



A significant resource for scientists

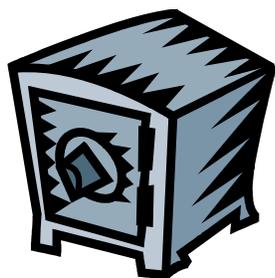
Evan Bolton, Ph.D.

NCBI/NLM/NIH

5<sup>th</sup> Meeting on U.S. Government Chemical  
Databases and Open Chemistry

August 25, 2011

[bolton@ncbi.nlm.nih.gov](mailto:bolton@ncbi.nlm.nih.gov)



# What is PubChem?

- An open archive
  - anyone can contribute
    - chemical structures
    - synonyms
    - comments
    - biological experiments
    - cross references
    - records versioned
    - URLs
  - links external resources
  - voluntary data push
  - automated updates
- A public resource
  - anyone can access
    - data downloadable
    - search, browse, retrieve
  - integrated
    - literature
    - sequences, protein 3-D
  - analysis capabilities
  - programmatic layers
    - PUG, PUG/SOAP
    - Entrez Utilities
    - URL-based interfaces

# What is PubChem?



- An open archive
  - anyone can contribute
    - chemical structures
    - synonyms
    - comments
    - biological experiments
    - cross references
    - records versioned
    - URLs
  - links external resources
  - voluntary data push
  - automated updates
- A public resource
  - anyone can access
    - data downloadable
    - search, browse, retrieve
  - integrated
    - literature
    - sequences, protein 3-D
  - analysis capabilities
  - programmatic layers
    - PUG, PUG/SOAP
    - Entrez Utilities
    - URL-based interfaces

# PubChem home page...

The screenshot shows the PubChem website interface. At the top, there is a navigation bar with links for [Databases](#), [Deposition](#), [Services](#), [Help](#), and [more](#). The main header features the PubChem logo. Below the logo, there are three search buttons: [BioAssay](#), [Compound](#), and [Substance](#). A search input field contains the text "aspirin", followed by a [GO](#) button and a link to [Advanced search](#). Below the search bar, there are links for [Chemical structure search](#) and [BioActivity analysis](#). A news section contains two items: [New PubChem in 2011 - Growth Continues ..](#) and [New Structures from AmicBase are now available in PubChem.](#), with a [more ...](#) link and an RSS icon. On the right side, there is a vertical menu of tools: [Bioactivity summary](#), [Bioactivity datatable](#), [Bioactivity structure-activity](#), [Chemical structure search](#), [3D conformer viewer](#), [Chemical structure clustering](#), [Deposition gateway](#), [Structure download](#), [Bioassay download](#), and [PubChem FTP](#). At the bottom, there is a footer with links for [Write to Helpdesk](#), [Disclaimer](#), [Privacy Statement](#), [Accessibility](#), and [Data Citation Guidelines](#), along with the text "National Center for Biotechnology Information" and "NLM | NIH | HHS".

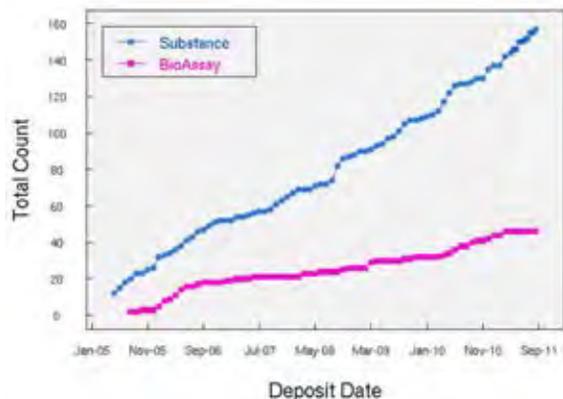
<http://pubchem.ncbi.nlm.nih.gov>

# PubChem contributors are many...

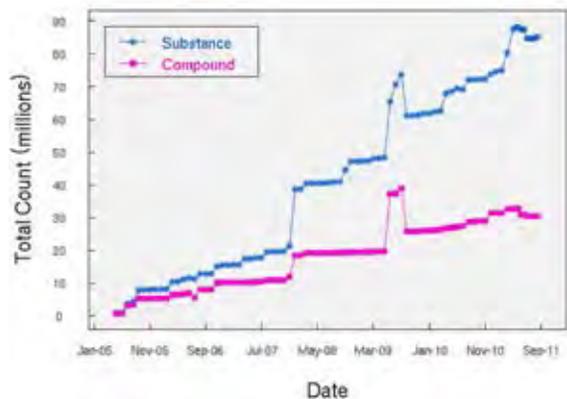


# PubChem contents are growing...

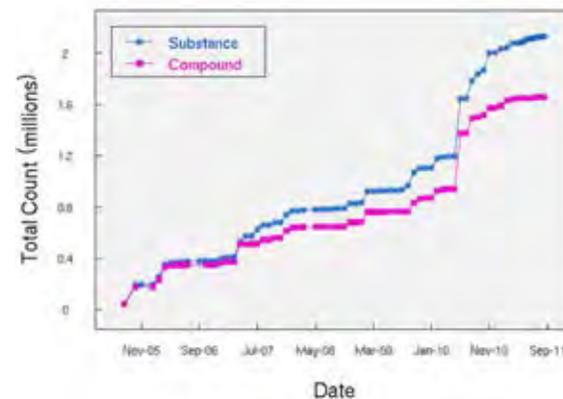
## Depositors



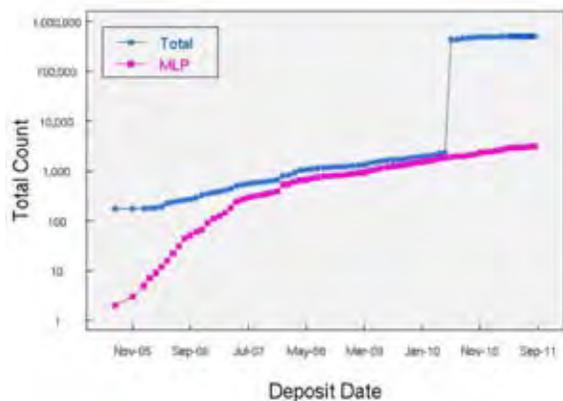
## Chemicals



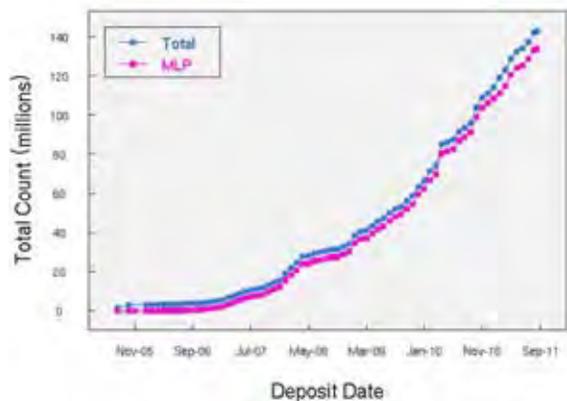
## Tested Chemicals



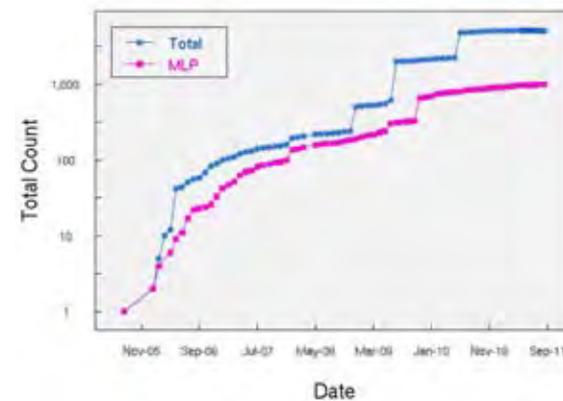
## Biological Assays



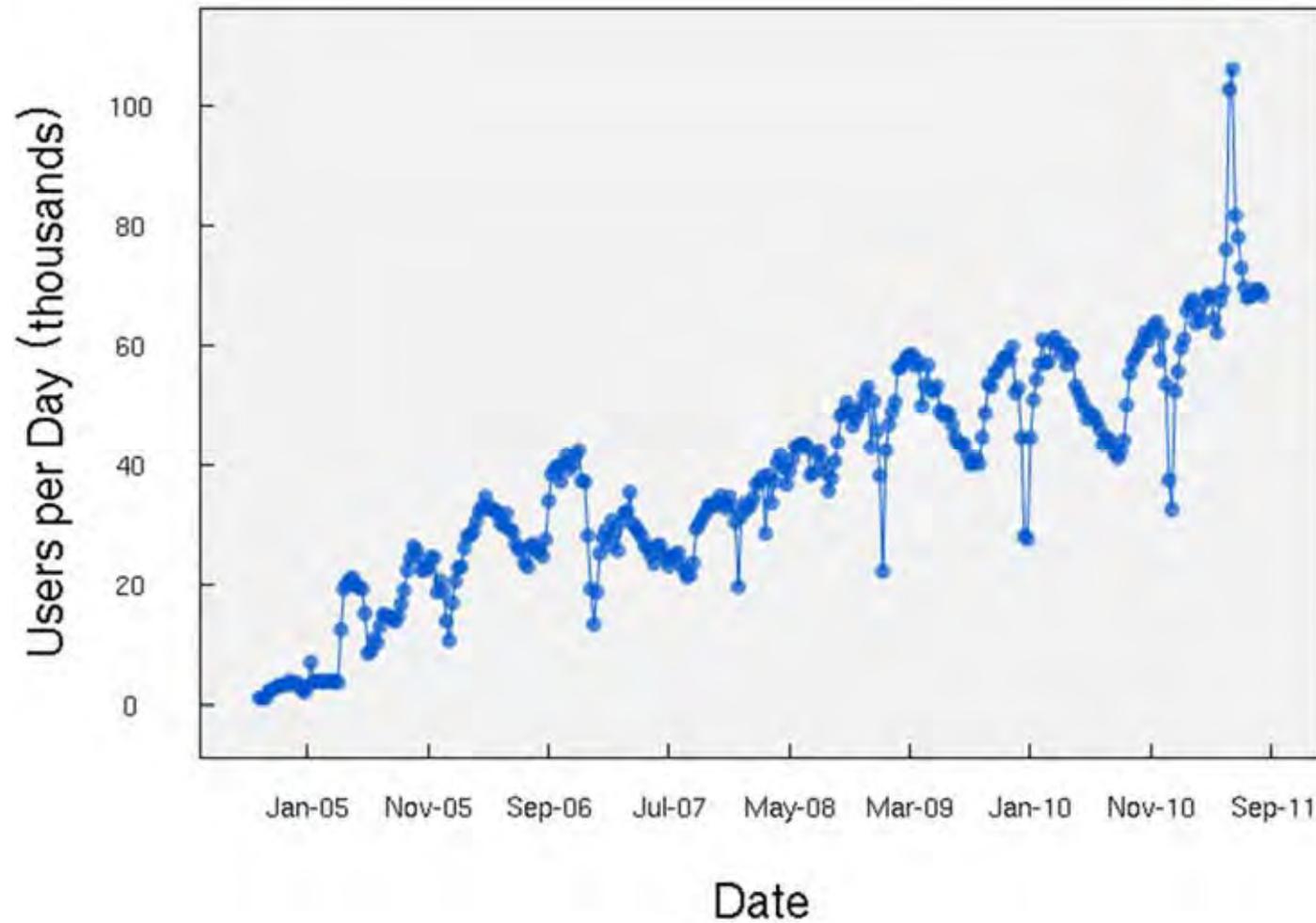
## Bioactivities



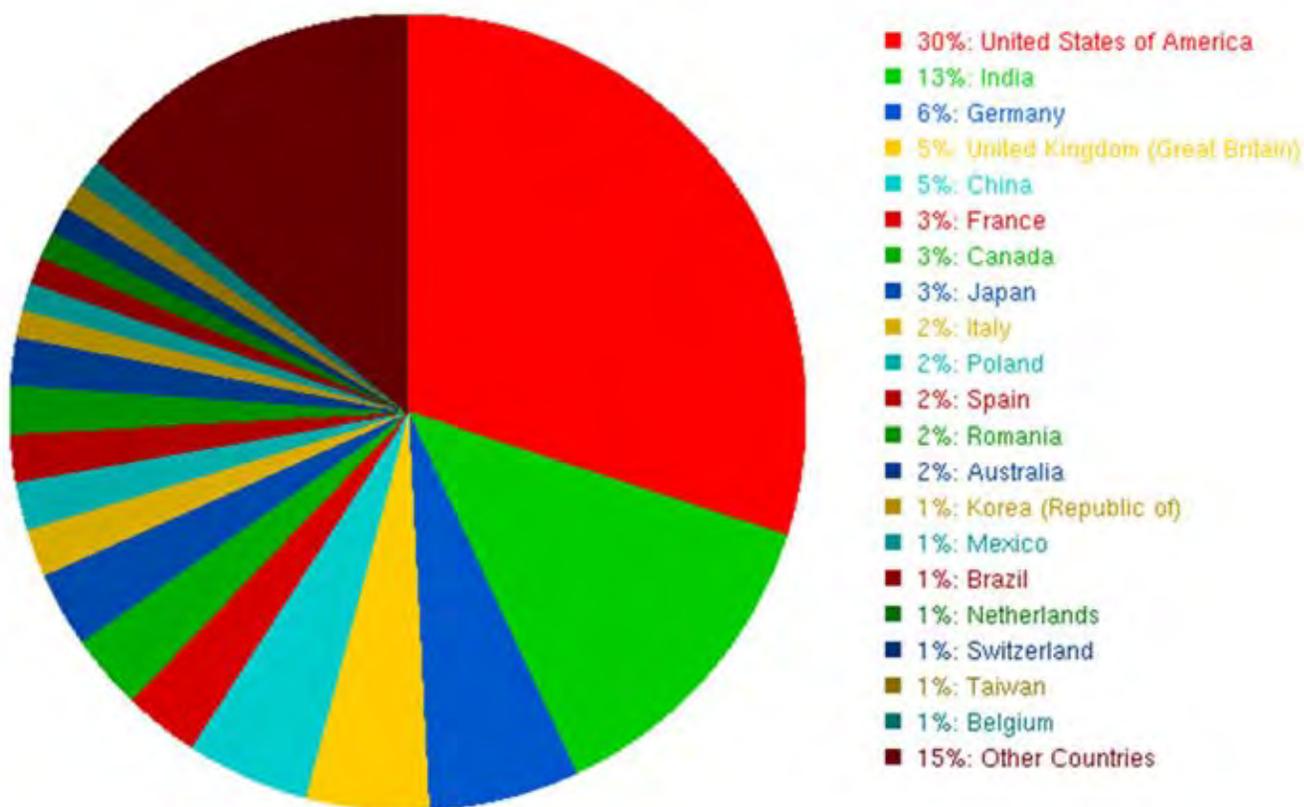
## Protein Targets



# PubChem is heavily used...

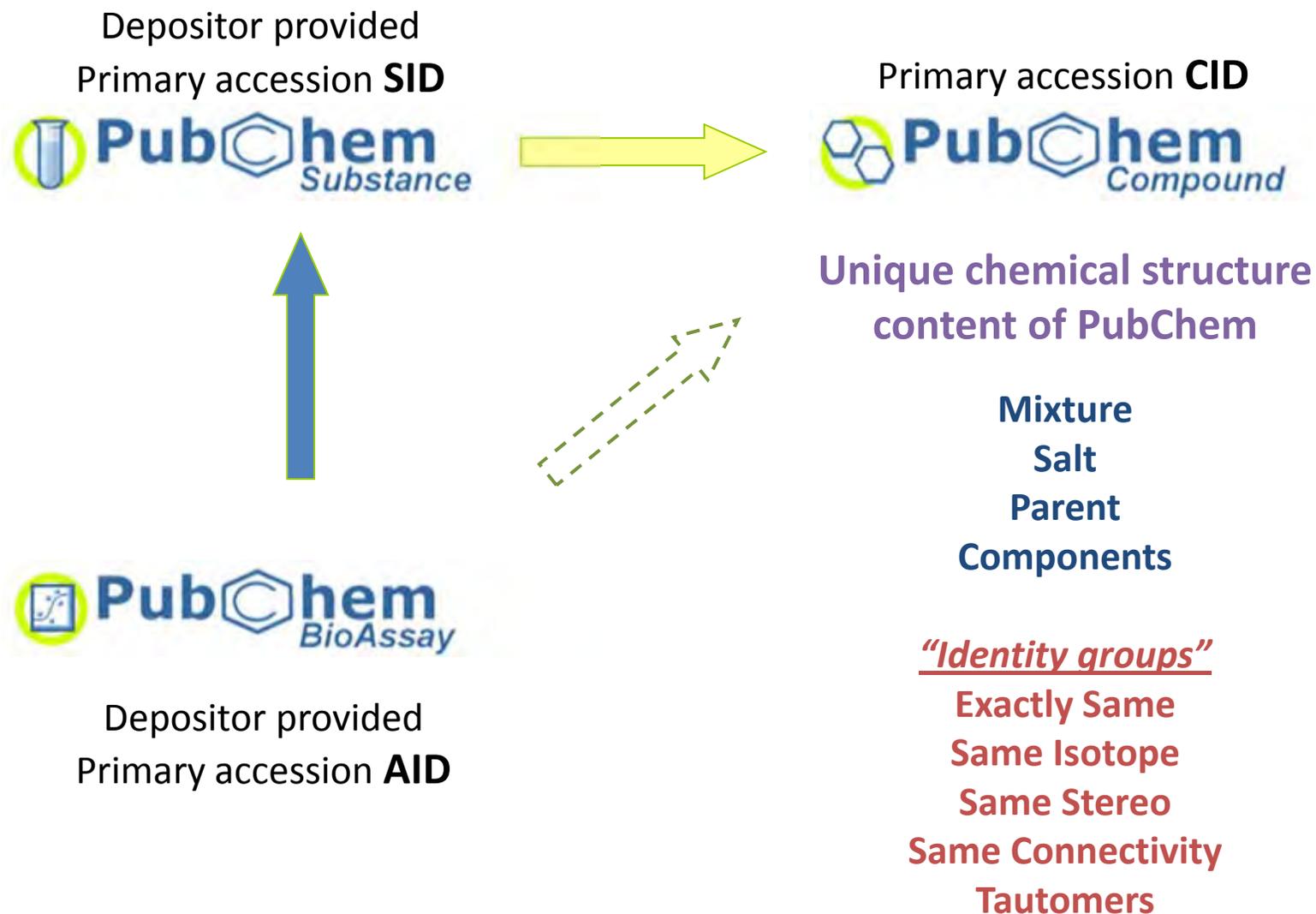


# PubChem is a global resource...



Interactive usage by country  
(Jul 15 2010 – Aug 15 2010)

# PubChem data relationships...



# The state of chemical information

(An aside)

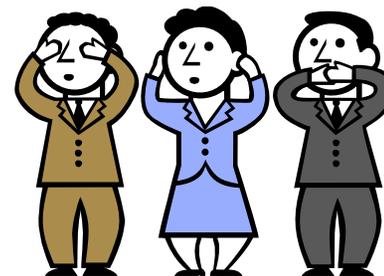


The **sad** state of  
chemical information

# Let's talk chemical information...

- **No “Global” rules or standards**

- based on individual organizational needs
- often based on individual preferences
- depictions of chemical structures



- **PubChem accepts data from many organizations**

- conflicting “business rules”
- previously unseen data representation schemes
- combinatorial ways of drawing the same structure



# What do you mean by that?

- “C” means?
  - form of carbon?
    - which one?
      - diamond?
      - graphite?
      - coal?
      - graphene?
      - charcoal?
      - carbon black?
      - nanotube?
  - methane?





