to define the language.

When a name is added an e-mail is sent to the Master Curators for review and approval. The additions are SAVED to the database by clicking on the SAVE button. A comment box will pop up for the further addition of comments which would be helpful to the curator when reviewing the suggested identifiers. Click OK when complete.

An e-mail is sent to the Master Curators for approval and checking. When they have done their checking you will receive a response regarding your changes if they have any additional comments to make.

Rejecting and Approving Identifiers

The process of rejecting or approving identifiers involves the selection of one or more identifiers and selection of a state for those names. A common selection is to reject an identifier since it is incorrectly associated with the structure.

There are four states:

Rejected - displays the identifier with a strike out line to indicate that it is not an appropriate match for the compound

Normal – returns a name to a normal state

Confirm – confirms that the identifier is appropriately matched with the compound

Redirect – offers the user to associate the compound with another ChemSpiderID (for example, when there are two tautomers or isomers that the users would like to connect).

State changes can be done on groups of identifiers at one time but clearly it is necessary to separately approve or reject in two separate operations. It is not necessary to save the state changes between these operations.

Guidelines for Removal and Approval of Identifiers

What we are trying to achieve with the actions of approval or rejection of identifiers is state changes which will assist the master curators in speeding up the process of database cleansing.

Master curators have the responsibility of moving curated identifiers to a final approval state of Confirmed or Deleted identifiers based on further research work, reversing the changes or leaving in their present state.

The intention is to remove the associations between structures and identifiers that cause confusion, mislead chemists in their understanding of the chemical structure and provide clarification.

There are various confusions requiring clarification. Specifically:

- All systematic names should match the structure as drawn. All stereochemistry in the name must be represented in the structure shown.
- Any systematic name should be adequate enough to unambiguously convert the name to the matching structure.
- CAS Registry numbers must be for the compound as shown. If the compound shown is the neutral base compound then the registry numbers should not be for the sodium salt or the chloride salt for example.
- Identifiers are not meant to be descriptors per se. For example, an identifier of “One of a series of hexamethylcyclohexanes” is not a good identifier.
- Duplicates can be subtly different but do need to be curated.
- There may be occasions to leave a name associated with a chemical but leave it in a normal state as it is not a valid name.

Associating Data Sources with a Record View

5. Associating Data Sources with a Record view

A record in ChemSpider can be related to one or more particular data sources. The data source itself might have multiple sources of information associated with it.

ASSOCIATED DATA SOURCES AND COMMERCIAL SUPPLIERS	
Data Source	External ID(s)
AKos	AKI-STT-00261475
ASINEX	BAS_00928841
ChemBank	DivK1c_000921, KBio1_000921, KBio2_002236, KBio2_004804, KBio2_007372, KBioSS_002236, NINDS_000921, Prestwick0_000369, Prestwick1_000369, SPBio_002418
ChemDB	3965746, 5699333
ChemIDplus	036505847

An example is shown above in the Associated Data Sources and Commercial Suppliers table for a particular compound. The left column gives the name of the data source and on the right hand side are the individual links (IDs).

This section will demonstrate how to link a chemical compound to an external data source. The link will usually comprise the particular URL of the source data.

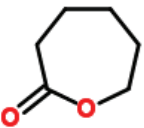
To add information related to a particular Data Source you should be a registered user and logged in.

Integrating via the Record View

In this example we will link the structure for Caprolactone which currently does not have a link to the Wikipedia article:

INHERENT PROPERTIES, IDENTIFIERS AND REFERENCES

2D 3D



ChemSpider ID: [9972](#)

