

Chemical Names, Chemical Structures

What's a Translator to Do?

ATA 64th Annual Conference – Miami, FL
26 October 2023

Matthew Schlecht, PhD
Word Alchemy Translation
wordalchemytranslation.com
mattschlecht@wordalchemytranslation.com



Introduction and Goals

Acquaint Translators & Interpreters With Chemical Names & Structures

Arise in many texts/contexts:

product labels, patents, lawsuits, medical
discussions and reports, clinical trials, etc.

What are some strategies for translating them?

Introduction and Goals

“What's in a name? That which we call a rose by any other name would smell just as sweet.”

W. Shakespeare, Romeo and Juliet, ~1595

Trouble is, in describing chemical substances

We need a name and/or pictographic representation

That name or representation should be unambiguous

Otherwise, we fail to communicate

Introduction and Goals

Chemical names & structures are a language

Alphabetic/syllabic and pictographic representations of substances

Alphabetic/syllabic representations vary from language to language

Pictographic representations are almost universally understood across languages

Introduction and Goals

Topics for today

There are names, and there are names

What are chemical names?

How to parse a chemical name/structure

Shortcuts for translating names

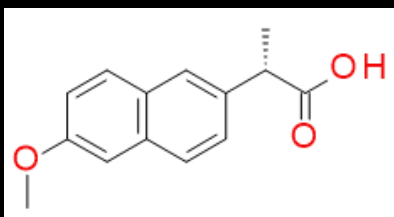
Resources for chemical names/structures

Chemical structures

pictographic representations are universally understood

Naproxen

a non-steroidal anti-inflammatory drug (NSAID)



[19]中华人民共和国专利局

[51]Int.Cl⁶

C07C 59/125



[12]发明专利说明书

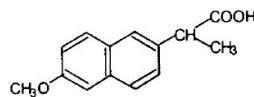
[21] ZL 专利号 92102703.6

[45]授权公告日 1997年4月23日

[11] 授权公告号 CN 1034661C

权 利 要 求 书

1、一种合成式(I)化合物(d1-萘普生)的方法:

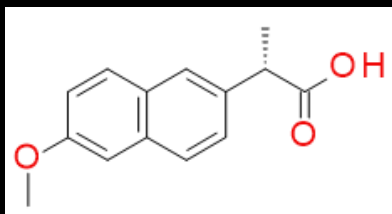


(I)

Chemical structures

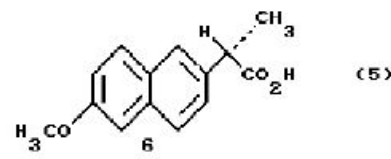
pictographic representations are universally understood

Naproxen



РОССИЙСКОЕ АГЕНТСТВО
ПО ПАТЕНТАМ И ТОВАРНЫМ ЗНАКАМ

(19) **RU** ⁽¹¹⁾ **2 086 532** ⁽¹³⁾ **C1**
(51) МПК⁶ **C 07 C 51/36, 57/30, 59/64,**
57/40



В этом предпочтительном синтезе первая стадия представляет собой типичный синтез эфира по Овильямсу, заключающемуся в реакции 2-оксинафталина (1) с метилирующим агентом, таким как диметилсульфат, с получением 2-метоксинафталина (2). С другой стороны, 2-метоксинафталин промышленно доступен и выпускается фирмой Sigma-Aldrich.

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ РОССИЙСКОЙ ФЕДЕРАЦИИ

(21), (22) Заявка: 4894655/04, 22.06.1990

(30) Приоритет: 22.06.1989 US 07/369875
02.03.1990 US 487465

(46) Дата публикации: 10.08.1997

(71) Заявитель:
Монсанто Компани (US)

(72) Изобретатель: Сун-Чи Чан Альберт[US]

(73) Патентообладатель:
Монсанто Компани (US)

Chemical structures

Often mixed in with text, as in lab notebook pages

222 June 2, 1980

Prep of α C=CC

			CR <u>261</u> 474 (1965)
HO <chem>C=CC</chem>	222.8mg	1.99	112.16
PCl_5	475.4mg	2.04	208.24
α <chem>C=CC</chem>	theoretical 259.4mg		130.61

Into a 25ml rb w/ football stirbar was weighed 475.4mg PCl_5 , this taken up in 10ml Et₂O + stirred at -10° (ice/ H°) under Ar during addn of 222.8mg α C=CC in 2X 1ml Et₂O, this stirred at -10° for 1hr, then to rt over 3hrs. Volatiles were removed in vacuo, the residue taken up in ether, ext w/ NaHCO_3 , back ext w/ Et₂O, dried & concentrated to yield 202.9mg yellow oil (78%) NMR-MS-IV-222-1, IR-MS-IV-222-1

There are names, and there are names

There are two general classes of chemical names:

Systematic names: derived from a set of rules, an algorithm to unambiguously link the name with the structure and the substance, and vice versa

Trivial names: all the rest of the names

There are names, and there are names

Rules for systematic names are set out
(and updated) by an international
organization for standard nomenclature:

IUPAC

International
Union of
Pure and
Applied
Chemistry

[<https://iupac.org/>]

There are names, and there are names

Chemical substances are most often referred to by a trivial name

Analogy: the reigning monarch of Great Britain is known as **King Charles III**

Full name: **Charles Philip Arthur George, Prince of Wales and Earl of Chester, Duke of Cornwall, Duke of Rothesay, Earl of Carrick and Baron Renfrew, Lord of the Isles, and Prince and Great Steward of Scotland**

Example: Sodium Hydroxide

Sodium hydroxide is known by many names:

Sodium Hydroxide

Lye

Caustic

Caustic Soda

Sodium Hydrate

NaOH

1310-73-2

It is essentially the same chemical substance but there are differences of register and context

Example: Sodium Hydroxide

Only the name “sodium hydroxide” is systematic because it relates directly to the structure of the formula



“1310-73-2” is the CAS (Chemical Abstracts Service) registry number, also unique.

A sort of social security number for a chemical substance

All other names are trivial (non-systematic) names but are still in widespread use

Example: Sodium Hydroxide

A true translation must reflect those differences

EN	sodium hydroxide
DE	Natriumhydroxid
ES	hidróxido de sodio
FR	hydroxyde de sodium
JA	水酸化ナトリウム <i>suisanka NATORIUMU</i>

Example: Sodium Hydroxide

A true translation must reflect those differences

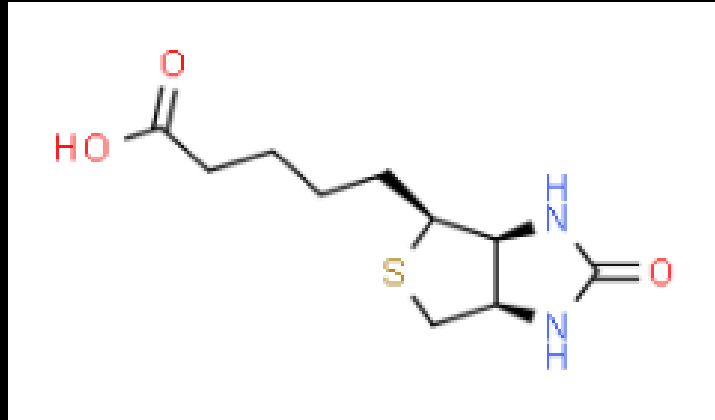
EN	lye
DE	Lauge
ES	lejía
FR	lessive
JA	あく; アク <i>aku; AKU</i>

Example: Sodium Hydroxide

A true translation must reflect those differences

EN	caustic soda
DE	Natronlauge, Ätznatron
ES	soda cáustica
FR	soude caustique
JA	苛性ソーダ <i>kasei SŌDA</i>

Example: Biotin

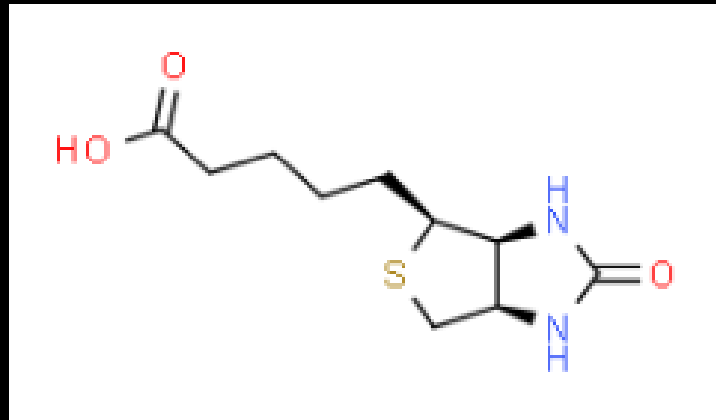


Biotin, also known as vitamin B₇ or vitamin H

It is a fairly simple, optically active, water-soluble molecule with a bicyclic core structure, a carboxylic acid side-chain, and three stereocenters
(*a stereocenter specifies 3-dimensional structure*)

