

August 25 - 29, 2019 | San Diego, CA

ACS National Meeting & Expo

Chemistry & Water



CINF: Chemical Nomenclature & Representation: Past, Present & Future

CINF 80: InChI open education resource (OER)

Presenter

Robert E Belford

Univ of Arkansas at Little Rck

Location: Grand Ballroom D, Omni San Diego Hotel

Date & Time: Monday, Aug 26 3:40 PM

Authors

Robert E Belford

Nathan Brown

Ehren C Bucholtz

Jordi Cuadros

Dr. Tanya Gupta

Vincent F Scalfani

Martin A Walker

Prof. Steven P Wathen

August 2017

InChI Trust created an Education Outreach & Training (EOT) Working Group

InChI TRUST Status and Future of the IUPAC InChI: Context and Use Cases InChI TRUST

InChI & Working Group Reports:
Training/Education/Outreach

Robert E. Belford
(rebelford@ualr.edu)

Should there be a working group
on
Training/Education/Outreach?

(Thursday Break-Out Group: 3:00-4:30 PM)

InChI TRUST What are the differences between Training/Education/Outreach? InChI TRUST
(my shot at this)

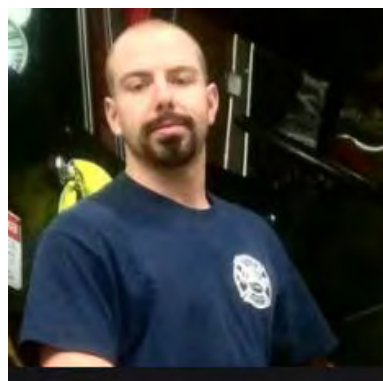
Training – Develop understanding of InChI technology and established use scenarios

Education – Develop understanding of how InChI fits into and can impact the practice of science

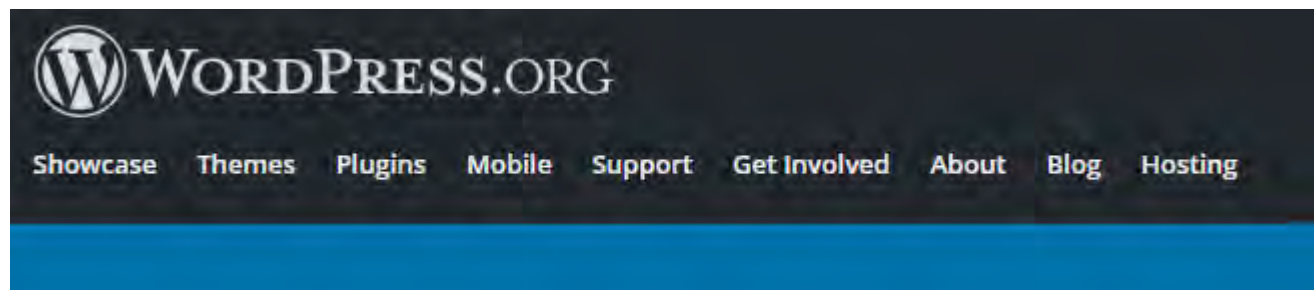
Outreach – Develop applications of InChI technology to new communities and establish new use scenarios.

InChI OER: Open Education Resource

February 2018 InChI Trust Provided Support to Develop InChI OER Website



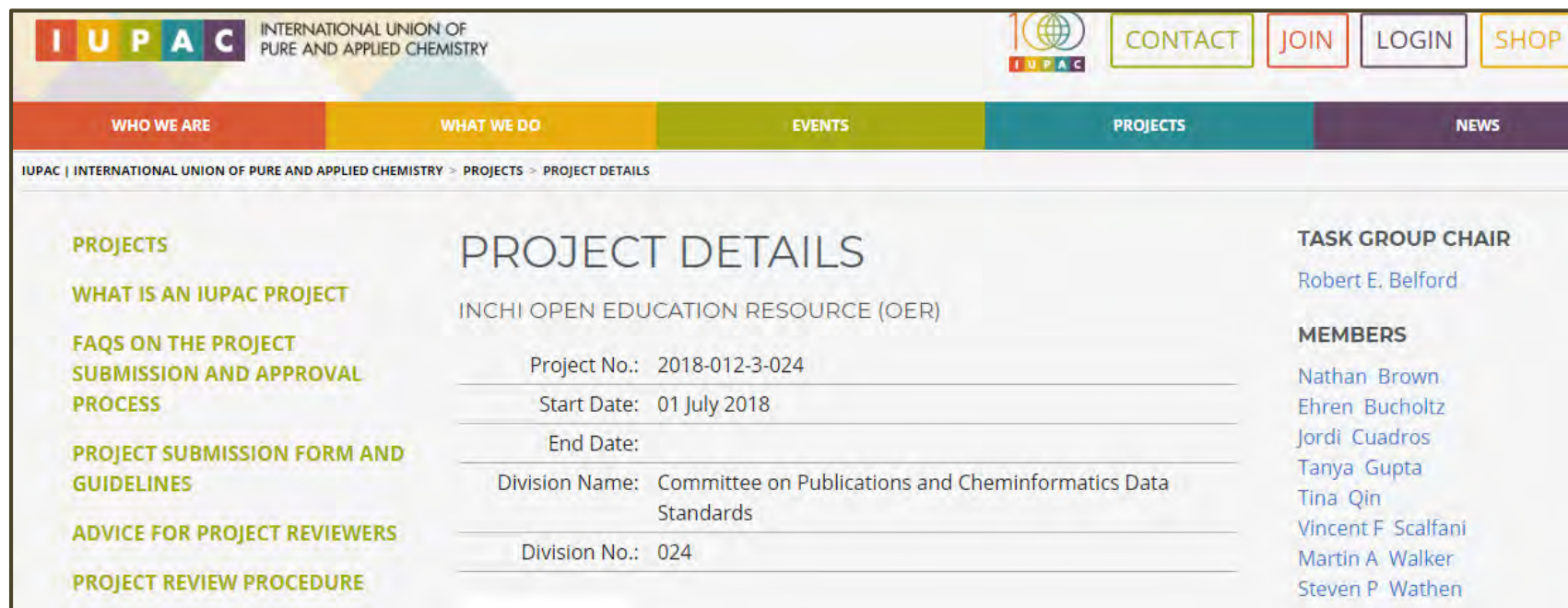
Andrew Cornell (UALR student) tackled the project



InChI OER: Open Education Resource

July 2018

IUPAC InChI formed an Open Education Resource (OER) Task Group



The screenshot shows the IUPAC website's project details page. The header includes the IUPAC logo and navigation links for CONTACT, JOIN, LOGIN, and SHOP. A menu bar highlights the PROJECTS section. The breadcrumb trail reads: IUPAC | INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY > PROJECTS > PROJECT DETAILS.

PROJECTS

- WHAT IS AN IUPAC PROJECT
- FAQS ON THE PROJECT
- SUBMISSION AND APPROVAL PROCESS
- PROJECT SUBMISSION FORM AND GUIDELINES
- ADVICE FOR PROJECT REVIEWERS
- PROJECT REVIEW PROCEDURE

PROJECT DETAILS

INCHI OPEN EDUCATION RESOURCE (OER)

Project No.: 2018-012-3-024

Start Date: 01 July 2018

End Date:

Division Name: Committee on Publications and Cheminformatics Data Standards

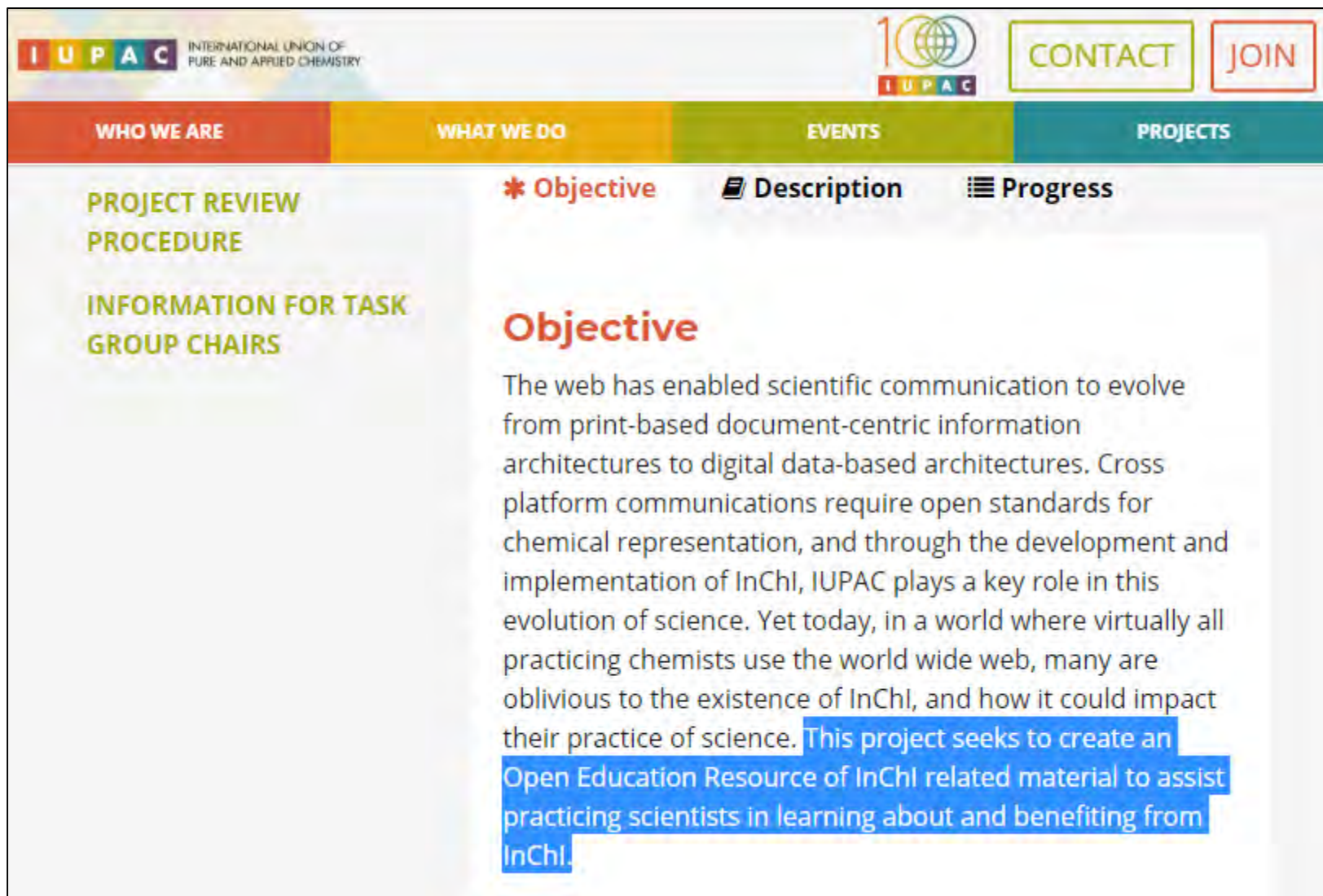
Division No.: 024

TASK GROUP CHAIR

[Robert E. Belford](#)