Empirical Formula: $C_6H_{10}O_2$

Molecular Weight: 114.1424

Nominal Mass: 114 Da

Average Mass: 114.1424 Da

Monoisotopic Mass: 114.06808 Da

Quick Links: [Permalink](#) [Similar](#) [Isomers](#) [Wikibox](#)

load save zoom

Systematic Name: oxepan-2-one

ASSOCIATED DATA SOURCES AND COMMERCIAL SUPPLIERS		FILTER
Data Source	External ID(s)	
Alfa Aesar	A10299	
Biological Magnetic Resonance Data Bank	bmse000493	
ChEBI	CHEBI:17915	
PubMed	10027218 , 10148987 , 10187756 , 10224025 , 10227470 , 10370214 , 10370219 , 10385224 , 10416661 , 10416662	
Royal Society of Chemistry	b718180a	
Sigma-Aldrich	167363_ALDRICH , 21510_FLUKA , 241296_ALDRICH	
Thomson Pharma	00277050 , 01775092	
University of Minnesota Biocatalysis and Biodegradation Database	c0059	
ZINC	ZINC00388417	

The link would normally be shown under Data Sources but a link to Wikipedia is absent.

To associate information with a particular structure record view, perform a search to find the record and click on EDIT in the Data Sources Table.

ASSOCIATED DATA SOURCES AND COMMERCIAL SUPPLIERS		FILTER	EDIT
Data Source	External ID(s)		

To ADD information simply click on ADD. A box will appear which allows for entry of an external ID, a name or number for example, the URL to that information (<http://en.wikipedia.org/wiki/Caprolactone>) and the associated data source, in this case Wikipedia. Click to add and then SAVE.

Lhasa Limited	502-44-3	Save Cancel
Biological Magnetic Resonance Data Bank	bmse000493	
Wikipedia	Caprolactone	

A comment box will appear. Fill in an appropriate comment, such as Wikipedia entry added, and Click OK.

As a result the header from the Wikipedia article will be inserted into the ChemSpider record and the Data Source Table will be populated with the link.

INHERENT PROPERTIES, IDENTIFIERS AND REFERENCES

2D 3D

ChemSpider ID: [Quick Links](#): [Permalink](#) [Similar](#) [Isomers](#) [Wikibox](#) [Embed](#) [Deprecate](#)

Empirical Formula: $C_6H_{10}O_2$

Molecular Weight: 114.1424

Nominal Mass: 114 Da

Average Mass: 114.1424 Da

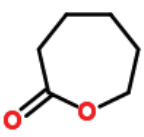
Monoisotopic Mass: 114.06808 Da

Cancel Article Monitor

load save zoom

Systematic Name: oxepan-2-one

SMILES: O=C1OCCCCC1



ASSOCIATED DATA SOURCES AND COMMERCIAL SUPPLIERS FILTER EDIT

Data Source	External ID(s)
PubMed	10027218 , 10148987 , 10187756 , 10224025 , 10227470 , 10370214 , 10370219 , 10385224 , 10413791 , 10416661
Royal Society of Chemistry	b718180a
Sigma-Aldrich	167363_ALDRICH , 21510_FLUKA , 241296_ALDRICH
Thomson Pharma	00277050 , 01775092
University of Minnesota Biocatalysis and Biodegradation Database	c0059
Wikipedia	Caprolactone
ZINC	ZINC00388417

Depositing New Structures

6. Depositing New Structures

This section will look at the process of submitting chemical structures and associated information to the ChemSpider database for the purpose of expanding content by public deposition.

