

化合物データベース (PubChem、ChEMBL、ChEBI) を使ってみる

化合物&メタボロームデータベースを知って・学んで・使う
(AJACSオンライン13)

<https://biosciencedbc.jp/event/ajacs/ajacs94.html>

2022/11/22

櫛田達矢 (理化学研究所バイオリソース研究センター (BRC))

理研バイオリソース研究センター (BRC)



実験動物 (>7,000)
例, 遺伝子改変マウス

実験植物 (>599,000)
例, シロイヌナズナ完全長cDNAクローン

細胞材料 (>7,000)
例, 疾患患者由来のヒトiPS細胞

RIKEN BRC, as one of the world leading scientific infrastructures, contributes to advancement of life science and human welfare in the 21st century.

Search for bioresource | Site search

Search from all bioresource databases | Search explanation

Gene name, Resource name, e.t.c. [Search]

Click here to search for individual resources

Mouse Plant Cell DNA Microbe

MSMMs (RBR00209) Mouse strains Experimental Animal Division	Tobacco Cell Line (rpc00039) Seed, cell lines, DNA materials Experimental Plant Division	HeLa/Fucci (RCB2812) Cell lines including stem cells Cell Engineering Division
pcDNA3 Venus-Akaluc (RDB15781) DNA clones, Genomic DNA etc. Gene Engineering Division	Prolixibacter denitrificans (JCM 18694) Microbial strains, DNA materials Microbe Division	Delivery of BRC Mail News Information services Integrated Bioresource Information Division

ミッション

理研BRCのバイオリソースを通して, ライフサイエンス研究, ヘルスケア, 育種, 有用物質生産に貢献する

遺伝子材料 (>126,000)
例, ヒトcDNAクローン

微生物材料 (>19,000)
例, 好気性細菌・放線菌・嫌気性細菌・乳酸菌・アーキア・極限環境細菌・酵母・糸状菌

<https://web.brc.riken.jp/>

アクティビティ

1. 理研BRCバイオリソースデータと公的データベースの統合 (ナレッジグラフ作成)
2. [BRCカタログシステム](#) (バイオリソースのキーワード検索サービス) の開発, 提供



Contents

1. PubChemの使い方を学ぶ
2. ChEMBLの使い方を学ぶ
3. ChEBIの使い方を学ぶ
4. まとめ

1. PubChemの使い方を学ぶ
2. ChEMBLの使い方を学ぶ
3. ChEBIの使い方を学ぶ
4. まとめ

PubChemとは

- <https://pubchem.ncbi.nlm.nih.gov>
- 運用機関：National Center for Biotechnology Information (NCBI, US)
- 説明：
 - 統合的な化学情報データベースです。化学物質とその生物活性に関する情報を収録した**PubChem Substance**、医薬品名やその効能、LogP値などの物性値や類似構造を持つ別の化合物の情報を収録した**PubChem Compound**、化合物を用いた実験の方法や結果を収録した**PubChem BioAssay**を統合しています。またThieme Chemistry の化学文献リンク、SpringerMaterials の化学および物理特性リンク、世界知的所有権機関（WIPO）の特許リンクなども統合しています。化合物の名称、分子式、構造などによる検索が可能です。

- 880件以上の機関, プロジェクト (コントリビューター) から提供された化合物及びそれに関連する情報 (例, バイオアッセイ, 生物活性, 遺伝子, タンパク質, 生物種, パスウェイ, セルライン) を収録.
- 2つの化合物データセット
 - Substance
 - コントリビューターから提供された全化合物の集合 (約**3億件**)
 - Compound
 - Substanceから抽出したユニークな構造を持つ化合物を整理した集合 (約**1.1億件**)

PubChemが扱う主な情報

PubChem Data Counts

Data Collection	Live Count	Description
Compounds	112,405,915	Unique chemical structures extracted from contributed PubChem Substance records
Substances	298,292,628	Information about chemical entities provided by PubChem contributors
BioAssays	1,506,836	Biological experiments provided by PubChem contributors
Bioactivities	301,290,840	Biological activity data points reported in PubChem BioAssays
Genes	103,988	Genes tested in PubChem BioAssays and those involved in PubChem Pathways and identified in PubChem Patents
Proteins	185,153	Proteins tested in PubChem BioAssays and those involved in PubChem Pathways and identified in PubChem Patents
Taxonomy	112,531	Organisms of proteins/genes tested in PubChem BioAssays and those involved in PubChem Pathways and identified in PubChem Patents
Pathways	239,263	Interactions between chemicals, genes, and proteins
Cell Lines	1,964	Cell Lines tested in PubChem BioAssays
Literature	34,813,640	Scientific publications with links in PubChem
Patents	42,395,312	Patents with links in PubChem
Data Sources	884	Organizations contributing data to PubChem

2022.11.12 確認

PubChemキーワード検索

SEARCH FOR

covid-19

①



Treating this as a text search. Learn more about [COVID-19 \(Coronavirus Disease 2019\) data in PubChem](#).

DATA SOURCE

COVID-19 Disease Map

Curation Efforts

The COVID-19 Disease Map shares resources and best practices to develop a disease map for COVID-19. The project is an international, community-driven effort. It aims to establish a knowledge repository on virus-host interaction mechanisms specific to the SARS-CoV-2. The COVID-19 Disease Map is an assembly of molecular interaction diagrams established based on literature evidence.

