

DEPARTAMENTO DE QUÍMICA INORGÁNICA

Tutorial sobre [PubChem](https://pubchem.ncbi.nlm.nih.gov)
<https://pubchem.ncbi.nlm.nih.gov>

Pablo González Herrero

PubChem es una base de datos pública de acceso gratuito perteneciente a **National Institutes of Health (NIH)**, la mayor agencia de investigación biomédica del mundo, gestionada por el Departamento de Estado de Salud de Estados Unidos. Se lanzó en 2004 y se ha convertido en una fuente de información clave para científicos, estudiantes y el público en general. Recopila principalmente moléculas orgánicas pequeñas, aunque también biomoléculas grandes, tales como nucleótidos, péptidos, lípidos, etc., e incluye la siguiente información sobre ellas:

- Estructura.
- Nombres e identificadores.
- Propiedades físicas y químicas.
- Datos espectroscópicos.
- Empresas comercializadoras.
- Datos sobre seguridad, peligrosidad y toxicidad.
- Actividad biológica y datos farmacológicos.
- Bibliografía y patentes.
- Fuentes de información.

The screenshot shows the PubChem website homepage. At the top, there is a navigation bar with the NIH logo and the text 'National Library of Medicine National Center for Biotechnology Information'. Below this is the 'PubChem' logo and a menu with 'About', 'Posts', 'Submit', and 'Contact'. The main content area features a large blue banner with the text 'Explore Chemistry' and 'Quickly find chemical information from authoritative sources'. A search bar is prominently displayed, with a search icon on the right. Below the search bar, there are several 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 search bar, there are four radio buttons: 'Use Entrez' (checked), 'Compounds', 'Substances', and 'BioAssays'. Below the radio buttons, there are four icons with labels: 'Draw Structure', 'Upload ID List', 'Browse Data', and 'Periodic Table'. At the bottom of the page, there is a statistics section with the following data: '111M Compounds', '278M Substances', '295M Bioactivities', '34M Literature', and '862 Data Sources'. There are also links for 'See More Statistics' and 'Explore Data Sources'.

1. Búsquedas sencillas en PubChem.

Para buscar un compuesto en **PubChem**, se debe introducir en el campo de búsqueda de la página principal alguno de los siguientes términos:

- **Nombre completo de la sustancia.** Puede introducirse el nombre IUPAC de moléculas orgánicas en inglés o nombres propios o triviales.
- Fórmula molecular.
- Notación **SMILES** (*Simplified Molecular-Input Line-Entry System*).
- Identificador **InChI** (*International Chemical Identifier*).
- Identificador **CAS** (*Chemical Abstracts Service*).

Por ejemplo, si introducimos el término **caffeine**, PubChem nos ofrece todas las entradas que lo contengan, incluidos todos los derivados de la cafeína. Además, nos ofrece la posibilidad de encontrar las entradas sobre genes o proteínas que interaccionan con la cafeína, pasos de la biosíntesis, bioensayos y bibliografía relacionada:

The screenshot shows the PubChem website interface. At the top, there is a navigation bar with 'PubChem' logo and links for 'About', 'Posts', 'Submit', and 'Contact'. Below this is a search bar with 'caffeine' entered. A message below the search bar says 'Treating this as a text search.' The results section is titled 'BEST MATCH' and features a card for 'Caffeine; 58-08-2; 1,3,7-Trimethylxanthine; Guaranine; Thein; Methyltheobromine; Cafeina; Theine; ...'. The card includes a chemical structure icon, the compound CID (2519), molecular formula (C₈H₁₀N₄O₂), molecular weight (194.19g/mol), IUPAC name, isomeric SMILES, InChIKey, InChI string, and create date. Below the card are tabs for 'Summary', 'Similar Structures Search', 'Related Records', and 'PubMed (MeSH Keyword)'. A table below the card shows the number of results for various categories: Compounds (289), Substances (1,205), Genes (4), Proteins (34), Pathways (20), BioAssays (477), Literature (36,286), and Patents (2,869). The 'Pathways' category is highlighted. Below the table, there is a section for 'Searching pathways information contributed by multiple sources. Read More...'. This section shows '20 results' and a 'SORT BY' dropdown set to 'Relevance'. A 'Download' button is also present. The first result is 'caffeine biosynthesis I', with source 'BioCyc', external ID 'PWY-5037', pathway type 'conserved', and a compound count of 12. The second result is 'caffeine biosynthesis II (via paraxanthine)'. On the right side, there is a section 'ACTIONS ON RESULTS WITH ID TYPE:' with radio buttons for 'Pathways', 'Compounds', 'Genes', and 'Proteins'. The 'Pathways' option is selected. There is also a 'Save for Later' button with a star icon.