Uploading Structures

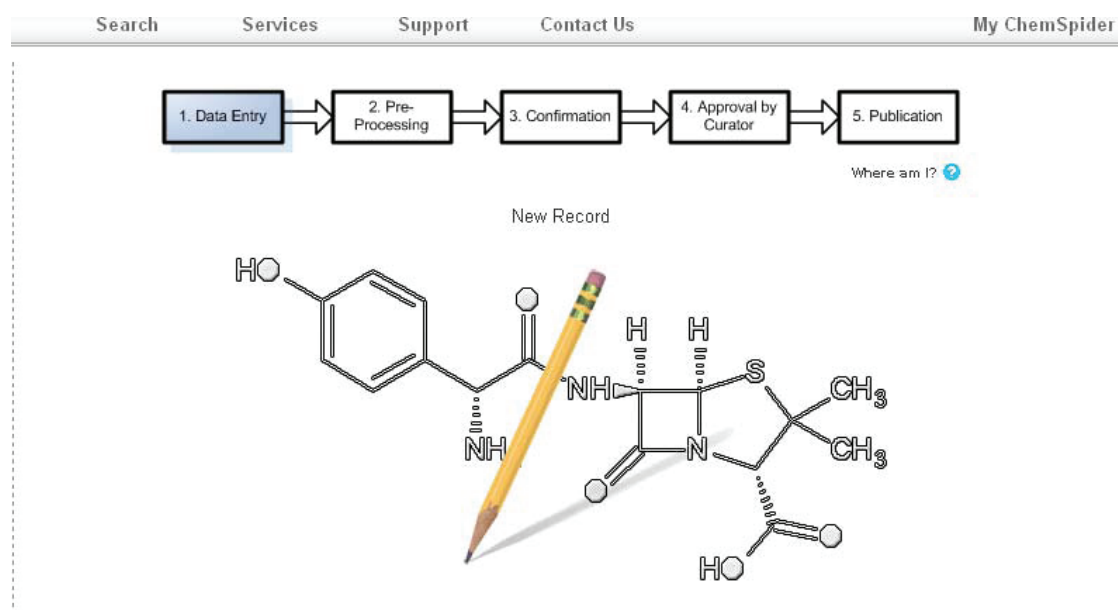
In order to upload structures you need to be a registered user and it is first necessary to login. Once logged in there is an option to click on My [ChemSpider](#) and select Deposit Structures and click on Input Structure. The main steps in the process are:

Step 1: Draw the structure in one of the online applets on [ChemSpider](#) or draw the structure using a desktop package and upload to [ChemSpider](#)

Step 3: Annotate the deposition with text & add identifiers

Step 4: Associate to an online article via URL link, DOI or PubMedID if data are available.

Step 5: Complete the deposition to [ChemSpider](#)



The Structure Drawing Applet will enable you to:

1. Auto-generate a structure from a chemical name, SMILES string or InChI (CONVERT)
2. Upload a structure as a molfile from another drawing package (LOAD)
3. Draw a structure directly in the Applet (DRAW/EDIT)
4. Convert an image of a structure to a "connection table"

There is currently two choices of structure drawing Applet, ACD/Labs SDA or Symex JDraw. Choose the one that is most comfortable for your needs.

Suitable desktop drawing packages are ISIS Draw, ChemDraw, Symyx Draw or ACD/ ChemSketch. For more complex structures it is recommended to draw the structure using a desktop package and save as a molfile and LOAD via the structure drawing Applet.

When the structure is displayed in the Applet, click on ACCEPT. If the structure already exists in the ChemSpider database there will be an indication below the structure and you have the option to click on the link to see the corresponding record. The link should be checked to see whether or not to continue submission.

Having drawn a structure and it has been accepted, you have options to continue with the submission or to cancel.

There are three specific types of information to be added to create the database record at this stage.:

1. Description and Tags – add general information about the chemical entity
2. Identifiers – Add chemical names, systematic names, CAS RNs
3. Links to websites – Add external IDs and associated URLs to websites and Digital Object Identifiers (DOIs)

Additional data such as spectral data can be added later following acceptance of the deposition.

The screenshot displays the submission workflow and the data entry interface. At the top, a horizontal flowchart shows five steps: 1. Data Entry, 2. Pre-Processing, 3. Confirmation, 4. Approval by Curator, and 5. Publication. Below this, navigation buttons include 'Cancel Submission' on the left and 'Finish Submission' on the right. A 'Where am I?' link is also present. The interface shows 'Record 1 of 1' and 'New Delete' options. A chemical structure of a substituted piperidine ring is shown. Below the structure are two main sections: 'External IDs (Hide Details)' and 'Description & Tags (Hide Details)'. The 'External IDs' section contains input fields for 'External ID:' and 'External URL:', with a 'Save' button to the right. A blue arrow points from the 'External ID:' field to the 'Description & Tags' section. The 'Description & Tags' section has a 'Description:' label.