Last Updated: 2022/08/01

Compounds
(1,700)

Substances
(76)

Genes
(656)

Proteins
(547)

Taxonomy
(2)

Pathways
(2,278)

BioAssays
(897)

Literature
(294,814)

Patents
(1,405)

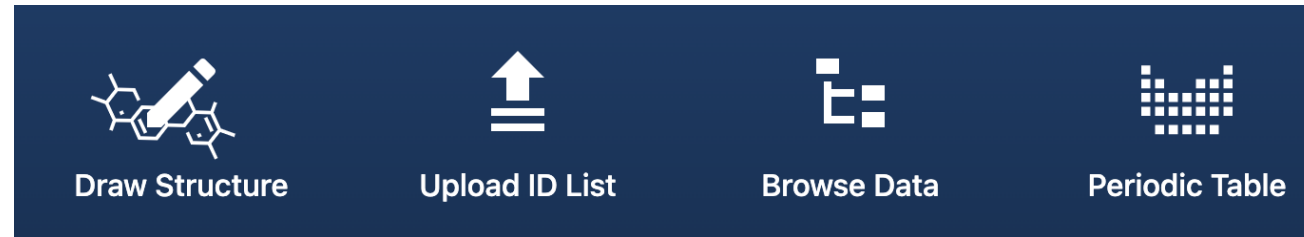
②

Searching chemical names and synonyms including IUPAC names and InChIKeys across the compound collection. Note that annotations text from compound summary pages is not searched. [Read More...](#)

① [トップ画面](#)の検索窓からキーワード検索が可能

② 検索結果がカテゴリ別に表示されるので、選んで検索結果を確認

- [PubChem トップ画面](#)には，検索インターフェイスを選ぶ，アイコンが表示されている。



- Draw Structure
 - 構造式エディタを使った検索が可能
- Upload ID List
 - CID, SID, GeneID, UniProt Accession番号を使ったバッチ検索が可能
- [Browse Data](#)
 - 木構造（例， Gene Ontology, NCBI Taxonomy）を使った絞り込み検索
- [Periodic Table](#)
 - 周期表を使った検索

Upload ID List を使った検索

UPLOAD ID LIST

A list of PubChem identifiers may be used as input to PubChem search where you can view or download records. Don't have a file handy? Download an [example list of PubChem CIDs](#).

STEP 1. Choose Identifier Type

Compound, e.g. CID like 2244

STEP 2. Provide a List of Identifiers

ENTER IDENTIFIERS SEPARATED BY COMMA OR SPACE

OR UPLOAD A FILE (ONE IDENTIFIER PER LINE)

ファイルを選択 PubChem_CID_list_example.csv

Preview of PubChem_CID_list_example.csv

Number of rows: 102
Number of columns: 1

1. 2244
2. 24847798
3. 24847966
4. 44219
5. 11980079
6. ...

選択されていません

SUCCESS

- ① トップ画面から「[Upload ID List](#)」アイコンを選択
- ② 「Choose Identifier Type」で検索対象のIDの種類を選択
- ③ 「Provide a List of Identifiers」で,
 - a. カンマもしくはスペース区切りのIDリストを入力
 - b. もしくは, 1行1IDで作成したテキストファイルをアップロードする.
 - c. サンプルファイルは[ここ](#)から入手可能
- ④ 「Search Pubchem With This List」ボタンをクリックし, 検索開始

Classification Browserを使った検索

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: phenylpropanates, or Gene Ontology: DNA repair). [More...](#)

②

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 2022/11/08.
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 **Compound** Display zero count nodes? **Yes** **No**

③ ④

Browse PubChem: PubChem Compound TOC Tree

- PubChem Compound TOC ? 63,629,784
 - Agrochemical Information ? 3,119
 - Associated Disorders and Diseases ? 31,783
 - Biologic Description ? 2,238,563
 - Biological Test Results ? 4,170,668
 - Chemical and Physical Properties ? 266,478
 - Classification ? 21,340,072
 - CAMEO Chemicals ? 4,799
 - CCSBase Classification ? 4,911
 - ChEBI Ontology ? 138,998
 - ChEMBL Target Tree ? 1,040,723
 - ChemIDplus ? 339,045
 - =====
 - KEGG: Lipid ? 1,956
 - ⑤ KEGG: Metabolite ? 521 ⑥
 - KEGG: Natural Toxins ? 281
 - KEGG: OTC drugs ? 306
 - KEGG: Peptide ? 22
 - KEGG: Pesticides ? 919

- ① トップページ画面から「Browse Data」アイコンを選択
- ② 「Select classification」で検索対象のデータセットを選択（例, PubChem Compound TOC）
- ③ 「Data type counts to display」でCompoundを選択
- ④ 「Display zero count nodes?」でNoを選択
- ⑤ 「Browse PubChem: PubChem Compound TOC Tree」で, カテゴリーを選択（例, KEGG: Metabolite）
- ⑥ 「数字」部分（例, 521）をクリックし, 検索開始

自前の化合物リストをPubChemに登録する

[Data Sources](#) を見ると、データタイプ別（例、化合物、バイオアッセイ）、国別のデータソースを確認することができる。

Data Sources

Find out *who* contributed *what* to PubChem.

Interested in becoming a PubChem contributor? Learn how to get started with a [PubChem submission](#).

