

# ChemIDplus, a public chemical information database at the U.S. National Library of Medicine (NLM)

Mitchell Miller, Ph.D. Scientific Thinking

George F. Hazard Jr. (Mike), Ph.D. NLM

Shannon Jordan, MPH NLM

Jenny Fang, MS NLM

Chuchu Lan ICF International

# Speaker Disclaimer

- ChemIDplus is a powerful system with a lot of data
- My role has been focused on the chemical structure databases and the components they interact with
- For more information about ChemIDplus and the data it contains, please contact Mike Hazard

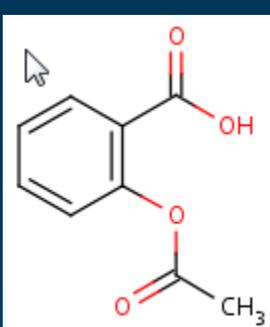
# Organization and Mission

- Division of Specialized Information Services
  - Environmental Health and Toxicology
  - HIV / AIDS
  - Disaster Information
  - Chemical Information

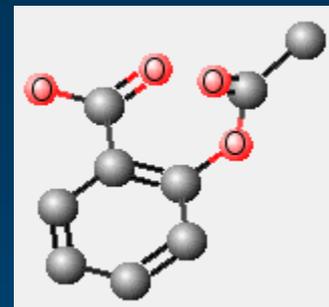
Chemical Dictionaries	Computer	Timeframe	Audience
Chemline	Mainframe	1974-1997	Professional
ChemID	Mainframe	1985-2000	Professional
ChemIDplus	Servers	1998-present	Public
Drug Information Portal	Servers	2008 - present	Public

# What data is in ChemIDplus?

- Overall records: >395,000
- 2D structures: > 302,000
- 3D structures: > 571,000 conformations
- Names/Synonyms: >1.5M
- Locators: > 1.3M
- Toxicity: > 139,000
- Classifications: > 200,000
- Properties (from Syracuse) ~ 25,000



# Structure Data



- Accelrys Direct (version 7) for storage and search
- Structures input singly or via SD file
- Structure derivatives generated immediately
  - 3D structures by Molecular Network's Corina
  - InChi strings, InChiKeys, SMILES by Direct
  - InChiKeys link to web searches
- Structures display as images and using Marvin applets and Chime
- 3D structures for display only

# ChemIDplus Locator Data

- Hyperlinks to over 125 resources
- Searchable and displayable
- Variety of types of links

Resource	Search Argument	Example
NLM PubMed Drugs@FDA	Chemical Name	Diazepam
NOAA CAMEO	Number	2577
EPA SRS	Registry Number	71-43-2
NIOSH Pocket Guide	HTML	neng0015.html

# ChemIDplus Interfaces

- Fully-featured application (including structure searching)
  - <http://chem.sis.nlm.nih.gov/chemidplus/>
- Lightweight version (simplified UI)
  - <http://chem.sis.nlm.nih.gov/chemidplus/chemidlite.jsp>
- Alternate lightweight interface: Drug Portal
  - <http://druginfo.nlm.nih.gov/drugportal/>
  - Focused on usability: suggestions, spell checked results
- Mobile version
  - <http://druginfo.nlm.nih.gov/m.drugportal/m.drugportal.jsp>

# ChemIDplus Advanced

- Name, formula, category, locator searching
- Structure searching
  - Substructure search
  - Similarity search
  - Salt / hydrate match
- Formula, toxicity and physical properties
- Server based
  - UNIX / Solaris, Linux / Oracle (12g)
  - Accelrys Direct, ChemAxon Marvin, Corina 3D
  - Tomcat: Java Servlets and JSPs
- First version released to public in 1998



Search Clear History Help

Display 5 results

### Substance Identification [i](#)

Name/Synonym  Equals

Data is available for 395,462 records.

### Toxicity [i](#)

Test: (any) between (any)

(mg/kg or ppm)

Species: (any)

Route: (any)

Effect: (any)

Toxicity data is available for 139,354 records.

### Physical Properties [i](#)

Melting Point

between

Either  Measurement Type

Physical property data was provided by [Syracuse Research Corporation](#) and is available for 25,461 records.