Example: Biotin

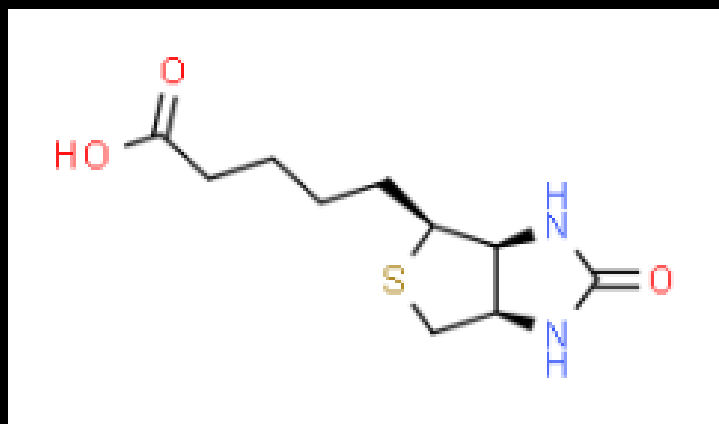
Trivial name



biotin [EN,DE]	БИОТИН (<i>biotin</i>) [RU]
biotina [ES]	비오틴 (<i>biotin</i>) [KR]
biotine [FR]	ビオチン (<i>biochin</i>) [JA]
biotiini [FI]	生物素 (<i>shēngwùsù</i>) [ZH]

Example: Biotin

Systematic name



5-[(3aR,6S,6aS)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl]pentanoic acid (IUPAC/EN).

5-[(3aR,6S,6aS)-2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl]pentansäure (IUPAC/DE)

ácido 5-[(3aS,4S,6aR)-2-oxohexahidro-1H-tieno[3,4-d]imidazol-4-yl]pentanoic (IUPAC/ES)

acide 5-[(3aR,6S,6aS)-2-oxohexahydro-1H-thiéno[3,4-d]imidazol-4-yl]pentanoïque (IUPAC/FR)

5-[(3aR,6S,6aS)-2-オキソヘキサヒドロ-1H-チエノ[3,4-d]イミダゾール-4-イル]ペンタン酸
(IUPAC/JA)

What are chemical names?

In fact, chemical names are a language
with many dialects

Chemical names are an agglutinative language

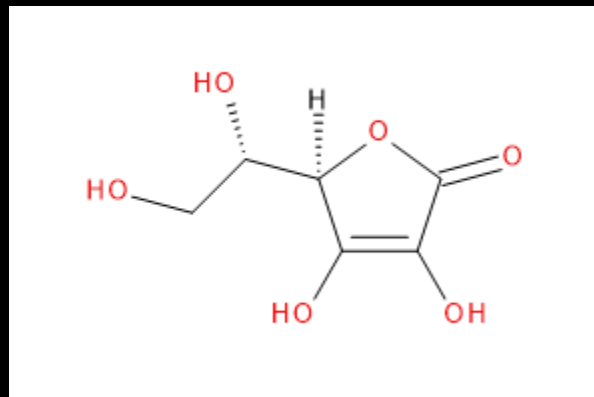
Chemical structures are akin to the pictograms of

Chinese characters 漢字/汉字 (*hànzì*), or

Egyptian hieroglyphics (*r n km.t*)

Example: Ascorbic Acid

Trivial names



ascorbic acid	[EN]	vitamin C
Ascorbinsäure	[DE]	Vitamin C
ácido ascórbico	[ES]	vitamina C
acide ascorbique	[FR]	vitamine C
アスコルビン酸 [ASUKORUBIN-san]	[JA]	ビタミンC [bitamin-c]

How to parse a chemical name

An organic chemical name has a core
or parent structure

Core/parent structure appears at or
near the right-hand of the name string

How to parse a chemical name

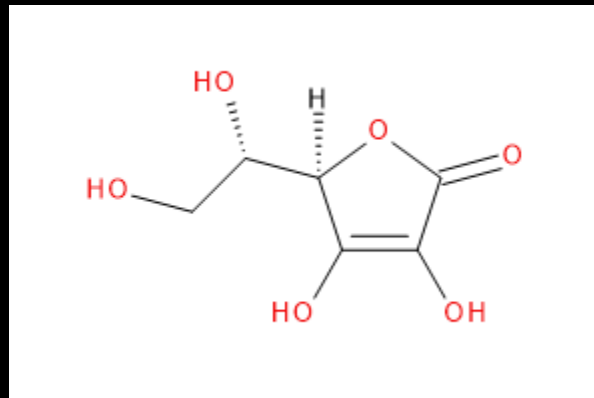
To this are added
prefixes, infixes, and/or suffixes

these indicate substituents (parts
added to the core/parent structure)

and numbers to give their assigned
positions

Example: Ascorbic Acid

Systematic name



(5R)-[(1S)-1,2-dihydroxyethyl]-3,4-dihydroxyfuran-2(5H)-one (IUPAC/EN).

(5R)-5-[(1S)-1,2-Dihydroxyethyl]-3,4-dihydroxy-5-hydrofuran-2-on (IUPAC/DE)

(5R)-3,4-dihidroxi-5-((1S)-1,2-dihidroxietyl)furano-2(5H)-ona (IUPAC/ES)

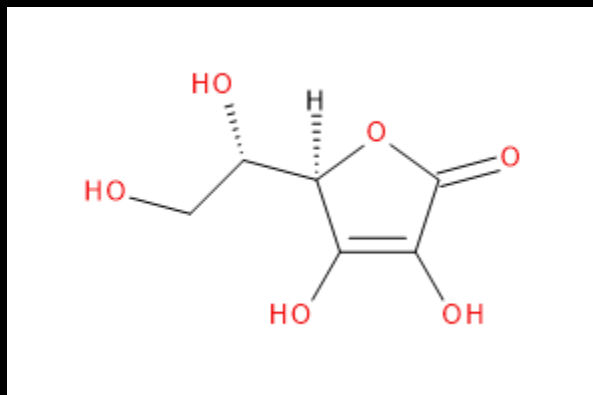
(5R)-5-((1S)-1,2-dihydroxyéthyl))-3,4-dihydroxy-5-hydrofuran-2-one (IUPAC/FR)

(5R)-3,4-ジヒドロキシ-5-((1S)-1,2-ジヒドロキシエチル)フラン-2(5H)-オン (IUPAC/JA)

(5R)-3,4-JIHIDOROKISI-5-((S)-1,2-JIHIDOROKISHIECHIRU)FURAN-2(5H)-ON

Example: Ascorbic Acid

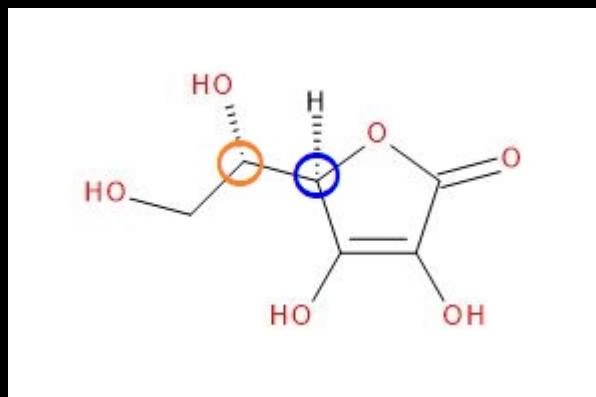
Breaking down the systematic name



(5R)-[(1S)-1,2-Dihydroxyethyl]-3,4-dihydroxyfuran-2(5H)-one

Example: Ascorbic Acid

Breaking down the systematic name

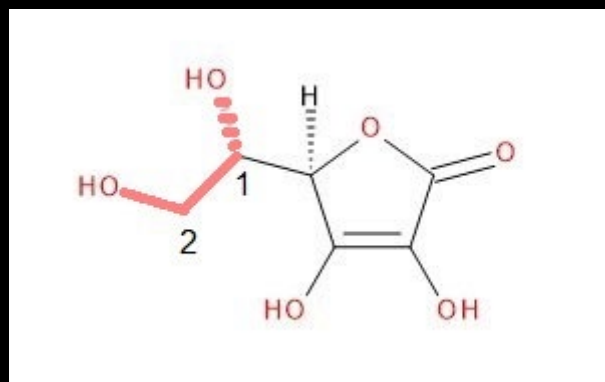


(5R)-[(1S)-1,2-dihydroxyethyl]-3,4-dihydroxyfuran-2(5H)-one

Indicators of the stereochemistry (3-dimensional aspects)
at core structure C-5 and substituent C-1.