16 sources

Download

FILTER BY 1 active x

Search Sources

SORT BY Last Updated Latest First

Data Type

- Live Substances (11)
- Annotations (6)
- Live BioAssays (1)
- Pathways (1)
- Classifications (1)
- On-Hold Substances (0)
- On-Hold BioAssays (0)

Source Category

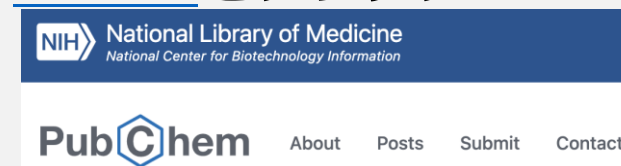
- Chemical Vendors (5)
- Research and Development (5)
- Governmental Organizations (5)
- Curation Efforts (4)
- Journal Publishers (0)
- Subscription Services (0)
- NIH Initiatives (0)

Source	Data Counts by Type	Last Updated
KEGG Research and Development, Curation Efforts Japan	41,103 Live Substances 4,544 Annotations 22 Classifications	2022/11/11
GlyYouCan Project Research and Development Japan	31,774 Live Substances 31,767 Annotations	2022/11/11
GlyCosmos Glycoscience Portal Research and Development Japan	149,903 Annotations	2022/11/11
GlycoNAVI Curation Efforts Japan	783 Annotations	2022/11/11
Watanabe Chemical Ind. Chemical Vendors Japan	2,846 Live Substances	2022/11/02

1.ncbi.nlm.nih.gov/source/11830#data=Annotations

データの登録方法

- トップ画面の[Submit](#)をクリック



- ユーザー登録もしくはログインする
- あらかじめ、オリジナルの化合物ID、化合物名および**InChI** や**SMILES** などの構造情報を記載したCSV形式のファイルを作成しておく。
- CSV形式をアップロードする。
- 個々の化合物に対して、PubChem Substance ID (SID)及びcompound ID (CID)が付与される。
- 同じCID*の化合物が持つ類似化合物、生物活性、相互作用などの情報を利用することができる。

*:同じCIDを持つ化合物は化合物構造が同一

- PUG REST

- Example

- <https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/cid/2244/PNG>
- <https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/name/aspirin/synonyms/XML>
- <https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/cid/1983/description/XML>
- <https://pubchem.ncbi.nlm.nih.gov/rest/pug/substance/name/glucose/sids/XML>
- <https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/cid/2244/sids/JSON>
- <https://pubchem.ncbi.nlm.nih.gov/rest/pug/assay/target/genesymbol/USP2/aids/TXT>
- <https://pubchem.ncbi.nlm.nih.gov/rest/pug/assay/aid/490,1000/targets/ProteinGI,ProteinName,GeneID,GeneSymbol/XML>
- <https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/cid/1000,1001/assaysummary/CSV>
- <https://pubchem.ncbi.nlm.nih.gov/rest/pug/assay/aid/504526/doseresponse/XML>
- <https://pubchem.ncbi.nlm.nih.gov/rest/pug/gene/synonym/EGFR/summary/JSON>
- <https://pubchem.ncbi.nlm.nih.gov/rest/pug/protein/accession/P00533,P01422/summary/JSON>

- PUG REST

- Example

- https://pubchem.ncbi.nlm.nih.gov/rest/pug/pathway/pwacc/Reactome:R-HSA-70171,BioCyc:HUMAN_PWY-4983/summary/JSON
- <https://pubchem.ncbi.nlm.nih.gov/rest/pug/taxonomy/taxid/9606,10090,10116/summary/JSON>
- <https://pubchem.ncbi.nlm.nih.gov/rest/pug/cell/synonym/HeLa/summary/JSON>
- <https://pubchem.ncbi.nlm.nih.gov/rest/pug/substance/sid/1917/classification/XML>
- <https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/cid/2244/dates/JSON>
- <https://pubchem.ncbi.nlm.nih.gov/rest/pug/substance/sourceall/ChEBI/xrefs/RegistryID/JSON>

1. PubChemの使い方を学ぶ
2. ChEMBLの使い方を学ぶ
3. ChEBIの使い方を学ぶ
4. まとめ

ChEMBLとは

- <https://www.ebi.ac.uk/chembl/>
- 運用機関：EMBL's European Bioinformatics Institute (EMBL-EBI, ENG)
- 説明
 - ChEMBLは創薬を目的とした生理活性をもつ化合物や小分子のデータベースです。化合物、核酸などの小分子と、そのターゲットやアッセイ情報を紐付けて収載しています。化合物や小分子の名称、分子式、分子量、標準SMILES、標準InChIなどの情報に加え、結合定数、ADMEデータなど生理活性に関する情報や、薬としての開発段階の情報などが付与されています。化合物を名称や構造で検索したり、ターゲットをタンパク質の名称で検索することができます。

[Integbioデータベースカタログ「ChEMBL」のページ](#)から引用

論文からマニュアルでキュレーションした高品質データベース

How is ChEMBL data curated?



15,072
Targets



2,331,700
Distinct compounds



19,780,369
Activities



85,431
Publications



198
Deposited Datasets

化合物数：2.3M件
(PubChem compound
の約1/50)

