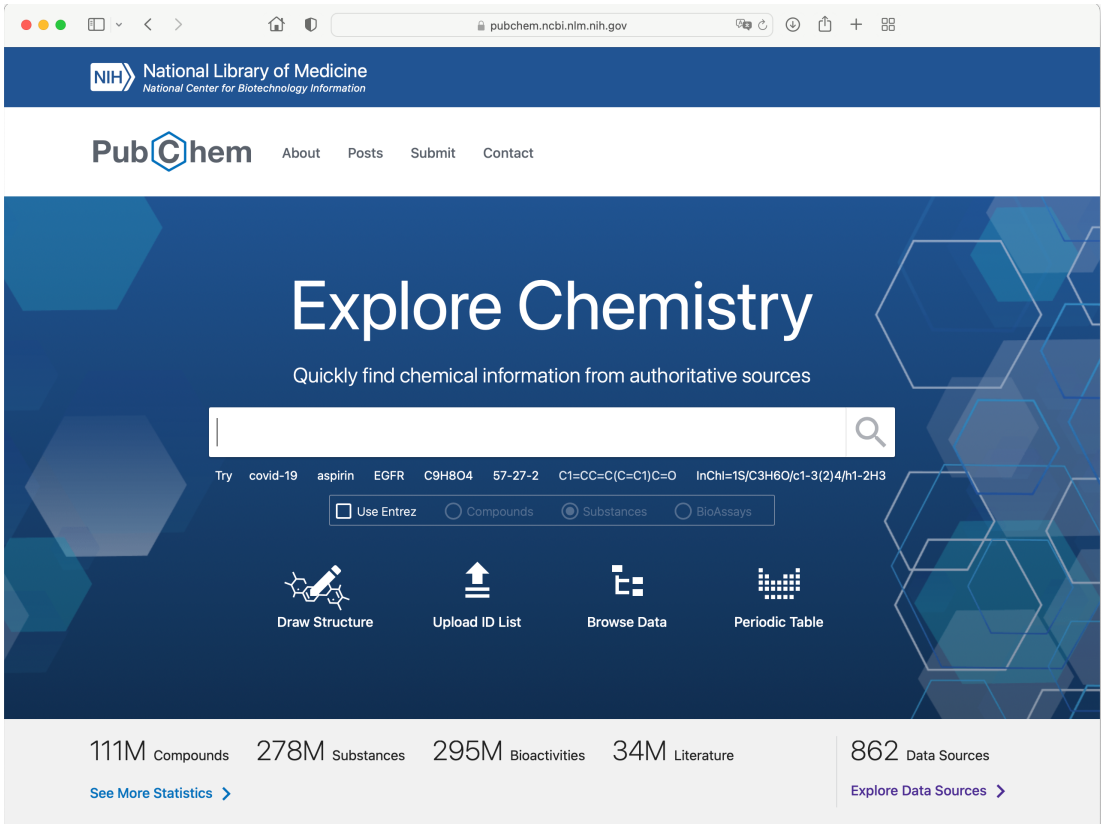
DEPARTAMENTO DE QUÍMICA INORGÁNICA

Tutorial sobre ***PubChem***
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.