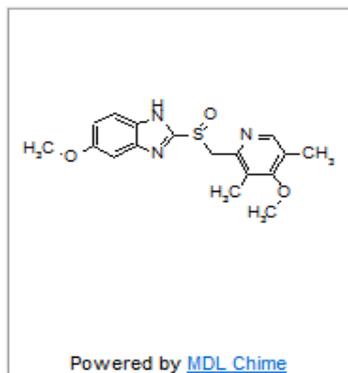
### Locator Codes [i](#)

(any)

AND

(any)

### Structure [i](#)



Powered by [MDL Chime](#)

### Structure Search Options [i](#)

- Substructure Search
- Similarity Search  %
- Exact (parent only)
- Flex (parent, salts, mixture)
- Flexplus (parent, all variations)

### Display structures using [i](#)

- Marvin
- Chime

Structure data is available for 302,722 records.

### Molecular Weight [i](#)

between

Molecular weight data is available for 302,722 records.

Go



Start New Query

Modify Query

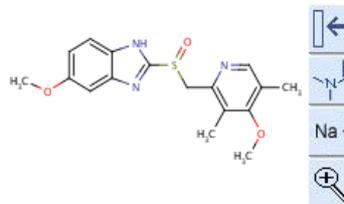
Show Query

Search History

Go To Record Number

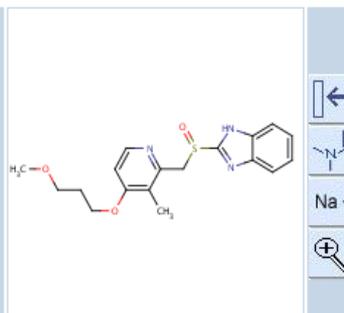
TOXNET Home

1 [Omeprazole \[USAN:INN:BAN:JAN\]](#)  
73590-58-6



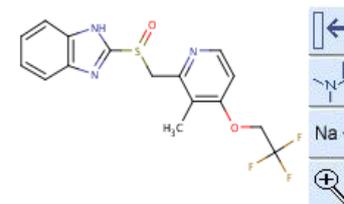
MW: 345.4211 - 100% similar

2 [Rabeprazole \[INN:BAN\]](#)  
117976-89-3



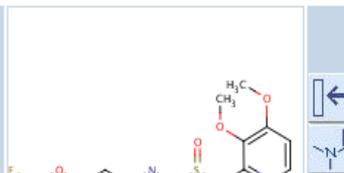
MW: 359.4479 - 91.3143% similar

3 [Lansoprazole \[USAN:INN:BAN\]](#)  
103577-45-3



MW: 369.366 - 84.1689% similar

4 [Pantoprazole \[USAN:INN:BAN\]](#)



# Spell Checker

- Over 80% of questions were text i.e. names
  - Misspellings and typos caused failed searches
    - e.g. Valeum for Valium
  - Chemical names foreign to some public users
- Chemical Spellchecker added

Your search for chemical name "VALEUM" did not retrieve any exact matches.

You may return to the [query page](#) and modify your search or examine any of the near matches below by clicking on it. Your search target may not be on the list.

[Valium](#)

[Valeans](#)

[Valeo](#)

[Valu-8](#)

[Valu-7](#)

[Vallene](#)

[Valium](#)



# ChemIDplus Lite

- First simplified User interface
- Names and identifier searching

# NLM Drug Information Portal



- Public in 2008
  - Over 36,000 drugs
  - Links to 19 NLM and federal resources
- New features for public users
  - Simple one box search interface
  - Autosuggest and balloon help
  - Descriptions added to spell check candidates
  - Embedded names found i.e. “10 percent valium”
  - Parent / salt / hydrate table cross-refers records

# Simple Interface

The screenshot shows the top navigation bar with a background of various pills. The main header includes the title 'Drug Information Portal' and the tagline 'Quick Access to Quality Drug Information'. The NLM logo and 'United States National Library of Medicine National Institutes of Health' are on the right. Below the header is a horizontal menu with six items: Home, News and Features, NLM Resources, NLM Research Resources, Resources by Audience / Class, and Other Resources. A 'Mobile Site - NEW!' link with a mouse cursor is positioned below the menu. The central area features a search box with a pill icon, the text 'Search DRUG INFORMATION PORTAL', and 'Information available for 36,313 drugs.' Below the search box are radio buttons for 'By Name' (selected) and 'By Category'. A 'Go' button is to the right of the search input field. A list of search options is provided below, each preceded by a right-pointing arrowhead.

