



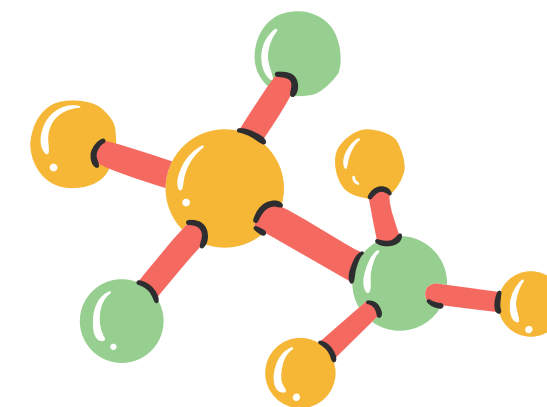
Universidad Nacional Autónoma de México  
FACULTAD DE QUÍMICA  
Departamento de Farmacia

# IX SIMPOSIO

**Tendencias actuales en la búsqueda  
y desarrollo de fármacos**

Talleres pre-simposio  
**Herramientas quimioinformáticas  
para el diseño de fármacos**

12 y 13 de junio de 20123





# **BASES DE DATOS**

## **PubChem**

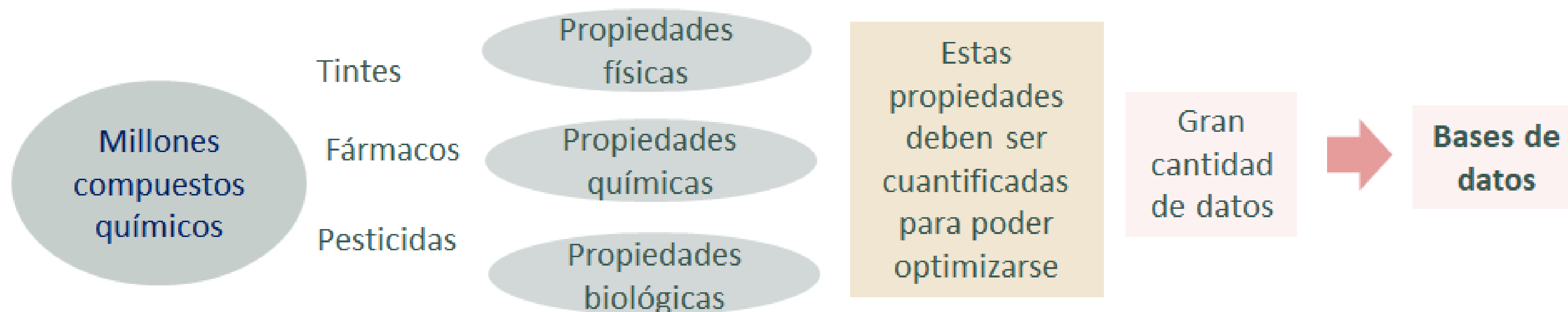


# Objetivos

- Familiarizarse con el tipo de información disponible en bases de datos públicas y los diferentes tipos de búsqueda que pueden realizarse, de tal manera que la búsqueda de información se realice en forma sencilla y eficiente
- Utilizar PUG-REST para acceder a información disponible en PubChem mediante programación.
- Ejemplificar la adquisición de diferente tipo de información disponible en PubChem.

# Base de datos

Estructura organizada de almacenamiento de información, normalmente asociada a un programa computacional.



Esta información no debe ser ignorada ni tratada como resultados aislados e independientes, ya que puede conducir a obtener conclusiones valiosas en muchas áreas de la química, en especial en el diseño de fármacos.



# Base de datos

- **Públicas**
- Institucionales / grupo (*in-house*)
  - Bibliotecas de química combinatoria
- Comerciales (MDDR, CRC Press, Prod. Nat)
- *De novo*: virtuales
  - Diseño de bibliotecas