MEMBERS

- [Nathan Brown](#)
- [Ehren Bucholtz](#)
- [Jordi Cuadros](#)
- [Tanya Gupta](#)
- [Tina Qin](#)
- [Vincent F. Scalfani](#)
- [Martin A. Walker](#)
- [Steven P. Wathen](#)

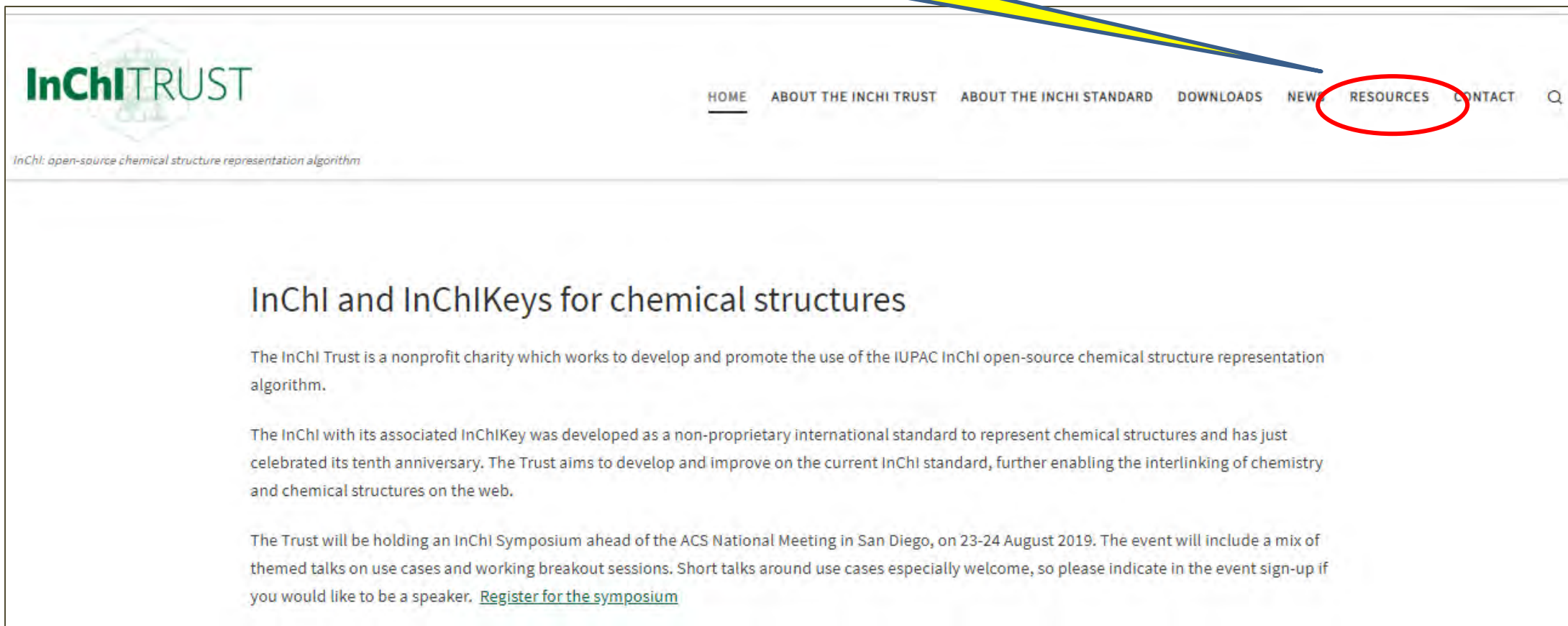


The screenshot shows the IUPAC website header with the logo and text "INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY". Navigation buttons for "CONTACT" and "JOIN" are visible. A menu bar contains "WHO WE ARE", "WHAT WE DO", "EVENTS", and "PROJECTS". The "PROJECTS" section is active, showing a sidebar with "PROJECT REVIEW PROCEDURE", "INFORMATION FOR TASK GROUP CHAIRS", and "Objective", "Description", and "Progress" tabs. The "Objective" tab is selected, displaying the following text:

Objective

The web has enabled scientific communication to evolve from print-based document-centric information architectures to digital data-based architectures. Cross platform communications require open standards for chemical representation, and through the development and implementation of InChI, IUPAC plays a key role in this evolution of science. Yet today, in a world where virtually all practicing chemists use the world wide web, many are oblivious to the existence of InChI, and how it could impact their practice of science. This project seeks to create an Open Education Resource of InChI related material to assist practicing scientists in learning about and benefiting from InChI.

Resources Link goes to InChI OER



InChI TRUST


[HOME](#) [ABOUT THE INCHI TRUST](#) [ABOUT THE INCHI STANDARD](#) [DOWNLOADS](#) [NEWS](#) **[RESOURCES](#)** [CONTACT](#) [Q](#)

InChI and InChIKeys for chemical structures

The InChI Trust is a nonprofit charity which works to develop and promote the use of the IUPAC InChI open-source chemical structure representation algorithm.

The InChI with its associated InChIKey was developed as a non-proprietary international standard to represent chemical structures and has just celebrated its tenth anniversary. The Trust aims to develop and improve on the current InChI standard, further enabling the interlinking of chemistry and chemical structures on the web.

The Trust will be holding an InChI Symposium ahead of the ACS National Meeting in San Diego, on 23-24 August 2019. The event will include a mix of themed talks on use cases and working breakout sessions. Short talks around use cases especially welcome, so please indicate in the event sign-up if you would like to be a speaker. [Register for the symposium](#)



[HOME](#)
[ABOUT THE INCHI TRUST](#)
[ABOUT THE INCHI STANDARD](#)
[DOWNLOADS](#)
[NEWS](#)
[RESOURCES](#)
[CONTACT](#)

InChI: open-source chemical structure representation algorithm

InChI Trust Open Education Resource (InChI-OER)

The InChI OER is an Open Education Resource (OER) devoted to the use of InChI, the International Chemical Identifier. Chemical nomenclature underpins chemical communication and with its release in 2005, the InChI initiative of the International Union of Pure and Applied Chemistry (IUPAC) supports the advancement of chemical nomenclature into the digital age. InChI is evolving to handle reactions, mixtures and other needs of 21st century scientific communication, and yet there is little educational material available on the use of InChI. This OER initiative, [IUPAC project 2018-012-3-024](#), is being created to provide a resource on InChI related resources to assist practicing scientists and educators in learning about and benefiting from the use of InChI.

This resource contains a tag taxonomy-based filter that defaults to Open Access content, which can be further refined by clicking multiple tags. Each post has a short description and an information box that include various download options. OER posts include educational content that has been uploaded to the site along with open access off site content. All material uploaded to this site is open access, and by clicking “non OER” content you can extend the filter to include information on off-site non open access material.

Search

Content Types

OER Post

Non OER

InChI Tags

InChI Algorithm and Description

| PUBLISHED | TITLE/LINK | CONTENT TYPE |
|-----------|---|--------------|
| 08/22/19 | <i>Introduction to the International Chemical Identifier (for Organic Chemistry Undergraduates)</i> | OER |
| | <i>InChILayersExplorer - An Spreadsheet to tech and learn the structure of an InChI</i> | OER |
| 07/29/19 | <i>RDKit InChI Calculation with Jupyter Notebook</i> | OER |
| | <i>InChI OER poster presented at ACS CERM Jun 2019</i> | OER |
| | <i>Batch Chemical IDs Conversion in Spreadsheets</i> | OER |
| | <i>InChI Student Worksheet</i> | OER |
| | <i>The Chemical Translation Service—a web-based tool to improve standardization of metabolomic report</i> | OER |
| | <i>InChI: a user's perspective</i> | OER |

Default Setting: OER Posts

Search

Content Types

OER Post

Non OER

InChI Tags

- InChI Algorithm and Description
- Algorithm
- Software
- Toolkits
- Resolvers
- InChI Modification
- Multiple Component
- RInChI
- Audience
- Researcher

CTRL Click - Select Multiple

Reset

| PUBLISHED | TITLE/LINK | CONTENT TYPE |
|-----------|---|--------------|
| 08/22/19 | <i>Introduction to the International Chemical Identifier (for Organic Chemistry Undergraduates)</i> | OER |
| | <i>InChILayersExplorer - An Spreadsheet to tech and learn the structure of an InChI</i> | OER |
| 07/29/19 | <i>RDKit InChI Calculation with Jupyter Notebook</i> | OER |
| | <i>InChI OER poster presented at ACS CERM Jun 2019</i> | OER |
| | <i>Batch Chemical IDs Conversion in Spreadsheets</i> | OER |
| | <i>InChI Student Worksheet</i> | OER |
| | <i>The Chemical Translation Service—a web-based tool to improve standardization of metabolomic report</i> | OER |
| | <i>InChI: a user's perspective</i> | OER |
| | <i>InChI As a Research Data Management Tool</i> | OER |
| | <i>The Birth of the InChI (video)</i> | OER |

Current Page= 1

Previous **1** 2 3 Next Last

Total Pages= 3

Tag Taxonomy Filter

Both OER and non-OER Content

Search

Content Types

OER Post

Non OER

InChI Tags

- InChI Algorithm and Description
 - Algorithm
 - Software
 - Toolkits
 - Resolvers
- InChI Application
 - Data Extraction
 - Search
- InChI Modification
 - Single Component

CTRL Click - Select Multiple

Reset

| PUBLISHED | TITLE/LINK | CONTENT TYPE |
|-----------|---|--------------|
| 08/22/19 | <i>Introduction to the International Chemical Identifier (for Organic Chemistry Undergraduates)</i> | OER |
| | <i>An Update on the Open Source InChI Project at Google Tech Talks (video)</i> | Non OER |
| | <i>InChILayersExplorer - An Spreadsheet to tech and learn the structure of an InChI</i> | OER |
| | <i>The Semantic Chemical Web: GoogleInChI and other Mashups</i> | Non OER |
| 07/29/19 | <i>RDKit InChI Calculation with Jupyter Notebook</i> | OER |
| | <i>InChI OER poster presented at ACS CERM Jun 2019</i> | OER |
| | <i>Batch Chemical IDs Conversion in Spreadsheets</i> | OER |
| | <i>IUPAC InChI at Google Tech Talks (video)</i> | Non OER |
| | <i>2012 San Diego ACS presentation: Registration system of mcule: InChI is the key (video)</i> | Non OER |
| | <i>InChI Student Worksheet</i> | OER |

Current Page= 1

Previous **1** 2 3 4 5 Next Last

Total Pages= 5

OER Content

| INFORMATION | |
|----------------------------|---|
| Content Type | OER |
| Uploaded By | Jordi Cuadros |
| Download Publication Files | http://www.inchi-trust.org/wp/wp-content/uploads/2019/06/InChILayersExplorer.xlsx |
| License | CC BY 3.0 Unported |
| Content Status | publish |
| Number of Comments | No Comments |
| Date Published | |
| Content Tags | Audience , Content type , Excel , File Type , Graduate , InChI Algorithm and Description , Researcher , Spreadsheet , Undergraduate |