ChEMBL – Data for Drug Discovery

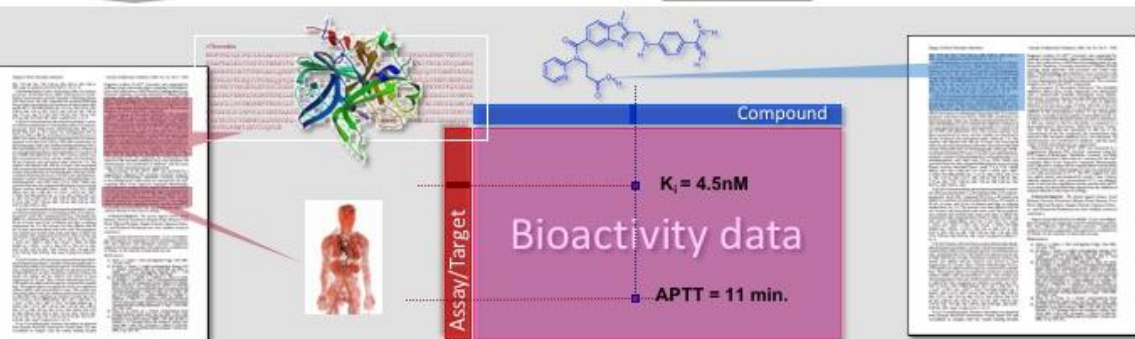
1. Scientific facts



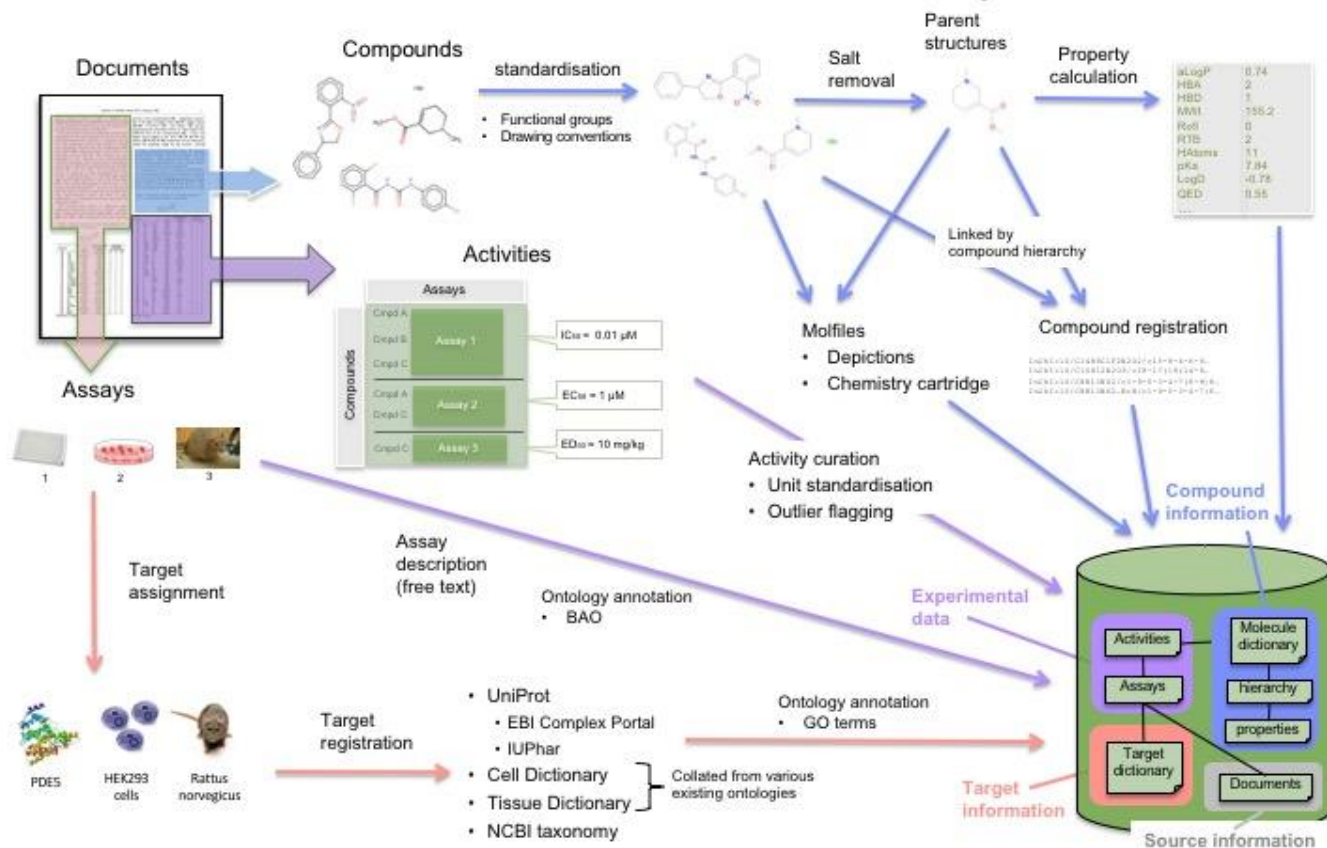
3. Insight, Tools and Resources for Translational Drug Discovery



2. Organization, curation and standardization of pharmacology data



ChEMBL Data and Curation Pipeline



Inclusion of information into the ChEMBL database.

Curation of data included in ChEMBL.

ChEMBLキーワード検索：創薬研究を意識した絞りこみ機能

The screenshot shows the ChEMBL search results for Aspirin. The interface includes a search bar at the top with the keyword 'aspirin' and a search icon. Below the search bar, there are navigation links for UniChem, ChEMBL-NTD, SureChEMBL, Malaria Inhibitor Prediction, Downloads, Web Services, and More. The search results are displayed in a table format, with columns for All Results (423), Compounds (49), Targets (7), Assays (244), Documents (123), Cells (0), and Tissues (0). A 'Filters' panel on the left allows users to refine the search results based on various criteria such as Max Phase, #RO5 Violations, Molecular Weight, AlogP, Level 1 and 2 ATC Code, Targets, and Bioactivities. The main content area displays the chemical structure of Aspirin (labeled 4) and a detailed list of its properties, including ID (CHEMBL25), Name (ASPIRIN), Max Phase (4 Approved), Molecular Formula (C9H8O4), Molecular Weight (180.16), ChEMBL Synonyms, Synonyms From Alternative Forms, Trade Names, and Molecule Type (Small molecule). A 'Rule Of Five: Yes' indicator is shown at the bottom, along with a 'Structure Search' button and a 'Share' button.