# Base de datos públicas

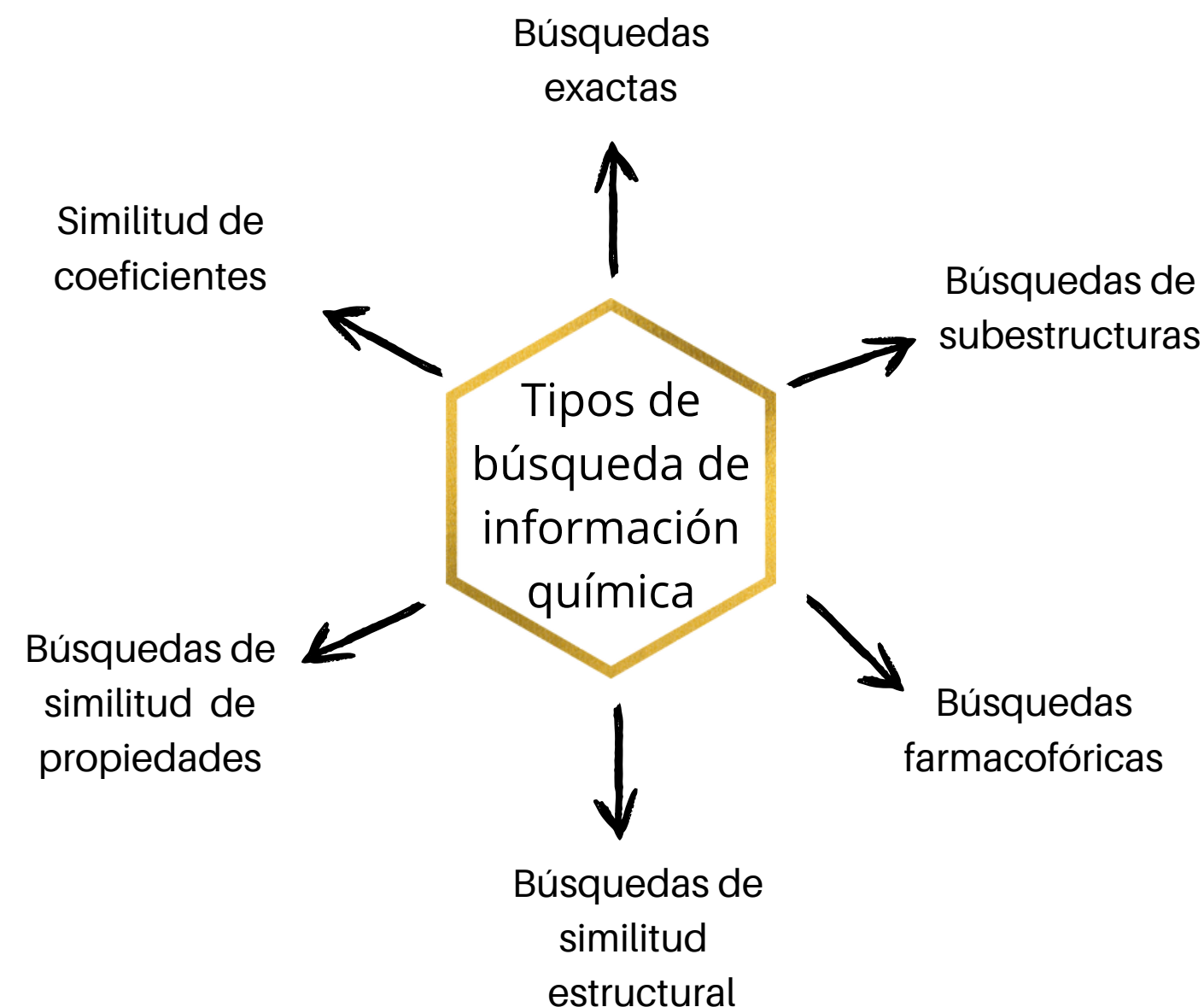
Nombre	Información	Información biológica	Link
ChEMBL	Búsqueda por <b>compuesto, blanco, ensayo, indicación terapéutica.</b>	Si	<a href="http://www.ebi.ac.uk/chembl">www.ebi.ac.uk/chembl</a>
Binding Database	Búsqueda por <b>compuesto, blanco o artículo.</b> Base de datos enfocada en las interacciones de moléculas pequeñas con proteínas que pueden ser blancos farmacológicos.	Si	<a href="http://www.bindingdb.org">www.bindingdb.org</a>
PubChem	Búsqueda por <b>compuesto, blanco, patente o ensayo.</b> Sustancias, estructuras y datos de bioactividad.	Si	<a href="https://pubchem.ncbi.nlm.nih.gov/">https://pubchem.ncbi.nlm.nih.gov/</a>

# Base de datos públicas

Nombre	Información	Información biológica	Link
DrugBank	Búsqueda por <b>compuesto, mecanismo de acción, indicación terapéutica y blanco</b> . Fármacos aprobados por la FDA (moléculas pequeñas y biotecnológicos), nutraceuticos y fármacos en fase experimental con datos farmacológicos, ADME-Tox.	No	<a href="http://www.drugbank.ca/">http://www.drugbank.ca/</a>
ChemSpider	Búsqueda por <b>compuesto</b> . Estructuras químicas de compuestos de interés y la predicción de muchos parámetros físico-químicos.	No	<a href="http://www.chemspider.com/">http://www.chemspider.com/</a>
ZINC	Búsqueda por <b>compuesto, blanco, tipo de base de datos</b> Más de 13 millones de compuestos disponibles comercialmente. Pueden usarse para cribado virtual.	No	<a href="http://zinc.docking.org/">http://zinc.docking.org/</a>
Chemical Universe Database (GBD)	166.4 billones de moléculas de hasta 17 átomos incluyendo C,N,O,S y halógenos formando una base de datos de todos los compuestos en el espacio químico	No	<a href="http://www.gdb.unibe.ch/">http://www.gdb.unibe.ch/</a>