- OnSite Content
 - Downloadable
- Offsite Content
 - Link to Original Source

Non-OER Content

| INFORMATION | |
|--------------------|--|
| Content Type | OER |
| DOI | https://doi.org/10.1186/s13321-018-0277-8 |
| License | cc 4.0 |
| Content Status | publish |
| Number of Comments | No Comments |
| Date Published | |
| Content Tags | Content type , HTML , InChI Modification , Multiple Component , PDF , Publication , RInChI |

- Offsite Content
 - Abstract
 - Link to Original Source
 - DOI

Tag Categories

- Audience
- Content Type
- Curricular Material
- File Type
- InChI Algorithm and Description
- InChI Application
- InChI Modification
- Language

InChI OER: Open Education Resource

Sample of Tag Taxonomy

- Audience
 - Graduate
 - Researcher
 - Undergraduate
- Curricular Material
 - biochemistry
 - Bioinformatics
 - Cheminformatics
 - Organic Chemistry

- Content type
 - Document
 - Poster
 - Presentation
 - Publication
 - Software
 - Application
 - Script
 - Spreadsheet
 - video

- File Type
 - Excel
 - Google Doc
 - Google Sheet
 - HTML
 - Matlab (.m)
 - PDF
- Language
 - Catalan
 - English

- InChI Algorithm and Description
 - Algorithm
 - Software
 - Resolvers
 - Toolkits
- InChI Application
 - Data Extraction
 - Search
- InChI Modification
 - Multiple Component
 - RInChI
 - Single Component

Representation of chemical structures

Authored by: WENDY A. WARR

Abstract:

At the root of applications for substructure and similarity searching, reaction retrieval, synthesis planning, drug discovery, and physicochemical property prediction is the need for a machine-readable representation of a structure. Systematic nomenclature is unsuitable, and notations and fragment codes have been superseded, except in certain specific applications. Connection tables are widely used, but there is no formal standard. Recently the International Union of Pure and Applied Chemistry (IUPAC) International Chemical Identifier (InChI) has started to attract interest. This review also summarizes the representation of chemical reactions and three-dimensional structures.

Link to Content

| INFORMATION | |
|--------------------|--|
| Content Type | Non OER |
| DOI | DOI: 10.1002/wcms.36 |
| Content Link | https://doi.org/10.1002/wcms.36 |
| Content Status | publish |
| Number of Comments | No Comments |
| Date Published | |
| Content Tags | Audience , English , Graduate , Publication , Researcher |

*Note: May or May Not be Open Access

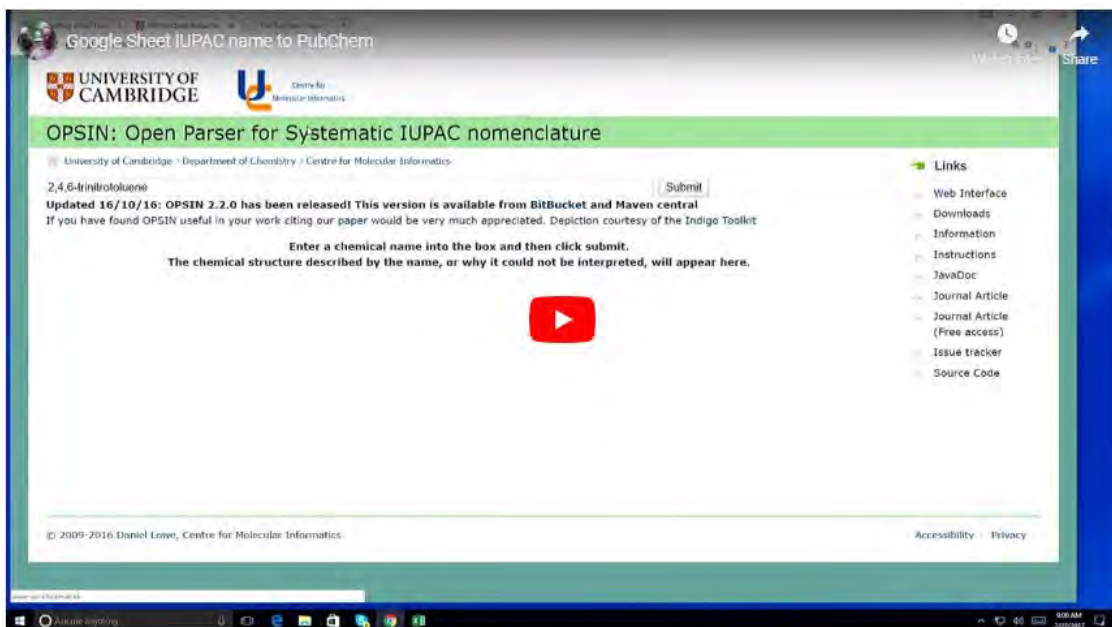
InChI OER: Open Education Resource On Site Content

IUPAC Name2PubChem

Authored by: Robert E. Belford

This submission shows you how to create a smart spreadsheet with Google Sheets that links an IUPAC name to a chemical's PubChem landing page. You may [click here](#) to get a copy of this sheet. This particular sheet uses the Centre for Molecular Informatics OPSIN ([Open Parser for Systematic IUPAC nomenclature](#)) web service to convert the name to an InChI key, which is then appended to a hyperlink to PubChem. You will note that some of the names do not work and this is because those names in the sample sheet are incorrect names. If you paste those names directly into the OPSIN web service, it will tell you were an error in parsing the name occurred.

The following video shows you how to create this Google Sheet and below it is the instructions and code needed. This application takes advantage of the canonical nature of the InChI and its key, and the fact that the key allows you to communicate over the web.



Step 1: Paste your IUPAC names into a column of your spreadsheet

Step 2: Convert IUPAC name to Standard InChI key

type the following script into the top cell of the column you want to place your keys into, and hit enter"

```
=IMPORTDATA("http://opsin.ch.cam.ac.uk/opsin/"&[SPREADSHEET CELL WITH IUPAC NAME]&".stdinchkey")
```

- the ampersand(&) concatenates the cell content to the URL
- the ampersand must be surrounded by quotation marks
- the URL must be in quotation marks

Click on the black box in the bottom right corner of cell and drag down, converting the entire column of names to keys.

Step 3: Hyperlink the key to PubChem

Type the following script into the top cell of the column you want to place your links into, and hit enter"

```
=HYPERLINK("https://pubchem.ncbi.nlm.nih.gov/compound/"&[SPREADSHEET CELL WITH INCHIKEY]&""
```

- the ampersand (&) concatenates the cell content to the URL
- the ampersand must be surrounded by quotation marks
- the URL must be in quotation marks

| INFORMATION | |
|--------------------|--|
| Content Type | OER |
| Uploaded By | Bob Belford |
| Content Link | https://docs.google.com/spreadsheets/d/1CJcwFQ7FF77HAGGKn8hXgEv1DXzL_zGWt4hix_JTIE/copy |
| Content Status | publish |
| Number of Comments | No Comments |
| Date Published | |
| Content Tags | Audience , Content type , Curricular Material , File Type , Google Sheet , Graduate , InChI Algorithm and Description , Researcher , Resolvers , Spreadsheet , Undergraduate |

InChI OER: Open Education Resource On Site Content

Two Ways to Upload Content (Require Log In)

1. Forms
 - (Authors)
2. WordPress Dashboard
 - (Editors)

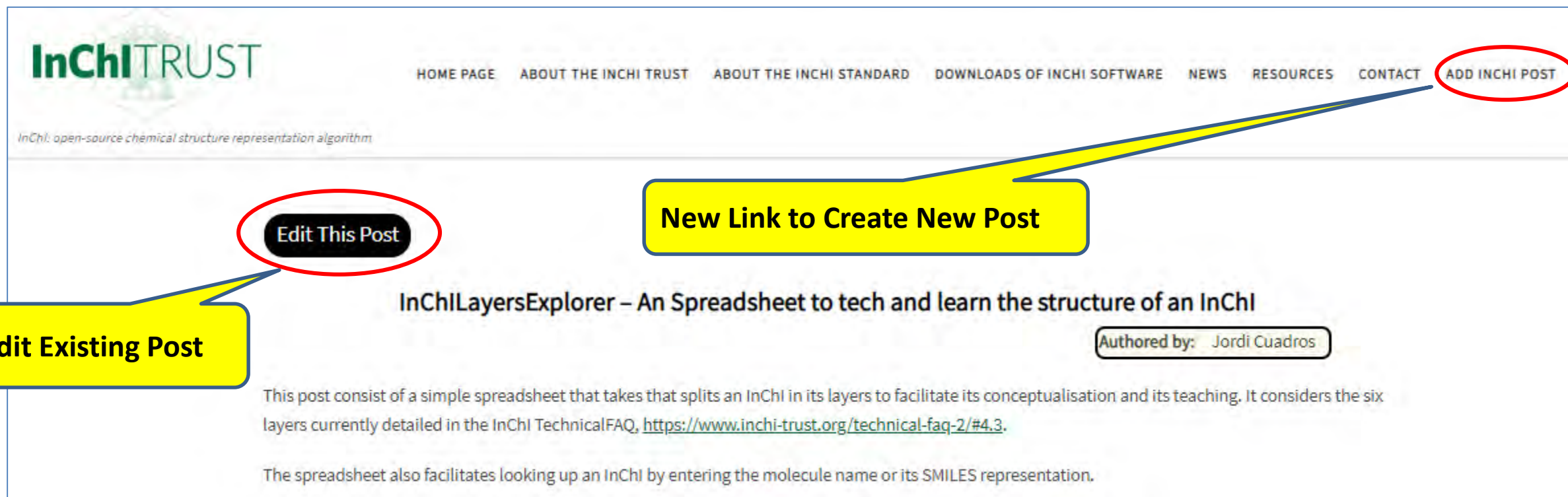
Two Types of Administrative Access

1. Author
 - Can Access Forms
2. Editor
 - Can Access Form
 - Can Access Dashboard
 - Approve Submissions
 - Add Tags to Taxonomy

InChI OER: Open Education Resource

Two Ways to UpLoad Content

1. Site Form (Must Log In)



The screenshot shows the InChI TRUST website interface. The navigation menu at the top includes: HOME PAGE, ABOUT THE INCHI TRUST, ABOUT THE INCHI STANDARD, DOWNLOADS OF INCHI SOFTWARE, NEWS, RESOURCES, CONTACT, and **ADD INCHI POST** (circled in red). A yellow callout box labeled "New Link to Create New Post" points to the "ADD INCHI POST" link. Below the navigation, a blog post titled "InChILayersExplorer – An Spreadsheet to tech and learn the structure of an InChI" is displayed. The author is "Jordi Cuadros". A black button labeled "Edit This Post" is circled in red, with a yellow callout box labeled "Edit Existing Post" pointing to it. The post content includes: "This post consist of a simple spreadsheet that takes that splits an InChI in its layers to facilitate its conceptualisation and its teaching. It considers the six layers currently detailed in the InChI TechnicalFAQ, <https://www.inchi-trust.org/technical-faq-2/#4.3>. The spreadsheet also facilitates looking up an InChI by entering the molecule name or its SMILES representation."