① 検索窓からキーワード検索が可能

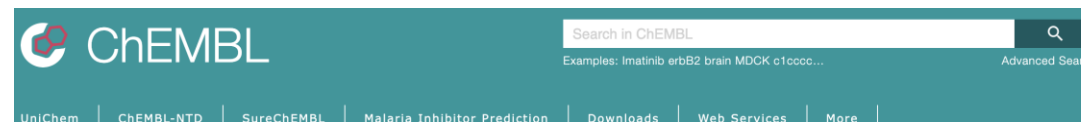
② 検索結果がカテゴリ別に表示されるので、選んで検索結果を確認

③ Filtersで検索結果の絞り込みが可能（例, RO5 Violations, Targets, Bioactivities）

④ 検索結果

⑤ Molecule Featuresを確認（例, Drug Type, Rule Of Five, 経口, 非経口など）

- ① トップ画面の検索窓からキーワード検索が可能
- ② 検索結果がカテゴリ別に表示されるので、選んで検索結果を確認
- ③ Filtersで検索結果の絞り込みが可能（例, RO5 Violations, Targets, Bioactivities）
- ④ 検索結果
- ⑤ Molecule Featuresを確認（例, Drug Type, Rule Of Five, 経口, 非経口など）





• ChEMBL Data Web Services

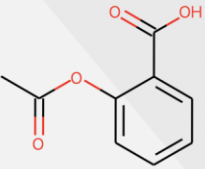
- https://www.ebi.ac.uk/chembl/api/data/molecule?molecule_properties_mw_freebase_lte=300
- <https://www.ebi.ac.uk/chembl/api/data/image/CHEMBL25>
- [https://www.ebi.ac.uk/chembl/api/data/similarity/CN1C\(=O\)C=C\(c2cccc\(Cl\)c2\)c3cc\(ccc13\)\[C@@\]\(N\)\(c4ccc\(Cl\)cc4\)c5cncn5C/80](https://www.ebi.ac.uk/chembl/api/data/similarity/CN1C(=O)C=C(c2cccc(Cl)c2)c3cc(ccc13)[C@@](N)(c4ccc(Cl)cc4)c5cncn5C/80)
- https://www.ebi.ac.uk/chembl/api/data/molecule?max_phase=4
- https://www.ebi.ac.uk/chembl/api/data/molecule?molecule_properties_num_ro5_violations=0
- https://www.ebi.ac.uk/chembl/api/data/molecule?biotherapeutic_isnull=false
- https://www.ebi.ac.uk/chembl/api/data/cell_line?cell_source_tissue_ends_with=carcinoma
- https://www.ebi.ac.uk/chembl/api/data/mechanism?mechanism_of_action_contains=Muscarinic%20acetylcholine%20receptor&action_type=ANTAGONIST
- https://www.ebi.ac.uk/chembl/api/data/target?target_components_accession=Q13936

Connectivity search

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

InChI
 InChIKey
 Source Compound ID
 UniChem Compound ID (UCI)

 DRAW MOL
  SEARCH











Searched Compound

UCI: 161671

InChI: InChI=1S/C9H8O4/c1-6(10)13-8-5-3-2-4-7(8)9(11)12/h...

InChI Key: BSYNRYMUTXBXSQ-UHFFFAOYSA-N

Search

Link to source	Source Compound ID	Source Name	Source ID ↑	Non matching layers
	CHEMBL25	ChEMBL	1	
	DB00945	DrugBank	2	
	AIN	PDBe (Protein Data Bank Europe)	3	
	4139	Guide to Pharmacology	4	
	24714725	PubChem ('Drugs of the Future' subset)	5	
	C01405	KEGG (Kyoto Encyclopedia of Genes and Genomes) Ligand	6	
	15365	ChEBI (Chemical Entities of Biological Interest).	7	
	13719	ChEBI (Chemical Entities of Biological Interest).	7	protonation

- <https://www.ebi.ac.uk/unichem/>

対象データベース

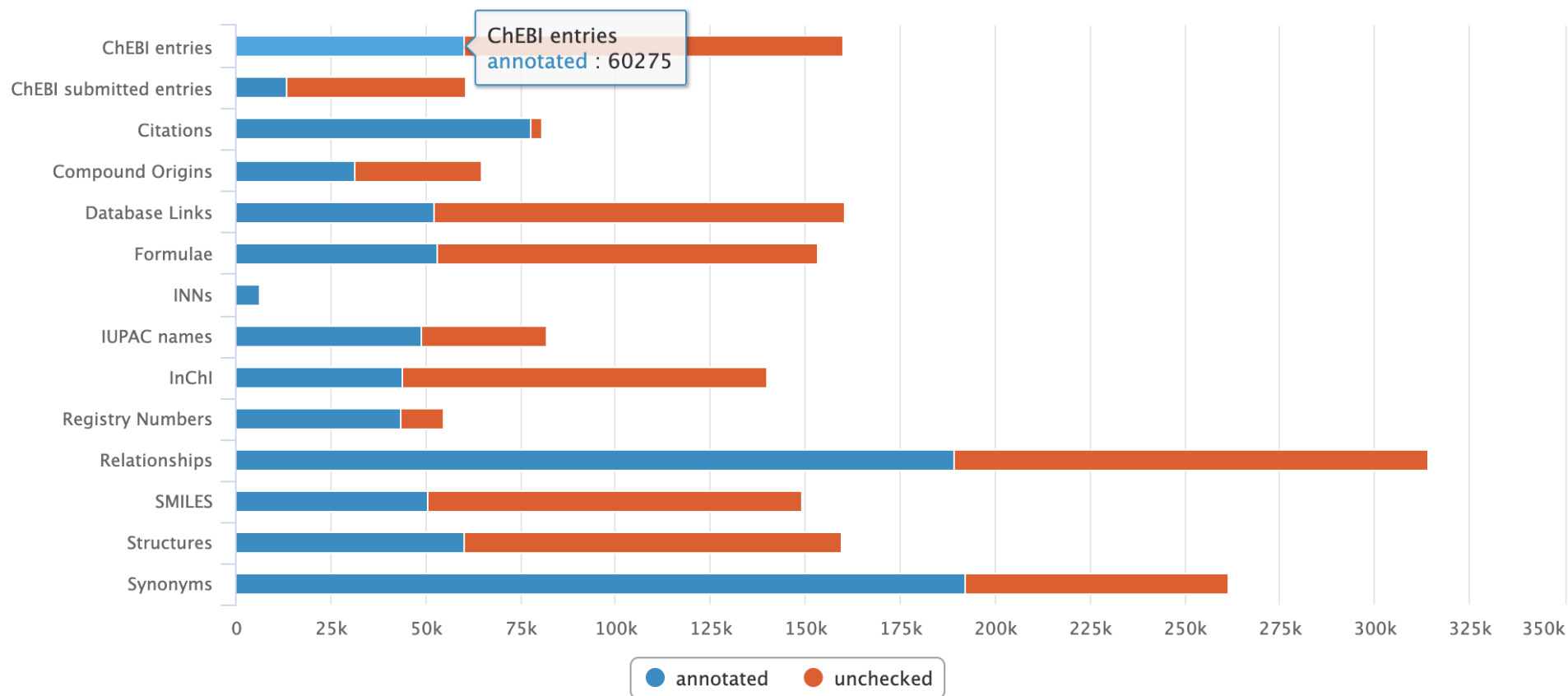
- **ChEMBL**, DrugBank, PDBe, **PubChem**, KEGG, **ChEBI**, Gene Expression Atlas, SureChEMBL, PharmGKB, Human Metabolome Database (HMDB), NIKKaji, BindingDB, LipidMaps, MetaboLights, Brenda, Rhea, SwissLipids, clinicaltrials など40データベース