# Minería de datos

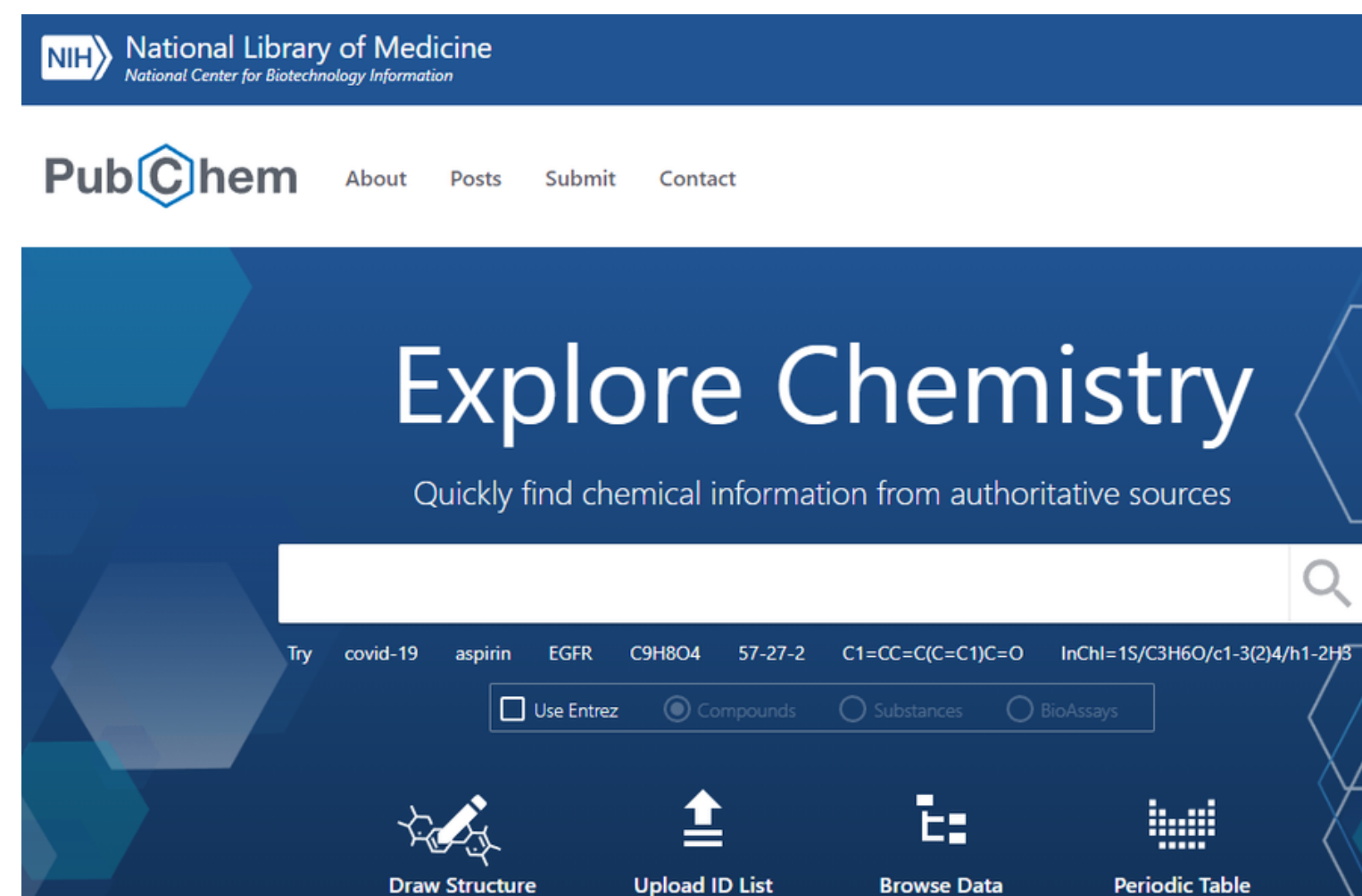
- Extracción de información de bases de datos
- Ayuda a comprender el contenido de un repositorio de datos e identificar relaciones numéricas, patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en determinado contexto.





# PubChem

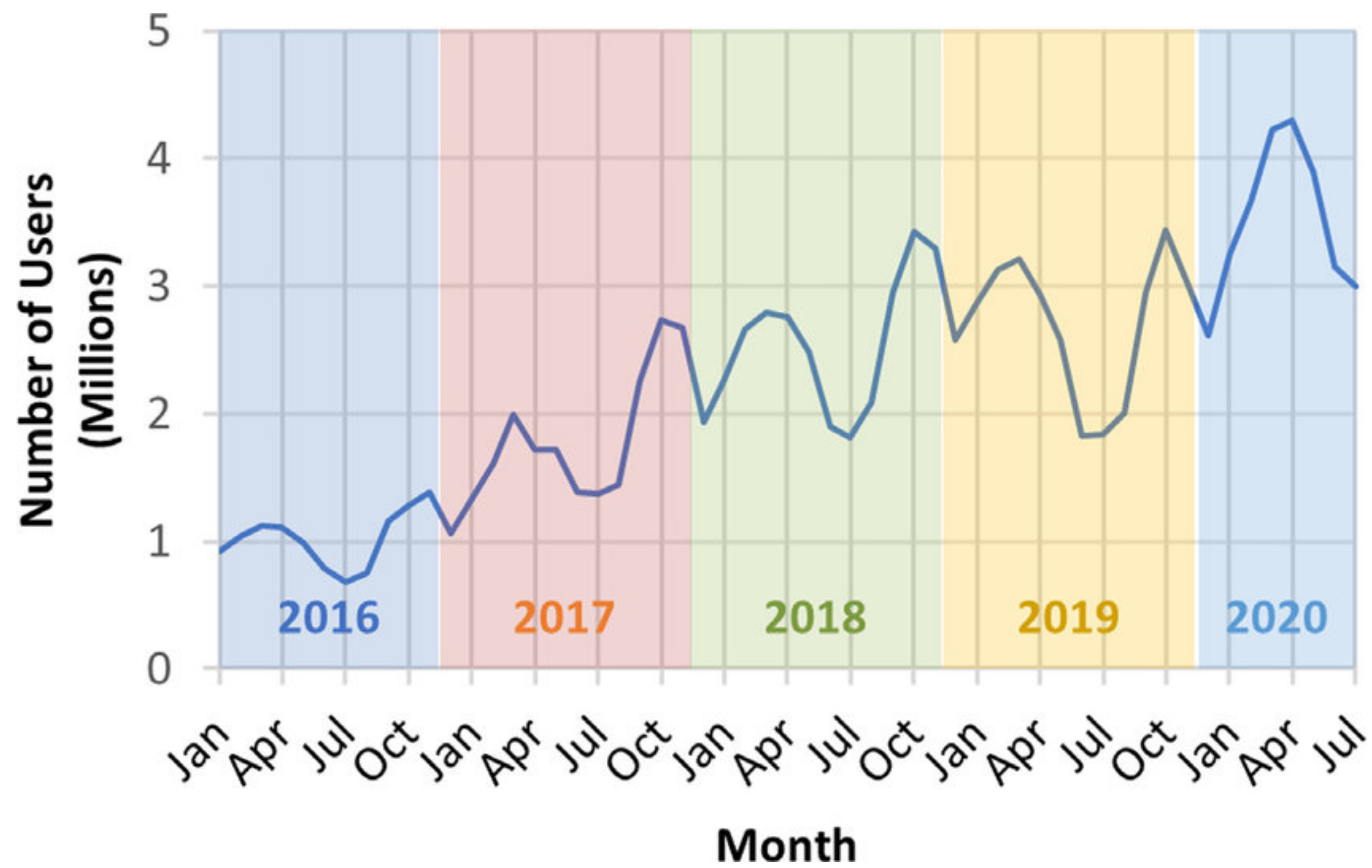
- Desarrollada y mantenida por Instituto Nacional de Salud de los Estados Unidos (NIH)
- Contiene varios tipos de entidades químicas:
  - Moléculas pequeñas
  - siRNAs y miRNAs
  - Carbohidratos
  - Lípidos
  - Péptidos