Submissions Require Approval

InChI OER: Open Education Resource

Two Ways to UpLoad Content

Form Fields

Public Page with Information Box



The form contains the following fields:

- SUBMIT** button with a tooltip: `{required, validate_format} [red_generic_field]`
- InChI Points Title**: Text input field.
- Publication Date**: Text input field with a calendar icon.
- Uploaded by**: Text input field with a tooltip: `Authors-Contributors-Roles` and an **@Add new** button.
- InChI Points Content**: Rich text editor with a tooltip: `Rich text editor with toolbar and options`.
- Content Link**: Text input field with a tooltip: `Content URL here` and an **@Add new** button.
- Content Use License**: Text input field with a tooltip: `Divide material must have a license that allows reuse with modification.` and an **@Add new** button.
- DOI**: Text input field with an **@Add new** button.
- Upload Content file**: File upload section with a **Choose file** button, `no file chosen` text, and an **@Add new** button.
- Content images**: Image upload section with a **Choose file** button, `no file chosen` text, and an **@Add new** button.

RDKit InChI Calculation with Jupyter Notebook

Author by: Vincent F. Scafolani

This RDKit InChI Calculation with Jupyter Notebook tutorial is useful to teach the basics of how to interact with InChI using a cheminformatics toolkit in a Jupyter Notebook. The notebook has the following learning objectives:

1. Setup RDKit with a Jupyter Notebook
2. Construct a molecule (RDKit molecular object) from a SMILES string
3. Display molecule images
4. Calculate an InChI for a molecule
5. Calculate InChIs for a list of molecules

| INFORMATION | |
|--------------------|---|
| Content Type | OER |
| Uploaded By | Vincent F. Scafolani |
| Content Link | https://gist.github.com/vfscafolani/f10c4718e9d2e48588c48874a854aa20 |
| License | CC-BY |
| Content Status | publish |
| Number of Comments | No Comments |
| Date Published | July 29, 2019 |
| Content Tags | Audience , Cheminformatics , Curricular Material , English , Language , Script , Software , Undergraduate |

InChI OER: Open Education Resource

Two Ways to Upload Content

1. Title (will also appear at end of URL)

SUBMIT

InChI Posts Title

IUPAC Name2PubChem

2. Uploaded by:

3. Author(s) – “add new” allows multiple authors

Uploaded By

Bob Belford

+ Add new

Authors-Contributors-Roles

Robert E. Belford

+ Add new

4. Body with WYSIWYG Editor. Note, code to embed video is
`[embed]https://youtu.be/oDxMUJ0dNWw[/embed]`

InChI Posts Content

Add Media Fields and Views Toolset Forms Access

Visual Text

Paragraph B I [List Icons] [Link Icon] [Image Icon] [Table Icon]

This submission shows you how to create a smart spreadsheet with Google Sheets that links an [IUPAC](#) name to a chemical's [PubChem](#) landing page. You may [click here](#) to get a copy of this sheet. This particular sheet uses the [Centre for Molecular Informatics OPSIN \(Open Parser for Systematic IUPAC nomenclature\)](#) web service to convert the name to an [InChI](#) key, which is then appended to a hyperlink to [PubChem](#). You will note that some of the names do not work and this is because those names in the sample sheet are incorrect names. If you paste those names directly into the [OPSIN](#) web service, it will tell you were an error in parsing the name occurred.

The following video shows you how to create this Google Sheet and below it is the instructions and code needed. This application takes advantage of the canonical nature of the [InChI](#) and its key, and the fact that the key allows you to communicate over the web.

`[embed]https://youtu.be/oDxMUJ0dNWw[/embed]`

P

InChI OER: Open Education Resource

Two Ways to UpLoad Content

5. Content Link if "Off Site", like Google Sheet, publication,

Content Link (Enter the link to the content here if applicable.)

+ Add new

Content Use License

Onsite material must have a license that allows reuse with modification.

+ Add new

DOI

+ Add new

6. Content Use License. All "Onsite" material must be OER

7. DOI if appropriate

8. File Upload for onsite OER content

Upload Content File

 No file chosen

9. Image Upload for onsite OER content

+ Add new

Content Images

 No file chosen

+ Add new

InChI OER: Open Education Resource

Two Tag Types

Content Types

OER Post

Non OER

InChI Tags

Audience

Graduate

Researcher

Undergraduate

Content type

Document

Poster

SUBMIT

10. Content Type

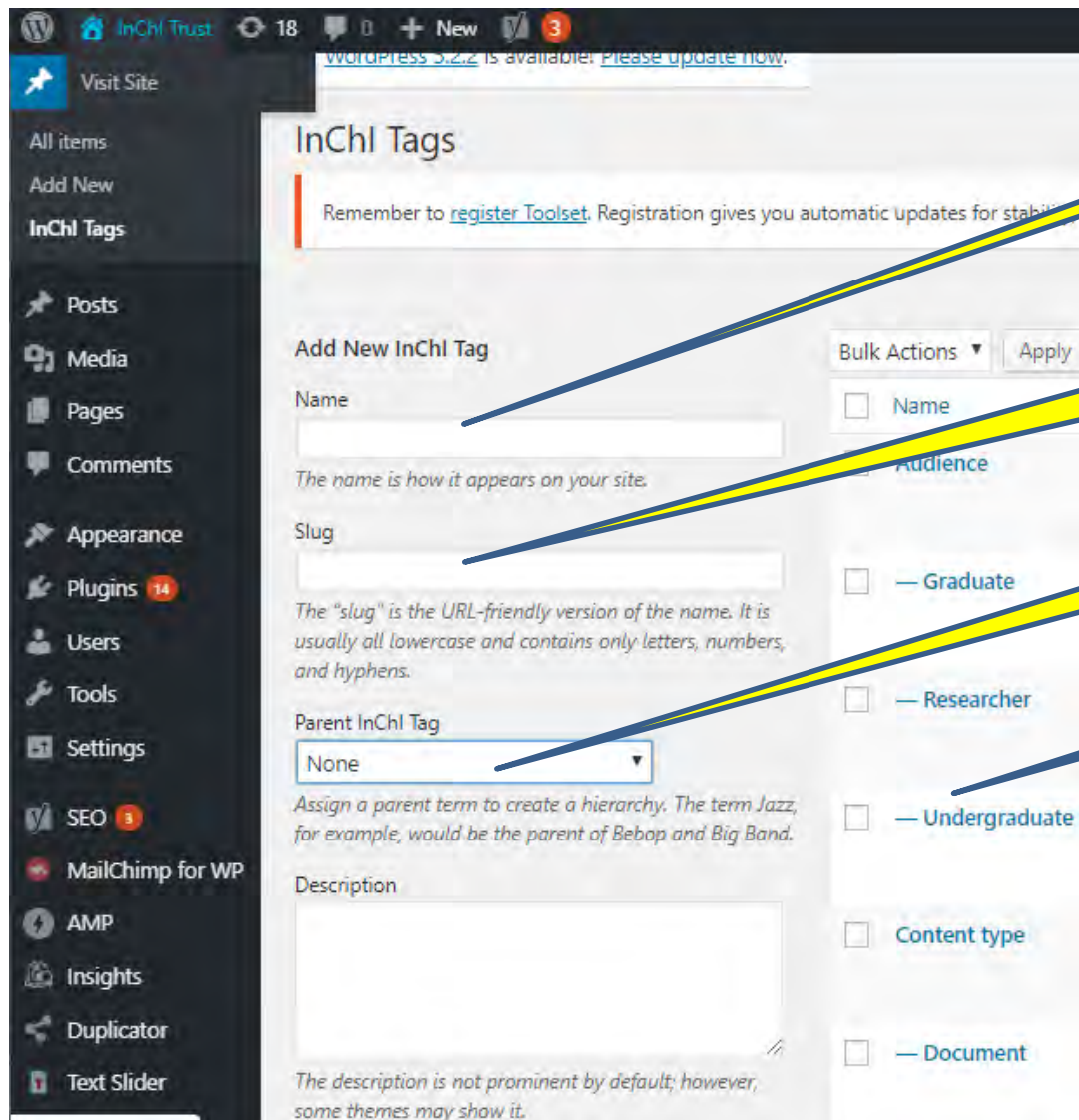
- On-Site Material must be OER
- Off-Site Material may be non-OER

11. InChI Tags

- Controlled Category
- Must tag parents, grandparents....
- Need to use WordPress Dashboard to add new tags

InChI OER: Open Education Resource

Adding Tags to Taxonomy



The screenshot shows the WordPress 'Add New InChI Tag' form. The form includes a 'Name' field, a 'Slug' field, a 'Parent InChI Tag' dropdown menu (set to 'None'), and a 'Description' text area. On the right side, there is a 'Bulk Actions' dropdown and a list of checkboxes for 'Audience' (Name, Graduate, Researcher, Undergraduate) and 'Content type' (Document). Annotations with yellow callout boxes point to the 'Name' field, 'Slug' field, 'Parent InChI Tag' dropdown, and one of the checkboxes in the 'Audience' list.

Tag Name

Tag Slug (Machine Name)-no blank spaces

Parent Tag (must be chosen)

Existing Tag

What is the LibreText HyperLibrary?