*Don Quixote de la Mancha and Sancho Panza,*   
1863, by [Gustave Doré](#)

Image from Wikipedia

[http://en.wikipedia.org/wiki/  
Don\\_Quixote](http://en.wikipedia.org/wiki/Don_Quixote)



Don Quixote, his horse Rocinante and his squire Sancho Panza after an unsuccessful attack on a windmill. By [Gustave Doré](#)

Image from Wikipedia

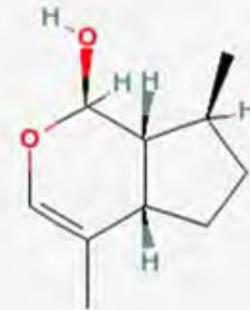
[http://en.wikipedia.org/wiki/Don\\_Quixote](http://en.wikipedia.org/wiki/Don_Quixote)

# What did you mean by that?

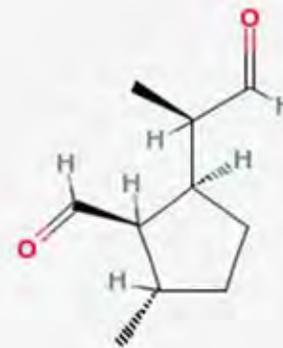
- Case Study:

(+)-Iridodial

**Defense chemicals from  
abdominal glands of 13  
rove beetle species of  
subtribe Staphylinina**

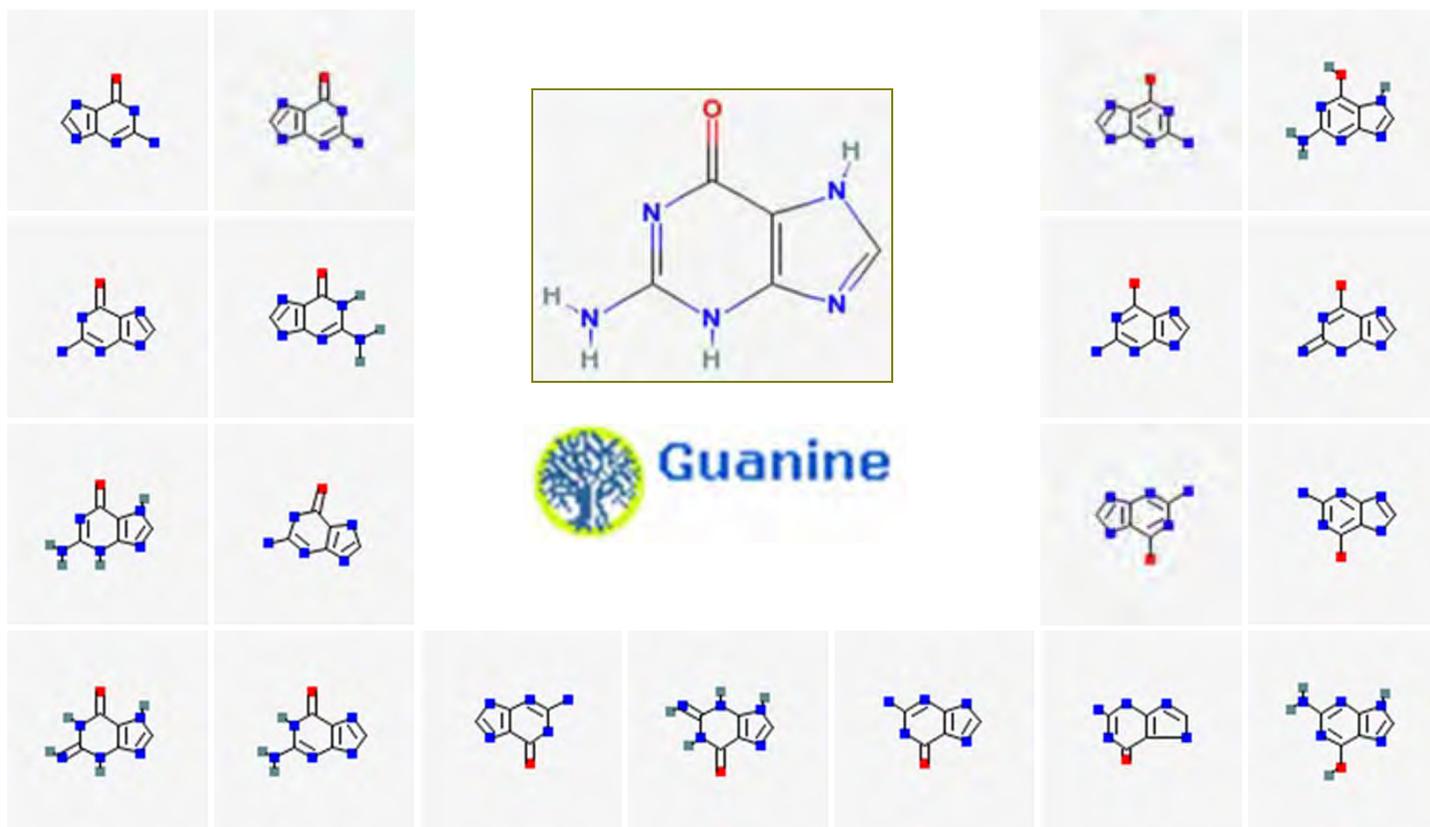


Ring Closed

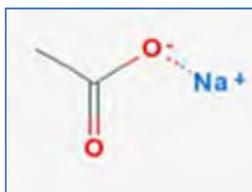


Ring Open

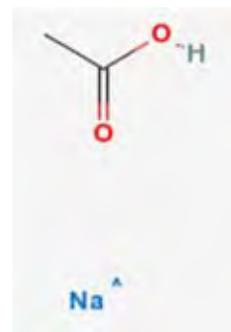
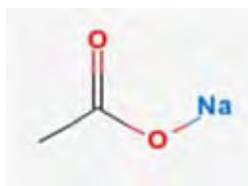
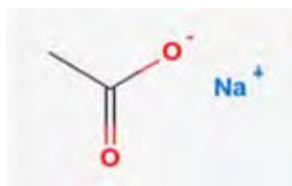
A chemical structure may be represented  
in many different ways



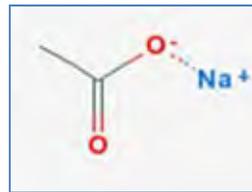
A chemical structure may be represented  
in many different ways



 Sodium Acetate

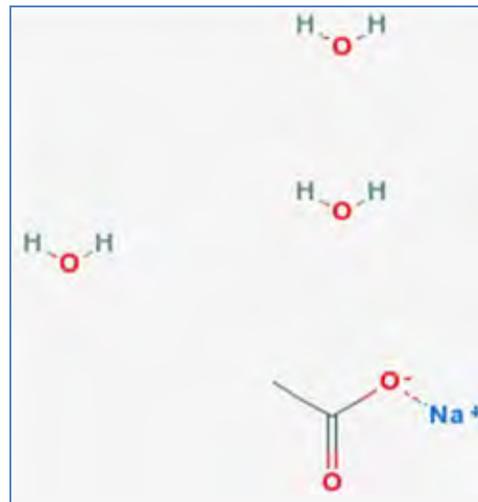


# What do you mean by “sodium acetate”?



## Sodium Acetate

The trihydrate sodium salt of acetic acid, which is used as a source of sodium ions in solutions for dialysis and as a systemic and urinary alkalizer, diuretic, and expectorant.

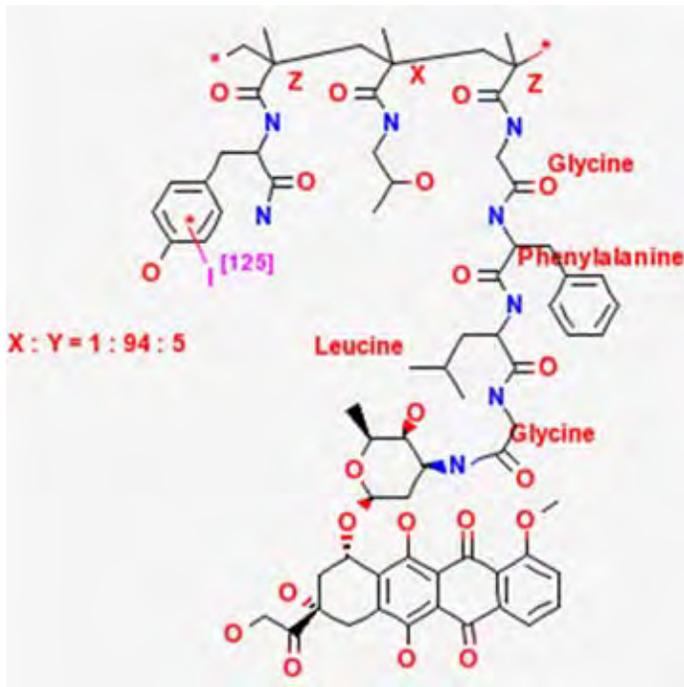


# Stereochemistry ← **Big** problem

- Import issues
  - Often obtained by perception of atom coordinates
    - Coordinates or stereo wedges may be ambiguous
  - Inconsistency between software packages for same file
- Export issues
  - Improper/inconsistent use of file format
- Format conversion adds/removes/changes stereo
- Relative stereochemistry improperly treated
- Depiction vs. machine readable
- Curated data may become corrupted!

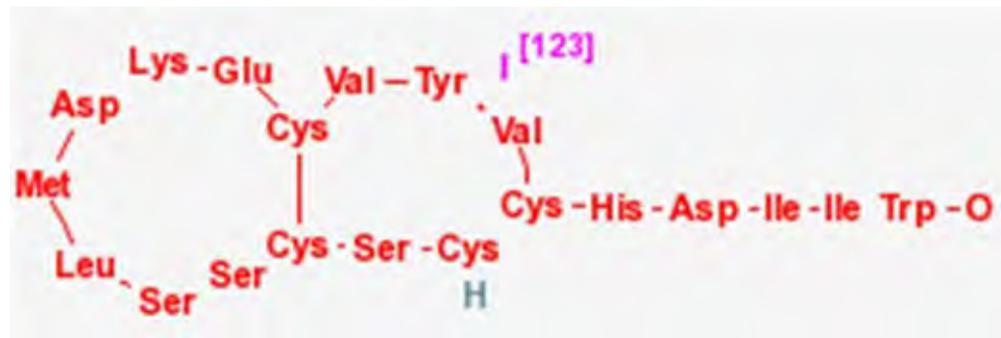
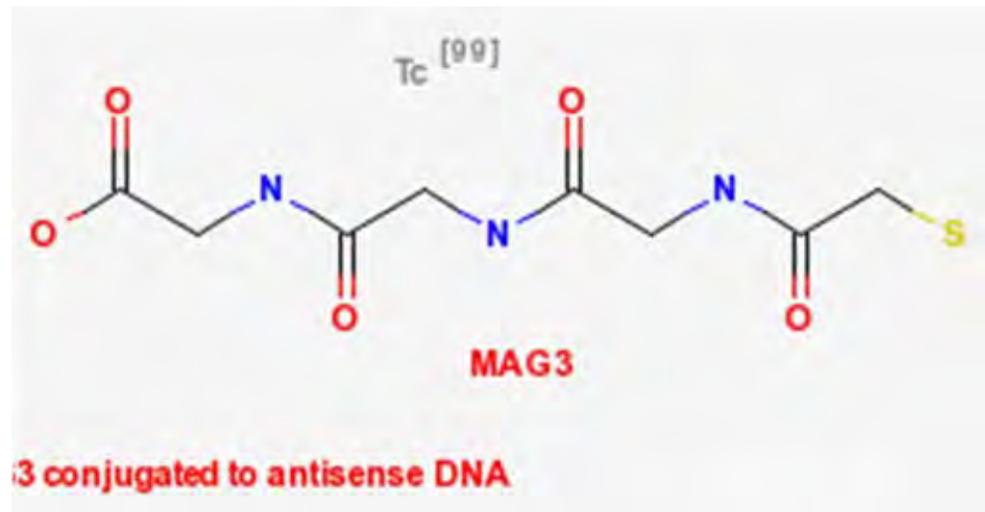


# Do we have a “defined” structure?



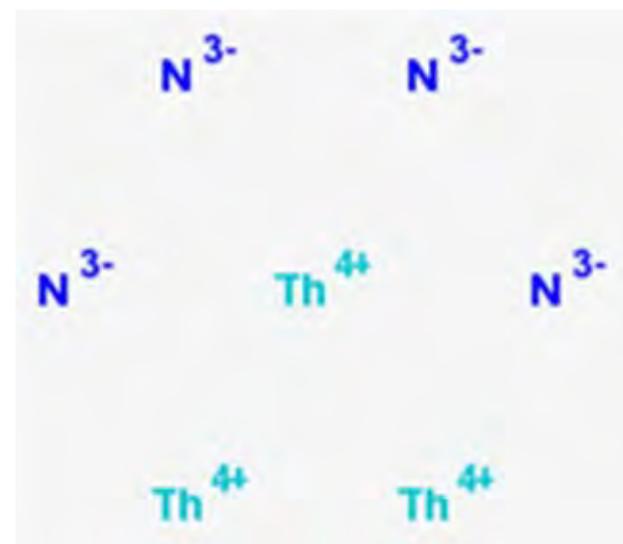
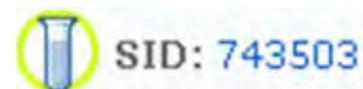
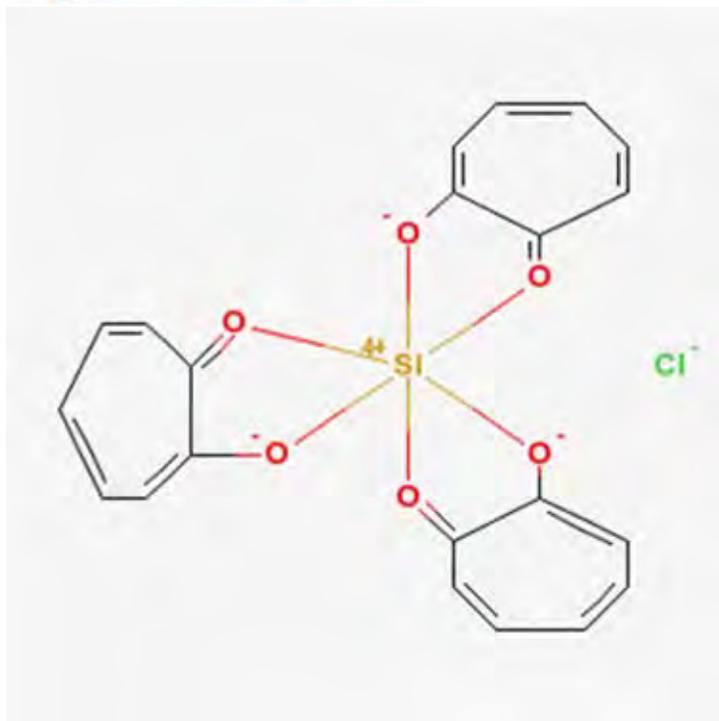
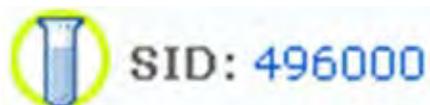
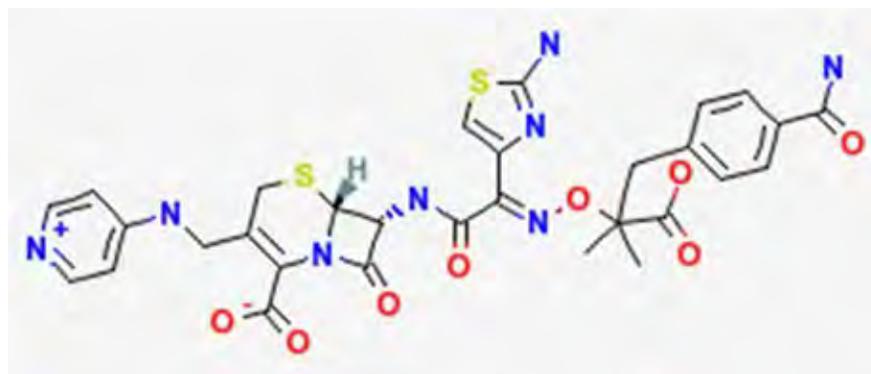
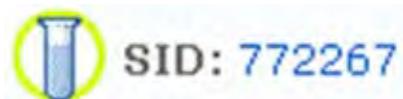
 SID: 5