# PubChem

- Contiene información química de más de 760 recursos
- 111 millones de estructuras de compuestos únicos
- Conectada con otras bases de datos
- Herramientas de búsqueda, análisis, descarga y visualización.
- Recurso clave en muchas áreas
  - Quimioinformática
  - Química farmacéutica
  - Diseño de fármacos

# PubChem



# PubChem

## Explorar información química en PubChem

1. Búsqueda por nombre químico
2. Búsqueda por estructura química
3. Búsqueda por nombre del gen/proteína/ruta bioquímica
4. Tabla periódica y elementos de PubChem
5. Acceso mediante programación

Todas las búsquedas hechas en el servidor de PubChem pueden ser automatizadas con el acceso mediante programación

## Quimioinformática en PubChem

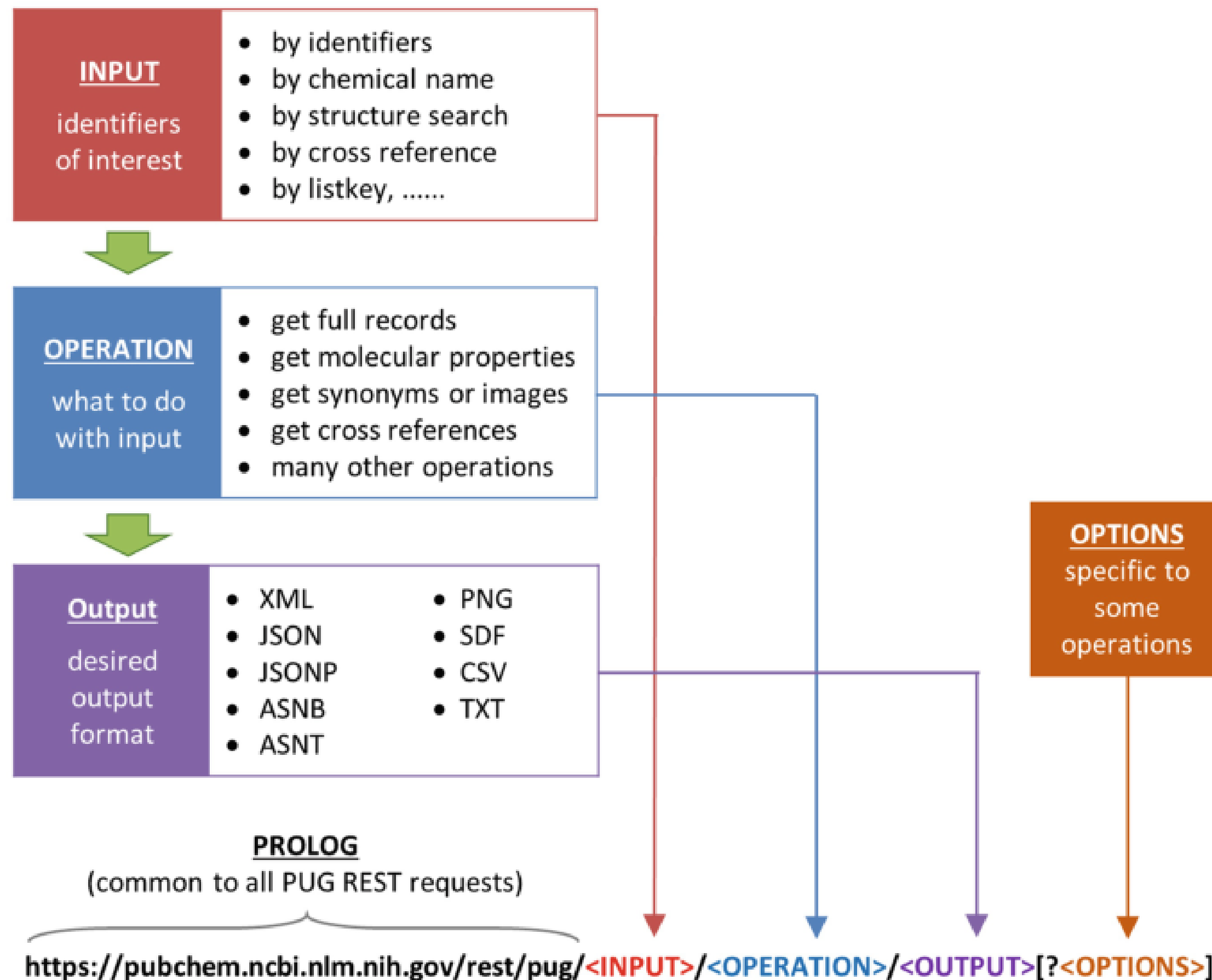
- Evaluación crítica de información química
- Representaciones químicas (InChI, SMILES, etc.)
  - Alternativas a búsquedas por nombre químico
  - Compartir/intercambiar/integrar datos químicos
- Búsqueda por nombre químico
- Búsqueda por estructura química
  - Identidad
  - Por similitud 2D/3D
  - Subestructura/superestructura
  - Fórmula molecular
- Análisis de relaciones estructura-actividad
- Automatización de recuperación de información química mediante código

# PubChem

## Acceso a PubChem mediante programación

- Rutas de acceso múltiples
- Dos tipos de acceso:
  - **PUG-REST** (para propiedades calculadas)  
<https://pubchem.ncbi.nlm.nih.gov/docs/pug-rest>
  - **PUG-View** (para información en texto)  
<https://pubchem.ncbi.nlm.nih.gov/docs/pug-view>
- Cuadernos de código (Jupyter) con código de ejemplo (Python/R) de libre acceso: <https://chem.libretexts.org/link?143689>