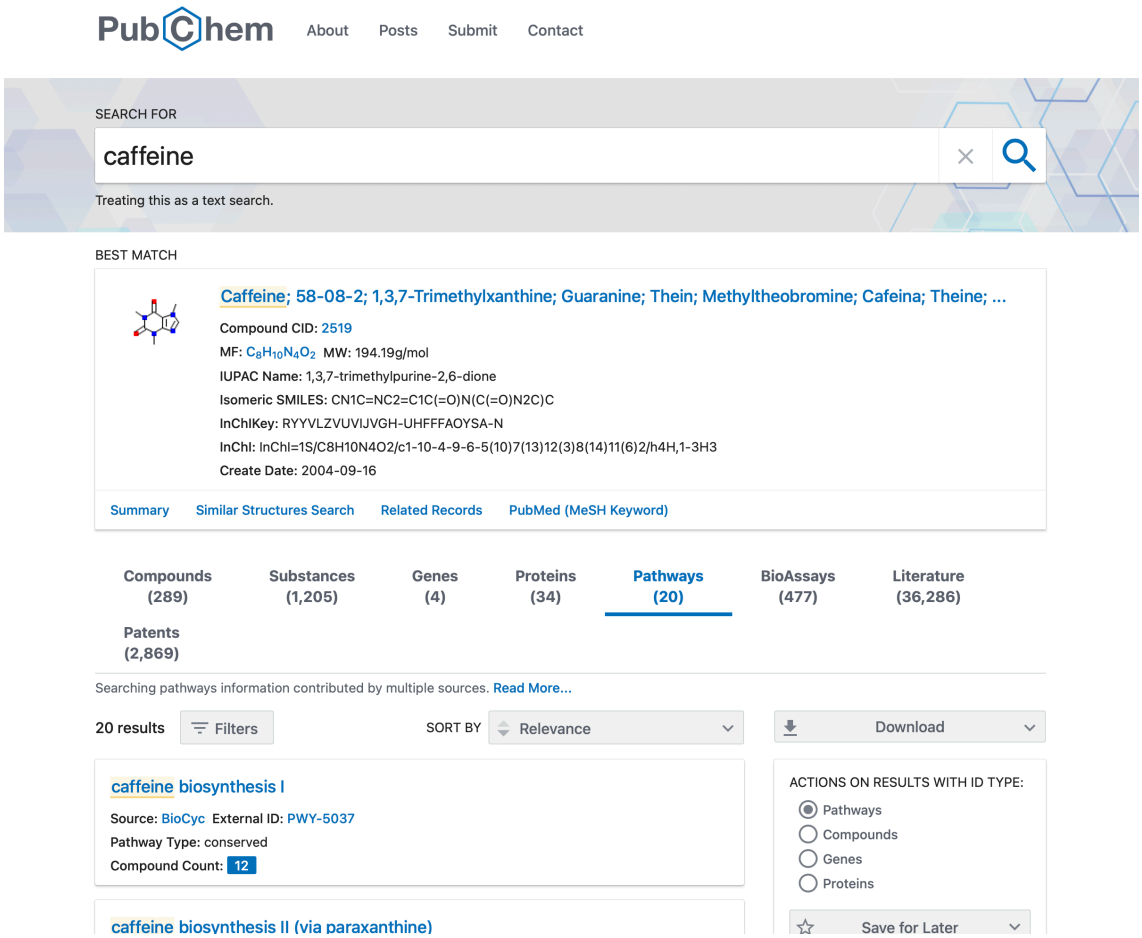


1. Búsquedas sencillas en PubChem.

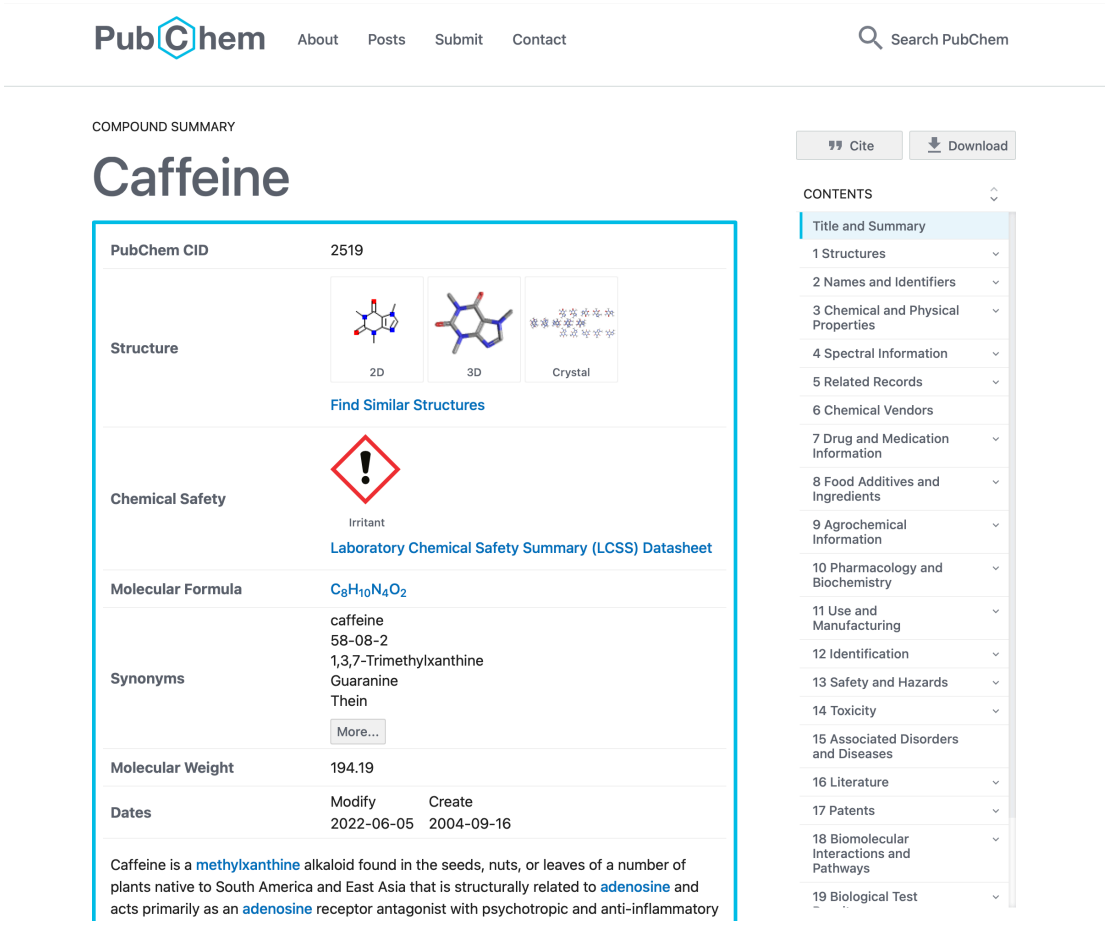
Para buscar un compuesto en **PubChem**, se debe introducir en el campo de búsqueda de la página principal alguno de los siguientes términos:

- **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*).

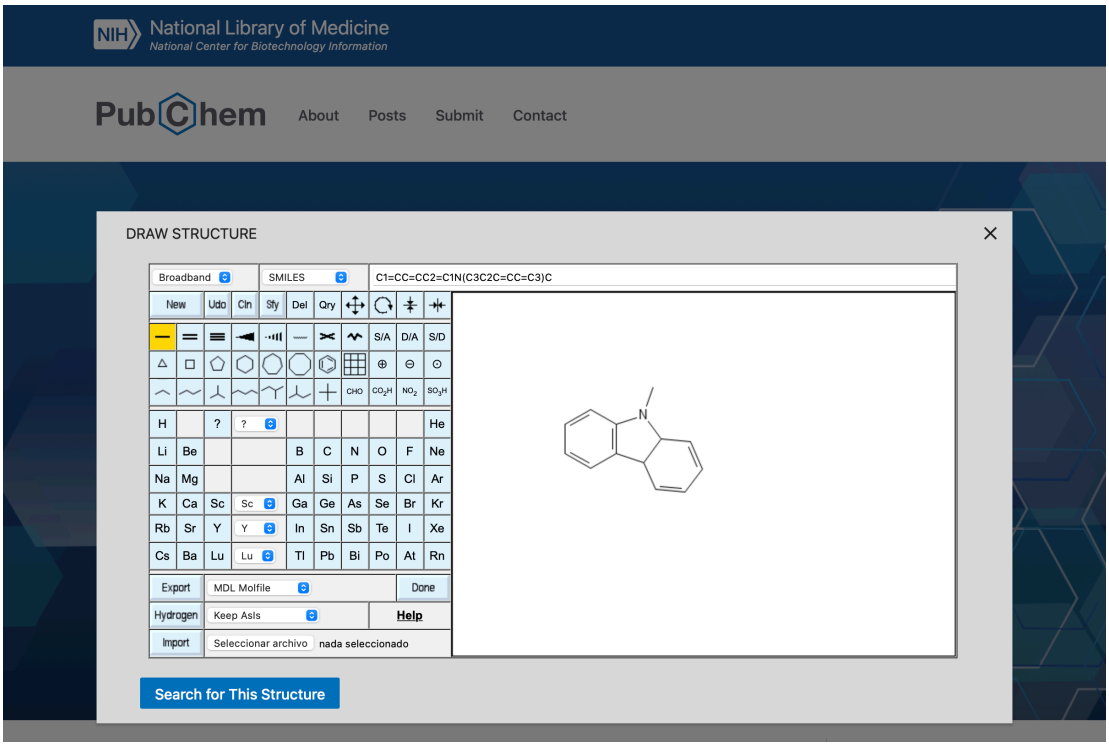
Por ejemplo, si introducimos el término **caffeine**, PubChem nos ofrece todas las entradas que lo contengan, incluidos todos los derivados de la cafeína. Además, nos ofrece la posibilidad de encontrar las entradas sobre genes o proteínas que interaccionan con la cafeína, pasos de la biosíntesis, bioensayos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