Example: Ascorbic Acid

Breaking down the systematic name

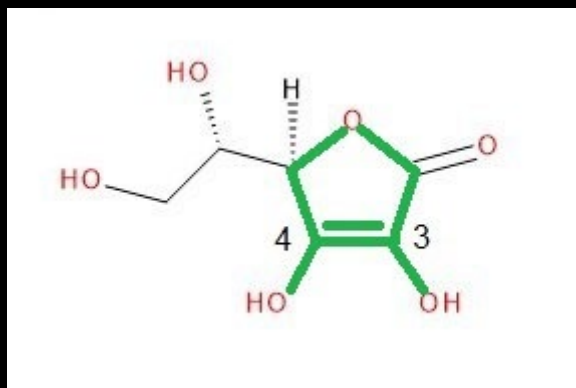


(5R)-[(1S)-1,2-Dihydroxyethyl]-3,4-dihydroxyfuran-2(5H)-one

Describes the 1,2-dihydroxyethyl substituent attached to C-5 of the core structure.

Example: Ascorbic Acid

Breaking down the systematic name

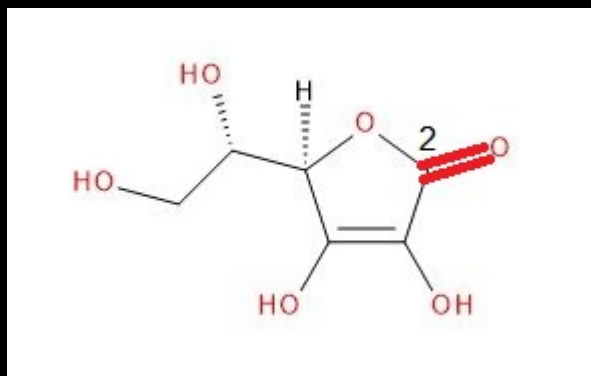


(5R)-[(1S)-1,2-Dihydroxyethyl]-3,4-dihydroxyfuran-2(5H)-one

Describes the **dihydroxyfuran** core structure, substituted with hydroxyl groups at C-3 and C-4.

Example: Ascorbic Acid

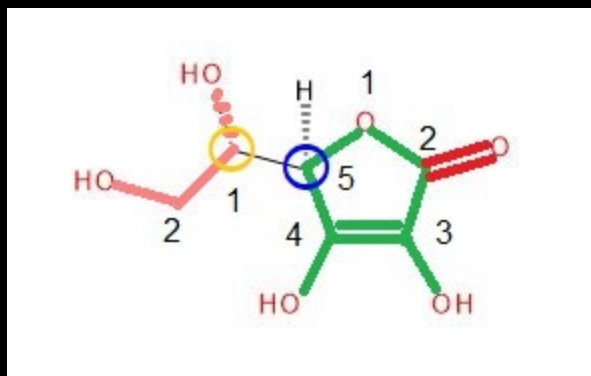
Breaking down the systematic name



(5R)-[(1S)-1,2-Dihydroxyethyl]-3,4-dihydroxyfuran-2(5H)-**one**

Describes the location of the **one** functional group in the core structure, at C-2.

with the ascorbic acid IUPAC name



(5R)-[(1S)-1,2-Dihydroxyethyl]-3,4-dihydroxyfuran-2(5H)-one

How to parse a chemical name

An inorganic chemical name will generally consist of two parts

In the case of a salt
a cation and an anion

How to parse a chemical name

In the earlier example of
sodium hydroxide:



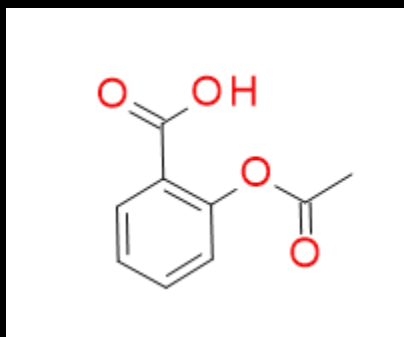
How to parse a chemical name

Most inorganic chemical names will follow this pattern of cation plus anion

Some simple organic chemical names also follow a two-part format

How to parse a chemical name

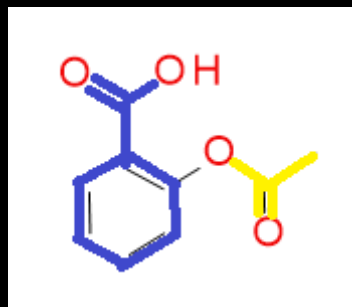
acetylsalicylic acid (aspirin, ASA)



IUPAC: 2-acetoxybenzoic acid

How to parse a chemical name

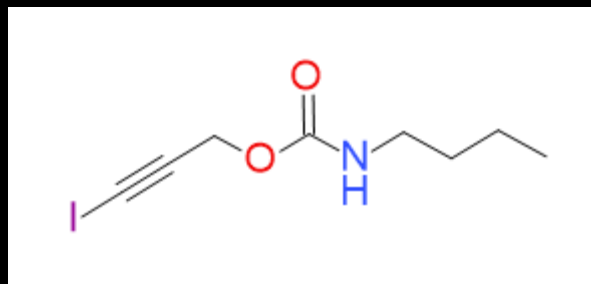
acetylsalicylic acid (aspirin, ASA)



IUPAC: 2-**acetyloxy**benzoic acid

How to parse a chemical name

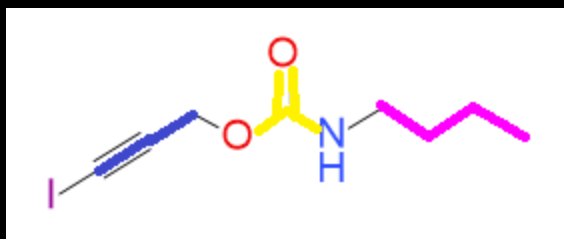
iodopropynyl butylcarbamate
(preservative used in cosmetics)



IUPAC: 3-iodoprop-2-ynyl N-butylcarbamate

How to parse a chemical name

iodopropynyl butylcarbamate
(preservative used in cosmetics)



IUPAC: 3-iodoprop-2-ynyl N-butylcarbamate

Shortcuts for translating names

Many chemical names are not found
in print or online glossaries

There are just too many of them, so use...

Online databases

Commercial chemicals sites

Artificial intelligence

Shortcuts for translating names

Online database: PubChem

Hosted by NIH, like PubMed

[<https://pubchem.ncbi.nlm.nih.gov/>]

The screenshot shows the PubChem website homepage. At the top left is the NIH logo and the text "National Library of Medicine National Center for Biotechnology Information". Below this is the PubChem logo and navigation links for "About", "Docs", "Submit", and "Contact". The main heading is "Explore Chemistry" with the subtitle "Quickly find chemical information from authoritative sources". A search bar is present with a magnifying glass icon. Below the search bar, there is a row of search suggestions: "Try covid-19 aspirin EGFR C9H8O4 57-27-2 C1=CC=C(C=C1)C=O InChI=1S/C3H6O/c1-3(2)4/h1-2H3". Below the suggestions are four radio buttons: "Use Entrez" (unchecked), "Compounds" (checked), "Substances" (unchecked), and "BioAssays" (unchecked). At the bottom, there are four icons with labels: "Draw Structure" (a pencil and a chemical structure), "Upload ID List" (an upward arrow), "Browse Data" (a grid icon), and "Periodic Table" (a grid of dots representing a periodic table).

Shortcuts for translating names

Online database: PubChem

PubChem: Draw Structure

The screenshot shows the 'DRAW STRUCTURE' interface in PubChem. The window title is 'DRAW STRUCTURE'. The interface includes a toolbar with various drawing tools, a periodic table with dropdown menus for element selection, and options for exporting to MDL Molfile, keeping AsIs, and importing a file. A blue button labeled 'Search for This Structure' is visible at the bottom left.