# PubChem



[Example]

`https://pubchem.ncbi.nlm.nih.gov/rest/pug/compound/cid/853/record/XML?record_type=3d`

# PubChem

## Materialos disponibles en:

- Sitio web del *Committee on Computers in Chemical Education* (CCCE)  
<http://olcc.ccce.divched.org/Fall2015OLCC>
- LibreTexts <https://libretexts.org/>

OLCC	Sitio web
2015	<a href="http://olcc.ccce.divched.org/Fall2015OLCC">http://olcc.ccce.divched.org/Fall2015OLCC</a> <a href="https://chem.libretexts.org/link?50589">https://chem.libretexts.org/link?50589</a>
2017	<a href="http://olcc.ccce.divched.org/Spring2017OLCC">http://olcc.ccce.divched.org/Spring2017OLCC</a> <a href="https://chem.libretexts.org/link?83678">https://chem.libretexts.org/link?83678</a>
2019	<a href="https://chem.libretexts.org/link?143689">https://chem.libretexts.org/link?143689</a>

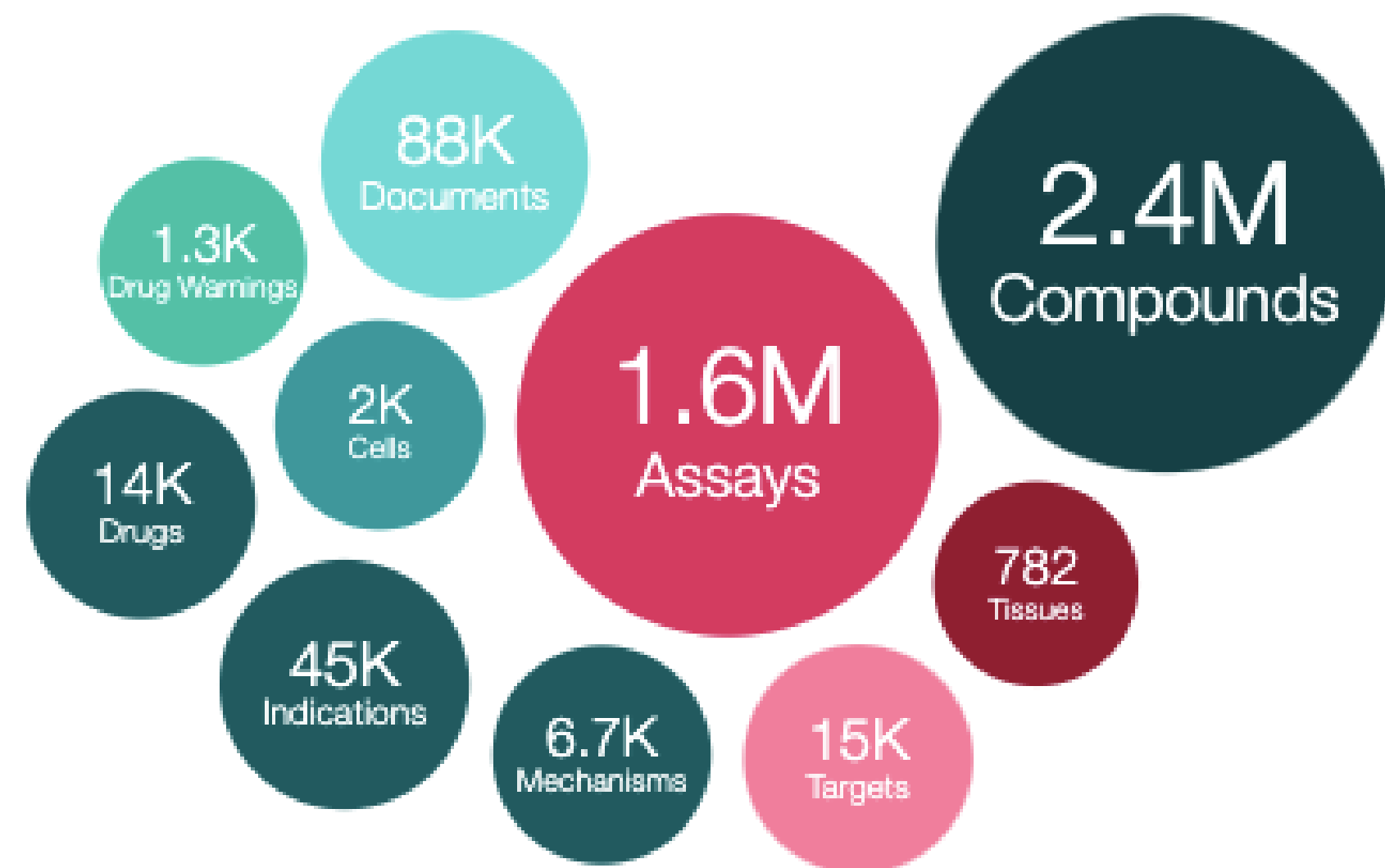
# Base de datos públicas

Nombre	Información	Información biológica	Link
ChEMBL	Búsqueda por <b>compuesto, blanco, ensayo, indicación terapéutica.</b>	Si	<a href="http://www.ebi.ac.uk/chembl">www.ebi.ac.uk/chembl</a>
Binding Database	Búsqueda por <b>compuesto, blanco o artículo.</b> Base de datos enfocada en las interacciones de moléculas pequeñas con proteínas que pueden ser blancos farmacológicos.	Si	<a href="http://www.bindingdb.org">www.bindingdb.org</a>
PubChem	Búsqueda por <b>compuesto, blanco, patente o ensayo.</b> Sustancias, estructuras y datos de bioactividad.	Si	<a href="https://pubchem.ncbi.nlm.nih.gov/">https://pubchem.ncbi.nlm.nih.gov/</a>