Drug Information Portal  
Quick Access to Quality Drug Information

United States National Library of Medicine National Institutes of Health

Home News and Features NLM Resources NLM Research Resources Resources by Audience / Class Other Resources

Mobile Site - **NEW!**

Search  
DRUG INFORMATION PORTAL  
Information available for 36,313 drugs.

By Name  By Category

Go

- ▶ Show examples.
- ▶ Show drug category descriptions.
- ▶ Show top "By Name" searches (previous seven days).
- ▶ Show top "By Category" searches (previous seven days).
- ▶ Show top dispensed prescriptions in the US Market, 2010.
- ▶ Show common drug name list.
- ▶ Show category name list.
- ▶ Show list of resources searched.

Home > Search Results



By Name  By Category

VALIUM

◀ Go back to previous page.

### Search Results

**Drug Name:** Diazepam [USAN:INN:BAN:JAN] [\[show more names\]](#) [\[show structure\]](#)

**Search Term:** VALIUM

**Description:** A benzodiazepine with anticonvulsant, anxiolytic, sedative, muscle relaxant, and amnesic properties and a long duration of action. Its actions are mediated by enhancement of GAMMA-AMINOBUTYRIC ACID activity. It is used in the treatment of severe anxiety disorders, as a hypnotic in the short-term management of insomnia, as a sedative and premedicant, as an anticonvulsant, and in the management of alcohol withdrawal syndrome. (From Martindale, The Extra Pharmacopoeia, 30th ed, p589).

**Categories:** [\[i\]](#) Adjuvants, anesthesia [\[show more categories\]](#)

### Summary

- ▶ [\[i\]](#) Summary of drug information (MedlinePlusDrug)
- ▶ [\[i\]](#) Summary of consumer health information (MedlinePlusTopics)
- ▶ [\[i\]](#) Summary of the effect on breastfeeding (LactMed)
- ▶ [\[i\]](#) Summary of Drug-Induced Liver Injury (LiverTox)
- ▶ [\[i\]](#) Manufacturers drug label (DailyMed)
- ▶ [\[i\]](#) Clinical trials (ClinicalTrials.gov)
- ▶ [\[i\]](#) Drug Identification and Image Display (Pillbox *beta*)

### Detailed Summary

- ▶ [\[i\]](#) Summary of reviewed biological and physical data (HSDB)
- ▶ [\[i\]](#) References from scientific journals (Medline/PubMed)
- ▶ [\[i\]](#) References from toxicological journals (TOXLINE)
- ▶ [\[i\]](#) Biological activities and chemical structures (PubChem)
- ▶ [\[i\]](#) Biological activities against HIV/AIDS and other viruses (NIAID ChemDB)
- ▶ [\[i\]](#) Toxicological and chemical resources (ChemIDplus)

### Additional Resources

- ▶ [\[i\]](#) Information from the US Food & Drug Administration (Drugs@FDA)
- ▶ [\[i\]](#) Information from the US Drug Enforcement Administration (DEA)
- ▶ [\[i\]](#) Search engine for other government resources (USA.gov)

◀ Go back to previous page.

Go

# Parent Child Table

- Cross-refers parent, salts and hydrate records
- Shows users the same locators and categories

Generic Name	Name unique to record	Unique Locators	Unique Categories
Chlordiazepoxide hydrochloride	Librium		
Chlordiazepoxide	Clordiazepossido	HSDB LiverTox	Anti-Anxiety Agents

# Some Collaborations

- FDA SRS System
- PubChem
- ClinicalTrials.gov
- NCI CADD group (Marc Nicklaus)
- EPA SRS
- Comparative Toxicogenomics Database
- Chemical Entities of Biological Interest

# Conclusion



- The value of the data is in the use people can make of it
  - Provide user interfaces that serve specific groups of users
  - Many want simple searching ... “Like Google”
  - Public chemical systems need “intelligence”
  - Validation, analysis, term mapping
  - **Avoid user overload!** ->
- Value of data is enhanced by sharing



# For more information

- ChemIDplus in general, the data, new collaborations
  - Mike Hazard
  - [hazardm@mail.nlm.nih.gov](mailto:hazardm@mail.nlm.nih.gov)
- Structure database management
  - Mitch Miller
  - [mitch.miller@thinkscience.us](mailto:mitch.miller@thinkscience.us)