Broadband		SMILES													
New	Undo	Clear	Style	Delete	Copy	Move	Refresh	Undo	Redo	Undo	Redo	Undo	Redo		
Line	Double	Triple	Arrow	Vertical	Horizontal	Diagonal	Wavy	S/A	D/A	S/D					
Triangle	Square	Pentagon	Hexagon	Heptagon	Octagon	Circle	Grid	+	-	0					
Wedge	Double	Triple	Wavy	Y	Z	+	CHO	CO ₂ H	NO ₂	SO ₃ H					
H		?	? v										He		
Li	Be				B	C	N	O	F	Ne					
Na	Mg				Al	Si	P	S	Cl	Ar					
K	Ca	Sc	Sc v		Ga	Ge	As	Se	Br	Kr					
Rb	Sr	Y	Y v		In	Sn	Sb	Te	I	Xe					
Cs	Ba	Lu	Lu v		Tl	Pb	Bi	Po	At	Rn					
Export	MDL Molfile v									Done					
Hydrogen	Keep AsIs v									Help					
Import	Choose File No file chosen														

Search for This Structure

Shortcuts for translating names

Online database: PubChem

PubChem: Draw Structure (2-acetyloxybenzoic acid)

DRAW STRUCTURE ×

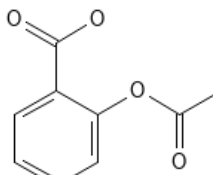
Broadband SMILES C1(=CC=CC=C1OC(C)=O)C(=O)O

New	Undo	Clear	Style	Delete	Query	↕	↻	⊕	⊖	⊙
—	=	≡	◀	⋯	⋯	⊗	⊕	S/A	D/A	S/D
△	□	⬡	⬢	⬣	⬤	⬥	⬦	⊕	⊖	⊙
∧	∨	∩	∪	∩	∪	+	CHO	CO ₂ H	NO ₂	SO ₃ H
H		?	? ▾							He
Li	Be				B	C	N	O	F	Ne
Na	Mg				Al	Si	P	S	Cl	Ar
K	Ca	Sc	Sc ▾		Ga	Ge	As	Se	Br	Kr
Rb	Sr	Y	Y ▾		In	Sn	Sb	Te	I	Xe
Cs	Ba	Lu	Lu ▾		Tl	Pb	Bi	Po	At	Rn

Export: MDL Molfile Done

Hydrogen: Keep AsIs Help

Import: Choose File No file chosen



The image shows the chemical structure of 2-acetyloxybenzoic acid, which consists of a benzene ring with a carboxylic acid group (-COOH) at the 1-position and an acetoxy group (-O-C(=O)-CH₃) at the 2-position.

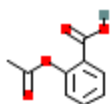
[Search for This Structure](#)

Shortcuts for translating names

Online database: PubChem

PubChem: Search for this structure

1 result



aspirin; ACETYLSALICYLIC ACID; 50-78-2; 2-Acetoxybenzoic acid; 2-(Acetyloxy)benzoic acid; ...

Compound CID: 2244

MF: $C_9H_8O_4$ MW: 180.16g/mol

IUPAC Name: 2-acetyloxybenzoic acid

Isomeric SMILES: CC(=O)OC1=CC=CC=C1C(=O)O

InChIKey: BSYNRYMUTXBXSQ-UHFFFAOYSA-N

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

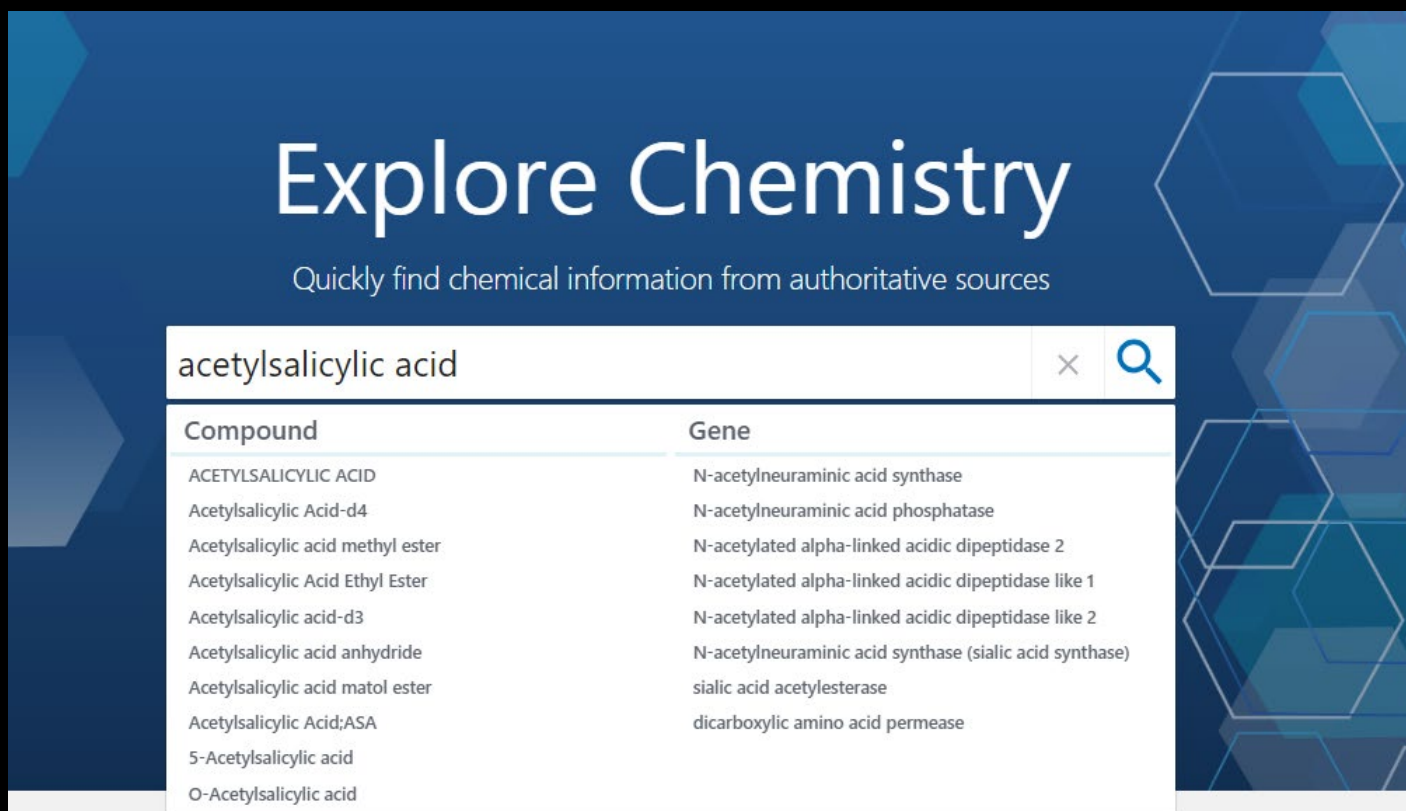
Create Date: 2004-09-16

Plus links to the **Nikkaji Number** (for Japanese), **Depositor-Supplied Synonyms** (tradenames and other trivial names) and all the chemical and pharmacological data you could ever want