- ChEMBLにクロスリファレンスの情報を提供

1. PubChemの使い方を学ぶ
2. ChEMBLの使い方を学ぶ
- 3. ChEBIの使い方を学ぶ**
4. まとめ

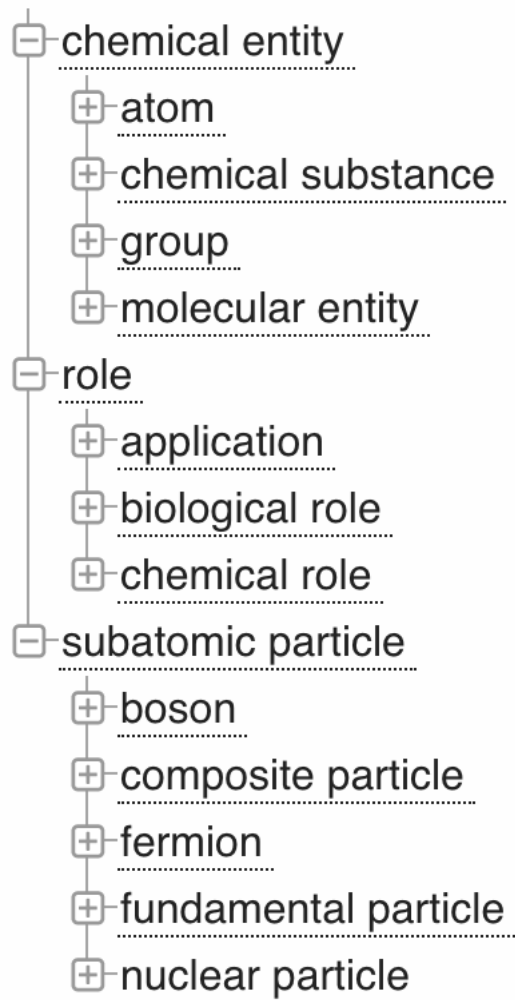
- <http://www.ebi.ac.uk/chebi/>
- 運用機関：EMBL's European Bioinformatics Institute (EMBL-EBI, ENG)
- 説明：
 - 生命プロセスに介在する小分子化合物の辞典です。ゲノムに直接コード化されている分子（例：核酸、タンパク、ペプチド）はこのデータベースの対象外です。登録されているデータは、多数の情報源のデータを組み合わせ、冗長性を除いています。主な情報源はIntEnz(EBI)、KEGG COMPOUND、PDBeChem(EBI)、ChEMBL(EBI)です。ChEBIオントロジー（化合物オントロジー）による分類情報と階層表示が特徴です。
 - 全てのエントリーにはデータの信頼性を示す★印がついています； 3つ星：ChEBIチームが手作業で注釈付けを行ったもの。2つ星：ChEMBLチームまたはデータ提供者が手作業で注釈付けを行ったもの。1つ星：情報源から自動的に取得した仮エントリーで、手作業での注釈付けが行われていないもの。

Status of ChEBI entries



- <https://www.ebi.ac.uk/chebi/statisticsForward.do>

- 化合物数 : 60,275 (PubChem Compoundの約1/2000, ChEMBLの約1/40)