 BIOLOGY
 BUSINESS
 CHEMISTRY
 ENGINEERING

 ¡SÍ! ESPAÑOL
 GEOSCIENCES
 HUMANITIES
 MATHEMATICS

 MEDICINE
 PHYSICS
 SOCIAL SCIENCES
 STATISTICS

 WORKFORCE

 STUDENTS

 INSTRUCTORS

 INSTITUTIONS

 VOLUNTEERS

398

Textbooks, Textmaps, and
LibreTexts

154

Courses using LibreTexts

223 million

Students served

\$31 million

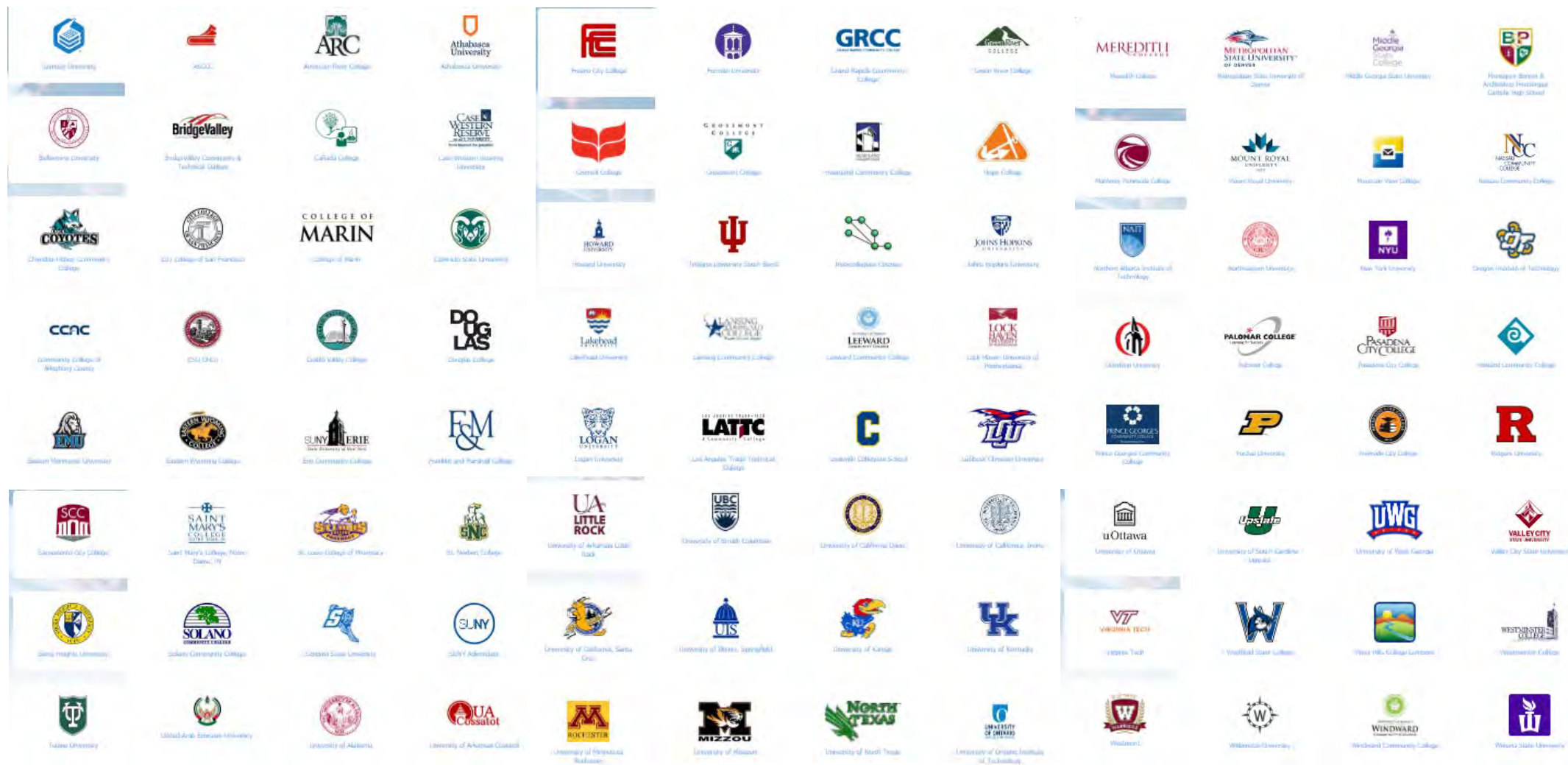
Total Amount Saved

**Interdisciplinary Open Education Resource (OER)
and an
Academic Online Authoring Environment**

InChI OER: Open Education Resource

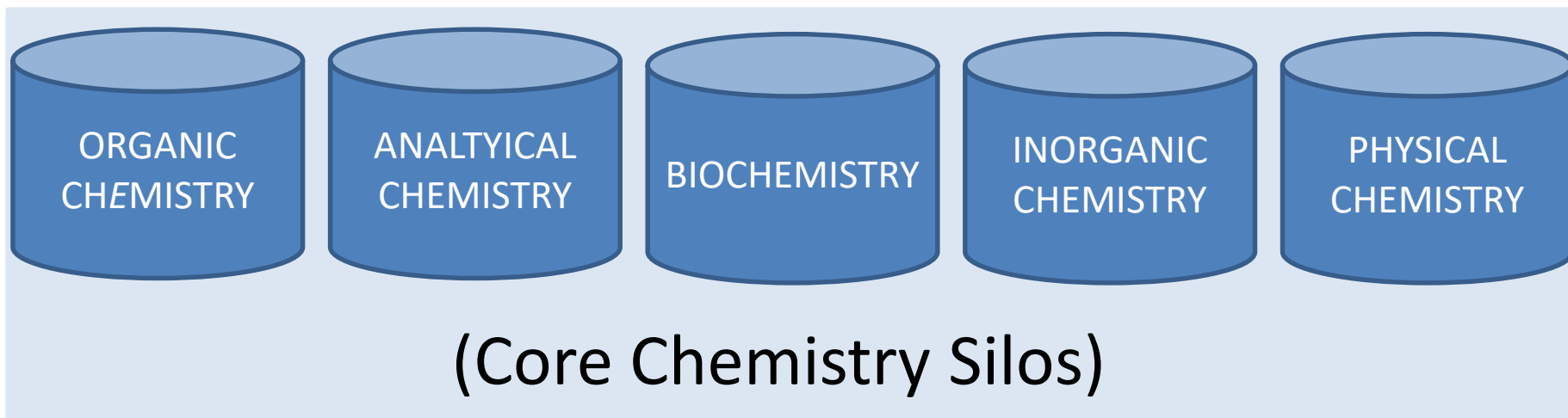
Integrating Content into the Classroom: LibreText

InChI: open-source chemical structure representation algorithm



InChI OER: Open Education Resource

Integrating Content into the Classroom: LibreText



Textbooks are Defined by Chemistry Subdisciplines

InChI OER: Open Education Resource

Integrating Content into the Classroom: LibreText



Ancillary
Material

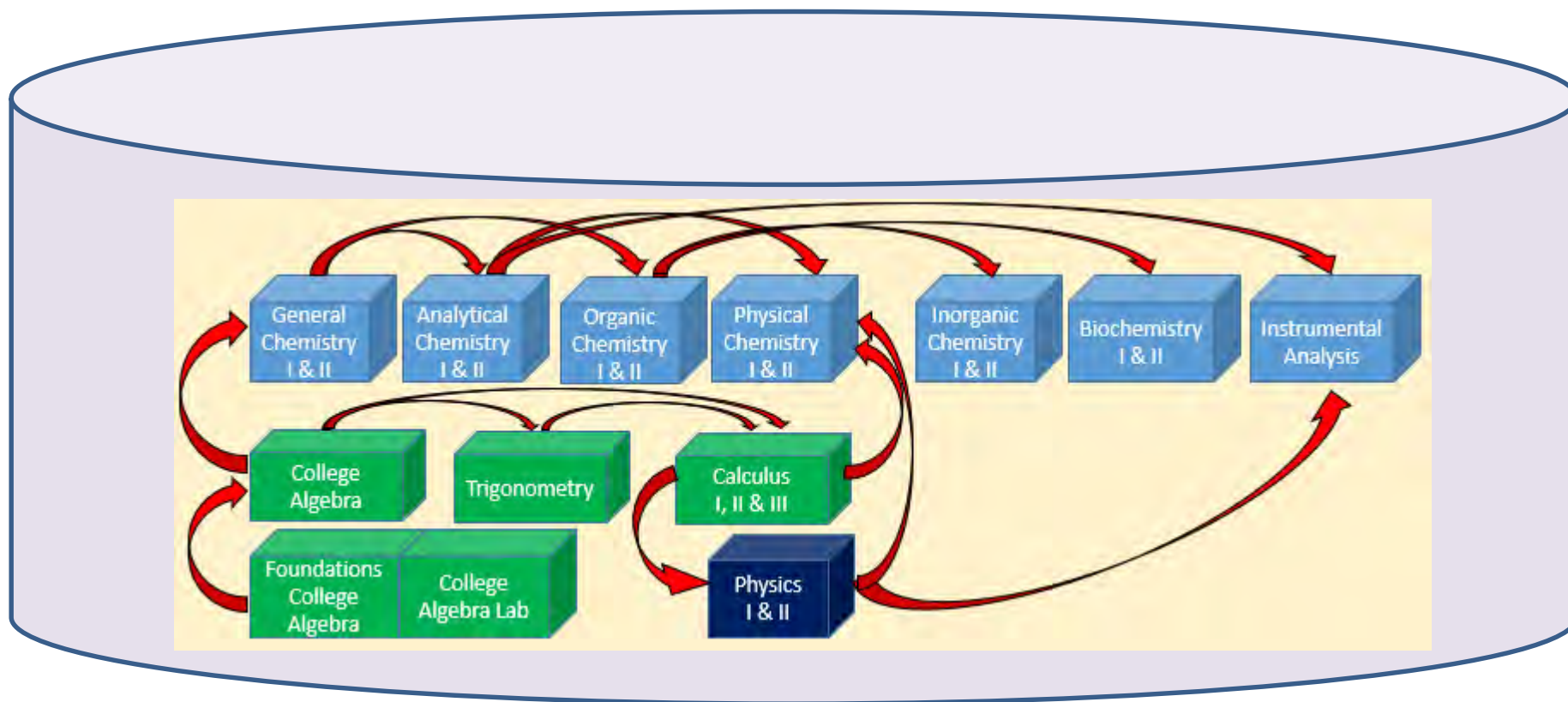
MAPS

BOOK



InChI OER: Open Education Resource

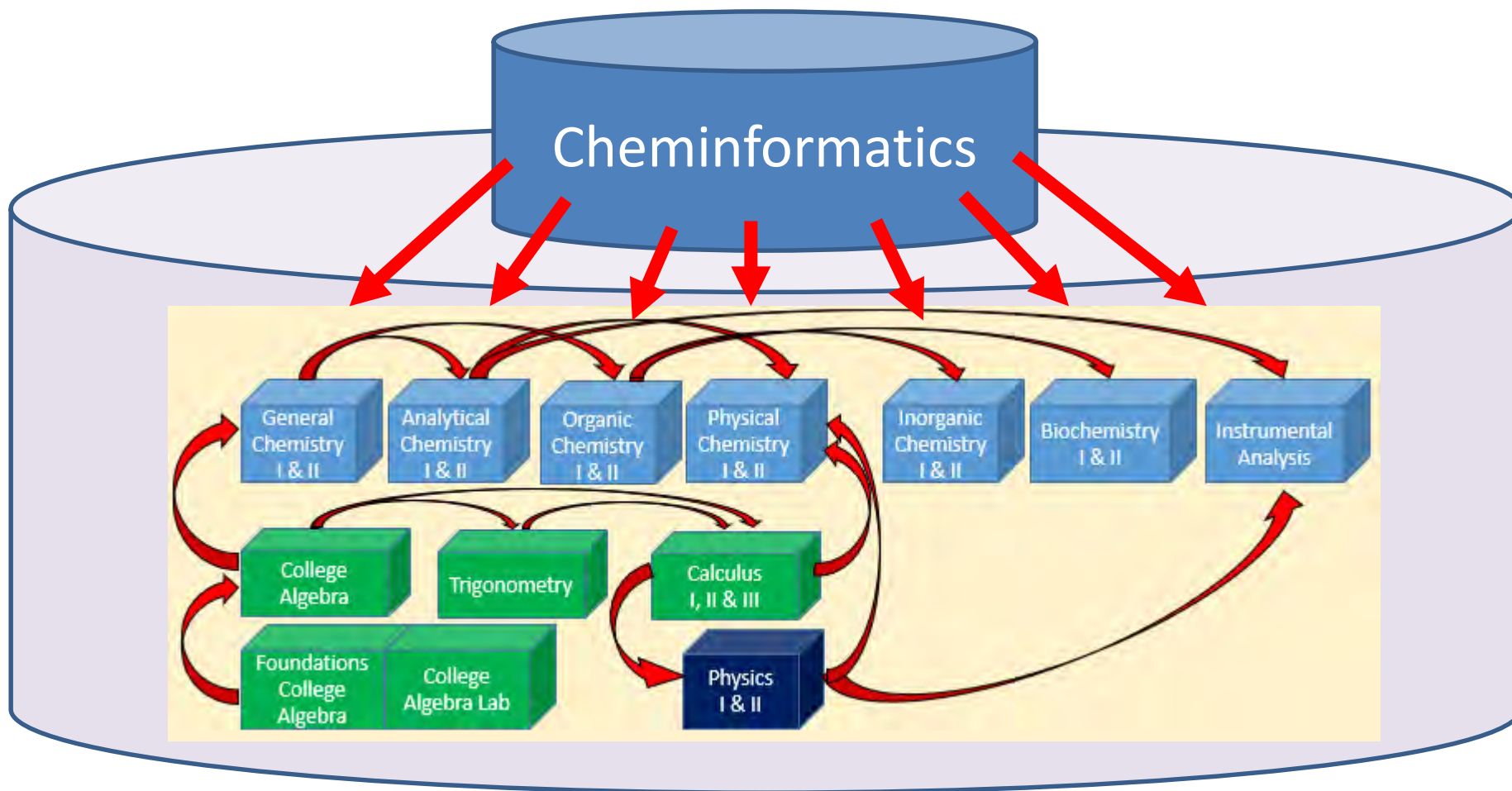
Integrating Content into the Classroom: LibreText