Shortcuts for translating names

Online database: PubChem

Same results from searching the name



Explore Chemistry

Quickly find chemical information from authoritative sources

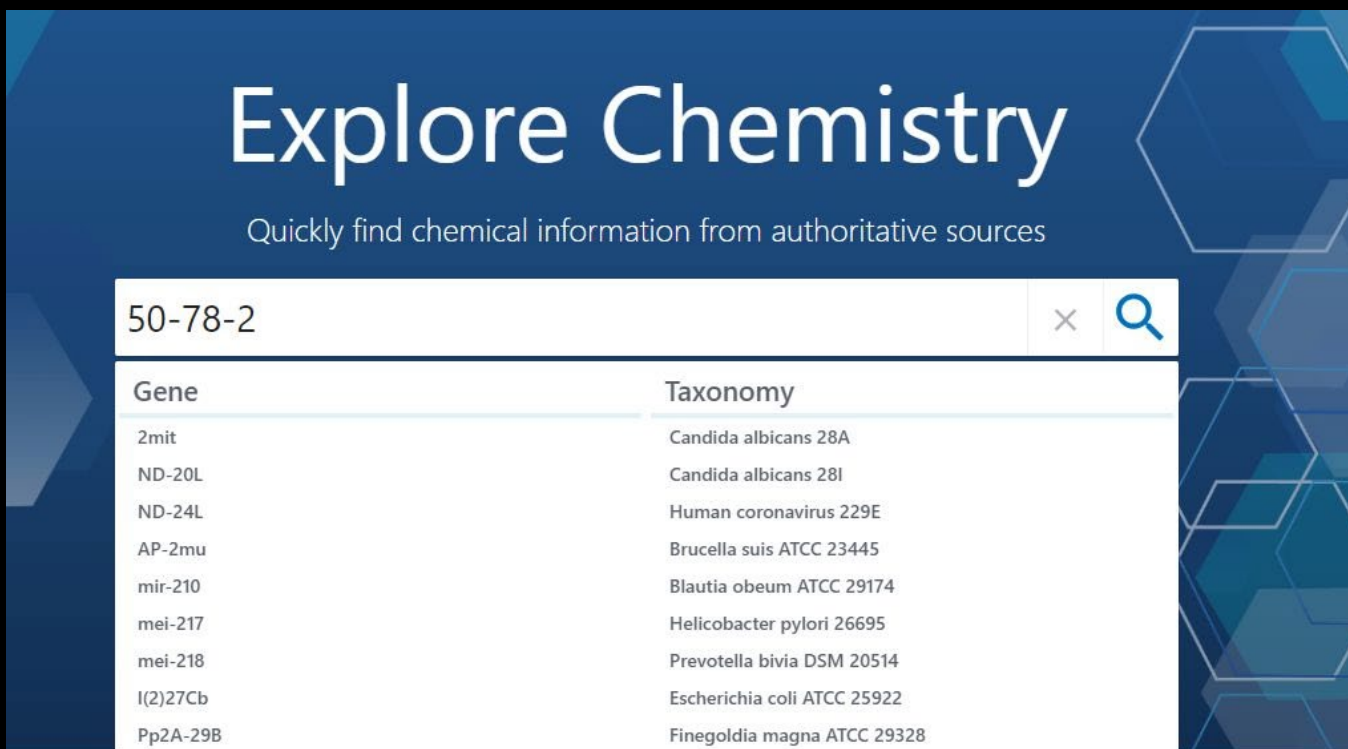
acetylsalicylic acid

Compound	Gene
ACETYLSALICYLIC ACID	N-acetylneuraminic acid synthase
Acetylsalicylic Acid-d4	N-acetylneuraminic acid phosphatase
Acetylsalicylic acid methyl ester	N-acetylated alpha-linked acidic dipeptidase 2
Acetylsalicylic Acid Ethyl Ester	N-acetylated alpha-linked acidic dipeptidase like 1
Acetylsalicylic acid-d3	N-acetylated alpha-linked acidic dipeptidase like 2
Acetylsalicylic acid anhydride	N-acetylneuraminic acid synthase (sialic acid synthase)
Acetylsalicylic acid matol ester	sialic acid acylesterase
Acetylsalicylic Acid;ASA	dicarboxylic amino acid permease
5-Acetylsalicylic acid	
O-Acetylsalicylic acid	

Shortcuts for translating names

Online database: PubChem

or searching the CAS registry number



Explore Chemistry

Quickly find chemical information from authoritative sources

50-78-2

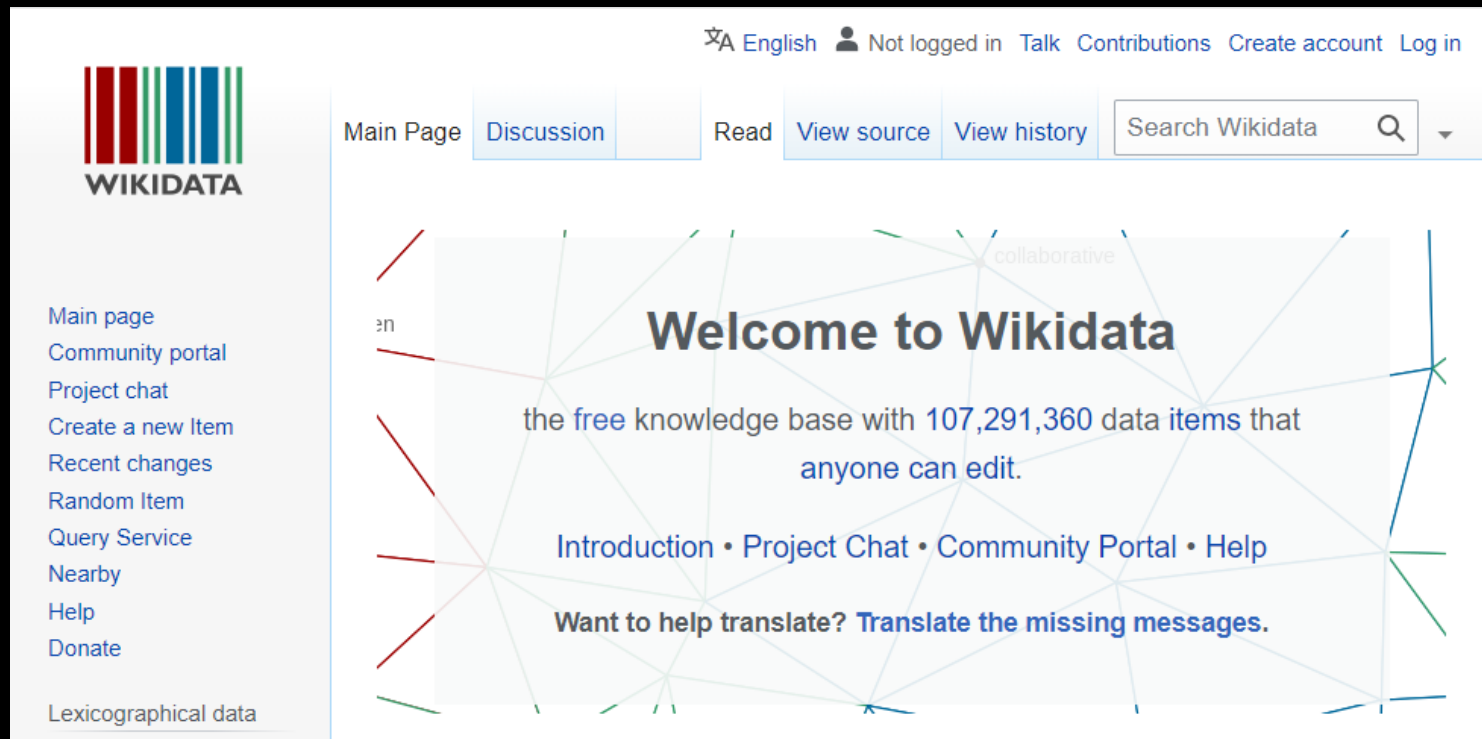
Gene	Taxonomy
2mit	Candida albicans 28A
ND-20L	Candida albicans 28I
ND-24L	Human coronavirus 229E
AP-2mu	Brucella suis ATCC 23445
mir-210	Blautia obeum ATCC 29174
mei-217	Helicobacter pylori 26695
mei-218	Prevotella bivia DSM 20514
I(2)27Cb	Escherichia coli ATCC 25922
Pp2A-29B	Finegoldia magna ATCC 29328

Shortcuts for translating names

Online database: WikiData

Hosted by Wikipedia

[https://www.wikidata.org/wiki/Wikidata:Main_Page]



The screenshot shows the Wikidata Main Page interface. At the top right, there are links for language selection (English), user status (Not logged in), and actions (Talk, Contributions, Create account, Log in). Below this is a navigation bar with buttons for 'Main Page', 'Discussion', 'Read', 'View source', and 'View history', along with a search box labeled 'Search Wikidata'. The main content area features a large graphic with a network of nodes and lines, overlaid with the text: 'collaborative', 'Welcome to Wikidata', 'the free knowledge base with 107,291,360 data items that anyone can edit.', and 'Introduction • Project Chat • Community Portal • Help'. At the bottom of the graphic, it says 'Want to help translate? Translate the missing messages.' On the left side, there is a sidebar with the Wikidata logo and a list of links: Main page, Community portal, Project chat, Create a new Item, Recent changes, Random Item, Query Service, Nearby, Help, Donate, and Lexicographical data.

Shortcuts for translating names

Online database: WikiData

WikiData: Search on “acetylsalicylic acid”

aspirin (Q18216)

medication used to treat pain and decrease the risk of heart disease [edit](#)

acetylsalicylic acid | ASA | o-acetoxybenzoic acid | o-carboxyphenyl acetate | salicylic acid acetate | o-(acetyloxy)benzoic acid
| 2-carboxyphenyl acetate | 2-acetoxybenzoic acid | o-acetylsalicylic acid | 2-acetoxybenzenecarboxylic acid | acetylsalicylic acid

[In more languages](#)
[Configure](#)

Language	Label	Description	Also known as
English	aspirin	medication used to treat pain and decrease the risk of heart disease	acetylsalicylic acid ASA o-acetoxybenzoic acid o-carboxyphenyl acetate salicylic acid acetate o-(acetyloxy)benzoic acid 2-carboxyphenyl acetate 2-acetoxybenzoic acid o-acetylsalicylic acid 2-acetoxybenzenecarboxylic acid acetylsalicylic acid
Spanish	ácido acetilsalicílico	fármaco de la familia de los salicilatos	Acido acetilsalicílico Cardirene
Traditional Chinese	阿司匹林	No description defined	阿士匹靈
Chinese	阿司匹林	非類固醇抗發炎藥	阿士匹靈 阿司匹靈 阿斯匹靈 乙酰水楊酸 2-乙酰氧基苯甲酸 邻乙酰氧基苯甲酸 O-乙酰水楊酸 乙酰柳酸

[All entered languages](#)

a total of more than 100 languages

Shortcuts for translating names

Online database: ChemSpider

Hosted by Royal Society of Chemistry (UK)

[<https://www.chemspider.com/>]

The screenshot shows the ChemSpider website interface. At the top, there is a navigation bar with links for Home, About us, Web APIs, Help, and Sign in. The Royal Society of Chemistry logo is in the top right corner. The main header features the ChemSpider logo and the tagline "Search and share chemistry". A search bar labeled "Search ChemSpider" is positioned on the right. Below the header, a purple banner contains a message: "For medical information relating to Covid-19, please consult the [World Health Organisation](#) or local healthcare provision." Underneath, there are tabs for "Simple", "Structure", "Advanced", and "History". The main content area is titled "Search ChemSpider" and includes a search box with the placeholder text "Matches any text strings used to describe a molecule." and a search button. Below the search box, it lists search criteria: "Systematic Name, Synonym, Trade Name, Registry Number, SMILES, InChI or CSID". The page is divided into two columns. The left column is titled "What is ChemSpider?" and contains the text: "ChemSpider is a free chemical structure database providing fast text and structure search access to over 100 million structures from hundreds of data sources." The right column is titled "Search by chemical names" and contains a bulleted list: "Systematic names", "Synonyms", "Trade names", and "Database identifiers".