 SID: 41



 SID: 8

# Is the structure reasonable?



# The (sad) state of chemical information

(End of aside)

# Automated structure processing...



- **Verification**

- Atom element
- Implicit hydrogen
- Functional group
- Valence

- **Standardization**

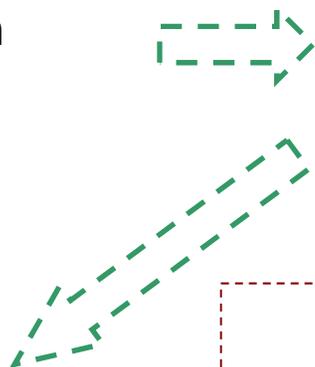
- Tautomer invariance
- Aromaticity detection
- Stereochemistry
- Explicit hydrogen

- **Calculation**

- Coordinates
- Properties
- Descriptors

- **Components**

- Isolate covalent units
- Neutralize (+/- proton)
- **Reprocess**
- Detect unique





# PubChem data access...

- Interfaces
  - text/numeric search
  - fielded/range search
  - precomputed similarities
    - 2-D, 3-D, identity groups
  - inter-database links
    - biomedical literature
    - MeSH ontology
      - biological roles
    - protein 3-D
    - pathways
  - external resource links
- Tools
  - bioactivity analysis
  - chemical clustering
  - chemical structure search
  - data download
  - FTP site
  - heatmap analysis
  - integrated 3-D layer
  - similarity computation
  - source summary
  - structure normalization

# PubChem data access...



- Interfaces
  - text/numeric search
  - fielded/range search
  - precomputed similarities
    - 2-D, 3-D, identity groups
  - inter-database links
    - biomedical literature
    - MeSH ontology
      - biological roles
    - protein 3-D
    - pathways
  - external resource links
- Tools
  - bioactivity analysis
  - chemical clustering
  - chemical structure search
  - data download
  - FTP site
  - heatmap analysis
  - integrated 3-D layer
  - similarity computation
  - source summary
  - structure normalization

# Entrez interface...

- Primary (text-based) search engine

Google-like approach... most likely answer is at the top...

Result record summaries

The screenshot shows the PubChem Compound search results page for 'aspirin'. The search bar at the top contains the query 'aspirin', which is highlighted with a red box. Below the search bar, there are options for 'Advanced Search', 'Preview/Index', 'History', 'Clipboard', and 'Details'. The results are displayed in a list format, with the first result being 'aspirin, ACETYL SALICYLIC ACID: 2-Acetoxybenzoic acid'. The second result is 'Calcascorbin, Calcicorbate, Febrosanal'. On the right side of the page, there is a 'Selected Compounds' table with a 'Compound Count' column. The table lists various categories and their corresponding counts, such as 'BioAssays, Active' (8), 'BioAssays, Tested' (14), 'Pharmacological Actions' (21), and 'Biological Properties' (65).

Selected Compounds	Compound Count
BioActivity Experiments	
BioAssays, Active	8
BioAssays, Tested	14
Protein 3D Structures	3
Crystal Structure Of Bovine Lactoperoxidase...	1
BioMedical Annotation	
Pharmacological Actions	21
Anti-inflammatory Agents, Non-Steroidal	20
BioSystems	2
Depositor Category	
Biological Properties	65
Chemical Vendors	51
Journal Publishers	22
NIH Molecular Libraries	8

User query

Rapid result subsets

# Entrez interface...

- Advanced search capability
  - makes it easy to rapidly create complex queries
  - helps with discoverability of indexes/filters

The screenshot displays the PubChem Compound Limits search interface within a Windows Internet Explorer browser window. The page features the NCBI and PubChem logos at the top. A search bar is present with the text "PubChem Compound" and a "Go" button. Below the search bar, there are several sections for advanced search criteria:

- Advanced Search:** Includes instructions on using field tags, Boolean operators (AND, OR, NOT), and square brackets for field tags.
- Search term limited to:** A dropdown menu set to "All Fields".
- Date:** Fields for "Date (YYYY-MM-DD)", "Create Date from", and "to".
- Chemical Properties:** A grid of fields for various properties such as MolecularWeight, MW, HydrogenBondDonorCount, HydrogenBondAcceptorCount, RotatableBondCount, TPSA, HeavyAtomCount, InteroAtomCount, TautomerCount, CovalentUnitCount, and Complexity.
- Stereochemistry:** Checkboxes for "No limit on chirality", "No limit on E/Z", "No chiral centers", "No E/Z centers", "Has chiral center(s)", "Has E/Z center(s)", "Fully unspecified chiral centers", "Fully unspecified E/Z centers", "Partially specified chiral centers", and "Partially specified E/Z centers".
- BioAssays:** Checkboxes for "Is any BioAssay", "Tested", "Probe", "Active", and "Inactive". Fields for "Test Concentration from" and "Active Concentration from" with units [uM] and [nM].
- Links:** Checkboxes for "PubMed", "PubMed via MeSH", "PubMed via Publisher", "Protein 3D Structure", "Protein Sequence", "BioSystems", "Nucleotide Sequence", "Pharmacological Action", and "MeSH".
- Elements:** A grid of checkboxes for chemical elements from H to U.
- Source:** A dropdown menu set to "Any".
- Category:** A dropdown menu set to "Any".

At the bottom of the page, there is a footer with the text "PubChem - Home | Search | Help | About | Feedback" and "PubChem | Compound | Substances | Warning".

# Entrez interface...

- “History” query result management
  - AND, OR, operations

PubChem Compound History - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/sites/entrez

NCBI PubChem Compound

My NCBI Welcome eebolton. [Sign Out]

Search PubChem Compound for aspirin Preview Go Clear

Limits Preview/Index History Clipboard Details

- Search History will be lost after eight hours of inactivity.
- Search numbers may not be continuous; all searches are represented.
- To save search indefinitely, click query # and select Save in My NCBI.
- To combine searches use #search, e.g., #2 AND #3 or click query # for more options.

Search	Most Recent Queries	Time	Result
#5 Search aspirin		08:23:08	71

Options

- AND
- OR
- NOT
- Delete
- Go
- Details
- Save in My NCBI

Clear

PubChem Help | Entrez Help | Write to the Help Desk

PubChem | Compound | Substance | BioAssay  
Chemical Structure Search | BioAssay Services | FTP Site | Deposit Data

NCBI | NLM | NIH | HHS | Privacy Statement | Freedom of Information Act | Disclaimer

http://www.ncbi.nlm.nih.gov/sites/entrez?querykey=5&dbase=pccompound&tab=Histc

# Entrez interface...

- Each database has lots of specialized indexes and filters
  - PubChem Compound
    - +50 indexes  
*e.g.*, aspirin[synonym]
    - +60 filters  
*e.g.*, ""has 3d conformer"[filter]



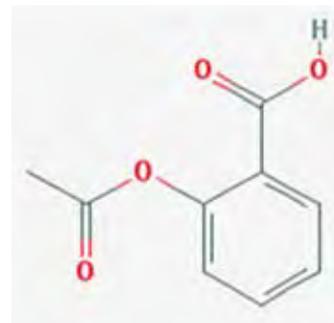
# Fielded queries to the rescue!

- Interested in chemical names?



Search just chemical name indexes

- “aspirin”
  - global keyword search - 69 hits
- “aspirin”[Synonym]
  - keyword search - 53 hits - many derivatives, mixtures, salts
- “aspirin”[CompleteSynonym]
  - exactly matches name - 1 hit



# Case study... “glucose”

- Search by global keyword ... 1,131 hits!



glucose - PubChem Compound Results - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/sites/entrez?db=pccompound&term=glucose

File Edit View Favorites Tools Help

glucose - PubChem Compound Results

NCBI PubChem Compound

My NCBI [Sign In] [Register]

Search PubChem Compound for glucose Go Clear Save Search

Advanced Search Preview/Index History Clipboard Details

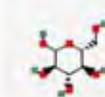
Display Summary Show 20 Sort By Send to

Tools: Links: Related Structures, BioAssays, BioSystems, Literature, Other Links

All: 1131 Rule of 5: 386

Items 1 - 20 of 1131 Page 1 of 57 Next

1: CID: 5793 Related Structures, BioAssays, BioSystems, Literature, Other Links

 dextrose; D-glucose; D-Glucopyranose  
IUPAC: (3R,4S,5S,6R)-6-(hydroxymethyl)oxane-2,3,4,5-tetrol  
MW: 180.155880 g/mol | MF: C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>  
Tested in BioAssays: All: 24, Active: 1; BioActivity Analysis

2: CID: 79025 Related Structures, BioSystems, Literature, Other Links

 alpha-D-glucose; alpha-Dextrose; alpha-D-Glucopyranose  
IUPAC: (2S,3R,4S,5S,6R)-6-(hydroxymethyl)oxane-2,3,4,5-tetrol  
MW: 180.155880 g/mol | MF: C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>

Selected Compounds Compound Count

- BioActivity Experiments
- BioAssays, Active 89
- BioAssays, Tested 205
- Protein 3D Structures 79
- Crystal Structure Of R120qNATIVE... 4
- BioMedical Annotation
- Pharmacological Actions 148
- Hypoglycemic Agents 84
- BioSystems 123
- Depositor Category
- Biological Properties 1,031
- Chemical Vendors 888
- Journal Publishers 267

Done Local intranet | Protected Mode: Off 100%

# Case study... “glucose”

- Search by “glucose[Synonym]” ... 975 hits!

The screenshot shows a Windows Internet Explorer browser window displaying the PubChem Compound search results for the query "glucose[synonym]". The browser address bar shows the URL "http://www.ncbi.nlm.nih.gov/sites/entrez". The search bar contains the text "glucose[synonym]" and the search button is labeled "Go". The results page shows a total of 975 hits, with the first 20 items displayed. The first two items are:

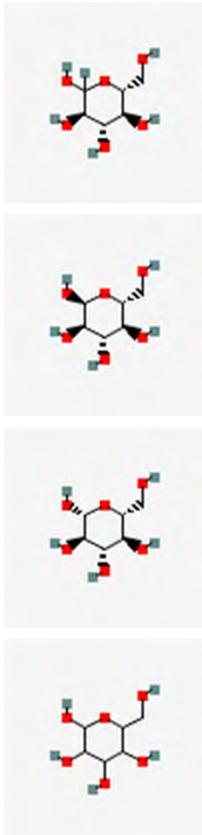
- Item 1: CID: 79025. Name: alpha-D-glucose, alpha-Dextrose, alpha-D-Glucopyranose. IUPAC: (2S,3R,4S,5S,6R)-6-(hydroxymethyl)oxane-2,3,4,5-tetrol. MW: 180.155880 g/mol | MF: C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>.
- Item 2: CID: 65533. Name: Glucose-1-phosphate, cori ester, alpha-D-glucose-1-phosphate. IUPAC: [(2R,3R,4S,5S,6R)-3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl] dihydrogen phosphate.

The right sidebar shows a table of "Selected Compounds" with a "Compound Count" column. The table lists various categories and their corresponding counts:

Selected Compounds	Compound Count
BioActivity Experiments	
BioAssays, Active	44
BioAssays, Tested	124
Protein 3D Structures	61
Processive Endocellulase Cell Of...	4
BioMedical Annotation	
Pharmacological Actions	36
Sweetening Agents	9
BioSystems	98
Depositor Category	
Biological Properties	880
Chemical Vendors	728
Journal Publishers	171

# Case study... “glucose”

- “glucose[CompleteSynonym]”... 4 hits!



glucose[CompleteSynonym] - PubChem Compound Results - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/sites/entrez

File Edit View Favorites Tools Help

glucose[CompleteSynonym] - PubChem Compound Re...

NCBI PubChem Compound

Search PubChem Compound for glucose[CompleteSynonym] Go Clear Save Search

Advanced Search Preview/Index History Clipboard Details

Display Summary Show 20 Sort By Send to

Tools: Links: Related Structures, BioAssays, BioSystems, Literature, Other Links

All: 4 Rule of 5: 4

Items 1 - 4 of 4 One page.

Selected Compounds	Compound Count
BioActivity Experiments	
BioAssays, Active	1
BioAssays, Tested	2
Protein 3D Structures	3
Diverse Substrates Recognition...	2
BioMedical Annotation	
Pharmacological Actions	4
Sweetening Agents	4
BioSystems	4
Depositor Category	
Biological Properties	4
Chemical Vendors	4
Journal Publishers	4

1: CID: 5793 Related Structures, BioAssays, BioSystems, Literature, Other Links

dextrose, D-glucose; D-Glucopyranose  
IUPAC: (3R,4S,5S,6R)-6-(hydroxymethyl)oxane-2,3,4,5-tetrol  
MW: 180.155880 g/mol | MF: C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>  
Tested in BioAssays: All: 24, Active: 1; BioActivity Analysis

2: CID: 79025 Related Structures, BioSystems, Literature, Other Links

alpha-D-glucose, alpha-Dextrose; alpha-D-Glucopyranose  
IUPAC: (2S,3R,4S,5S,6R)-6-(hydroxymethyl)oxane-2,3,4,5-tetrol  
MW: 180.155880 g/mol | MF: C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>

Done Local intranet | Protected Mode: Off 100%

# What is data quality?

## Ideal

- Validated
- Available
- Complete
- Succinct
  
- Useful
- Facile
- Seamless
  
- **Happy user**



## Usually found

- Best guess
- Something close
- Fragmented
- Verbose
  
- Might help
- Lots of work
- Issues
  
- **Frustrated user**

# What is data quality?

## Ideal

- Validated
- Available
- Complete
- Succinct
  
- Useful
- Facile
- Seamless
  
- **Happy user**



## Usually found

- Best guess
- Something close
- Fragmented
- Verbose
  
- Might help
- Lots of work
- Issues
  
- **Frustrated user**

# How many names in PubChem?

**49.0 million!**

Provided more than once:

**11.5 million** {23.5% of 49.0M}

Unique chemical names:

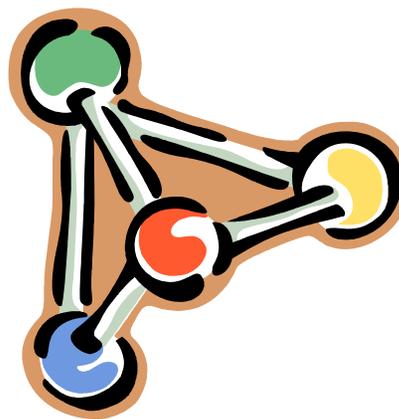
**4.65 million** {40.9% of 11.5M}



# Chemical name “consistency” filtering

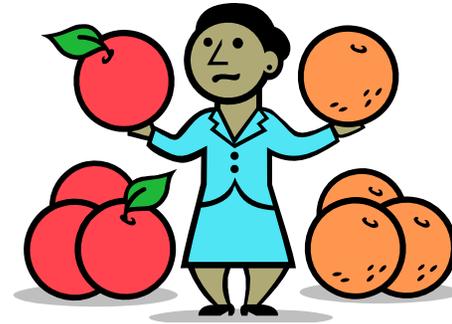
Ensure name-chemical associations are consistent at some level of structural “sameness”

- **Same structure**
- **Same stereo isomer**
  - varies by isotope
- **Same parent structure**
  - varies by charge/salt
- **Same parent stereo isomer**
  - varies by charge/salt/isotope
- **Same connectivity**
  - varies by isotope/stereo
- **Same parent connectivity**
  - varies by charge/salt/isotope/stereo



# Can match a name to one “chemical”?

- Yes! And often!
- One vote per depositor
  - First check that depositor is consistent
- But what consistency ratio?
  - 2 out of 3 is okay!
  - 3 out of 4 is okay!
  - 3 out of 5 is okay!



**60%**

# Affect of filtering on chemical names

- 4.65M unique chemical names
- Assign Synonym to a single “CID” ← Sliding quality scale
  - 4.61M (99.1%) names with “consistent” structure

## Observation:

Very few cases where inconsistency is found!