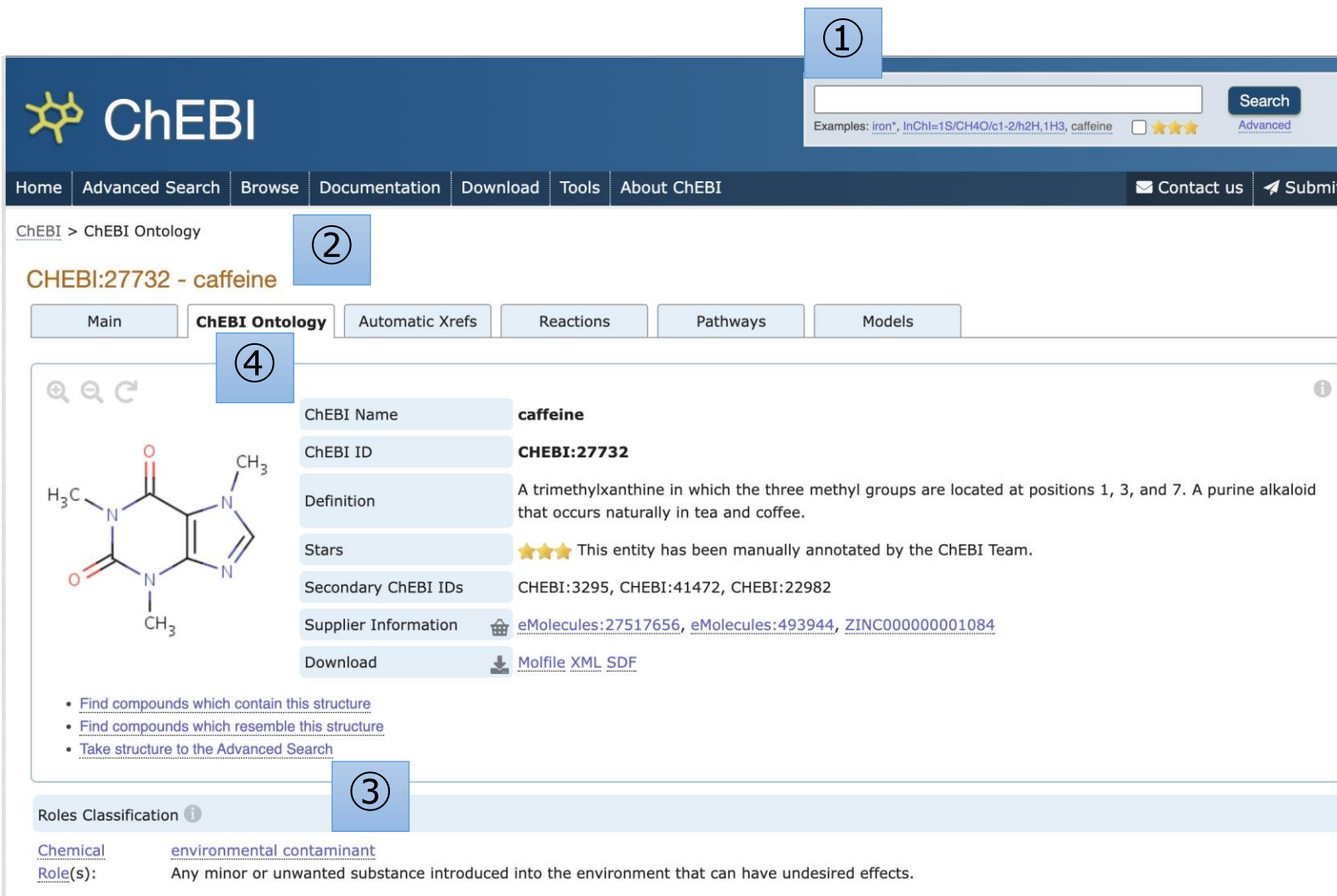
- 化合物の分子構造のツリー ([chemical entity](#), [subatomic particle](#))
 - (例, 有機化合物, アルカロイド, プリン)
- 化合物のロールのツリー ([role](#))
 - [application](#) (例, 医薬品, 化粧品, 染料, 香料, 燃料)
 - [biological role](#) (例, 抗体, 毒物, 阻害剤, 代謝物, アンタゴニスト)
 - [chemical role](#) (例, 酸, 酸化防止剤, 触媒, 供与体)

ChEBIのデータ構造 1/2

- role概念の下位に化合物が置かれる
 - 例, 酵素阻害剤の下位に, [caffeine](#)
 - 例, 植物代謝物の下位に, [caffeine](#)
 - 例, アデノシン受容体アンタゴニストの下位に, caffeine
 - 例, 向精神薬の下位に, [caffeine](#)
 - 例, 利尿剤の下位に, [caffeine](#)



①



ChEBI

Search

Examples: iron*, InChI=1S/CH4O/c1-2/h2H,1H3, caffeine

Home | Advanced Search | Browse | Documentation | Download | Tools | About ChEBI | Contact us | Submit

ChEBI > ChEBI Ontology

CHEBI:27732 - caffeine

Main | ChEBI Ontology | Automatic Xrefs | Reactions | Pathways | Models

④

ChEBI Name: **caffeine**

ChEBI ID: **CHEBI:27732**

Definition: A trimethylxanthine in which the three methyl groups are located at positions 1, 3, and 7. A purine alkaloid that occurs naturally in tea and coffee.

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

Secondary ChEBI IDs: CHEBI:3295, CHEBI:41472, CHEBI:22982

Supplier Information: [eMolecules:27517656](#), [eMolecules:493944](#), [ZINC000000001084](#)

Download: [Molfile](#) [XML](#) [SDF](#)

• [Find compounds which contain this structure](#)

• [Find compounds which resemble this structure](#)

• [Take structure to the Advanced Search](#)

Roles Classification ⓘ

Chemical: [environmental contaminant](#)

Role(s): Any minor or unwanted substance introduced into the environment that can have undesired effects.