Home About us Web APIs Help Sign in

ROYAL SOCIETY OF CHEMISTRY

ChemSpider

Search and share chemistry

Search ChemSpider

For medical information relating to Covid-19, please consult the [World Health Organisation](#) or local healthcare provision.

Simple Structure Advanced History

Search ChemSpider

Matches any text strings used to describe a molecule.

Search

Systematic Name, Synonym, Trade Name, Registry Number, SMILES, InChI or CSID

What is ChemSpider?

ChemSpider is a free chemical structure database providing fast text and structure search access to over 100 million structures from hundreds of data sources.

Search by chemical names

- Systematic names
- Synonyms
- Trade names
- Database identifiers

Shortcuts for translating names

Online database: ChemSpider

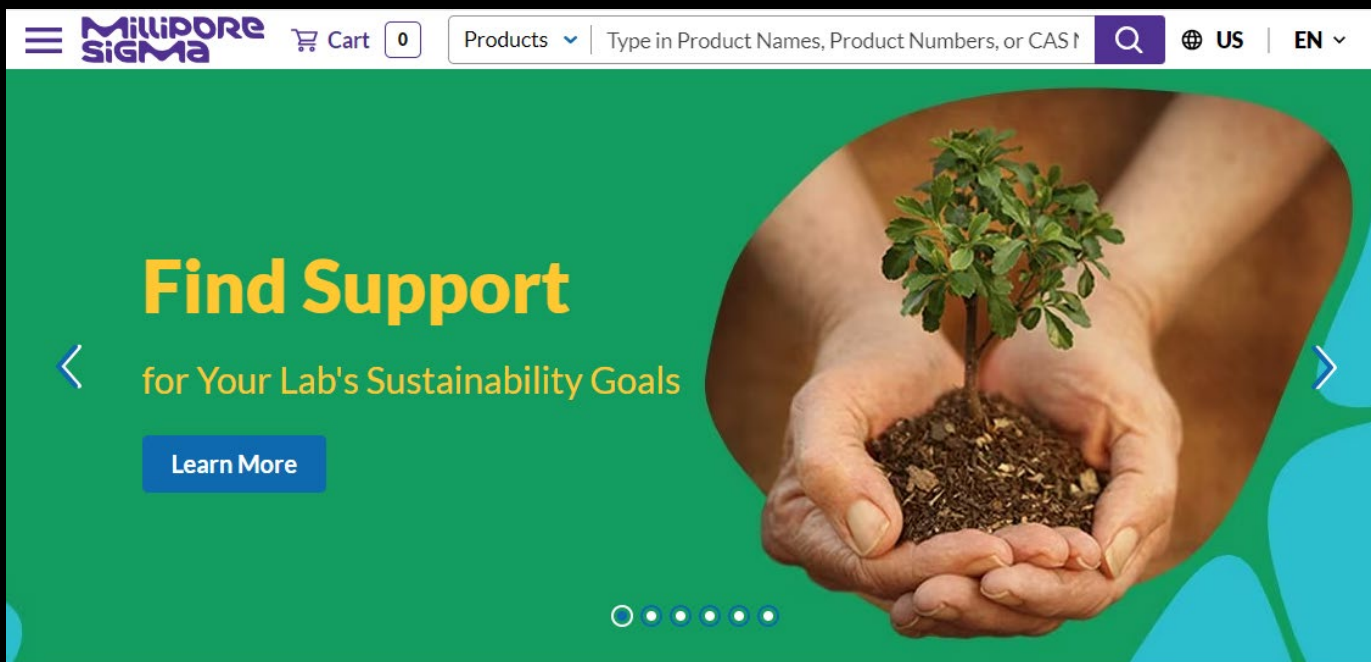
also has a structure search function

The screenshot displays the 'Structure search' interface on ChemSpider. At the top, there are three tabs: 'Draw structure' (selected), 'Convert structure', and 'Load structure'. Below the tabs, a message reads 'Use our editor to draw your structure'. A 'CLEAN' button is located in the top right corner. The main workspace is titled 'Ketcher' and 'Elemental'. It features a toolbar with various icons for drawing and editing, and a vertical list of elements: A, H, C, N, O, S, F, P, Cl, Br, and I. The central area is a large empty canvas for drawing the chemical structure.

Shortcuts for translating names Commercial chemicals sites Sigma-Aldrich

[<https://www.sigmaaldrich.com/US/en>]

set up for US region and EN language



The screenshot shows the top navigation bar of the Sigma-Aldrich website. On the left is the Millipore Sigma logo. Next to it is a shopping cart icon with a '0' inside. To the right is a search bar with a dropdown menu labeled 'Products' and a search icon. Further right are region and language selectors for 'US' and 'EN'. Below the navigation bar is a large green banner with a circular image of hands holding a small tree. The text on the banner reads 'Find Support for Your Lab's Sustainability Goals' with a 'Learn More' button. At the bottom of the banner are five small white circles, with the second one from the left being filled in.

Millipore
SIGMA

Cart 0

Products Type in Product Names, Product Numbers, or CAS#

US EN

Find Support

for Your Lab's Sustainability Goals

Learn More

Shortcuts for translating names

Commercial chemicals sites

Sigma-Aldrich

to set up for another language, click on the globe icon
this brings up the “Change Location” drop-down lists
e.g., now have selected “Europe” and “Germany” for German language

CHANGE LOCATION

Please select a region and location.

Region	Location
Europe	Germany
North America	Gibraltar
Africa	Greece
South America	Holy See
Asia & Australia	Hungary

In the United States and Canada we operate as MilliporeSigma. Please confirm that you are not located in one of these locations and you intend to visit our sites outside these locations by checking the box to the left and clicking Continue.

[Continue](#)

Shortcuts for translating names

Commercial chemicals sites

Sigma-Aldrich

enter "acetylsalicylic acid" in the search box for the name in German

The screenshot shows the Merck website interface. At the top, the Merck logo is on the left, and a search bar contains the text "acetylsalicylic acid". To the right of the search bar are language options "DE" and "DE". Below the search bar is a navigation menu with items: "Anwendungen", "Produkte", "Serviceangebote", "Dokumente", "Support", "Konto", "Bestellungssuche", "Schnelleinkauf", and "Warenkorb".

On the left side, there is a sidebar with the following sections:

- Erweiterte Suche**
- Struktursuche**
- Search Within**
 - Produkte
 - Baustein-Explorer
 - Technische Dokumente
 - Inhalt der Website
 - Artikel
 - Gene
 - Chromatograms
- Versand**
 - Ships Today (9)
- Produktkategorie**

The main content area displays the search results for "Acetylsalicylic Acid". At the top of this section is the title "Acetylsalicylic Acid". Below it is a filter box labeled "Angewandte Filter" containing the search term "Schlüsselwort: 'acetylsalicylic acid'". Below the filter box, it says "Anzeige 1-12 of 12 Ergebnisse for 'acetylsalicylic acid' innerhalb Produkte" and a button "Nach Relevanz sortieren".

The first result is for "Acetylsalicylsäure". It includes a chemical structure image of acetylsalicylic acid. Below the structure, the following information is provided:

- Synonym(e):** 2-Acetoxybenzoesäure, ASA, Aspirin
- Lineare Formel:** 2-(CH₃CO₂)C₆H₄CO₂H
- CAS No.:** 50-78-2
- Molekulargewicht:** 180.16
- EC No.:** 200-064-1
- Beilstein No.:** 779271

At the bottom left of the result, there is a link "Alle Fotos (6)".

Shortcuts for translating names Commercial chemicals sites TCI (Tokyo Chemical Industry)

[<https://www.tcichemicals.com/US/en>]

set up for US region and EN language

The screenshot shows the TCI website interface for the United States. At the top right, there are links for "United States" and "Contact Us". The main navigation bar includes "Products", "Services", "References", and "Support", each with a dropdown arrow. A search bar is located in the center, with the placeholder text "Enter CAS RN, Product Name or #, Keyword" and a search icon. Below the search bar are three search options: "Structure Search", "Advanced/Multi Search", and "Product Document Search". To the right of the search bar are "Sign In" and a shopping cart icon showing "\$0.00".