Si pulsamos sobre el nombre del resultado principal, se muestran todos los datos disponibles. Se puede navegar por los datos usando la tabla de contenido de la derecha. Para muchas sustancias, están disponibles los espectros de RMN de ¹H y ¹³C.

COMPOUND SUMMARY

Caffeine

Cite Download

CONTENTS

- Title and Summary
- 1 Structures
- 2 Names and Identifiers
- 3 Chemical and Physical Properties
- 4 Spectral Information
- 5 Related Records
- 6 Chemical Vendors
- 7 Drug and Medication Information
- 8 Food Additives and Ingredients
- 9 Agrochemical Information
- 10 Pharmacology and Biochemistry
- 11 Use and Manufacturing
- 12 Identification
- 13 Safety and Hazards
- 14 Toxicity
- 15 Associated Disorders and Diseases
- 16 Literature
- 17 Patents
- 18 Biomolecular Interactions and Pathways
- 19 Biological Test

PubChem CID 2519

Structure



2D



3D



Crystal

[Find Similar Structures](#)

Chemical Safety



Irritant

[Laboratory Chemical Safety Summary \(LCSS\) Datasheet](#)

Molecular Formula $C_8H_{10}N_4O_2$

Synonyms

- caffeine
- 58-08-2
- 1,3,7-Trimethylxanthine
- Guaranine
- Thein

[More...](#)

Molecular Weight 194.19

Dates

Modify	Create
2022-06-05	2004-09-16

Caffeine is a [methylxanthine](#) alkaloid found in the seeds, nuts, or leaves of a number of plants native to South America and East Asia that is structurally related to [adenosine](#) and acts primarily as an [adenosine](#) receptor antagonist with psychotropic and anti-inflammatory

Las búsquedas en PubChem se pueden realizar también introduciendo una **estructura**. Para ello, hay que pulsar en **Draw Structure**, y accedemos a una aplicación de dibujo:

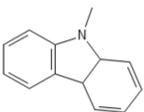
NIH National Library of Medicine
National Center for Biotechnology Information

PubChem About Posts Submit Contact

DRAW STRUCTURE ✕

Broadband
SMILES
C1=CC=CC2=C1N(C3C2C=CC=C3)C

New	Undo	Clear	Style	Delete	Copy	Redo	Refresh	Zoom In	Zoom Out
Line	Double	Triple	Wedge	Arrow	Circle	Square	Hexagon	Octagon	Triangle
Circle	Hexagon	Octagon	Triangle	Grid	Atom	Atom	Atom	Atom	Atom
Alkyl	Alkyl	Alkyl	Alkyl	Alkyl	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 Seleccionar archivo nada seleccionado

[Search for This Structure](#)

Esta aplicación tiene muchas funcionalidades, entre otras:

- Generar automáticamente la notación SMILES o InChI.
- Exportar la estructura dibujada en distintos formatos.
- Importar archivos de estructuras en formatos SDF/MOL, SMILES/SMART o InChI.

2. Navegación por las distintas clasificaciones

Una de las utilidades más importantes de **PubChem** es que clasifica las sustancias de acuerdo con una serie de categorías organizadas jerárquicamente, entre las que se puede navegar y descubrir sustancias con determinadas propiedades, funciones biológicas, actividad farmacológica, etc. Se accede a esta navegación pulsando en **Browse Data**.