Enter the External ID and associated URL under External IDs. Enter the chemical names & synonyms under Identifiers. Click on ADD after entering each identifier.

If the entry is in error, the identifier can be deleted by clicking on the red cross. The cross itself is not indicative of inaccurate data.

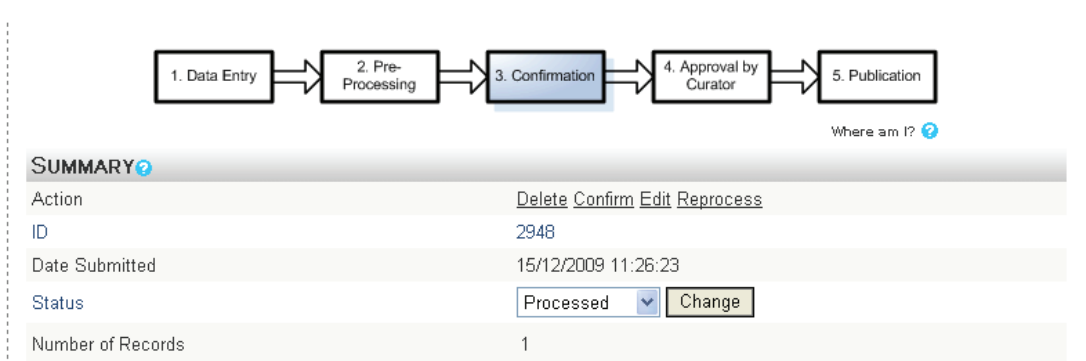
The screenshot shows a web interface for database curation. It features three main sections: 'Identifiers (Hide Details)', 'Links (Hide Details)', and 'Properties (Hide Details)'. In the 'Identifiers' section, there is a text input field containing 'isolaurepan' and an 'Add' button to its right. Below this, the same identifier 'isolaurepan' is displayed with a red 'X' icon to its right. A blue arrow points from the top left towards this red 'X'. In the 'Links' section, there are three buttons: '[Add Publication]', '[Add From PubMed]', and '[Add URL(s)]'. A blue arrow points from the bottom right towards the '[Add Publication]' button. The 'Properties' section is partially visible at the bottom.

Click on Add Publication to give more details of the published article to be linked to.

When the data, as represented above, is ready for submission then click on FINISH SUBMISSION. This completes the entry of the single structure and associated details. The submission is not yet complete. If you are only submitting a single structure in this session then click FINISH and SUBMIT.

If you have other structures to submit then click on NEW and enter a new structure and the associated information. Submissions are NOT complete until Finish and Submit is selected. This means that submissions can be started session to session and retained for editing but not fully submitted until FINISH and SUBMIT is selected.

An e-mail is sent to the depositors inbox, the data are 'pre-processed' which involves checking for structure integrity, property prediction and creation of an SDF file. If there are considered to be no errors then select CONFIRM.



The data are then sent to the master curator for approval. Their process is:

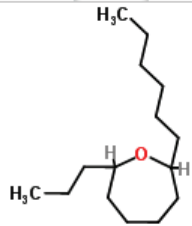
- Check the structure
- Check the associated data
- Check for appropriate labels
- Approve and publish or reject with a reason

Rejected data is sent with annotations to the depositor for their attention, editing and resubmission. When approved an e-mail is sent to the depositor with a link to the published record.

An example of the published record is below.