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 ^{1}H y ^{13}C.



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:

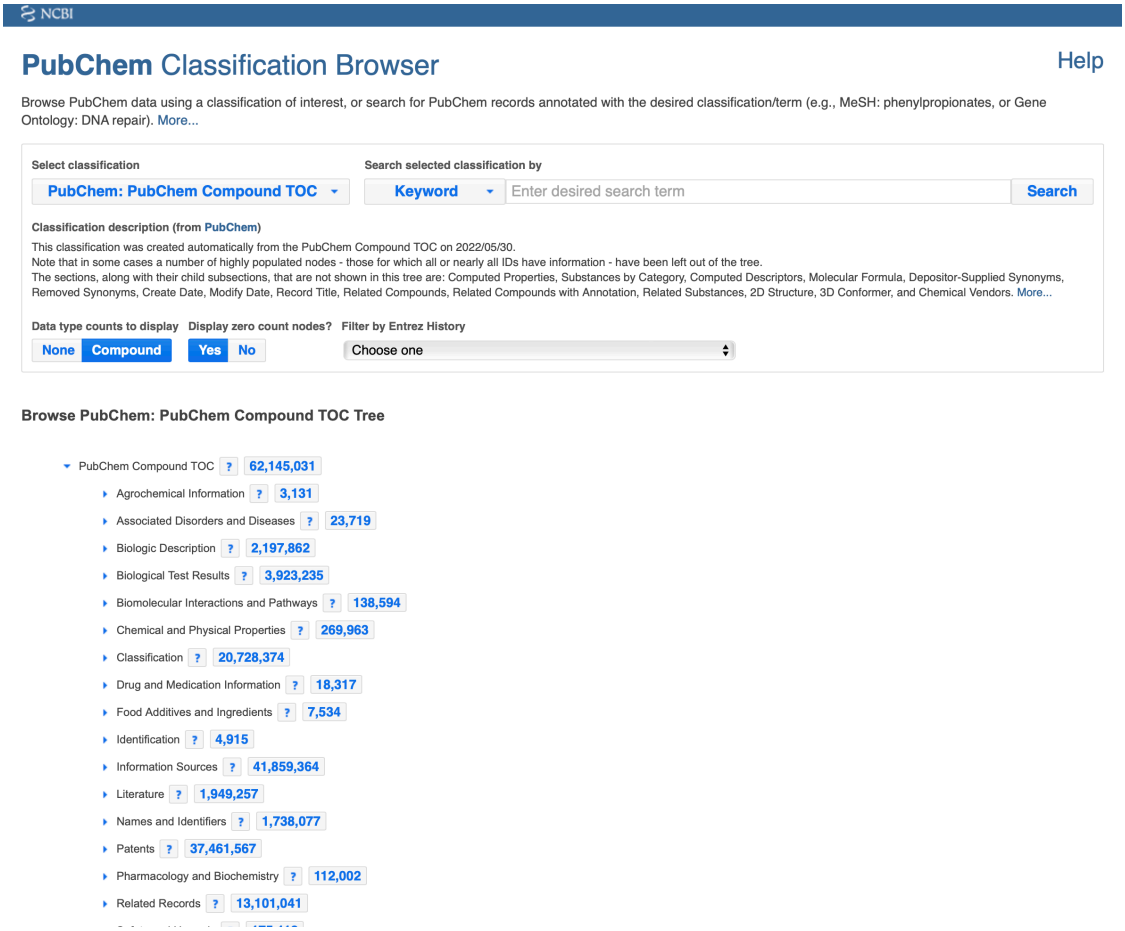


Esta aplicación tiene muchas funcionalidades, entre otras:

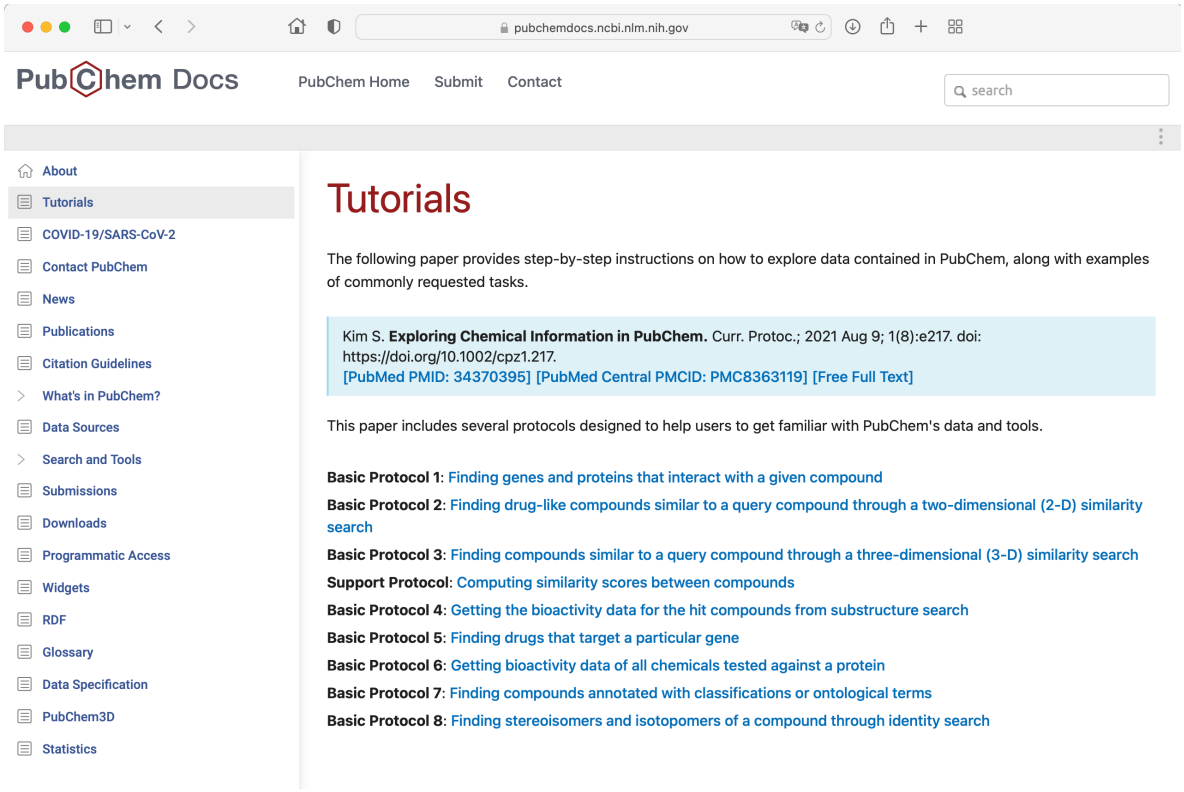
- Generar automáticamente la notación SMILES o InChI.
- Exportar la estructura dibujada en distintos formatos.
- Importar archivos de estructuras en formatos SDF/MOL, SMILES/SMART o InChI.

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**.



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
search
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