# ChEMBL

- Desarrollado por el Instituto Europeo de Bioinformática (EBI)
- Base de datos moleculares curada manualmente.
- Actualizada regularmente y es de libre acceso
- Tipo de información:
  - Dianas moleculares
  - Moléculas bioactivas



Current Release: ChEMBL 33

Provided under a [Creative Commons Attribution-ShareAlike 3.0 Unported license](#)  
Last Update on 2023-05-31T00:00:00 | [Release notes](#)



15,398

Targets



2,399,743

Distinct compounds



20,334,684

Activities



88,630

Publications



215

Deposited Datasets

[Contact Us](#)

*Expert opinion on Drug Design. 2017;12(08):1-11*

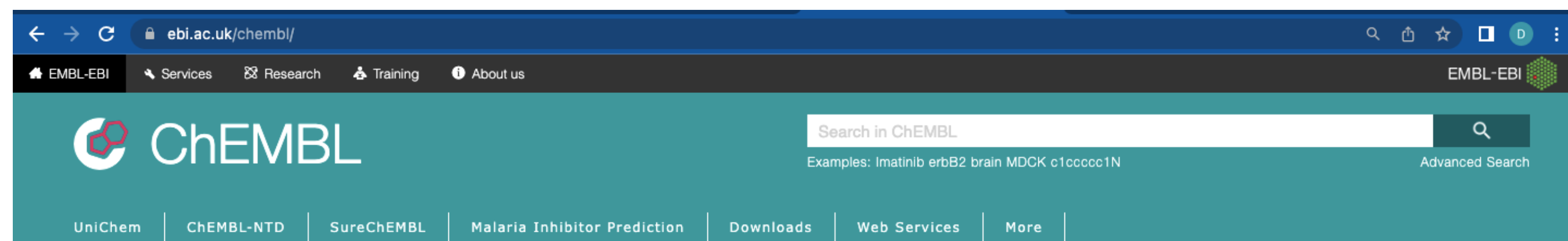
- Su estructura permite realizar:
  - Minería de datos relacional
  - Cribado virtual
  - Modelado predictivo
  - Entre otras técnicas en el campo del descubrimiento de fármacos en quimioinformática

# ChEMBL

Servidor web proporciona acceso a un amplio rango de datos sobre:

- Moléculas
- Blancos
- Mecanismos de acción
- Mediciones de bioactividad
- Documentos.
- Tejidos
- Células

En total hay 34 puntos finales distintos.

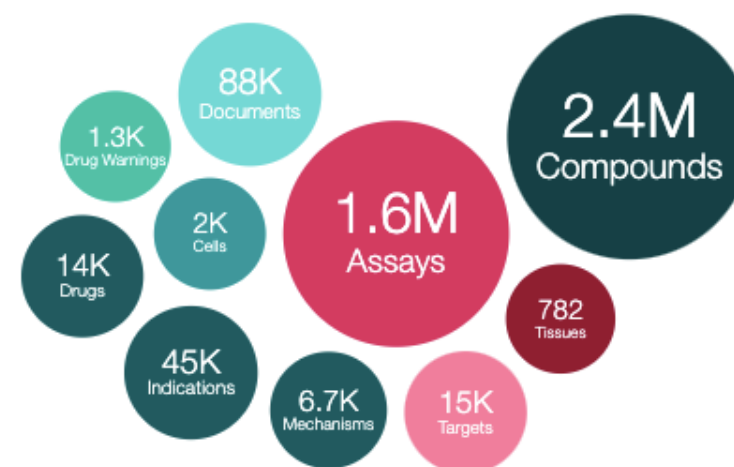


ChEMBL is a manually curated database of bioactive molecules with drug-like properties. It brings together chemical, bioactivity and genomic data to aid the translation of genomic information into effective new drugs.

## Explore ChEMBL

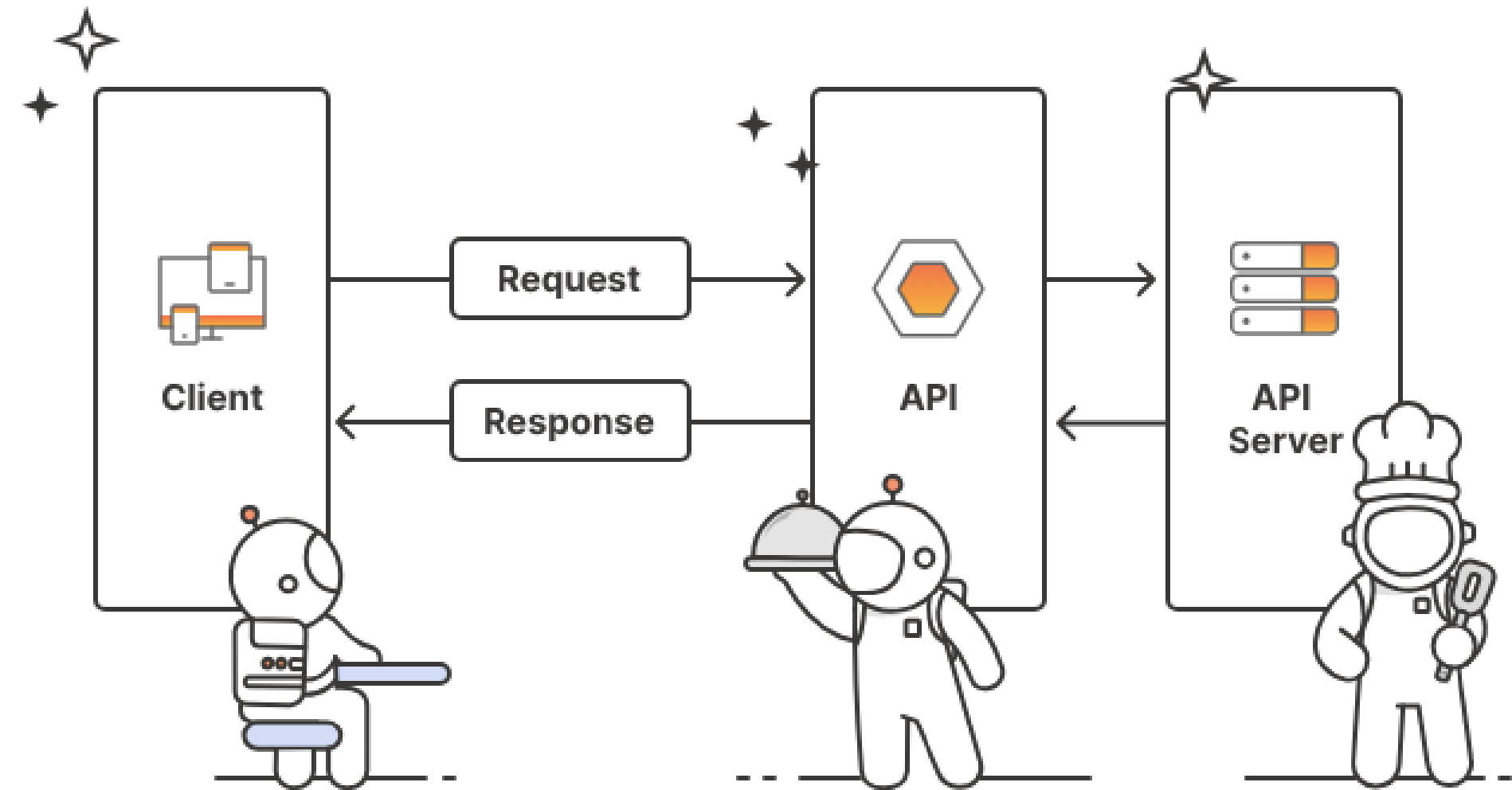
**Description:** Shows a summary of the ChEMBL entities and quantities of data for each of them.