① [トップ画面](#)の検索窓からキーワード検索が可能

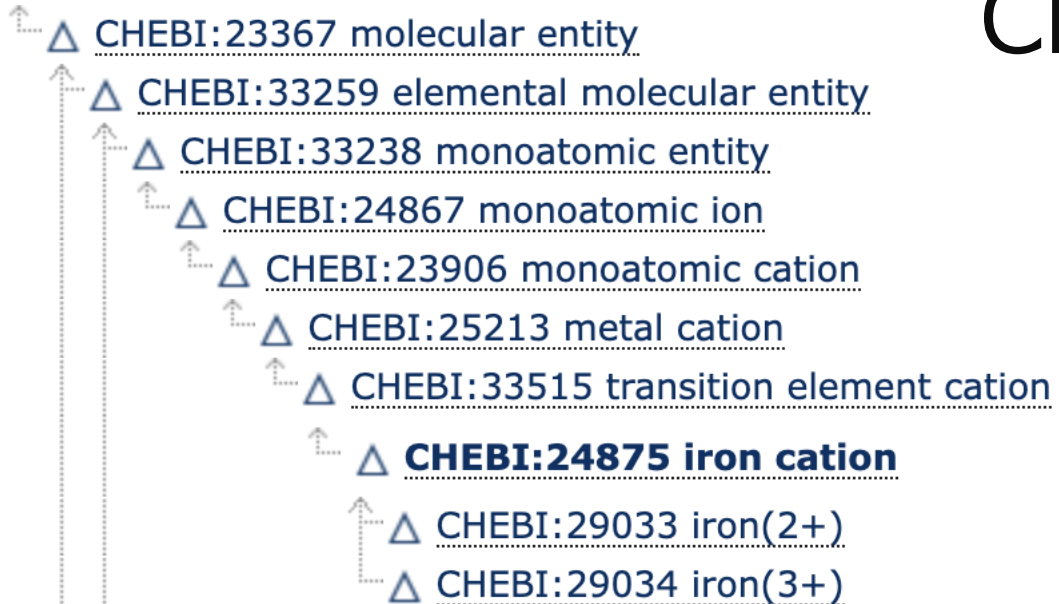
② 検索結果

③ ロールのクラス分類

④ オントロジーツリーを表示する場合は、このタブを選択

ChEBIを使用する利点 1/2

1. 化合物に対して、roleの概念（化合物の性質）の情報をオントロジー構造（木構造）から、推論することができる。
2. 分析機器の解析結果である様々な粒度の化合物名に対して、ChEBIオントロジーの階層を使った化合物アノテーションが有効な場合がある。
 - a. ICP発光分光分析装置 (ICP-OES) を使ったイオノーム分析では、鉄陽イオンを Fe²⁺やFe³⁺の総量として測定している場合がある。この時、鉄陽イオンに対して、iron(2+) ([CHEBI:29033](#))及びiron(3+) ([CHEBI:29034](#))の上位概念である iron cation ([CHEBI:24875](#))を使ってアノテーションすることが可能である。
 - b. 質量分析の結果で、詳細な分子構造（例、構造異性体）や立体構造（例、鏡像異性体）が不明な化合物に対して、オントロジーの上位概念に用いてアノテーションをすることが可能である（例、phosphatidylcholine 32:1 ([CHEBI:66849](#)））。
- データ駆動型材料開発～オントロジーとマイニング、計測と実験装置の自動制御～、監修:船津公人, 第2章 材料データベースとオントロジー, 2.2.3.オープン時代における化合物データベースとオントロジー, 櫛田達矢著, (株)エヌ・ティー・エス, [ISBN:978-4-86043-759-6 C3058](#)



2.a



2.b

1. PubChemの使い方を学ぶ
2. ChEMBLの使い方を学ぶ
3. ChEBIの使い方を学ぶ
4. まとめ

PubChem, ChEMBL, ChEBI比較

	運用機関	化合物数	統計量のページ	ライセンス
PubChem	NCBI	Compounds: 112,405,915 Substances: 298,292,628	https://pubchemdocs.ncbi.nlm.nih.gov/statistics	https://fairsharing.org/FAIRsharing.qt3w7z を参照
ChEMBL	EBI	2,331,700	https://www.ebi.ac.uk/chembl/	CC BY-SA 3.0
ChEBI	EBI	60,275	https://www.ebi.ac.uk/chebi/statisticsForward.do	CC BY 4.0

2022.11.12 確認

KEGG COMPOUND: C07481 (Caffeine)

で見た外部リンクの状況

	データセット	KEGG COMPOUNDとのダイレクトリンク	ID	Standard InChI
KEGG	COMPOUND	-	C07481	InChI=1S/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3
	DRUG	あり	D00528	InChI=1S/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3
PubChem	PubChem substance	あり	SID 9684 *	ND
	PubChem compound	なし	CID 2519 +	InChI=1S/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3
ChEMBL		あり	CHEMBL113	InChI=1S/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3
ChEBI		あり	CHEBI:27732	InChI=1S/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3

* と + 間にはダイレクトリンクあり

参考資料

- **バイオDBとウェブツール ラボで使える最新70選** 知る・学ぶ・使う、バイオDX時代の羅針盤, 小野浩雅／編, 実験医学増刊 Vol.40 No.17, [ISBN 978-4-7581-0406-7](https://doi.org/10.1002/exmi.1417), 2022年10月20日発行
 - **PubChem** — 化合物とそれに関与する遺伝子, パスウェイなどの情報をフリーで提供【櫛田達矢】
 - **ChEMBL** — 手動でキュレーションされた生物活性分子のデータベース【山田一作】
- 統合TV
 - [PubChem を使って化合物の情報を検索する](#) (2021.12.14)
 - [化合物データベース](#) (2018.12.14)



謝辞

- この資料は、山田一作 博士（公益財団法人 野口研究所）の講習会 [AJACS筑波4「化合物データベース」](#)（2018年7月10日）の [講義資料](#) を参考に作成しました。山田博士に感謝いたします。