# Depositors agree... but stereo an issue

	One Vote, 60%	% of Total
CID	3,671,623	79.7%
STE	4,591	0.1%
PCID	40,209	0.9%
PSTE	6	0.0%
CON	887,314	19.3%
PCON	4,643	0.1%

- CID – same exact structure
  - no variation
- STE – same structure stereo form
  - variable isotopic form
- CON – same structure connectivity
  - variable stereo/isotopic form
- PCID – same exact parent structure
  - variable salt/charge state form
- PSTE – same parent structure stereo form
  - variable salt/charge state/isotopic form
- PCON – same parent structure connectivity
  - variable salt/charge state/isotopic/stereo form

# PubChem data access...



- Interfaces
  - text/numeric search
  - fielded/range search
  - precomputed similarities
    - 2-D, 3-D, identity groups
  - inter-database links
    - biomedical literature
    - MeSH ontology
      - biological roles
    - protein 3-D
    - pathways
  - external resource links
- Tools
  - bioactivity analysis
  - chemical clustering
  - chemical structure search
  - data download
  - FTP site
  - heatmap analysis
  - integrated 3-D layer
  - similarity computation
  - source summary
  - structure normalization

# Compound Summary

Aspirin - PubChem Public Chemical Database - Windows Internet Explorer

http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=2244

File Edit View Favorites Tools Help

Aspirin - PubChem Public Chemical Database

NCBI PubChem Compound

PubMed | Entrez | Structure | PubChem | Help

## Aspirin - Compound Summary (CID 2244)

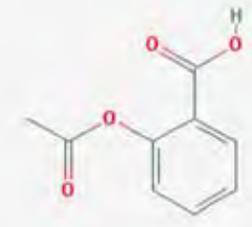
The prototypical analgesic used in the treatment of mild to moderate pain. It has anti-inflammatory and antipyretic properties and acts as an inhibitor of cyclooxygenase which results in the inhibition of the biosynthesis of prostaglandins. Aspirin also inhibits platelet aggregation and is used in the prevention of arterial and venous thrombosis. (From Martindale, The Extra Pharmacopoeia, 30th ed, p5)

Table of Contents

- BioMedical Annotation
  - Medication Information
  - Pharmacological Action
  - Pharmacological Classification
  - Chemical Classification
  - Safety and Toxicology
  - Literature Links
- BioAssay Results
- BioSystems and Pathways
- Protein Structures
- Synonyms
- Properties
- Descriptors
- Compound Information
- Substance Information
  - Category
- Exports

Structure & Quick Link Bar

2D 3D



Pc3D Viewer Download

Compound ID	2244	[?]
Molecular Weight	180.15742 [g/mol]	[?]
Molecular Formula	C <sub>9</sub> H <sub>8</sub> O <sub>4</sub>	[?]
XLogP3	1.2	[?]
H-Bond Donor	1	[?]
H-Bond Acceptor	4	[?]

Local intranet | Protected Mode: Off

# BioMedical Annotation

The screenshot shows a Windows Internet Explorer browser window displaying the PubChem Public Chemical Database page for Aspirin. The browser's address bar shows the URL <http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?dd=2244>. The page title is "Aspirin - PubChem Public Chemical Database".

The main content area is titled "BioMedical Annotation: (Total:1)". Below this, the word "Aspirin" is displayed in a large blue font. Underneath, there is a section for "Medication Information (Total 158)" with a "Show more ..." link. This section contains a list of medication entries:

- Drug Facts [Novartis Consumer Health, Inc.]
  - Indication & Usage
  - Warnings Overdosage
  - Dosage & Administration
- ASPIRIN 325mg [Genuine First Aid LLC]
- Chewable Aspirin 81mg [Genuine First Aid LLC]
- Drug Facts [Novartis Consumer Health, Inc.]

Below the medication information, there is a section for "Pharmacological Action" which lists several categories of actions:

- Anti-Inflammatory Agents, Non-Steroidal - Anti-inflammatory agents that are not steroids. In addition to anti-inflammatory actions, they have analgesic, antipyretic, and ...
- Antipyretics - Drugs that are used to reduce body temperature in fever.
- Cyclooxygenase Inhibitors - Compounds or agents that combine with cyclooxygenase (PROSTAGLANDIN-ENDOPEROXIDE SYNTHASES) and thereby prevent its substrate-enzyme ...
- Fibrinolytic Agents - Fibrinolysin or agents that convert plasminogen to FIBRINOLYSIN.
- Platelet Aggregation Inhibitors - Drugs or agents which antagonize or impair any mechanism leading to blood platelet aggregation, whether during the phases of activation and ...

At the bottom, there is a section for "Pharmacological Classification" with sub-links for "Chemical Actions and Uses", "Pharmacologic Actions", "Molecular Mechanisms of Pharmacological Action", and "Enzyme Inhibitors".

The browser's status bar at the bottom indicates "Local intranet | Protected Mode: Off" and a zoom level of "100%".

# BioMedical Annotation

PubChem Compound Links for MeSH (Select 68016861) - PubChem Compound Results - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/sites/entrez?Db=pubchem&from=mesh&Cmd=Link&Name=...

NCBI PubChem Compound

Search PubChem Compound for [ ] Go Clear

Advanced Search Preview/Index History Clipboard Details

Display Summary Show 20 Sort By Send to

Tools: Links: Related Structures, BioAssays, BioSystems, Literature, Other Links

All: 85 Rule of 5: 56

Items 1 - 20 of 85 Page 1 of 5 Next

1: CID: 44119558 Related Structures, Literature, Other Links

Naigesic; Nalfon. OR4881  
IUPAC: calcium 2-(3-phenoxyphenyl)propanoate dihydrate  
MW: 558.632400 g/mol | MF: C<sub>30</sub>H<sub>30</sub>CaO<sub>8</sub>

2: CID: 24195826 Related Structures, BioAssays, Literature, Other Links

Salicylate, sodium; Monosodium salicylate; Sodium salicylic acid  
IUPAC: 2-hydroxybenzoic acid; sodium  
MW: 161.110510 g/mol | MF: C<sub>7</sub>H<sub>6</sub>NaO<sub>3</sub>  
Tested in BioAssays: All: 7, Active: 0, BioActivity Analysis

	Count
BioActivity Experiments	
BioAssays, Active	48
BioAssays, Tested	63
Protein 3D Structures	13
Crystal Structure Of A Ternary	2
BioMedical Annotation	
Pharmacological Actions	85
Cyclooxygenase Inhibitors	74
BioSystems	3
Depositor Category	
Biological Properties	84
Chemical Vendors	81
Journal Publishers	69
NIH Molecular Libraries	53

Aspirin - PubChem Public Chemical Database - Windows Internet Explorer

BioMedical Annotation (Total: 1)

Aspirin

Medication Information (Total: 150)

- Drug Facts [Novartis Consumer Health, Inc.]  
Indications & Usage Warnings/Overdose  
Dosage & Administration
- ASPIRIN 325mg [Genuine First Aid LLC]
- Chewable Aspirin 81mg [Genuine First Aid LLC]
- Drug Facts [Novartis Consumer Health, Inc.]

Pharmacological Action

Anti-Inflammatory Agents, Non-Steroidal - Anti-inflammatory agents that are not steroids. In addition to...  
Antipyretics - Drugs that are used to reduce body temperature in fever.  
**Cyclooxygenase Inhibitors - Drugs that combine with cyclooxygenase (PROSTAGLANDIN ENDOPEROXIDE SYNTHASE) and thereby prevent its substrate-enzyme...**  
Fibrinolytic Agents - Enzymes or agents that convert plasminogen to FIBRINOLYSIN.  
Platelet Aggregation Inhibitors - Drugs or agents which antagonize or repair any mechanism leading to blood platelet aggregation, whether during the phases of activation and ...

Pharmacological Classification

Chemical Actions and Pharmacologic Actions  
Molecular Weights  
Molecular Weight of Pharmacological Action

Cyclooxygenase Inhibitors

# Safety and Toxicology ... Literature

Aspirin - PubChem Public Chemical Database - Windows Internet Explorer

http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=2244

File Edit View Favorites Tools Help

Aspirin - PubChem Public Chemical Database

## Safety and Toxicology

- [HSDB](#) - Peer-reviewed summary of toxicity and biomedical effects
- [NIOSH Pocket Guide](#) - NIOSH Material Data Safety Sheets
- [NIOSH ICSC](#) - NIOSH International Chemical Safety Cards
- [CCRIS](#) - Carcinogenicity, tumor promotion, tumor inhibition, and mutagenicity tests
- [EINECS](#) - European Inventory of Existing Commercial Chemical Substances
- [GENETOX](#) - Genetic toxicology information
- [Haz-Map](#) - Occupational toxicology information
- [TOXLINE](#) - Citations to the toxicological literature
- [LactMed](#) - Information on chemicals that breastfeeding mothers may be exposed
- [ClinicalTrials.gov](#) - Registry of federal and private clinical trials
- [NTP DBS](#) - Toxicological assay results

**Literature** Choose by Subheadings:

<a href="#">administration and dosage</a>	<a href="#">adverse effects</a>	<a href="#">agonists</a>
<a href="#">analogs and derivatives</a>	<a href="#">analysis</a>	<a href="#">antagonists and inhibitors</a>
<a href="#">blood</a>	<a href="#">cerebrospinal fluid</a>	<a href="#">chemical synthesis</a>
<a href="#">chemistry</a>	<a href="#">classification</a>	<a href="#">contraindications</a>
<a href="#">diagnostic use</a>	<a href="#">economics</a>	<a href="#">history</a>
<a href="#">immunology</a>	<a href="#">isolation and purification</a>	<a href="#">metabolism</a>
<a href="#">pharmacokinetics</a>	<a href="#">pharmacology</a>	<a href="#">physiology</a>
<a href="#">poisoning</a>	<a href="#">radiation effects</a>	<a href="#">standards</a>
<a href="#">supply and distribution</a>	<a href="#">therapeutic use</a>	<a href="#">therapy</a>
<a href="#">toxicity</a>	<a href="#">urine</a>	

Done Local intranet | Protected Mode: Off 100%

# Biological Assay Results

Aspirin - PubChem Public Chemical Database - Windows Internet Explorer

http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=2244

File Edit View Favorites Tools Help

Summary SGE Farm PubChem Yahoo! Google Monitoring ORCA OEDocs Str Wiki Suggested Sites Web Slice Gallery

Aspirin - PubChem Public Chemical Database

## BioAssay Results:

Tested in BioAssays: All: 1269 Active: 15 Inactive: 544

BioActivity Summary:

This Compound with Similar Compounds with Similar Conformers

Total Data (row): 1536 .. see all Download

#	AID	SID	Activity	Activity Concentration	Value [µM]	BioAssay	Protein Target	DataTable
1	443491	103164874	IC50 PubChem standard value	0.162	Inhibition of potato LOX-5 assessed as inhibition of hydroperoxide production after 5 mins by EIA			
2	240795	103164874	IC50 PubChem standard value	0.3	In vitro inhibitory concentration against COX-1 enzyme	RecName: Full=Prostaglandin G/H synthase 1; AltName: Full=Cyclooxygenase-1; Short=COX-1; AltName: Full=Prostaglandin H2 synthase 1; Short=PGH synthase 1; Short=PGHS-1; Short=PHS 1; AltName: Full=Prostaglandin-endoperoxide synthase 1; Flags: Precursor[gi:548481]		
3	288821	103164874	IC50 PubChem standard value	0.3	Inhibition of ovine COX1 by enzyme immuno assay	RecName: Full=Prostaglandin G/H synthase 1; AltName: Full=Cyclooxygenase-1; Short=COX-1; AltName: Full=Prostaglandin H2 synthase 1; Short=PGH synthase 1; Short=PGHS-1; Short=PHS 1; AltName: Full=Prostaglandin-endoperoxide synthase 1; Flags: Precursor[gi:548481]		
4	344873	103164874	IC50 PubChem standard value	0.3	Inhibition of ovine COX1 by enzyme immuno assay	RecName: Full=Prostaglandin G/H synthase 1; AltName: Full=Cyclooxygenase-1; Short=COX-1; AltName: Full=Prostaglandin H2 synthase 1; Short=PGH synthase 1; Short=PGHS-1; Short=PHS 1; AltName: Full=Prostaglandin-endoperoxide synthase 1; Flags: Precursor[gi:548481]		
5	352495	103164874	IC50 PubChem standard value	0.3	Inhibition of ovine COX1 by enzyme immuno assay	RecName: Full=Prostaglandin G/H synthase 1; AltName: Full=Cyclooxygenase-1; Short=COX-1; AltName: Full=Prostaglandin H2 synthase 1; Short=PGH synthase 1; Short=PGHS-1; Short=PHS 1; AltName: Full=Prostaglandin-endoperoxide synthase 1; Flags: Precursor[gi:548481]		

■ Chemical Probe 
 ■ Active 
 ■ Inactive 
 ■ Inconclusive 
 ■ Unspecified

Local intranet | Protected Mode: Off 100%

# Pathway and Protein Information

Aspirin - PubChem Public Chemical Database - Windows Internet Explorer

http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?od=2244

File Edit View Favorites Tools Help

Summary SGE Farm PubChem Yahoo! Google Monitoring ORCA OEDocs Str Wiki Suggested Sites Web Slice Gallery

Aspirin - PubChem Public Chemical Database

**BioSystems and Pathways:**

**Bile secretion** Conserved  
Bile is a vital secretion, essential for digestion and absorption of fats and fat-soluble vitamins in the small intestine. Moreover, bile is an important route of elimination for excess cholesterol ...

**Bile secretion** Organism-specific  
Bile is a vital secretion, essential for digestion and absorption of fats and fat-soluble vitamins in the small intestine. Moreover, bile is an important route of elimination for excess cholesterol ...

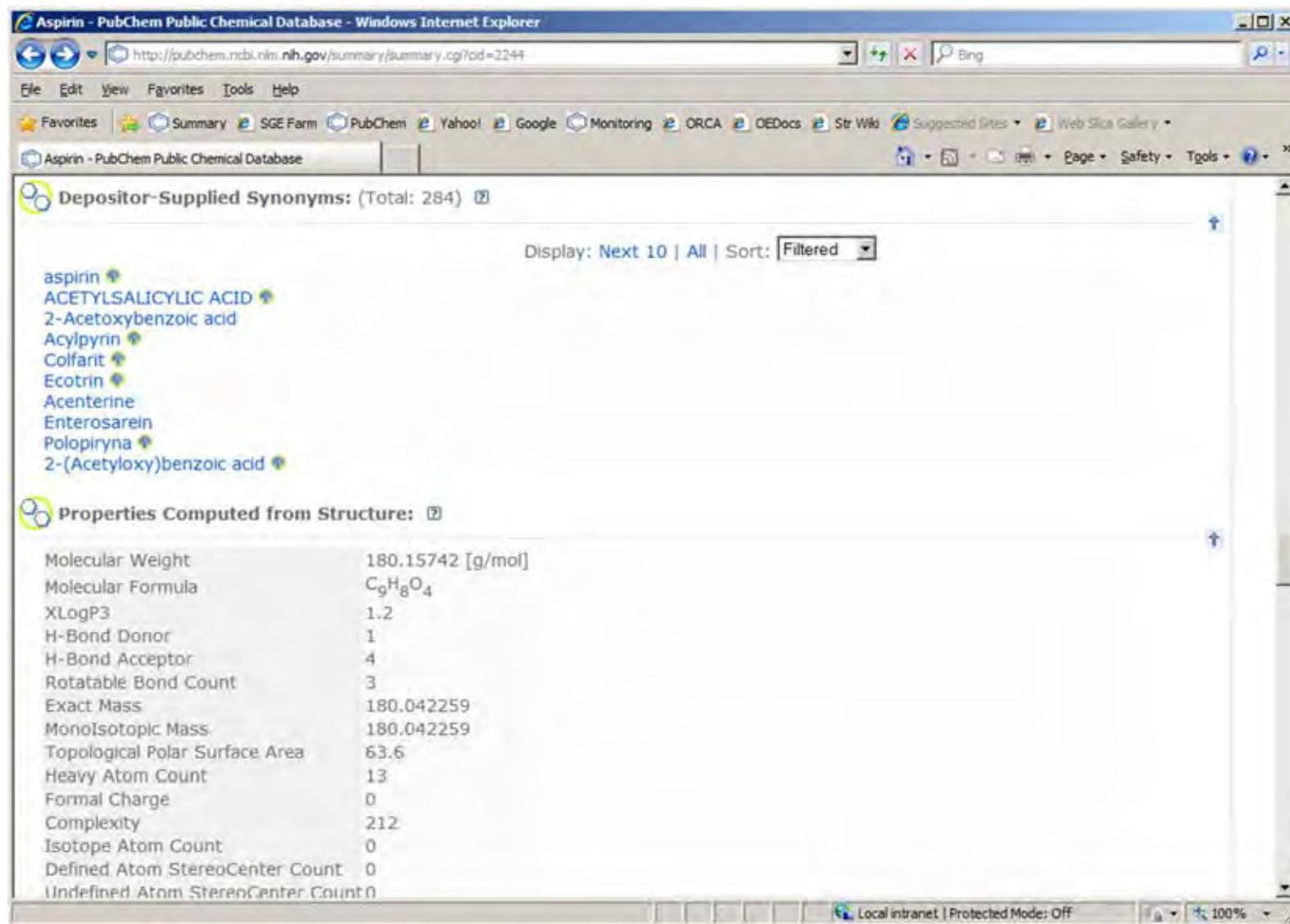
**Bile secretion** Organism-specific  
Bile is a vital secretion, essential for digestion and absorption of fats and fat-soluble vitamins in the small intestine. Moreover, bile is an important route of elimination for excess cholesterol ...  
more...