**Instructions:** Click on a bubble to explore a specific ChEMBL entity in more detail.



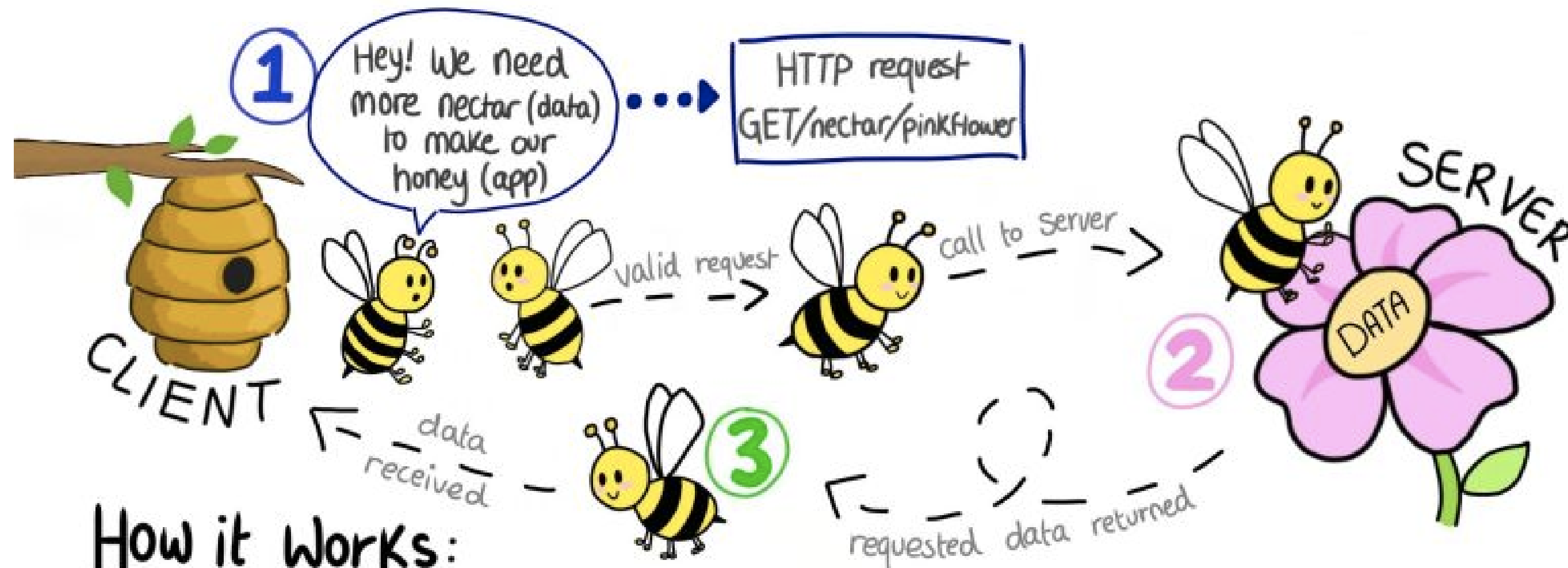
# API

- Interfaz de programación de aplicación.
- Conjunto de reglas y protocolos que permite que distintas aplicaciones se comuniquen y compartan información entre sí.
- Son como puentes que permiten que la información fluya entre sistemas y aplicaciones.



# What is an API?

An application programming interface allows two programs to communicate. On the web, APIs sit between an application and a web server, and facilitate the transfer of data.