UALR Zero Textbook Cost (ZTC) Initiative

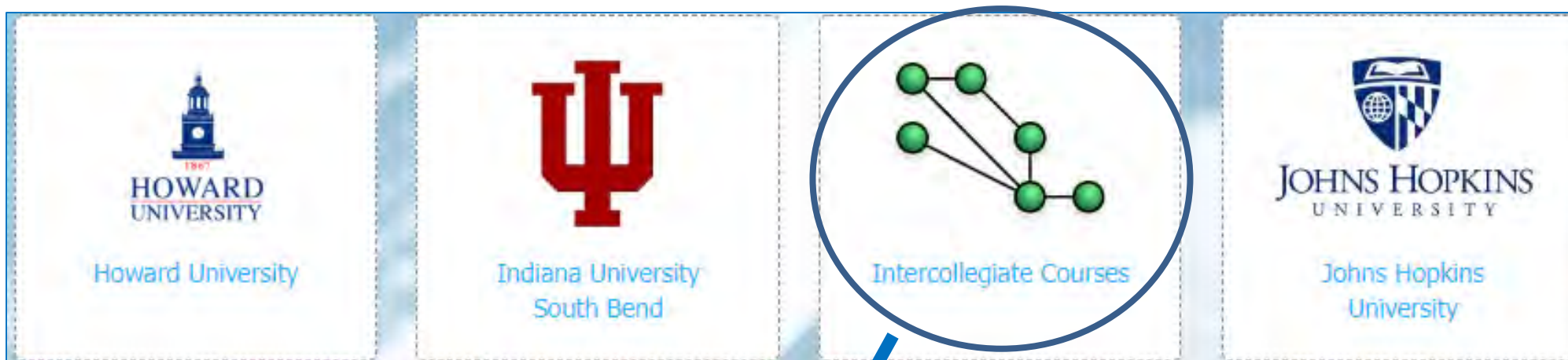
InChI OER: Open Education Resource

Integrating Content into the Classroom: LibreText



InChI OER: Open Education Resource

Integrating Content into the Classroom: LibreText



- UA Little Rock
- Centre College
- Ramon Llull



Cheminformatics OLCC (2019)

- St Louis College of Pharmacy
- Otterbein University

InChI OER: Open Education Resource

Integrating Content into the Classroom: LibreText

Cheminformatics OLCC (2019)

Last updated: Aug 16, 2019, 2:28:44 by Bowen S. Seford Page restriction: Public Page ID: 143569



Greetings, welcome to the homepage of the Fall 2019 Cheminformatics OLCC.

NOTE: THIS SITE IS UNDER CONSTRUCTION

This course is sponsored by the ACS Division of Chemical Education's Committee on Computers in Chemical Education. This course is designed for either graduate students, or upper division undergraduate students. In this course students will learn how molecules are represented on computers, and use PubChem and a variety of open source software to learn some basic coding and to access and analyze data obtained through a variety of APIs.



1. Introduction



2. Representing Small Molecules on Computers



3. Database Resources in Cheminformatics



4. Searching Databases for Chemical Information

No image available

5. Molecular Similarity

No image available

6. Ligand-based Virtual Screening

No image available

7. Structure-based Virtual Screening

No image available

8. Machine-learning Basics


InChI OER: Open Education Resource

Integrating Content into the Classroom: LibreText


Home >> Course Shells >> Intercollegiate Courses >> Cheminformatics OLCC (2019)








2. Representing Small Molecules on Computers

Last updated: Jun 23, 2019, 10:45 AM by Delmar Larsen Page restriction: Public Page ID: 144246

★ < 1.9: Assignments | 2.1: Introduction >     

Greetings...

+
✕
Topic hierarchy 

| | |
|--|---|
| <p> 2.1: Introduction Introduction to graph theory, connection tables, line notation and general issues of molecular representation.</p> <p> 2.3: Molecular Graph Issues This is sort of an overview of molecular structural issues that the simplified connection table can not cover. This is a review section for chemistry students and we will see how many of these issues are dealt with once we look at actual molecular representations. So this section is background material.</p> <p> 2.5: Structural Data Files Actual Structural Data Files. We will use Hack-a-Mol to explore molfiles and discuss several other file types.</p> <p> 2.7 Programming Topics</p> | <p> 2.2: Connection Tables Introduction to the concept of the Simplified Connection Table; the atom block and the bond block.</p> <p> 2.4: Line Notation Introduction to line notation, the focus is on SMILES and InChI.</p> <p> 2.6: Chemical Resolvers, Molecular Editors and Visualization</p> |
|--|---|

InChI OER: Open Education Resource

Integrating Content into the Classroom: LibreText

2.4: Line Notation

Last updated: Aug 11, 2019, 1:28 PM by Robert E. Belford Page restriction: Public Page ID: 154856



◀ 2.3: Molecular Graph Issues

2.5: Structural Data Files ▶



Contributed by [Robert Belford](#)

Professor (Chemistry) at [University of Arkansas at Little Rock](#)

LEARNING OBJECTIVES:

- Explain what SMILES, SMARTS and SMIRKS are.
- Explain what InChI and InChIKey are.
- Review SMILES specification rules.
- Compare and contrast SMILES and InChI.
- Demonstrate how to interpret SMILES, SMARTS, InChI strings into their corresponding chemical structures.

Introduction

Line notations represent structures as a linear string of characters. They are widely used in Cheminformatics because computers can easily process linear strings of data. Examples of line notations include the Wiswesser Line-Formula Notation (WLN)¹, Sybyl Line Notation (SLN)^{2,3} and Representation of structure diagram arranged linearly (ROSDAL)^{4,5}. Currently, the most widely used linear notations are the Simplified Molecular-Input Line-Entry System (SMILES)⁶⁻⁹ and the IUPAC Chemical Identifier (InChI)¹⁰⁻¹³, which are described below. In this class we will focus on SMILES and InChI line notation.

SMILES

The **Simplified Molecular-Input Line-Entry System** (SMILES)⁶⁻⁹ is a line notation for describing chemical structures using short ASCII strings. SMILES is like a connection table in that it identifies the nodes and edges of a molecular graph. SMILES was developed in the late 1980s and implemented by Daylight Chemical Information Systems (Santa Fe, NM), but it is still widely used today. A detailed information on SMILES can be found in [Chapter 3¹⁴](#) of the Daylight Theory Manual as well as the [SMILES tutorial¹⁵](#).

– Table of contents

1. Introduction

2. SMILES

2.1. 2.1.1. [SMILES Specification Rules](#)

2.1.2. [Isomeric SMILES](#)

2.1.3. [Limitations of SMILES](#)

2.2. [SMARTS](#)

2.3. [SMIRKS](#)

3. InChI

3.1. [InChI: A Layered Notation](#)

3.1.1. [Standard InChI](#)

3.1.2. [NonStandard InChI](#)

3.1.3. [Drawbacks of InChI](#)

3.2. [InChI Keys](#)

3.2.1. [InChIKeys and web Searching](#)

3.3. [InChI OER](#)

3.3.1. [InChI Layers Explorer](#)

4. [References and Further Reading](#)

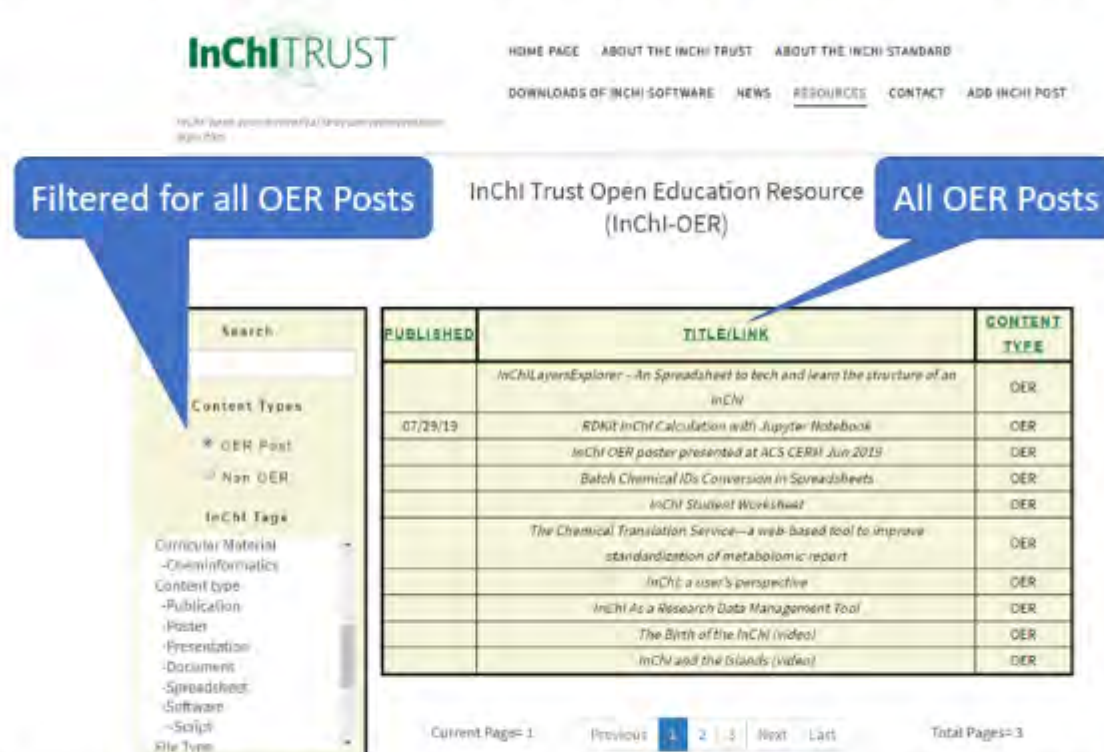
5. [Contributors](#)

InChI OER: Open Education Resource

Integrating Content into the Classroom: LibreText

InChI OER

The InChI Trust runs an Open Education Resource (OER) where you can find material on InChI <https://www.inchi-trust.org/oer/>. The InChI OER is a repository where anyone can upload and tag material on InChI, or link to and tag existing material on the use of InChI. Once material is posted within the OER it can be searched through a filter system.