INHERENT PROPERTIES, IDENTIFIERS AND REFERENCES

2D 3D



ChemSpider ID: [23078757](#) Quick Links: [Permalink](#) [Similar](#) [Isomers](#) [Wikibox](#)

Empirical Formula: [C15H30O](#)

Molecular Weight: 226.4

Nominal Mass: 226 Da

Average Mass: 226.3981 Da

Monoisotopic Mass: 226.229666 Da

load save zoom

Systematic Name: 2-hexyl-7-propyl-oxepane

SMILES: [CCCCC1CCCC\(O1\)CCC](#)

InChI: [InChI=1/C15H30O/c1-3-5-6-7-11-15-13-9-8-12-14\(16-15\)10-4-2/h14-15H,3-13H2,1-2H3](#)

InChIKey: [DYPBWZZUWVXKCZ-UHFFFAOYAH](#)

Std. InChI: [InChI=1S/C15H30O/c1-3-5-6-7-11-15-13-9-8-12-14\(16-15\)10-4-2/h14-15H,3-13H2,1-2H3](#)

Std. InChIKey: [DYPBWZZUWVXKCZ-UHFFFAOYSA-N](#)

ASSOCIATED DATA SOURCES AND COMMERCIAL SUPPLIERS FILTER

Data Source	External ID(s)
Jan Davies	10.1039/b919673k

Organic & Biomolecular Chemistry

The international home of synthetic, physical and biomolecular organic chemistry.



Access to this content has been provided by your institution

- PDF
- HTML article
- ESI (Electronic supplementary information)
- Search for citing articles

Communication

Org. Biomol. Chem., 2010, 8, 39 - 42, DOI: [10.1039/b919673k](#)

Highly efficient synthesis of medium-sized lactones via oxidative lactonization: concise total synthesis of isolaurepan

Makoto Ebine, Yuto Suga, Haruhiko Fujiwara and Makoto Sasaki

A catalytic amount of TEMPO in the presence of $\text{PhI}(\text{OAc})_2$ effected oxidative lactonization of 1,6- and 1,7-dioles, directly affording seven- and eight-membered lactones, respectively, in good yields.



SUPPLEMENTAL INFORMATION

Disclaimer

Description

Lactone

Links & References

Makoto Ebine. [Highly efficient synthesis of medium-sized lactones via oxidative lactonization: concise total synthesis of isolaurepan](#), *Organic & Biomolecular Chemistry*, 2009 [DOI: [10.1039/b919673k](#)]

Uploading Spectra to ChemSpider

7. Uploading Spectra to ChemSpider

ChemSpider has been enriched with NMR, MS, IR and Raman Spectra. This was facilitated by integration of an open source spectral viewing component, JSpecView. 2D NMR spectra can also be submitted but today we can only host them as images.

In order to upload spectra to ChemSpider and associate them to the record view you need to be a registered user and logged in to the website.

You need to locate the record for the substance by conducting a chemical name or structure search. Remember that if you search by chemical name that you should confirm that the structure returned is the exact structure with which you wish to associate spectral data. Check for stereochemistry specifically. In the record view use the ADD Spectrum option, click on the link and a box will appear which enables you to select the spectrum type from a list. Browse your computer for the location of your files which should be in JCAMP format (.Dx or .JDX file extensions).

The screenshot shows a web form titled "Please enter spectrum information here". It contains the following fields and options:

- Spectrum Type:** A dropdown menu with a list of options. The option "HNMR" is currently selected and highlighted in blue.
- Spectrum File:** A text input field with a "Browse..." button to its right.
- Associated Hyperlink:** A text input field.
- Comments:** A large text area for entering additional information.

Below the form, there is a checkbox with the text "Please check this box if you agree with the policies of Open Data. Please read here if you want to know more about our data policies." and an "Add" button. At the bottom of the page, there are links for "Privacy", "Advertising", and "Help".

If you would like to associate the spectrum with a hyperlink back to an associated webpage insert the URL in the hyperlink field. Additional comments can be added in the comments section. These could comprise colour of solution, liquid or solid state, details of sample or details of experimental parameters.

When completed click on ADD Spectrum and the spectrum should be loaded and displayed in the JSpecView applet.