A promotional banner for "Organotellurium-Mediated Living Radical Polymerization (TERP) Reagents" is featured. It includes a "Learn More" button and a list of bullet points:

- Living Radical Polymerization using organic tellurium chain transfer agents
- Capable of polymerizing monomers containing amino, carboxyl, hydroxyl groups, and more
- Suitable for synthesizing block copolymers and hyperbranched polymers

Below the banner is a section titled "Explore Chemical Reagents" with five categories, each represented by an icon and a right-pointing arrow:

- Chemistry
- Life Science
- Materials Science
- Analytical Chemistry
- Custom & Bulk

Shortcuts for translating names

Commercial chemicals sites

TCI (Tokyo Chemical Industry)

select other country/region and language (many more than shown below)

Select Country or Region & Language ✕

Americas	Europe	Asia Pacific	Africa & Middle East
Argentina (English)	Austria (Deutsch)	Australia (English)	Egypt (English)
Brazil (English)	Austria (English)	China [Hong Kong] (English)	Israel (English)
Canada (English)	Belgium (Nederlands)	China [Mainland] (简体中文)	Other Africa & Middle East
Canada (Français)	Belgium (Français)	India (English)	
Mexico (English)	Belgium (English)	Indonesia (English)	
United States (English)	Bulgaria (English)	Japan (日本語)	
Other Americas	Croatia (English)	Japan (English)	
	Cyprus (English)	Korea (South) (한국어)	
	Czech Republic (English)	Korea (South) (English)	

Shortcuts for translating names

Artificial intelligence (AI)

Can AI help here?

Can I trust AI?

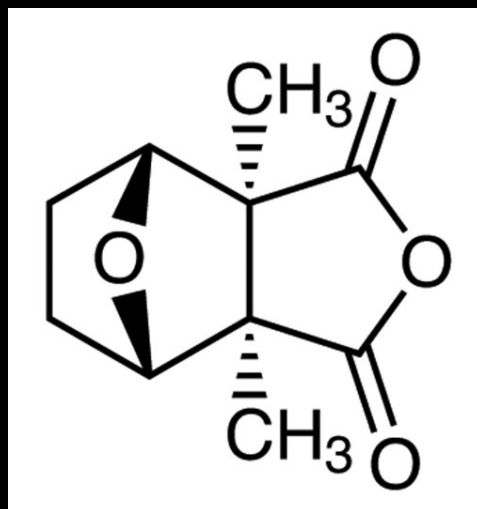
Maybe

Sometimes

Shortcuts for translating names

Artificial intelligence (AI)

test case: cantharidin



IUPAC name:

(1S,2R,6S,7R)-2,6-dimethyl-4,10-dioxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione

Shortcuts for translating names

Artificial intelligence (AI)

Can ChatGPT nomenclature?



give me the IUPAC systematic name for cantharidin



The IUPAC systematic name for cantharidin is:



(1aS,2R,4aR,8aR,8bR,9R,10R,11aS)-1,2,8a,8b,9,10,11,11a-octahydro-1a,2,4a,8a,10,11a-hexahydro-8,10-dihydroxy-2,9-dimethyl-7-oxo-1H,6H,7H,11H-[1,3]dioxacyclodecino[2,3,4-ij]isoquinolin-1-yl ethyl carbonate.

Please note that this is a highly detailed and systematic representation of the chemical structure of cantharidin, and it may not be commonly used in everyday discussions. The more common name for this compound is simply "cantharidin."

Not even close to **(1S,2R,6S,7R)-2,6-dimethyl-4,10-dioxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione**

Shortcuts for translating names Artificial intelligence (AI)

ChatGPT works much better
in the other direction

M

give me the common name for the chemical substance that has the IUPAC systematic name "2,6-dimethyl-4,10-dioxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione"



The common name for the chemical substance with the IUPAC systematic name "2,6-dimethyl-4,10-dioxatricyclo[5.2.1.0^{2,6}]decane-3,5-dione" is "**Cantharidin**."



However, ChatGPT failed when I included the stereochemical descriptors (**1S,2R,6S,7R**)

Shortcuts for translating names Artificial intelligence (AI)

In patent translations
might AI save time by rendering a lengthy list
of substituents or compounds faster
than a translator could type them?

Test: compare GoogleTranslate,
Bing Translate, DeepL, and ChatGPT

Shortcuts for translating names

Patent: JP03006475B2
“Photoresist composition”

(19)日本國特許庁 (J P)	(12) 特 許 公 報 (B 2)	(11)特許番号 特許第3006475号 (P3006475)
(45)発行日 平成12年2月7日(2000.2.7)		(24)登録日 平成11年11月26日(1999.11.26)
(51)Int.Cl. ⁷	識別記号	F I
G 0 3 F 7/039	6 0 1	G 0 3 F 7/039 6 0 1
	7/004 5 0 1	7/004 5 0 1
H 0 1 L 21/027		H 0 1 L 21/30 5 0 2 R
請求項の数1(全 9 頁)		
(21)出願番号 特願平8-74506	(73)特許権者 000004237 日本電気株式会社 東京都港区芝五丁目7番1号	
(22)出願日 平成8年3月28日(1996.3.28)	(72)発明者 岩佐 繁之 東京都港区芝五丁目7番1号 日本電気 株式会社内	
(65)公開番号 特開平9-265177	(74)代理人 100082935 弁理士 京本 直樹 (外2名)	
(43)公開日 平成9年10月7日(1997.10.7)		
審査請求日 平成8年3月28日(1996.3.28)	審査官 山鹿 勇次郎	
最終頁に続く		
(54)【発明の名称】	フォトレジスト組成物	

Shortcuts for translating names

The test text

【0018】

本発明の構成要素の光により酸を発生する光酸発生剤は、例えば、シクロヘキシルメチル（2-オキソシクロヘキシル）スルホニウム、トリフルオロメタンスルホナート、ジシクロヘキシル（2-オキソシクロヘキシル）スルホニウムトリフルオロスルホナート、ジシクロヘキシルスルホニルシクロヘキサノン、ジメチル（2-オキソシクロヘキシル）スルホニウムトリフルオロメタンスルホナート、トリフェニルスルホニウムトリフルオロメタンスルホナート、ジフェニルヨードニウムトリフルオロメタンスルホナートが挙げられる。

[0018]

A photoacid generator which generates an acid when exposed to light as a component of the present invention, for example...

Shortcuts for translating names

Google Translate

The screenshot shows the Google Translate web interface. At the top left is the Google Translate logo. Below it are four buttons: 'Text', 'Images', 'Documents', and 'Websites'. The source language is set to 'Japanese' and the target language is 'English'. The input text on the left is a Japanese chemical name: シクロヘキシルメチル (2-オキソシクロヘキシル) スルホニウム、トリフルオロメタンスルホナート、ジシクロヘキシル (2-オキソシクロヘキシル) スルホニウムトリフルオロスルホナート、ジシクロヘキシルスルホニルシクロヘキサノン、ジメチル (2-オキソシクロヘキシル) スルホニウムトリフルオロメタンスルホナート、トリフェニルスルホニウムトリフルオロメタンスルホナート、ジフェニルヨードニウムトリフルオロメタンスルホナート. The output text on the right is the English translation: Cyclohexylmethyl(2-oxocyclohexyl)sulfonium, trifluoromethanesulfonate, dicyclohexyl(2-oxocyclohexyl)sulfonium trifluorosulfonate, dicyclohexylsulfonylcyclohexanone, dimethyl(2-oxocyclohexyl)sulfonium trifluoromethanesulfonate, triphenylsulfonium trifluoromethane Sulfonate, diphenyliodonium trifluoromethane sulfonate. There is a close button (X) next to the input text.

Shortcuts for translating names

BingTranslate

The screenshot shows the Microsoft Bing Translate interface. At the top, there is a search bar with the text "Search the web" and a magnifying glass icon. Below the search bar, there are navigation links for "Text", "Translator", "Conversation", "Apps", "For business", and "Help". The main content area is divided into two columns. The left column has a dropdown menu set to "Japanese" and contains a long Japanese text string: シクロヘキシルメチル (2-オキシシクロヘキシル) スルホニウム、トリフルオロメタンスルホナート、ジシクロヘキシル (2-オキシシクロヘキシル) スルホニウムトリフルオロスルホナート、ジシクロヘキシルスルホニルシクロヘキサノン、ジメチル (2-オキシシクロヘキシル) スルホニウムトリフルオロメタンスルホナート、トリフェニルスルホニウムトリフルオロメタンスルホナート、ジフェニルヨードニウムトリフルオロメタンスルホナート. A small "x" icon is next to the text. The right column has a dropdown menu set to "English" and contains the translated text: cyclohexylmethyl (2-oxocyclohexyl) sulfonium, trifluoromethanesulfonate, dicyclohexyl (2-oxocyclohexyl) sulfonium trifluorosulfonate, dicyclohexylsulfonylcyclohexanone, dimethyl (2-oxocyclohexyl) sulfonium, trifluoromethanesulfonate, Triphenylsulfonium, Trifluoromethanesulfonate, Diphenyliodonium, Trifluoromethanesulfonate. A double-headed arrow icon is positioned between the two columns.