The screenshot shows the InChI Trust Open Education Resource (InChI-OER) website. A search filter on the left is set to 'OER Post'. A table of posts is displayed with columns for 'PUBLISHED', 'TITLE/LINK', and 'CONTENT TYPE'. A blue callout box points to the filter, and another points to the table.

| PUBLISHED | TITLE/LINK | CONTENT TYPE |
|-----------|---|--------------|
| | <i>InChI Layers Explorer - An Spreadsheet to teach and learn the structure of an InChI</i> | OER |
| 07/29/19 | <i>RDKit InChI Calculation with Jupyter Notebook</i> | OER |
| | <i>InChI OER poster presented at ACS CERI Jun 2019</i> | OER |
| | <i>Batch Chemical IDs Conversion in Spreadsheets</i> | OER |
| | <i>InChI Student Worksheet</i> | OER |
| | <i>The Chemical Translation Service—a web-based tool to improve standardization of metabolomic report</i> | OER |
| | <i>InChI: a user's perspective</i> | OER |
| | <i>InChI As a Research Data Management Tool</i> | OER |
| | <i>The Birth of the InChI (video)</i> | OER |
| | <i>InChI and the Islands (video)</i> | OER |

Figure 2.4.7: InChI OER tag filter and associated content. The default setting is to show all OER site material, clicking non-OER will extend the filter to include off site material like publications which have records that have been submitted to the OER.

InChI Layers Explorer

In this activity we will use the InChI OER to obtain an Excel spreadsheet that breaks an InChI into layers, and start to analyze how cheminformatics functionality can be integrated into common tools like spreadsheets. Go to the [InChI OER](#) and in the filter click "Spreadsheet" (middle of figure 2.4.8). This filters the content to items that are tagged "spreadsheet" and also removes any tag that is not associated with one of those content items. Now move down to tag category "File Type" and while holding the <ctrl> key, click Excel (right figure 2.4.8). You now get a list of excel spreadsheets (figure 2.4.9).

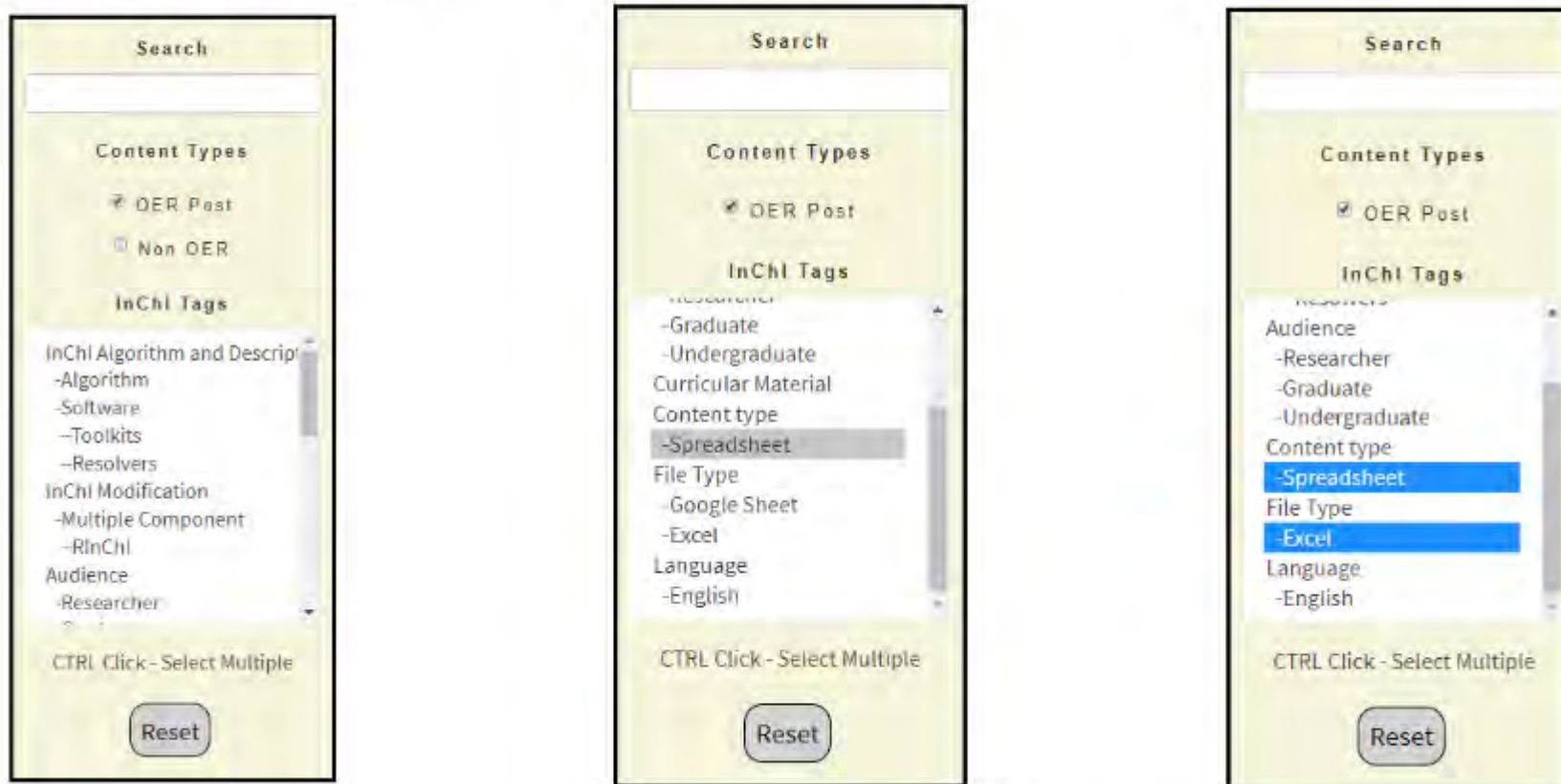


Figure 2.4.8: InChI OER Tag Filter.

On the left is the default setting and all content loaded to the site is displayed in the window (right side of figure 2.4.7). In the middle the filter for spreadsheets is activated, and you can see there are two types that have been uploaded, Google Sheets and Excel sheets. On the right both Spreadsheet and Excel have been activated, and so only spreadsheets in Excel are displayed and the content view is reduced to those items that are tagged both "Spreadsheet" and "Excel" (Figure 2.4.9)

| <u>PUBLISHED</u> | <u>TITLE/LINK</u> | <u>CONTENT TYPE</u> |
|------------------|---|---------------------|
| | <i>InChILayersExplorer - An Spreadsheet to tech and learn the structure of an InChI</i> | OER |
| | <i>Batch Chemical IDs Conversion in Spreadsheets</i> | OER |
| | <i>Identifier conversion on an Excel spreadsheet</i> | OER |

Figure 2.4.9: At the time this page was created there were three items uploaded to the InChI OER that were tagged as Excel Spreadsheets.

Click on the InChILayersExplorer and you go to it's content page. This page will have a description of the content and a green information box (Figure 2.4.10), and in the information box is a "Download Publication Files", that allows you to obtain the spreadsheet.

| INFORMATION | |
|----------------------------|---|
| Content Type | OER |
| Uploaded By | Jordi Cuadros |
| Download Publication Files | http://www.inchi-trust.org/wp/wp-content/uploads/2019/06/InChILayersExplorer.xlsx |
| License | CC BY 3.0 Unported |
| Content Status | publish |
| Number of Comments | No Comments |
| Date Published | |
| Content Tags | Audience , Content type , Excel , File Type , Graduate , InChI Algorithm and Description , Researcher , Spreadsheet , Undergraduate |

Figure 2.4.10: Green Information box for the InChILayersExplorer

ACTIVITY 2.4.1

Using the InChI Layers Explorer show the difference between the InChI for (R)-thalidomide and (S)-thalidomide. Note, the goal of this activity is not to answer the question, to gain an understanding on how the InChI Layers Explorer works, which is in effect a "smart spreadsheet" that communicates with databases via webservice functions. One of the skills we hope you can gain from this class is enough familiarity with how code works so if you see new code, you can hack in and figure how it works. Be sure to enable the spreadsheet after you download it. This spreadsheet communicates with the NCI Chemical Resolver (section 2.7.)

1. Type (R)-thalidomide in the yellow region (type over CoA), OK, it fails, now try the (S) isomer, and it still fails, so now try thalidomide without specifying an isomer. OK, so you have the InChI for thalidomide, but there is nothing in the stereochemical layer, as you have not specified the stereochemistry. This spreadsheet uses the Chemical Identifier Resolver of the NIH which will be covered in section 2.6.2.1.1, which can be accessed directly at <https://cactus.nci.nih.gov/chemical/structure> and is shown in figure 2.4.11. Now let's start by searching for (R)-thalidomide directly in the resolver (figure 2.4.11).

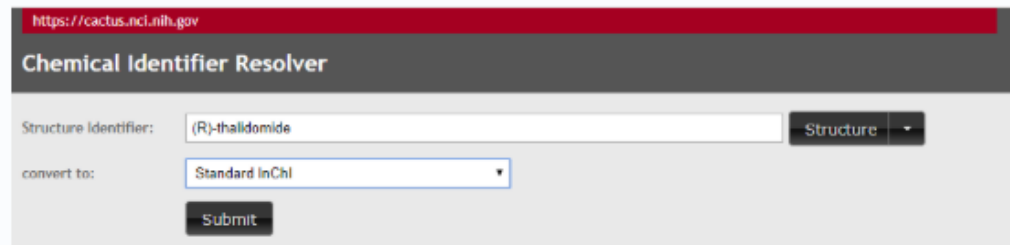


Figure 2.4.11: NCI/CADD Chemical Resolver set up to find standard InChI for (R)-thalidomide

As you may have guessed, neither (R) or (S) works, but "thalidomide" does (incidentally, you have to hit submit, not Structure), and so this resolver will not provide information on the isomers of thalidomide. So now do a web search of (R)-thalidomide, and paste in its key (UEJJHQACJXSKW-SECBINFHSA-N), and note the stereochemical layer [t9-/m1/s1] is the only part that is different. Now repeating for (S)-thalidomide.

You should get the following results

Table 2.4.3

| Compound | InChI Key | Stereochemical layer |
|-----------------|----------------------------|----------------------|
| thalidomide | UEJJHQACJXSKW-UHFFFAOYSA-N | none |
| (R)-thalidomide | UEJJHQACJXSKW-SECBINFHSA-N | /t9-/m1/s1 |
| (S)-thalidomide | UEJJHQACJXSKW-VIFPVBQESA-N | /t9-/m0/s1 |

Notes on Activity 2.4.1

1. Builds up on a Story line

Stereochemistry

Isomers are different molecules with the same atomic constituency, that is they have the same number of atoms for each element and the atom tables are essentially identical, (the numbering of the atoms may be different, but the two atom tables are isomorphic). There are two basic types of isomers, constitutional isomers and stereoisomers. Constitutional isomers are also called structural isomers and have different bond connectivity for the same atoms. This means they have different (non-isomorphic) bond tables, and so the Simple Connection Table (SCT) has no problem distinguishing constitutional isomers. Stereoisomers have the same (isomorphic) SCTs, that is, both the atom and the bond table are essentially the same, (the atom numbering may differ, but this is reflected in the bond connections and so the SCTs are essentially the same (isomorphic). What distinguishes the atoms of stereochemical isomers is the atomic arrangement in space, not the connections.