## How it Works:

**1 Request**  
API call is initiated by the Client application via a HTTP request

**2 Receive**  
Our worker bee acts as an API, going to a Flower (server) to collect nectar (data)

**3 Response**  
The API transfers the requested data back to the requesting application, usually in JSON format

[https://www.ebi.ac.uk/chembl/g/#search\\_results/all/  
query=aspirina](https://www.ebi.ac.uk/chembl/g/#search_results/all/query=aspirina)

ChEMBL aspirina

All Results 448 Compounds 55 Targets 0 Assays 252 Documents 141 Cells 0 Tissues 0

## Compounds

Show Full Query

55 Compounds  
0 Selected - Select All  
Browse Activities

Table Cards Graph Heatmap

Filters

Type

- Small molecule 53
- Unknown 2

Max Phase

- Predclinical 50
- Approved 3
- Phase 3 1

Records per page: 24 Select All

Showing 1-24 out of 55 records

CHEMBL25  
Name: ASPIRIN  
Max Phase: 4.0  
Full Mwt: 180.16  
Alogp: 1.31

CHEMBL1697753  
Name: ASPIRIN DL-LYSINE  
Max Phase: 3.0  
Full Mwt: 326.35  
Alogp: 1.31

CHEMBL1900528  
Name: TIOPIRUM  
Max Phase: 4.0  
Full Mwt: 392.52  
Alogp: 2.35

[https://www.ebi.ac.uk/chembl/compound\\_report\\_car  
d/CHEMBL25/](https://www.ebi.ac.uk/chembl/compound_report_card/CHEMBL25/)

Structure Search

Chemical structure of Aspirin (CHEMBL25): CC(=O)Oc1ccccc1C(=O)O

ID: CHEMBL25  
Name: ASPIRIN  
Max Phase: Approved  
Molecular Formula: C9H8O4  
Molecular Weight: 180.16

CHEMBL Synonyms:

- ACETYL SALICYLATE, Acetylsalicylic Acid, ACETYL SALICYLIC ACID, ACETYL SALICYLIC ACID (WHO-IP), ACETYL SALICYLIC ACID ACETUM, ACIDUM ACETYL SALICYLICUM, ACIDUM ACETYL SALICYLICUM (WHO-IP), Aspirin, ASPIRIN, (S)-2-(4-oxo-3-phenylbut-3-en-1-yl)benzoic acid, NSC-27323, NSC-485186, ACETYL SALICYLATE LYSINE, ACETYL SALICYLIC ACID LYSINATE, ASPROIC, ASPROICOL, ASPIRIN DL-LYSINE, ASPIRIN LYSINE, ASPIRIN LYSINE, DL-LYSINE ACETYL SALICYLATE, DL-LYSINE ACETYL SALICYLATE, EDICALIN, FLECTADOL, LASPIL, L-LYSINE ACETYL SALICYLATE, Lysine Acetylsalicylate, LYSINE ACETYL SALICYLATE, SOLPILIN, VIKOPILIN, VITALGINE

Synonyms From Alternative Forms:

- 8-HOUR Bayer, Acetylsalicylic Acid, ACETYL SALICYLIC ACID, Acetylsalicylic Acid, ALKA-RAPID, ANADIN ALL NIGHT, ANGETTES 75, ASPIRIN, ASPIRIN CLR, Bayer Extra Strength Aspirin for Migraine Pain, DAVANOL, DESPILIN CY, DESPILIN DIRECT, DURLAZA, Easpirin, ENPILIN, Equi-Pain, GONCAREX, LYSOL, MAX STRENGTH ASPIRIN CLR, MEASURIN, PICOPIRIN 80, RU-SEALS 300, RU-SEALS 600, RU-SEALS 75, RU-SEALS CARDIO 75, RYNOCEL, PLATET, PLATET 330, POSTHE 300, POSTHE 75, Salicylic Acid Acetate, VIKALORE

Trade Names:

Molecule Type: Small molecule

Name And Classification

- Representations
- Sequences
- Alternative Forms
- Molecule Features
- Drug Indications
- Drug Mechanisms
- Similar Compounds
- Metabolism
- Activity Charts
- Literature
- Target Predictions
- Calculated Properties
- Cross References
- UniChem Cross References
- UniChem Connectivity Layer Cross References

`new_client.molecule.filter(pref_name__iexact='aspirin')`

Cantidad de entradas asociadas a la etiqueta de 'aspirina': (2, 34).  
Las columnas presentes en la tabla generada son: Index(['atc\_classifications', 'availability\_type', 'biotherapeutic', 'black\_box\_warning', 'chebi\_par\_id', 'chirality', 'cross\_references', 'dosed\_ingredient', 'first\_approval', 'first\_in\_class', 'helm\_notation', 'indication\_class', 'inorganic\_flag', 'max\_phase', 'molecule\_chembl\_id', 'molecule\_hierarchy', 'molecule\_properties', 'molecule\_structures', 'molecule\_synonyms', 'molecule\_type', 'natural\_product', 'oral', 'parenteral', 'polymer\_flag', 'pref\_name', 'prodrug', 'structure\_type', 'therapeutic\_flag', 'topical', 'usan\_aten', 'usan\_aten\_definition', 'usan\_aten\_text', 'usan\_year', 'withdrawn\_flag'], dtype='object')

	atc_classifications	availability_type	biotherapeutic	black_box_warning	chebi_par_id	chirality	cross_references	dosed_ingredient	first_approval	first_in_class	...
0	[B01AC06, N02BA01, N02BA51, A01AD05, N02BA71]	2	None	0	15366	2	[{'xref_id': 'aspirin', 'xref_name': 'aspirin'...}]	True	1950	0	...
1	[B01AC06, N02BA01, N02BA51, A01AD05, N02BA71]	2	None	0	15366	2	[{'xref_id': 'aspirin', 'xref_name': 'aspirin'...}]	True	1950	0	...

2 rows x 34 columns

`new_client.molecule..filter(chembl_id='CHEMBL25').only(['molecule_chembl_id', 'pref_name', 'molecule_structures'])`

	molecule_chembl_id	molecule_structures	pref_name
0	CHEMBL25	{'canonical_smiles': 'CC(=O)Oc1ccccc1C(=O)O', ...}	ASPIRIN
1	CHEMBL25	{'canonical_smiles': 'CC(=O)Oc1ccccc1C(=O)O', ...}	ASPIRIN

¡Gracias por la atención!