**Protein Structures:** (Total: 5)

	<b>MMDB ID: 75951 PDB ID: 3IAZ</b> Structural Basis Of The Prevention Of Nsaid-Induced Damage Of The Gastrointestinal Tract By C-Terminal Half (C-Lobe) Of Bovine Colostrum Protein Lactoferrin: Binding And Structural Studies Of The C-Lobe Complex With Aspirin <b>Taxonomy: Bos taurus</b>
	<b>MMDB ID: 70578 PDB ID: 3GCL</b> Mode Of Ligand Binding And Assignment Of Subsites In Mammalian Peroxidases: Crystal Structure Of Lactoperoxidase Complexes With Acetyl Salicylic Acid, Salicylhydroxamic Acid And Benzylhydroxamic Acid <b>Taxonomy: Bos taurus</b>
	<b>MMDB ID: 54234 PDB ID: 2QQT</b> Crystal Structure Of The Complex Of Bovine Lactoperoxidase With Acetyl Salicylic Acid At 2.5 A Resolution <b>Taxonomy: Bos taurus</b>

Local intranet | Protected Mode: Off 100%

# Synonyms and Computed Properties



The screenshot shows a web browser window displaying the PubChem page for Aspirin. The browser title is "Aspirin - PubChem Public Chemical Database - Windows Internet Explorer". The address bar shows the URL: <http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=3244>. The browser's Favorites bar includes links to Summary, SGE Farm, PubChem, Yahoo!, Google, Monitoring, ORCA, OEDocs, Str Wiki, Suggested Sites, and Web Slice Gallery. The main content area is divided into two sections:

**Depositor-Supplied Synonyms: (Total: 284)**

Display: Next 10 | All | Sort: Filtered

- aspirin
- ACETYLSALICYLIC ACID
- 2-Acetoxybenzoic acid
- Acylpyrin
- Colfarit
- Ecotrin
- Acenterine
- Enterosarein
- Polopiryna
- 2-(Acetyloxy)benzoic acid

**Properties Computed from Structure:**

Molecular Weight	180.15742 [g/mol]
Molecular Formula	C <sub>9</sub> H <sub>8</sub> O <sub>4</sub>
XLogP3	1.2
H-Bond Donor	1
H-Bond Acceptor	4
Rotatable Bond Count	3
Exact Mass	180.042259
MonoIsotopic Mass	180.042259
Topological Polar Surface Area	63.6
Heavy Atom Count	13
Formal Charge	0
Complexity	212
Isotope Atom Count	0
Defined Atom StereoCenter Count	0
Undefined Atom StereoCenter Count	0

The browser's status bar at the bottom indicates "Local intranet | Protected Mode: Off" and a zoom level of 100%.

# Compound and Substance Information

Aspirin - PubChem Public Chemical Database - Windows Internet Explorer  
http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=2244

File Edit View Favorites Tools Help

Summary SGE Farm PubChem Yahoo! Google Monitoring ORCA OEDocs Str Wiki Suggested Sites Web Slice Gallery

Aspirin - PubChem Public Chemical Database

### Compound Information:

**CID 2244**  
Create Date: 2004-09-16

**Related Compounds:**  
Same, Connectivity: [4 Links](#)

**Similar Compounds:** [1629 Links](#)  
**Similar Conformers:** [3079 Links](#) [View Conformers](#)

### Substance Information:

**Substances:**  
All: [976 Links](#)  
Same structure: [149 Links](#)  
Mixture: [827 Links](#)

**Category:** [for same structure substances]  
Biological Properties: [80 Links](#)

- BIDD ( 1 )  
SID 93166276 - External ID: BIDD:GT0118
- BindingDB ( 1 )  
SID 49846175 - External ID: 22360
- ChEBI ( 1 )  
SID 8143164 - External ID: CHEBI:15365
- ChemBank ( 25 )
- CHEMBL ( 1 )  
SID 103164874 - External ID: CHEMBL25
- ChemSpider ( 32 )
- Circadian Research, Kay Laboratory, University of California at San Diego (UCSD) ( 1 )  
SID 92303504 - External ID: A 5376
- Comparative Toxicogenomics Database ( 1 )

Local intranet | Protected Mode: Off 100%

# Streamlined access to depositor websites

The image displays two overlapping browser windows. The left window shows the ChEBI website page for acetylsalicylic acid (CHEBI:15365). The right window shows a list of external links to depositor websites, with a red circle highlighting the link to the University of California San Diego (UCSD) and a blue arrow pointing from it to the ChEBI entry.

**ChEBI Entry: acetylsalicylic acid (CHEBI:15365)**

**Chemical Structure:** CC(=O)Oc1ccccc1C(=O)O

**ChEBI Name:** acetylsalicylic acid

**ChEBI ID:** CHEBI:15365

**Definition:** Benzoic acid carrying an acetoxy group at the 2-position.

**Stars:** ★★★ This entity has been manually annotated by the ChEBI team.

**Secondary ChEBI IDs:** CHEBI:101280, CHEBI:40705, CHEBI:2890, CHEBI:22188, CHEBI:22189

**See structure as:**  Image  Applet  Download Molecule

**Find compounds which contain this structure**

**Find compounds which resemble this structure**

**more structures >>>**

**InChI:** InChI=1S/C9H8O4/c1-6(10)13-8-5-3-2-4-7(8)9(11)12/h2-6H,1H3,(H,11,12)

**InChIKey:** InChIKey=BSYNRYMUTXBXSQ-UHFFFAOYSA-N

**SMILES:** CC(=O)Oc1ccccc1C(=O)O

**Formula:** C<sub>9</sub>H<sub>8</sub>O<sub>4</sub>

**Source:** KEGG COMPOUND

**External Links (from right window):**

- Substances: 8
- AE: 976 Links
- Same structure: 149 Links
- History: 327 Links
- Category: [for same structural substances] 8
- Biological Properties: 80 Links
- BIDD (1)
- SD 92166278 - External ID: BIDD 070118
- BindingDB (1)
- SD 49846179 - External ID: 22360
- CHEBI (1)
- SD 8143194 - External ID: CHEBI:15365
- ChEMBL (25) @
- ChEMBL (1)
- SD 10311111 - External ID: CHEBI:15365
- ChEMBL (1)
- University Laboratory, University of California at San Diego (UCSD) (1)
- SD 45376 - External ID: A 5376
- Toxicogenomics Database (1)
- SD 2788943 - External ID: D001241
- CrossyGate (1)
- SD 8191516 - External ID: 2244
- DrugBank (1)
- SD 4680803 - External ID: D60945
- OTY/NCI (2)
- SD 87798 - External ID: 27223
- SD 476108 - External ID: 40A188
- ICCB-Longwood/MSRB Screening Facility, Harvard Medical School (6) @
- LeadScope (1)
- SD 8984366 - External ID: LS-143
- NextBio (1)



# Entrez interface...

- Primary (text-based) search engine

The screenshot shows a web browser window titled "aspirin - PubChem Compound Results". The address bar displays the URL: <http://www.ncbi.nlm.nih.gov/sites/entrez?db=pccompound&term=aspirin>. The browser's menu bar includes File, Edit, View, Favorites, Tools, and Help. The toolbar shows various icons for navigation and search. The main content area features the NCBI logo and the PubChem Compound logo. A search bar contains the text "PubChem Compound" and "for aspirin". Below the search bar are buttons for "Go", "Clear", and "Save Search". There are also buttons for "Advanced Search", "Preview/Index", "History", "Clipboard", and "Details". The display options are set to "Summary", "Show 20", and "Sort By". A "Tools" section includes icons for printing and linking, with links to "Related Structures", "BioAssays", "BioSystems", "Literature", and "Other Links". The search results show "All: 69" items, with a "Rule of 5: 15" filter. The first two results are displayed: 1. CID: 2244, aspirin, ACETYLSALICYLIC ACID: 2-Acetoxybenzoic acid, with IUPAC, MW, and MF information. 2. CID: 6247, Calcascorbin, Calscorbate, Febrosanal, with IUPAC, MW, and MF information. A "Selected Compounds" table is visible on the right side of the page.

Selected Compounds	Compound Count
BioActivity Experiments	
BioAssays, Active	8
BioAssays, Tested	14
Protein 3D Structures	3
Crystal Structure Of Bovine Lactoperoxidase...	1
BioMedical Annotation	
Pharmacological Actions	21
Anti-inflammatory Agents, Non-Steroidal	20
BioSystems	2
Depositor Category	
Biological Properties	65
Chemical Vendors	51
Journal Publishers	22
NIH Molecular Libraries	8

# BioActivity Analysis Tool

aspirin - PubChem Compound Results - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/pc...  
aspirin - PubChem Compound

NCBI PubChem Compound

Search PubChem Compound for aspirin

Display: Summary Show 20 Sort By Send to

Tools **BioAssays** Links: Related Structures, BioAssays, Literature, Other Links

All: 71 BioAssays: 20 Protein3D: 4 Rule of 5: 23

Items 1 - 20 of 71 Page 1 of 4 Next Recent Activity

1: CID: 2244 Related Structures, BioAssays, Literature, Other Links

 aspirin, Acetylsalicylic acid  
IUPAC: 2-acetoxybenzoic acid  
MW: 180.157420 g/mol | MF: C<sub>9</sub>H<sub>8</sub>O<sub>4</sub>  
Tested in BioAssays: All: 284 Active: 2: BioActivity Analysis  
Cyclooxygenase Inhibitors... more

2: CID: 6247 Related Structures, Literature

 Calcisambolol, Calcisambolol, Febrosanal  
IUPAC: calcium 2-acetoxybenzoate  
MW: 398.376960 g/mol | MF: C<sub>19</sub>H<sub>14</sub>CaO<sub>8</sub>  
Cyclooxygenase Inhibitors... more

3: CID: 24847966 Related Structures

PubChem BioActivity Analysis: Summary - Windows Internet Explorer

http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi

File Edit View Favorites Tools Help

NCBI PubChem BioAssay

PubChem BioAssay

BioActivity Analysis: 1864 Bioassays (1930 components), 69 Compounds and 1829 Protein Targets

Summary DataTable Structure-Activity

Revise BioAssay and Compound Selection

BioAssay Summary: Total Bioassays (1864) (1930 components) MLP (722) Defined Protein Target (628) Summary/Confirmatory (357) Primary (264)

Compound Summary: Total Compounds (69) Active Compounds (8)

Structure 1 - 20 of 68

BioAssays Targets Compounds

Total Pages: 97 Display: 20 Page 1

Selected BioAssays

Active	Inactive	Tested	AC <sub>50</sub> [μM]	AC <sub>50</sub> [nM]	AC Range	BioAssay [Outcome Type]	Protein Target
<input type="checkbox"/>	<input type="checkbox"/>	1	0.0018			qMTS Validation Assay for Inhibitors of Human's cyclooxygenase (COX) [Confirmatory]	Human cyclooxygenase-2 [ncic500000000]
<input type="checkbox"/>	<input type="checkbox"/>	8	0.0018			qMTS Assay for Inhibitors of Human's cyclooxygenase (COX) [Confirmatory]	Human cyclooxygenase-2 [ncic500000000]
<input type="checkbox"/>	<input type="checkbox"/>	1	0.0224			qMTS Assay for Inhibitors of Human's Lysine Methyltransferase 2 [Confirmatory]	Human lysine methyltransferase 2 [ncic500000000]

Local intranet | Protected Mode: Off

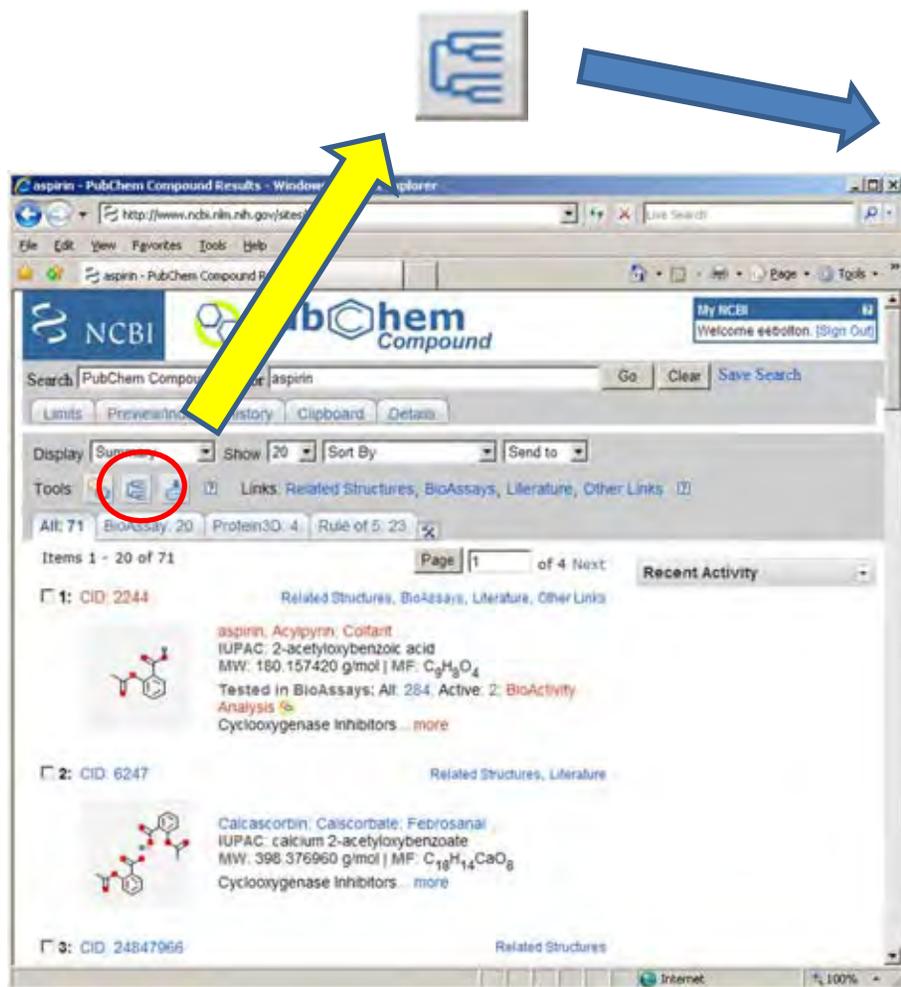
# Download Facility

The image illustrates the download facility on the PubChem website. It consists of two screenshots from a Windows Internet Explorer browser.

**Left Screenshot: PubChem Compound Results**  
The browser window shows the search results for "aspirin". The search bar contains "aspirin". Below the search bar, there are options for "Limits", "Preview/Index", "Library", "Clipboard", and "Details". The "Tools" section includes a download icon (a blue box with a white arrow pointing down) which is circled in red. A yellow arrow points from this icon to the right screenshot. The main content area displays a list of results, with the first result being "aspirin, Acetylsalicylic acid" (CID: 2244). The second result is "Calcascorbil, Calcisorbate, Febrosanal" (CID: 6247). The third result is "CID: 24847966".

**Right Screenshot: PubChem Download Service**  
The browser window shows the "PubChem Download Service" page. The URL is [http://pubchem.ncbi.nlm.nih.gov/pc\\_fetch/pc\\_fetch.cgi?db=pccompound&doc](http://pubchem.ncbi.nlm.nih.gov/pc_fetch/pc_fetch.cgi?db=pccompound&doc). The page title is "PubChem Download Service". The main content area includes a "Download" button and a "Choose a format" dropdown menu. The dropdown menu is open, showing options: "SDF", "Text ASN.1", "Binary ASN.1", "XML", "SDF", "Image", "Small Image", "SMILES", and "InChI". Below the dropdown menu, there are options for "Choose a compression type", "Retrieve 3D records/images?", and "Save this job in XML format (e.g. for PUG)". A "Save Job" button is located at the bottom of the form. The footer contains links for "Write to Helpdesk", "Disclaimer", "Privacy statement", and "Accessibility".