You may ask why is this important? One example often used in textbooks is the biological significance of two stereoisomers of thalidomide, a chemical used as an antidepressant for pregnant mothers in the 1960s. In synthesizing the chemical the "drug" was actually a mixture of both isomers, one of which was an effective medication and the other of which caused horrific birth defects. This was clearly an "unintended consequence" and one of the most important functions of cheminformatics is to help scientists identify unintended effects of potential drugs by looking at a multiplicity of bioassays, including toxicological screening assays.



Figure 2.3.2: Birth defect caused by the mutagenic isotope of thalidomide, which was prescribed by the mother's doctor.

Section 2.3: was a review on stereoisomers

Notes on Activity 2.4.1

1. Builds up on a Story line

Lets look at thalidomide, the molecule responsible for the birth defects in figure 2.3.2. Figure 2.3.7 shows four ways of drawing this structure.

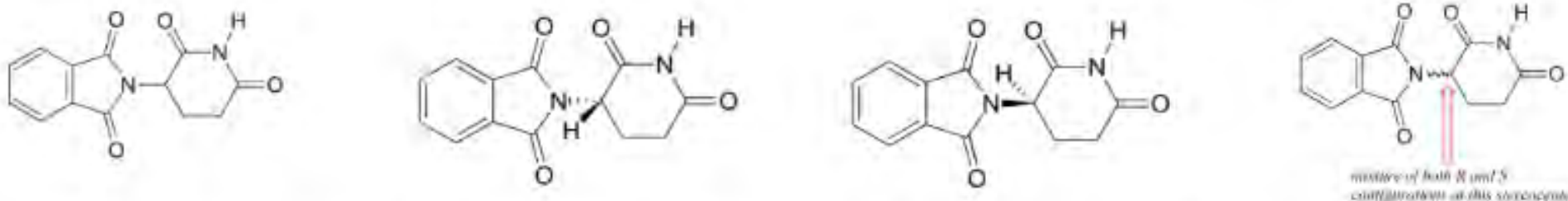


Figure 2.3.7 : Four ways of drawing thalidomide, each means something different. Sequentially (left to right) these are; undefined, left-hand (S), right-hand(R) and a mixture. See ICP rules (below) to understand R and S notation.

Story line was continued in section of drawing structures
(can help anchor concepts in long term memory)

InChI OER: Open Education Resource

Integrating Content into the Classroom: LibreText

Notes on Activity 2.4.1

2. Involves links to material not yet covered as students learn to trouble shoot things

1. Type (R)-thalidomide in the yellow region (type over CoA), OK, it fails, now try the (S) isomer, and it still fails, so now try thalidomide without specifying an isomer. OK, so you have the InChI for thalidomide, but there is nothing in the stereochemical layer, as you have not specified the stereochemistry. These spreadsheet uses the Chemical Identifier Resolver of the NIH which will be covered in [section 2.6.2.1.1](#), which can be accessed directly at <https://cactus.nci.nih.gov/chemical/structure> and is shown in figure 2.4.11. Now let's start by searching for (R)-thalidomide directly in the resolver (figure 2.4.11).

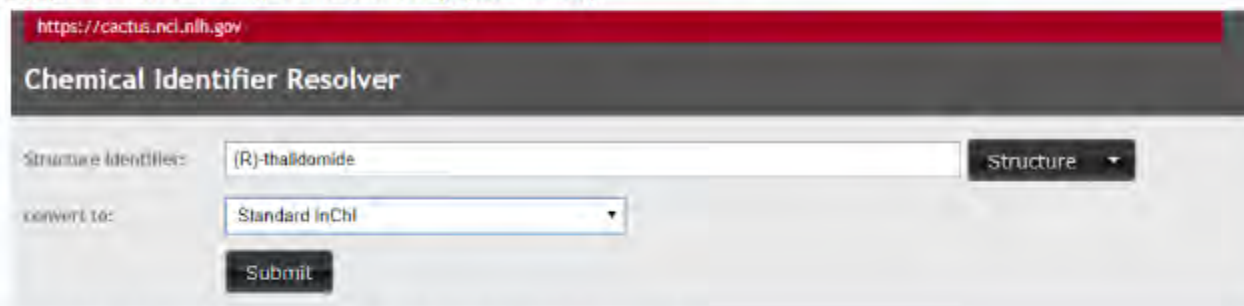


Figure 2.4.11: NCI/CADD Chemical Resolver set up to find standard InChI for (R)-thalidomid

Section 2.6 covers chemical resolvers
(cues students to go and seek things out)

Notes on Activity 2.4.1

3. Activity provides skills that will be used in future problems

Note, if you click on the merged cells that generates the InChI (Rows 7-8) you see the following code.

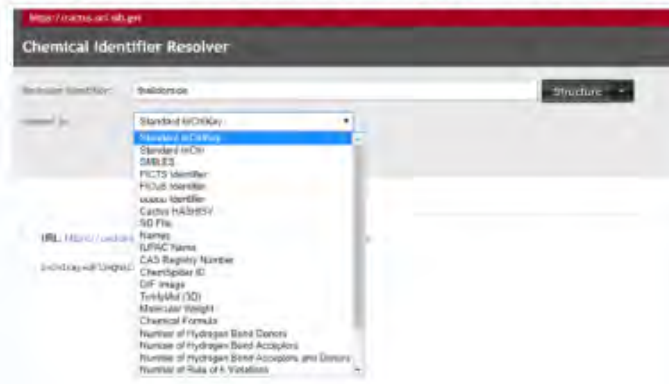
| | |
|----|--|
| 1 | Enter an InChI (or a compound name, synonym, SMILES or InChIKey) |
| 2 | thalidomide |
| 3 | |
| 4 | |
| 5 | |
| 6 | InChI |
| 7 | =IFERROR(IF(MID(A2,1,6)="InChI=",A2,WEBSERVICE("https://cactus.nci.nih.gov/chemical/structure/"&ENCODEURL(A2)&"/stdinchi")), "") |
| 8 | |
| 9 | |
| 10 | IF(logical_test, [value_if_true], [value_if_false]) |

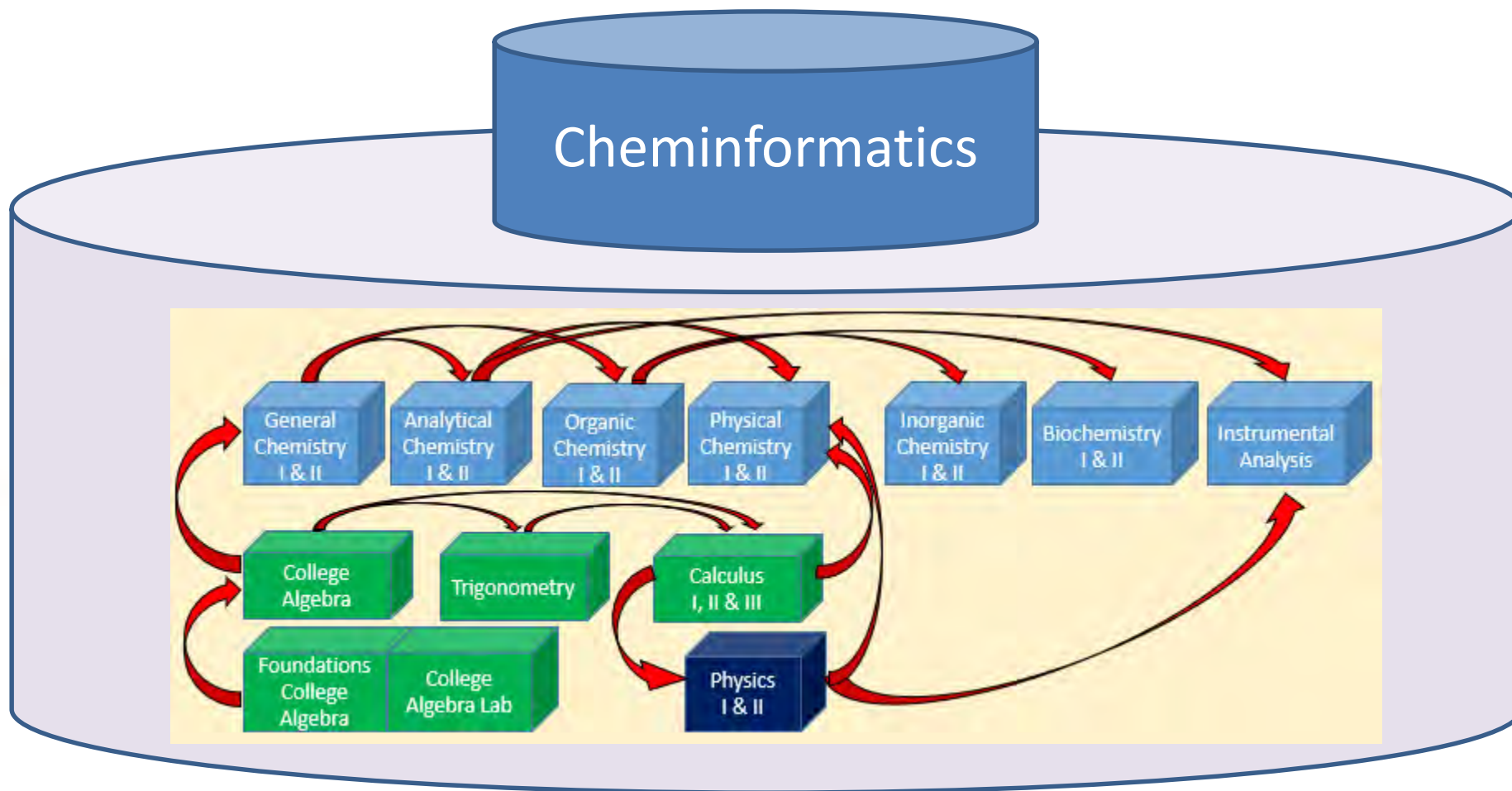
Figure 2.4.11: Code in spreadsheet that uses WEBSERVICE function to get InChI from NCI/CADD chemical resolver

Now open up a browser tab and paste in the following URL:

<https://cactus.nci.nih.gov/chemical/structure/thalidomide/stdinchi>

Now go back to the NCI Chemical Resolver and click the dropdown box of the "convert to" field (figure2.4.12 and try another option, say "TwirlyMol(3D)).





<https://www.inchi-trust.org/>

[https://chem.libretexts.org/Courses/Intercollegiate_Courses/Cheminformatics_OLCC_\(2019\)](https://chem.libretexts.org/Courses/Intercollegiate_Courses/Cheminformatics_OLCC_(2019))