Shortcuts for translating names

DeepL

The screenshot shows the DeepL Translator web interface. At the top, there is a navigation bar with the DeepL logo, a 'Translator' dropdown menu, and links for 'DeepL Pro', 'For Business', 'Why DeepL?', and 'API'. A 'Start free trial' button is also present. Below the navigation bar, there are three main service buttons: 'Translate text' (31 languages), 'Translate files' (.pdf, .docx, .pptx), and 'DeepL Write' (AI-powered edits). The main interface is set to translate from Japanese to English (US). The input text on the left is a list of Japanese chemical names, and the output on the right is their corresponding English names.

Japanese ↔ English (US) Glossary

シクロヘキシルメチル (2-オキソシクロヘキシル) スルホニウム、トリフルオロメタンスルホナート、ジシクロヘキシル (2-オキソシクロヘキシル) スルホニウム、トリフルオロスルホナート、ジシクロヘキシルスルホニルシクロヘキサノン、ジメチル (2-オキソシクロヘキシル) スルホニウムトリフルオロメタンスルホナート、トリフェニルスルホニウムトリフルオロメタンスルホナート、ジフェニルヨードニウムトリフルオロメタンスルホナート

Cyclohexylmethyl (2-oxocyclohexyl)sulfonium, trifluoromethanesulfonate, dicyclohexyl (2-oxocyclohexyl)sulfonium trifluorosulfonate, Dicyclohexylsulfonylcyclohexanone, dimethyl (2-oxocyclohexyl)sulfonium trifluoromethanesulfonate, triphenylsulfonium trifluoromethanesulfonate, diphenyliodonium trifluoromethanesulfonate

Shortcuts for translating names

ChatGPT

M

please translate the following from Japanese into English as a continuous list without numbering: シクロヘキシルメチル (2-オキシシクロヘキシル) スルホニウム、トリフルオロメタンスルホナート、ジシクロヘキシル (2-オキシシクロヘキシル) スルホニウムトリフルオロスルホナート、ジシクロヘキシルスルホニルシクロヘキサノン、ジメチル (2-オキシシクロヘキシル) スルホニウムトリフルオロメタンスルホナート、トリフェニルスルホニウムトリフルオロメタンスルホナート、ジフェニルヨードニウムトリフルオロメタンスルホナート



Certainly, here is the continuous list of the chemical compounds translated from Japanese to English without numbering:

Cyclohexylmethyl (2-oxocyclohexyl) sulfonium, trifluoromethanesulfonate, dicyclohexyl (2-oxocyclohexyl) sulfonium trifluorosulfonate, dicyclohexylsulfonylcyclohexanone, dimethyl (2-oxocyclohexyl) sulfonium trifluoromethanesulfonate, triphenylsulfonium trifluoromethanesulfonate, diphenyliodonium trifluoromethanesulfonate.

Shortcuts for translating names

Comparison

GT

Cyclohexylmethyl(2-oxocyclohexyl)sulfonium, trifluoromethanesulfonate, dicyclohexyl(2-oxocyclohexyl)sulfonium trifluorosulfonate, dicyclohexylsulfonylcyclohexanone, dimethyl(2-oxocyclohexyl)sulfonium trifluoromethanesulfonate, triphenylsulfonium trifluoromethane Sulfonate, diphenyliodonium trifluoromethane sulfonate

Bing

cyclohexylmethyl (2-oxocyclohexyl) sulfonium, trifluoromethanesulfonate, dicyclohexyl (2-oxocyclohexyl) sulfonium trifluorosulfonate, dicyclohexylsulfonylcyclohexanone, dimethyl (2-oxocyclohexyl) sulfonium trifluoromethanesulfonate, Triphenylsulfonium Trifluoromethanesulfonate, Diphenyliodonium Trifluoromethanesulfonate

DeepL

Cyclohexylmethyl (2-oxocyclohexyl)sulfonium, trifluoromethanesulfonate, dicyclohexyl (2-oxocyclohexyl)sulfonium trifluorosulfonate, Dicyclohexylsulfonylcyclohexanone, dimethyl (2-oxocyclohexyl)sulfonium trifluoromethanesulfonate, triphenylsulfonium trifluoromethanesulfonate, diphenyliodonium trifluoromethanesulfonate

ChatGPT

Cyclohexylmethyl (2-oxocyclohexyl) sulfonium, trifluoromethanesulfonate, dicyclohexyl (2-oxocyclohexyl) sulfonium trifluorosulfonate, dicyclohexylsulfonylcyclohexanone, dimethyl (2-oxocyclohexyl) sulfonium trifluoromethanesulfonate, triphenylsulfonium trifluoromethanesulfonate, diphenyliodonium trifluoromethanesulfonate.

Shortcuts for translating names

Conclusions for AI assistance

Errors are mainly inappropriate capitalization and introduction of spaces, otherwise accurate

Use of AI saves time, even if minor editing is necessary

DeepL appears to require the least editing, but all four are reasonably good.

Conclusions

Chemical names can be tricky

Chemical structures - only for the brave

Best strategies:

- Search on an English name/CAS number (or structure) in an institutional chemical database, or commercial chemical database after specifying language preference
- For the adventurous, draw the structure and let AI generate the name

Conclusions

Best strategies:

Searching chemical names across languages

Institutional chemicals databases:

NIH PubChem [<https://pubchem.ncbi.nlm.nih.gov/>]

WikiData [https://www.wikidata.org/wiki/Wikidata:Main_Page]

ChemSpider [<https://www.chemspider.com/>]

Commercial chemicals databases:

Sigma-Aldrich [<https://www.sigmaaldrich.com/US/en>]

TCI (Tokyo Chemical Industry) [<https://www.tcichemicals.com/US/en>]

Conclusions

Best strategies:

AI/MT can save time with lists of chemical substances and substituents:

Google Translate [<https://translate.google.com/>]

Bing Translate [<https://www.bing.com/translator>]

DeepL [<https://www.deepl.com/translator>]

ChatGPT [<https://chat.openai.com/>]

Resources

IUPAC Gold Book Compendium of Chemical Terminology

<https://goldbook.iupac.org/>

IUPAC Blue Book Nomenclature of Organic Chemistry

<https://iupac.qmul.ac.uk/BlueBook/PDF/>

**How to name organic compounds using the IUPAC rules
(Univ. Ill. Urbana-Champaign)**

http://www.chem.uiuc.edu/GenChemReferences/nomenclature_rules.html

**Short Summary of IUPAC Nomenclature of Organic Compounds
(Angelo State University, 9 pages)**

https://www.angelo.edu/faculty/kboudrea/organic/IUPAC_Handout.pdf

**IUPAC Nomenclature of Alkanes - Naming Organic Compounds
(The Organic Chemistry Tutor, YouTube, 11 min)**

https://www.youtube.com/watch?v=TYU_JluleME&ab_channel=TheOrganicChemistryTutor

**IUPAC Nomenclature of Alkanes and Cycloalkanes | Organic Chemistry
(Chad's Prep, YouTube, 35 min)**

https://www.youtube.com/watch?v=hcpWpluvXgc&ab_channel=Chad%27sPrep

Resources

NIH PubChem [<https://pubchem.ncbi.nlm.nih.gov/>]

WikiData [https://www.wikidata.org/wiki/Wikidata:Main_Page]

ChemSpider [<https://www.chemspider.com/>]

Sigma-Aldrich [<https://www.sigmaaldrich.com/US/en>]

TCI (Tokyo Chemical Industry) [<https://www.tcichemicals.com/US/en>]

GoogleTranslate [<https://translate.google.com/>]

Bing Translate [<https://www.bing.com/translator>]

DeepL [<https://www.deepl.com/translator>]

ChatGPT [<https://chat.openai.com/>]

Chemical Names, Chemical Structures What's a Translator to Do?

ATA 64th Annual Conference – Miami, FL
26 October 2023

Matthew Schlecht, PhD
Word Alchemy Translation
wordalchemytranslation.com
mattschlecht@wordalchemytranslation.com



Word
錬金術 炼金术
Alchimie **Alchemy** АЛХИМИЯ
Alquimia