# Structure Clustering Tool



aspirin - PubChem Compound Results - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/sites/

NCBI PubChem Compound

Search PubChem Compound for aspirin

Display Summary Show 20 Sort By Send to

Tools  Links: Related Structures, BioAssays, Literature, Other Links

All: 71 BioAssay: 20 Protein3D: 4 Rule of 5: 23

Items 1 - 20 of 71 Page 1 of 4 Next Recent Activity

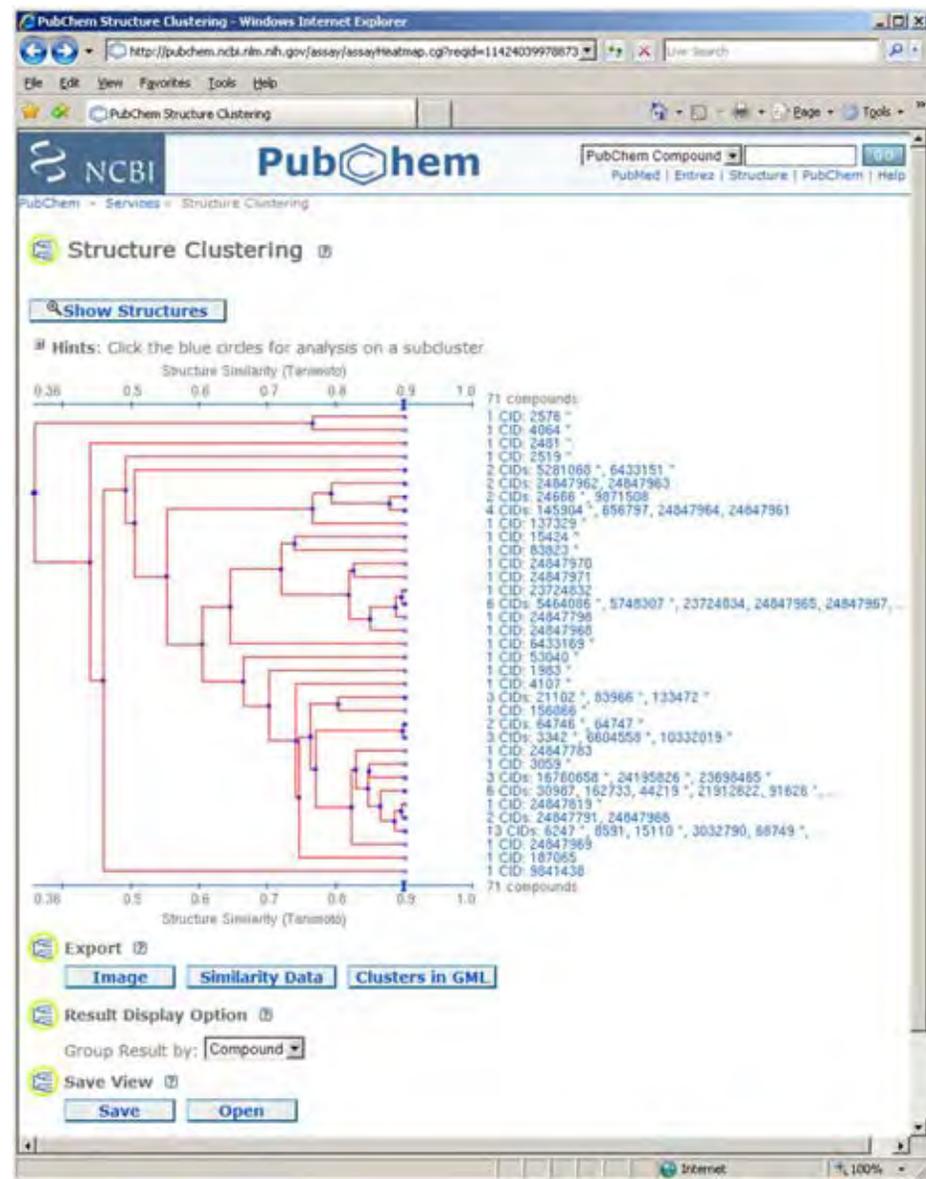
1: CID: 2244 Related Structures, BioAssays, Literature, Other Links

 aspirin, Acetylsalicylic acid  
IUPAC: 2-acetoxybenzoic acid  
MW: 180.157420 g/mol | MF: C<sub>9</sub>H<sub>8</sub>O<sub>4</sub>  
Tested in BioAssays: All: 284 Active: 2 BioActivity Analysis  
Cyclooxygenase Inhibitors... more

2: CID: 6247 Related Structures, Literature

 Calcisarcosine, Calcisarcosate, Febrosanal  
IUPAC: calcium 2-acetoxybenzoate  
MW: 398.376960 g/mol | MF: C<sub>18</sub>H<sub>14</sub>CaO<sub>6</sub>  
Cyclooxygenase Inhibitors... more

3: CID: 24847966 Related Structures



PubChem Structure Clustering - Windows Internet Explorer

http://pubchem.ncbi.nlm.nih.gov/assay/assay/heatmap.cgi?regid=11424039970873

NCBI PubChem

PubChem Compound PubMed | Entrez | Structure | PubChem | Help

Structure Clustering

Show Structures

Hints: Click the blue circles for analysis on a subcluster

Structure Similarity (Tanimoto)

0.36 0.5 0.6 0.7 0.8 0.9 1.0

71 compounds

- 1 CID: 2578 \*
- 1 CID: 4064 \*
- 1 CID: 2481 \*
- 1 CID: 2519 \*
- 2 CIDs: 5281068 \*, 6433151 \*
- 2 CIDs: 24847962, 24847963
- 2 CIDs: 24666 \*, 4971508
- 4 CIDs: 145304 \*, 656797, 24847964, 24847961
- 1 CID: 137329 \*
- 1 CID: 15424 \*
- 1 CID: 83823 \*
- 1 CID: 24847970
- 1 CID: 24847971
- 1 CID: 23724532
- 6 CIDs: 5464086 \*, 5748307 \*, 23724834, 24847965, 24847967, ...
- 1 CID: 24847796
- 1 CID: 24847968
- 1 CID: 6433163 \*
- 1 CID: 53040 \*
- 1 CID: 1963 \*
- 1 CID: 4197 \*
- 3 CIDs: 21102 \*, 83566 \*, 133472 \*
- 1 CID: 156866 \*
- 2 CIDs: 64746 \*, 64747 \*
- 3 CIDs: 3342 \*, 6604558 \*, 10332019 \*
- 1 CID: 24847783
- 1 CID: 3059 \*
- 3 CIDs: 16780658 \*, 24195826 \*, 23698485 \*
- 6 CIDs: 30987, 162733, 44219 \*, 21912622, 91826 \*, ...
- 1 CID: 24847819 \*
- 2 CIDs: 24847791, 24847968
- 13 CIDs: 6247 \*, 8591, 15110 \*, 3032790, 86749 \*, ...
- 1 CID: 24847969
- 1 CID: 187065
- 1 CID: 8841438

71 compounds

Structure Similarity (Tanimoto)

0.36 0.5 0.6 0.7 0.8 0.9 1.0

Export

Image Similarity Data Clusters in GML

Result Display Option

Group Result by: Compound

Save View

Save Open

# Structure Clustering Tool

aspirin - PubChem Compound Results - Windows Internet Explorer

http://www.ncbi.nlm.nih.gov/sites/

NCBI PubChem Compound

Search PubChem Compound for aspirin

Display Summary Show 20 Sort By Send to

Tools  Links: Related Structures, BioAssays, Literature, Other Links

All: 71 BioAssay: 20 Protein3D: 4 Rule of 5: 23

Items 1 - 20 of 71 Page 1 of 4 Next Recent Activity

1: CID: 2244 Related Structures, BioAssays, Literature, Other Links

 aspirin, Acetylsalicylic acid  
IUPAC: 2-acetoxybenzoic acid  
MW: 180.157420 g/mol | MF: C<sub>9</sub>H<sub>8</sub>O<sub>4</sub>  
Tested in BioAssays: All: 284 Active: 2 BioActivity Analysis  
Cyclooxygenase Inhibitors... more

2: CID: 6247 Related Structures, Literature

 Calcisorbite, Calcisorbate, Febrosanal  
IUPAC: calcium 2-acetoxybenzoate  
MW: 398.376960 g/mol | MF: C<sub>18</sub>H<sub>14</sub>CaO<sub>8</sub>  
Cyclooxygenase Inhibitors... more

3: CID: 24847966 Related Structures

PubChem Structure Clustering - Windows Internet Explorer

http://pubchem.ncbi.nlm.nih.gov/assay/assay/heatmap.cgi?reqid=66813699651579

NCBI PubChem

PubChem Compound

Structure Clustering

Show IDs Only

Hints: Click the blue circles for analysis on a subcluster

Structure Similarity (Tanimoto)

0.36 0.5 0.6 0.7 0.8 0.9 1.0 71 compounds

2 CID: 2576, 4064

1 CID: 2481

Select the blue circles or the lines on their left to click

1 CID: 2513

2 CID: 5281068, 6433151

64 CID: 24847962, 24847963, 248695, 9271506, 145504

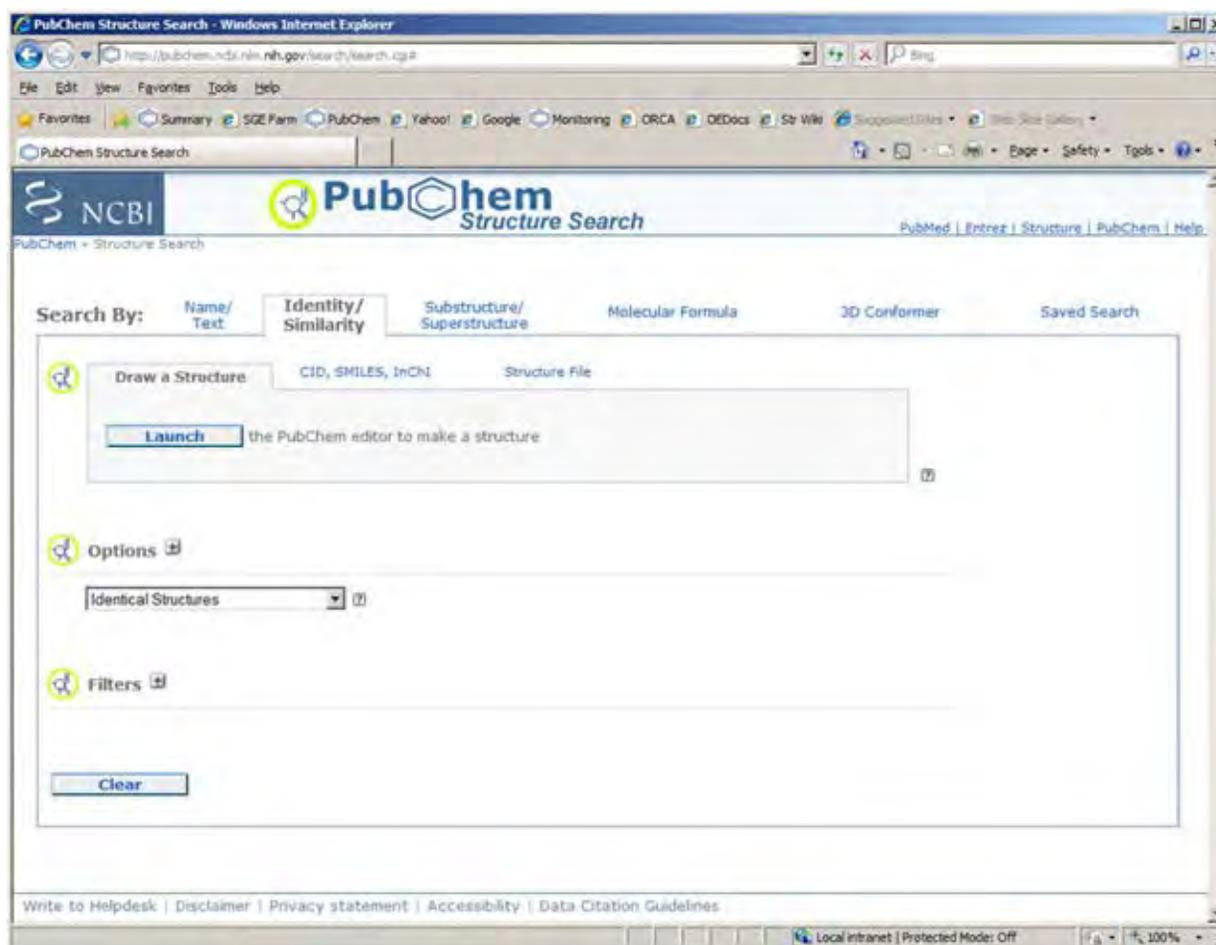
1 CID: 9941438

Structure Similarity (Tanimoto)

0.36 0.5 0.6 0.7 0.8 0.9 1.0 71 compounds

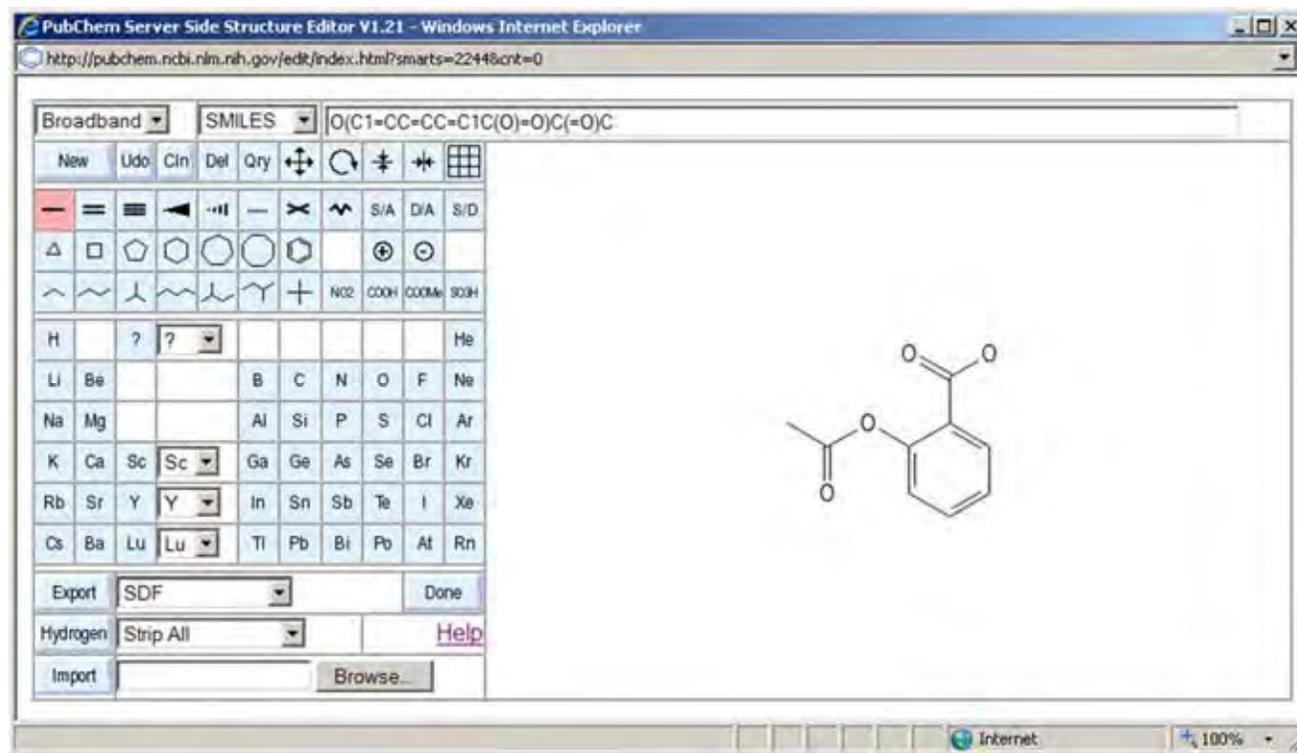
# Chemical structure search

- Structure query interface
  - One tab for each query type



# Chemical structure sketcher

- Ability to dynamically enter complex structural queries without a plugin



Ihlenfeldt WD, Bolton EE, Bryant SH. The PubChem chemical structure sketcher.  
*J Cheminform.* 2009 Dec 17;1(1):20. [PMID: 20298522]

# Score Matrix Service

- Pair-wise scores in matrix format
  - Similarity scores between compounds
- Allows users to obtain PubChem scores for arbitrary CID lists
- Enables further (external) analysis

The screenshot shows the PubChem Score Matrix Service web interface in a Windows Internet Explorer browser. The page title is "PubChem Score Matrix Service" and the URL is "http://pubchem.ncbi.nlm.nih.gov/score\_matrix/score\_matrix.cgi". The interface includes a "Submit Job" button and a "Save Job" button. The "Score Type" is set to "2D Similarity (substructure keys)". The "CID List" section has three input fields: "Enter CIDs", "Upload a file with CIDs...", and "Choose from Entrez History". The "Second CID List" section has three input fields: "If none, will get self-scores for primary list", "Enter CIDs", "Upload a file with CIDs...", and "Choose from Entrez History". The "Format" is set to "CSV" and the "Compression" is set to "Gzip (.gz)".

PubChem Score Matrix Service

Submit Job Submit this job to PubChem's score matrix service

Score Type Method used to compute scores  
2D Similarity (substructure keys) Choose score type

CID List Primary list of CIDs  
Enter CIDs  
Upload a file with CIDs...  
Choose from Entrez History

Second CID List Secondary list of CIDs  
If none, will get self-scores for primary list  
Enter CIDs  
Upload a file with CIDs...  
Choose from Entrez History

Format Format of the matrix file  
CSV Choose format

Compression Compression of the matrix file  
Gzip (.gz) Choose compression

Save Job Save this job in XML format (e.g. for PUG)

Write to Helpdesk | Disclaimer | Privacy statement | Accessibility

# PubChem data submission

How do users put data into PubChem?

**PubChem Deposition Gateway**

This site is for users who want to deposit new data into PubChem and/or test data upload procedures. To search PubChem you may use the [PubChem home page](#).

To test data upload procedures create a test account using the link below. To deposit new data into PubChem use the link below to create a new deposition account or to log into a previously created account.

New User: ?	Existing User: ?
<a href="#">Create Test Account</a> ?	Username: <input type="text"/>
<a href="#">Create Deposition Account</a> ?	Password: <input type="password"/>
	<a href="#">Log In</a> <a href="#">Forgot Password?</a>

[Write to Helpdesk](#) | [Disclaimer](#) | [Privacy statement](#) | [Accessibility](#)

<http://pubchem.ncbi.nlm.nih.gov/deposit>

# Standardization Service

- Performs PubChem chemical structure “standardization”
  - Provides CID if structure is in PubChem
- Allows users to examine PubChem methodology affects on their data

The screenshot shows the PubChem Standardization Service web interface in a Windows Internet Explorer browser. The page title is "PubChem Standardization Service" and the URL is "http://pubchem.ncbi.nlm.nih.gov/standardize/standardize.cgi". The interface includes a "Submit Job" button and a "Save Job" button. Under the "Input" section, there are three radio buttons: "Start with SMILES", "Start with InChI", and "Start with an SDF file" (with a "Browse" button). Under the "Output" section, there are three radio buttons: "Get back SMILES", "Get back InChI", and "Get back SDF". A "Save this job in XML format (e.g. for PUG)" option is also present. The footer contains links for "Write to Helpdesk", "Disclaimer", "Privacy statement", and "Accessibility".