NCBI

PubChem Classification Browser

Help

Browse PubChem data using a classification of interest, or search for PubChem records annotated with the desired classification/term (e.g., MeSH: phenylpropionates, or Gene Ontology: DNA repair). [More...](#)

Select classification: **PubChem: PubChem Compound TOC** Search selected classification by: **Keyword** Enter desired search term **Search**

Classification description (from PubChem)
This classification was created automatically from the PubChem Compound TOC on 2022/05/30.
Note that in some cases a number of highly populated nodes - those for which all or nearly all IDs have information - have been left out of the tree.
The sections, along with their child subsections, that are not shown in this tree are: Computed Properties, Substances by Category, Computed Descriptors, Molecular Formula, Depositor-Supplied Synonyms, Removed Synonyms, Create Date, Modify Date, Record Title, Related Compounds, Related Compounds with Annotation, Related Substances, 2D Structure, 3D Conformer, and Chemical Vendors. [More...](#)

Data type counts to display: **None** **Compound** Display zero count nodes? **Yes** **No** Filter by Entrez History: **Choose one**

Browse PubChem: PubChem Compound TOC Tree

- PubChem Compound TOC **62,145,031**
 - Agrochemical Information **3,131**
 - Associated Disorders and Diseases **23,719**
 - Biologic Description **2,197,862**
 - Biological Test Results **3,923,235**
 - Biomolecular Interactions and Pathways **138,594**
 - Chemical and Physical Properties **269,963**
 - Classification **20,728,374**
 - Drug and Medication Information **18,317**
 - Food Additives and Ingredients **7,534**
 - Identification **4,915**
 - Information Sources **41,859,364**
 - Literature **1,949,257**
 - Names and Identifiers **1,738,077**
 - Patents **37,461,567**
 - Pharmacology and Biochemistry **112,002**
 - Related Records **13,101,041**

Debido a las inmensas posibilidades que ofrece **PubChem**, no es posible una descripción pormenorizada en el presente texto. Existen varios manuales de uso, que se pueden encontrar pulsando en **About** y, después, en **Tutorials**, que dan una idea del alcance de esta base de datos:

pubchemdocs.ncbi.nlm.nih.gov

PubChem Docs PubChem Home Submit Contact

- About
- Tutorials**
- COVID-19/SARS-CoV-2
- Contact PubChem
- News
- Publications
- Citation Guidelines
- What's in PubChem?
- Data Sources
- Search and Tools
- Submissions
- Downloads
- Programmatic Access
- Widgets
- RDF
- Glossary
- Data Specification
- PubChem3D
- Statistics

Tutorials

The following paper provides step-by-step instructions on how to explore data contained in PubChem, along with examples of commonly requested tasks.

Kim S. **Exploring Chemical Information in PubChem**. *Curr. Protoc.*; 2021 Aug 9; 1(8):e217. doi: <https://doi.org/10.1002/cpz1.217>.
[\[PubMed PMID: 34370395\]](#) [\[PubMed Central PMCID: PMC8363119\]](#) [\[Free Full Text\]](#)

This paper includes several protocols designed to help users to get familiar with PubChem's data and tools.

- Basic Protocol 1:** [Finding genes and proteins that interact with a given compound](#)
- Basic Protocol 2:** [Finding drug-like compounds similar to a query compound through a two-dimensional \(2-D\) similarity search](#)
- Basic Protocol 3:** [Finding compounds similar to a query compound through a three-dimensional \(3-D\) similarity search](#)
- Support Protocol:** [Computing similarity scores between compounds](#)
- Basic Protocol 4:** [Getting the bioactivity data for the hit compounds from substructure search](#)
- Basic Protocol 5:** [Finding drugs that target a particular gene](#)
- Basic Protocol 6:** [Getting bioactivity data of all chemicals tested against a protein](#)
- Basic Protocol 7:** [Finding compounds annotated with classifications or ontological terms](#)
- Basic Protocol 8:** [Finding stereoisomers and isotopomers of a compound through identity search](#)