# Automated structure processing...



- **Verification**

- Atom element
- Implicit hydrogen
- Functional group
- Valence

- **Standardization**

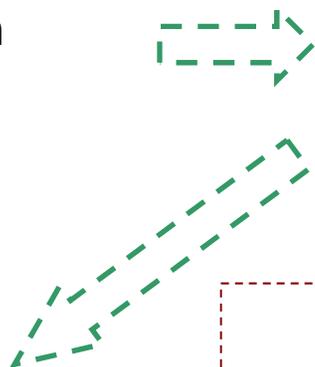
- Tautomer invariance
- Aromaticity detection
- Stereochemistry
- Explicit hydrogen

- **Calculation**

- Coordinates
- Properties
- Descriptors

- **Components**

- Isolate covalent units
- Neutralize (+/- proton)
- **Reprocess**
- Detect unique



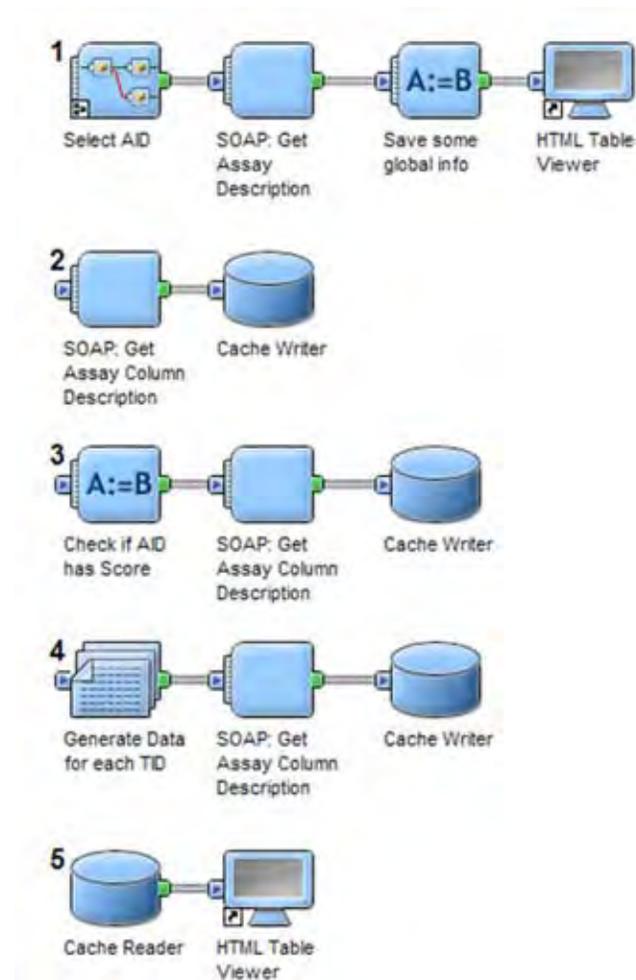
# Standardization Service

- Performs PubChem chemical structure “standardization”
  - Provides CID if structure is in PubChem
- Allows users to examine PubChem methodology affects on their data

The screenshot shows the PubChem Standardization Service web interface in a Windows Internet Explorer browser. The page title is "PubChem Standardization Service" and the URL is "http://pubchem.ncbi.nlm.nih.gov/standardize/standardize.cgi". The interface includes a "Submit Job" button and a "Save Job" button. Under the "Input" section, there are three radio buttons: "Start with SMILES", "Start with InChI", and "Start with an SDF file" (with a "Browse" button). Under the "Output" section, there are three radio buttons: "Get back SMILES", "Get back InChI", and "Get back SDF". A "Save Job" button is also present, with the text "Save this job in XML format (e.g. for PUG)" next to it. The footer contains links for "Write to Helpdesk", "Disclaimer", "Privacy statement", and "Accessibility".

# Power User Gateway (PUG)

- Programmatic interface to many PubChem services
- Allows scripted access to PubChem
- Enables one to save a query/view
- SOAP interface
  - Accessible by Pipeline Pilot, Taverna, Java, PERL, Python, VB.net, C#.net, etc.



# InChI Compound-based Lookup

The screenshot displays the PubChem website interface. At the top left is the NCBI logo, and next to it is the PubChem Compound logo. On the top right, there is a 'My NCBI' section with links for '[Sign In]' and '[Register]'. The search bar contains the text 'Search PubChem Compound for "InChI=1S/C9H8O4/c1-6(10)13-8-5-3-2-4-7(8)9(11)12" Go Clear Save Search'. Below the search bar are buttons for 'Limits', 'Preview/Index', 'History', 'Clipboard', and 'Details'. The display settings show 'Display Summary', 'Show 20', and 'Sort By'. The 'Tools' section includes icons for a search, a hand, a printer, and a help icon, followed by the text 'Links: Related Structures, BioAssays, Literature, Other Links'. Below this are buttons for 'All: 1', 'BioAssay: 1', 'Protein3D: 1', and 'Rule of 5: 1'. The main content area shows a search result for '1: CID: 2244' with a chemical structure image of aspirin and the following text: 'aspirin; Acylpyrin; Colfarit ... IUPAC: 2-acetyloxybenzoic acid MW: 180.157420 g/mol | MF: C<sub>9</sub>H<sub>8</sub>O<sub>4</sub> Tested in BioAssays: All: 242, Active: 2; BioActivity Analysis Cyclooxygenase Inhibitors... more'. On the right side, there is a 'Recent Activity' box with a search icon and the text 'InChI=1S/C9H8O4/c1-6(10)13-8-5-3-2-4-7(8)9(11)12 (1) PubChem Compound'. At the bottom of the page, there are links for 'PubChem Help | Entrez Help | Write to the Help Desk' and 'PubChem | Compound | Substance | BioAssay Chemical Structure Search | BioAssay Services | FTP Site | Deposit Data'.

[PubChem Help](#) | [Entrez Help](#) | [Write to the Help Desk](#)

[PubChem](#) | [Compound](#) | [Substance](#) | [BioAssay](#)  
[Chemical Structure Search](#) | [BioAssay Services](#) | [FTP Site](#) | [Deposit Data](#)

# InChIKey Compound-based Lookup



NCBI **PubChem** Compound

My NCBI [Sign In] [Register]

Search PubChem Compound for "BSYNRYMUTXBXSQ-UHFFFAOYSA-N" Go Clear Save Search

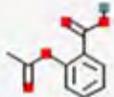
Limits Preview/Index History Clipboard Details

Display Summary Show 20 Sort By Send to

Tools Links: Related Structures, BioAssays, Literature, Other Links

All: 1 BioAssay: 1 Protein3D: 1 Rule of 5: 1

1: CID: 2244 Related Structures, BioAssays, Literature, Other Links

 aspirin, Acylpyrin, Colfarit ...  
IUPAC: 2-acetyloxybenzoic acid  
MW: 180.157420 g/mol | MF: C<sub>9</sub>H<sub>8</sub>O<sub>4</sub>  
Tested in BioAssays: All: 242, Active: 2; BioActivity Analysis  
Cyclooxygenase Inhibitors... more

Recent Activity

- Turn Off Clear
- Q "BSYNRYMUTXBXSQ-UHFFFAOYSA..." (1)
- Q "InChI=1S/C9H8O4/c1-6(10)..." (1) PubChem Compound

[PubChem Help](#) | [Entrez Help](#) | [Write to the Help Desk](#)

[PubChem](#) | [Compound](#) | [Substance](#) | [BioAssay](#)  
[Chemical Structure Search](#) | [BioAssay Services](#) | [FTP Site](#) | [Deposit Data](#)

[NCBI](#) | [NLM](#) | [NIH](#) | [HHS](#) | [Privacy Statement](#) | [Freedom of Information Act](#) | [Disclaimer](#)

# Integrated 3-D Layer



Aspirin - Compound Summary (CID 2244)

The prototypical analgesic used in the treatment of mild to moderate pain. It has anti-inflammatory and antipyretic properties and acts as an inhibitor of cyclooxygenase which results in the inhibition of the biosynthesis of prostaglandins. Aspirin also inhibits platelet aggregation and is used in the prevention of arterial and venous thrombosis. (From Martindale, The Extra Pharmacopoeia, 30th ed, p5)®

- Table of Contents
  - Drug and Chemical Information®
    - Pharmacological Action
    - Pharmacological Classification
    - Chemical Classification
    - Safety and Toxicology
    - Literature Links
    - Literature Mining
  - BioActivity Results
  - Synonyms
  - Properties
  - Descriptors
  - Compound Information
  - Substance Information
  - Category
  - Exports

Structure & Quick Link Bar

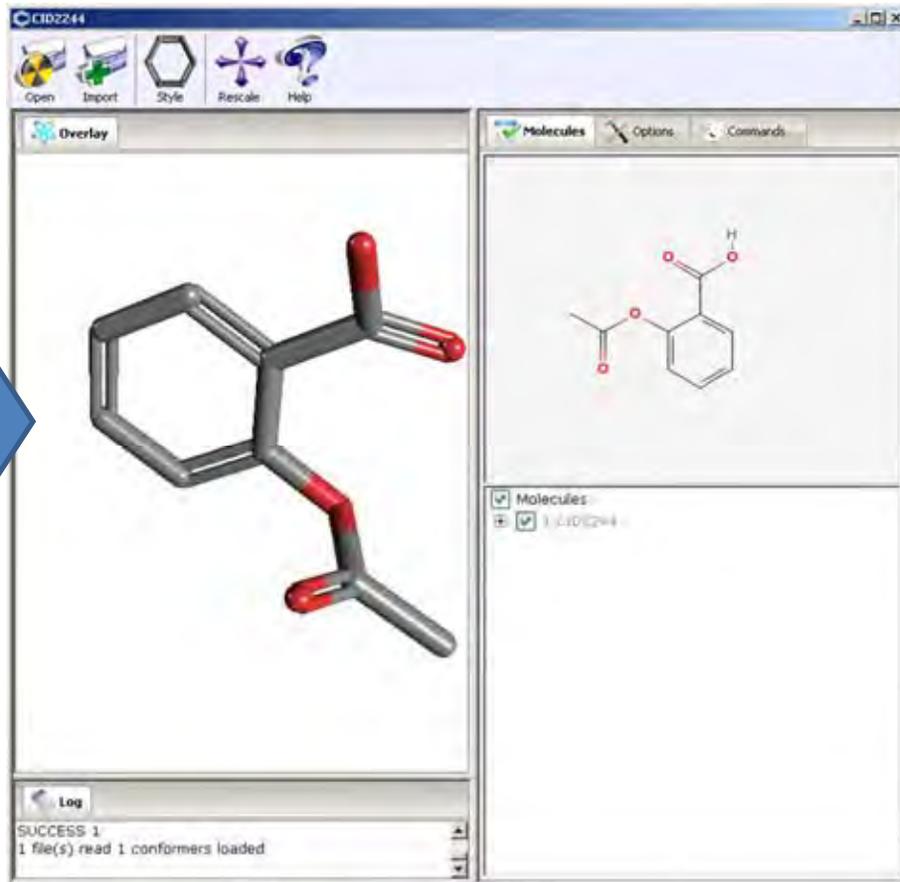
2D 3D

Pc3D Viewer

Compound ID: 2244  
Molecular Weight: 180.15742 [g/mol]  
Molecular Formula: C<sub>9</sub>H<sub>8</sub>O<sub>4</sub>  
XLogP3: 1.2  
H-Bond Donor: 1  
H-Bond Acceptor: 4

Links

Protein Structure: (5)



CID2244

Open Import Style Rescale Help

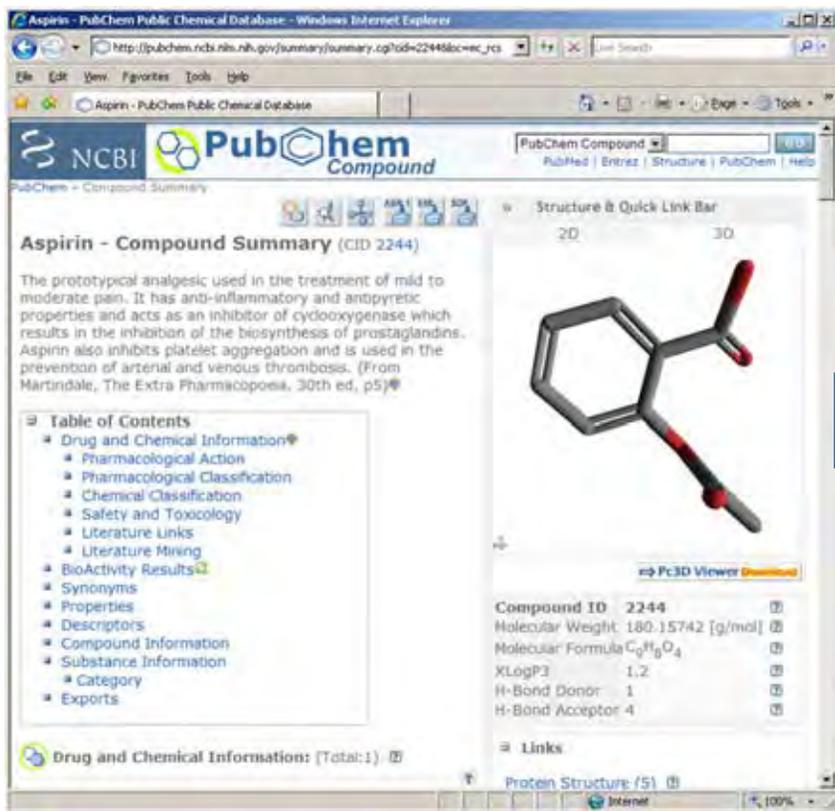
Overlay Molecules Options Commands

Molecules: 1: CID2244

Log

SUCCESS: 1  
1 file(s) read 1 conformers loaded

# Integrated 3-D Layer



Aspirin - Compound Summary (CID 2244)

The prototypical analgesic used in the treatment of mild to moderate pain. It has anti-inflammatory and antipyretic properties and acts as an inhibitor of cyclooxygenase which results in the inhibition of the biosynthesis of prostaglandins. Aspirin also inhibits platelet aggregation and is used in the prevention of arterial and venous thrombosis. (From Martindale, The Extra Pharmacopoeia, 30th ed, p5)®

- Table of Contents
  - Drug and Chemical Information®
    - Pharmacological Action
    - Pharmacological Classification
    - Chemical Classification
    - Safety and Toxicology
    - Literature Links
    - Literature Mining
  - BioActivity Results
  - Synonyms
  - Properties
  - Descriptors
  - Compound Information
  - Substance Information
  - Category
  - Exports

Structure & Quick Link Bar

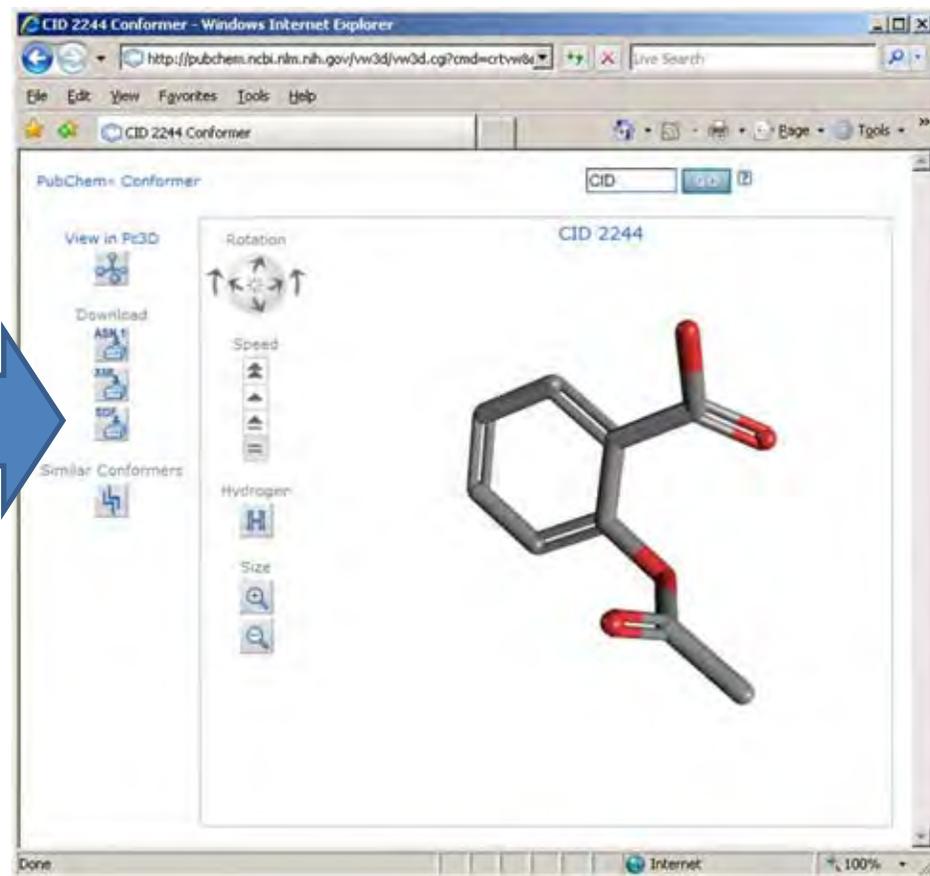
2D 3D

CC(=O)OC1=CC=CC=C1

Compound ID: 2244  
Molecular Weight: 180.15742 [g/mol]  
Molecular Formula: C<sub>9</sub>H<sub>8</sub>O<sub>4</sub>  
XLogP3: 1.2  
H-Bond Donor: 1  
H-Bond Acceptor: 4

Links

Protein Structure (5)



CID 2244 Conformer

View in Pc3D

Download

- ASN
- X3D
- EDL

Similar Conformers

Rotation

Speed

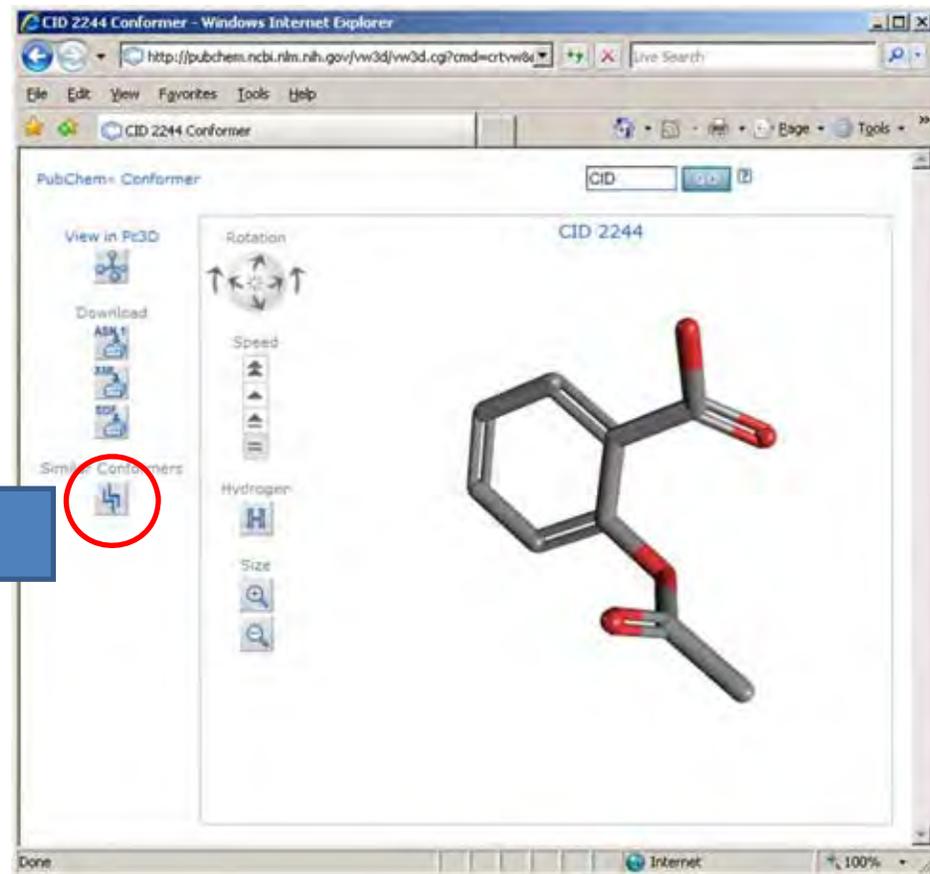
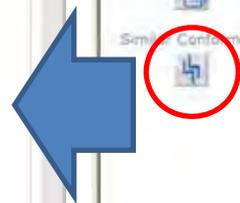
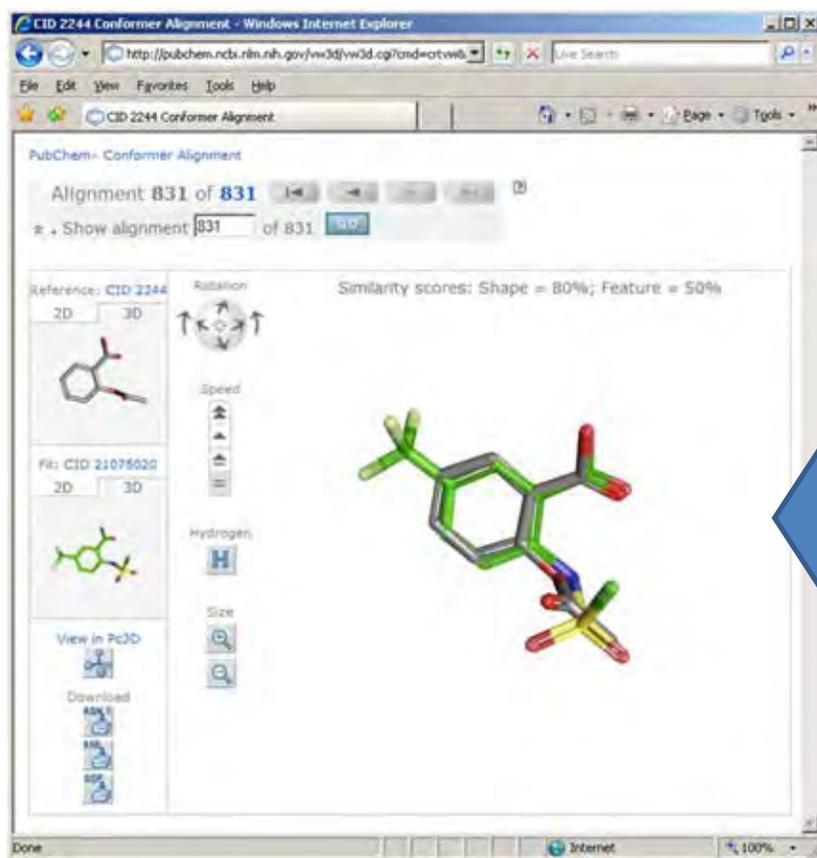
Hydrogen

Size

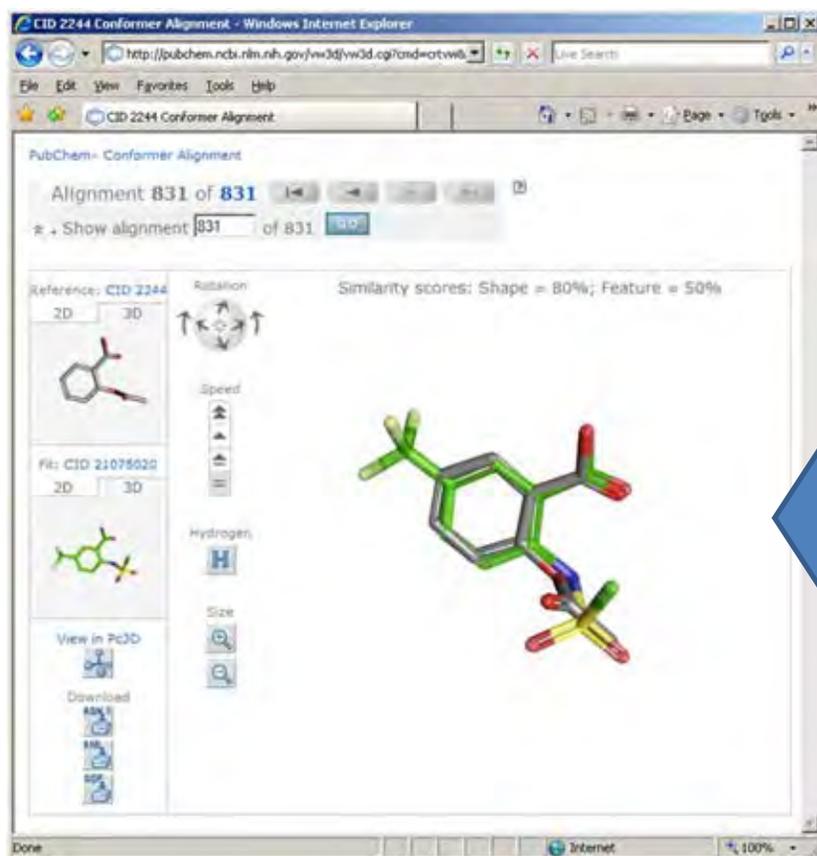
CID 2244

CC(=O)OC1=CC=CC=C1

# Integrated 3-D Layer



# Integrated 3-D Layer



Aspen - PubChem Public Chemical Database - Windows Internet Explorer

http://pubchem.ncbi.nlm.nih.gov/summary/summary.cgi?cid=2244&loc=ec\_rct&P...

Compound Information:

CID 2244

Create Date: 2004-09-16

Related Compounds: Same Connectivity: 2 Links

Similar Compounds: 1416 Links

Similar Conformers: 828 Links **View Conformers**

Substance Information:

Substances: All: 711 Links Same structure: 116 Links Mixture: 595 Links

Category: [for same structure substances]

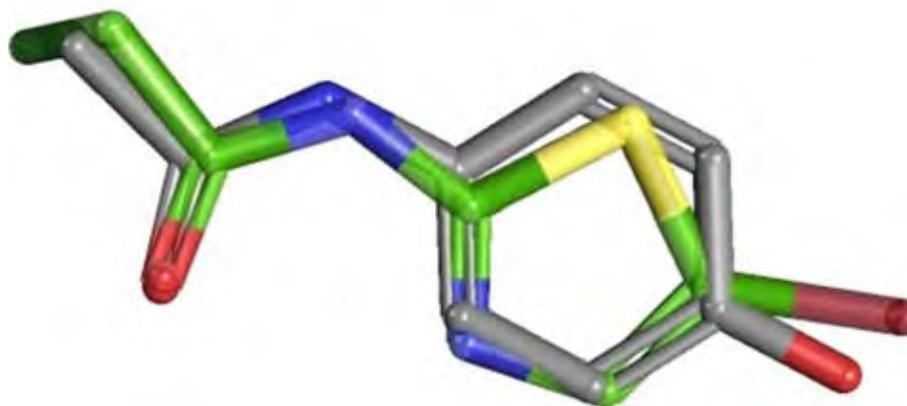
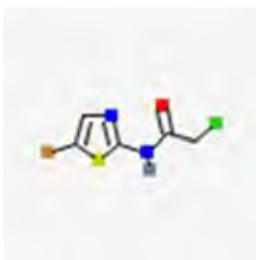
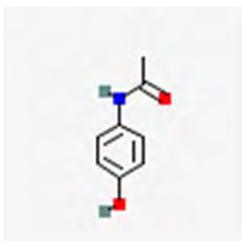
Biological Properties: 39 Links

- BindingDB ( 1 )  
SID 49846175 - External ID: 22360
- ChEBI ( 1 )  
SID 8143164 - External ID: ChEBI:15365
- ChemBank ( 25 )
- Comparative Toxicogenomics Database ( 1 )  
SID 53788943 - External ID: D001241
- DiscoveryGate ( 1 )  
SID 8151516 - External ID: 2244
- DrugBank ( 1 )  
SID 46505803 - External ID: D800945
- DTP/NCI ( 2 )  
SID 87798 - External ID: 27223  
SID 476106 - External ID: 406106
- LeadScope ( 1 )  
SID 49854366 - External ID: LS-143

http://pubchem.ncbi.nlm.nih.gov/ww3d/vw3d.cgi?and=otw6...

# What is a “Similar Conformer”?

- When two conformers have ...
  - ... similar shapes ( $ST \geq 0.80$ )
  - ... similar features ( $CT \geq 0.50$ )
  - ... BUT only shape optimized



Similarity scores: Shape = 92%; Feature = 54%

# PubChem Publications...

Databases | Deposition | Services | Help | **more**

**Links**

- ▶ About PubChem
- ▶ Announcements
- ▶ **Publications**
- ▶ PubChem 3D release notes
- ▶ Add PubChem searches to your site
- ▶ Pc3D viewer application
- ▶ Structure group home
- ▶ 3D macromolecular structures
- ▶ Conserved domains
- ▶ BioSystems

BioAssay ? ?

Advanced search

Chemical structure search | BioActivity analysis

**New** PubChem in 2011 - Growth Continues ..

**New** Structures from **AmicBase** are now available in PubChem.

more ...

Bioactivity summary

Bioactivity datatable

Bioactivity structure-activity

Chemical structure search

3D conformer viewer

Chemical structure clustering

Deposition gateway

Structure download

Bioassay download

PubChem FTP

{Click}

# PubChem Publications...

PubChem Publications - Windows Internet Explorer  
http://pubchem.ncbi.nlm.nih.gov/publications.html

NCBI Pubchem

Search PubChem Substance for [ ]

### PubChem Publications

[PUBCHEM HOME](#) [SUBSTANCE](#) [COMPOUND](#) [BIOASSAY](#) [HELP](#) [NEWS](#) [FTP](#) [PUBLICATIONS](#) [CITING](#)

#### Citing the Resources

[DATA USAGE & CITATION GUIDELINES >>](#)

Please see the [PubChem Data Usage and Citation Guidelines](#) page for information about how to cite individual or multiple records from a PubChem database. If you are referencing the overall PubChem Substance, Compound, or BioAssay database, the following articles can be cited:

**PubChem Substance and PubChem Compound**

- Bolton E, Wang Y, Thiessen PA, Bryant SH. PubChem: Integrated Platform of Small Molecules and Biological Activities, Chapter 12 IN *Annual Reports in Computational Chemistry*, Volume 4, American Chemical Society, Washington, DC, 2008 Apr. [FREE Full text available in PubChem FTP](#)

**PubChem BioAssay**

- Wang Y, Bolton E, Dracheva S, Karapetyan K, Shepemaker BA, Suzek TO, Wang J, Xiao J, Zhang J, Bryant SH. An overview of the PubChem BioAssay resource. *Nucleic Acids Res.* 2010 Jan;38(Database issue):D255-66. Epub 2009 Nov 19. doi:10.1093/nar/gkp965. [PubMed PMID: 19933261] [FREE Full text available in PubMed Central](#)

#### Additional Journal Articles About PubChem

[VIEW ALL PUBCHEM ARTICLES IN FORMID >>](#)

- Zhou Z, Wang Y, Bryant SH. Multi-conformation 3D QSAR study of benzenesulfonyl-pyrazol-ester compounds and their analogs as cathepsin B inhibitors. *J Mol Graph Model.* 2011 Jul 7; Epub ahead of print. doi:10.1016/j.jmgm.2011.06.013. [PubMed PMID: 21798778]
- Kim S, Bolton EE, Bryant SH. PubChem3D: Biologically relevant 3-D similarity. *J Cheminform.* 2011 Jul 22;3(1):26. Epub ahead of print. doi:10.1186/1758-2946-3-26. [PubMed PMID: 21781288] [Free Full Text on J Cheminform](#)
- Kim S, Bolton EE, Bryant SH. PubChem3D: Shape compatibility filtering using molecular shape quadrupoles. *J Cheminform.* 2011 Jul 20;3(1):25. Epub ahead of print. doi:10.1186/1758-2946-3-25. [PubMed PMID: 21774909] [Free Full Text on J Cheminform](#)
- Bolton EE, Kim S, Bryant SH. PubChem3D: Similar conformers. *J Cheminform.* 2011 May 9;3(1):13. doi:10.1186/1758-2946-3-13. [Epub ahead of print] [PubMed PMID: 21554721] [Free Full Text on J Cheminform](#)

http://www.acscomp.org/Publications/ARCC/volume4/chapter12.html Local intranet [Protected Mode: Off] 100%

# PubChem3D Thematic Series

Journal of Cheminformatics

Welcome Evan Bolton (Log off)  
Chemistry Central home | Feedback | Support | My details

Home | Browse articles | Supplement | **Series** | Search | Submit article | My Journal of Cheminformatics | About Journal of Cheminformatics

articles

research article | book report | commentary | database | editorial | introduction | meeting report | methodology | oral presentation | poster presentation | preliminary communication | review | software | most accessed articles

Thematic Series [All thematic series](#)

PubChem3D

PubChem is a public and open archive for the biological activities of small molecules. PubChem provides search and analysis capabilities to assist users to locate desired information. This is especially important given that the PubChem contents are both vast (with tens of millions of unique chemical structures and over hundred million biological test results) and diffuse (with an uneven distribution of biological annotation of contained compounds). Many of the analysis tools PubChem provides utilize the notion of chemical similarity. PubChem3D adds a new integrated layer to the existing capabilities of the PubChem platform. This includes a theoretical 3-D description for PubChem Compound records and the notion of 3-D similarity, to augment 2-D similarity approaches. This thematic series covers: the technology and methodology behind the PubChem3D project, enabling it to function and scale; the validation and analysis of the PubChem3D content and approach; and the various PubChem3D components, with examples of how they can be of utility to researchers.

Research article [Open Access](#) **Highly accessed**

**PubChem3D: Biologically relevant 3-D similarity**  
Sunghwan Kim, Evan E Bolton, Stephen H Bryant  
*Journal of Cheminformatics* 2011, 3:26 (22 July 2011)  
[Abstract] [Provisional PDF] [PubMed] [Related articles]

2D	= 0.56
5 <sup>3D</sup> sim	= 0.73
C <sup>3D</sup> sim	= 0.28
Combo <sup>3D</sup> sim	1.01

Frequency (billions)

Tanimoto

$C^{3Dsim}$  (0.07 ± 0.05)  
 $5^{3Dsim}$  (0.54 ± 0.10)  
 $Combo^{3Dsim}$  (0.62 ± 0.13)

<http://www.jcheminf.com/series/PubChem3D>

# Summary

- PubChem is a chemical biology resource
  - open and public to all
  - continues to grow rapidly
  - many tools to get at the information you need
  - uses Google-like approach of likely answers first
- Fundamental problems exist in chemical information exchange
  - stereo corruption a major issue
  - chemical name filtering helps remove noise

# PubChem Crew ...

**Steve Bryant**

**Jie Chen**

**Tiejun Chen**

**Lewis Geer**

**Asta Gindulyte**

**Volker Hahnke**

**Lianyi Han**

**Jane He**

**Siqian He**

**Kenneth Karapetian**

**Sunghwan Kim**

**Qingliang Li**

**Ben Shoemaker**

**Tugba Suzek**

**Paul Thiessen**

**Jiyao Wang**

**Yanli Wang**

**Jewen Xiao**

**Bo Yu**

**Jian Zhang**

**Jun Zhang**

