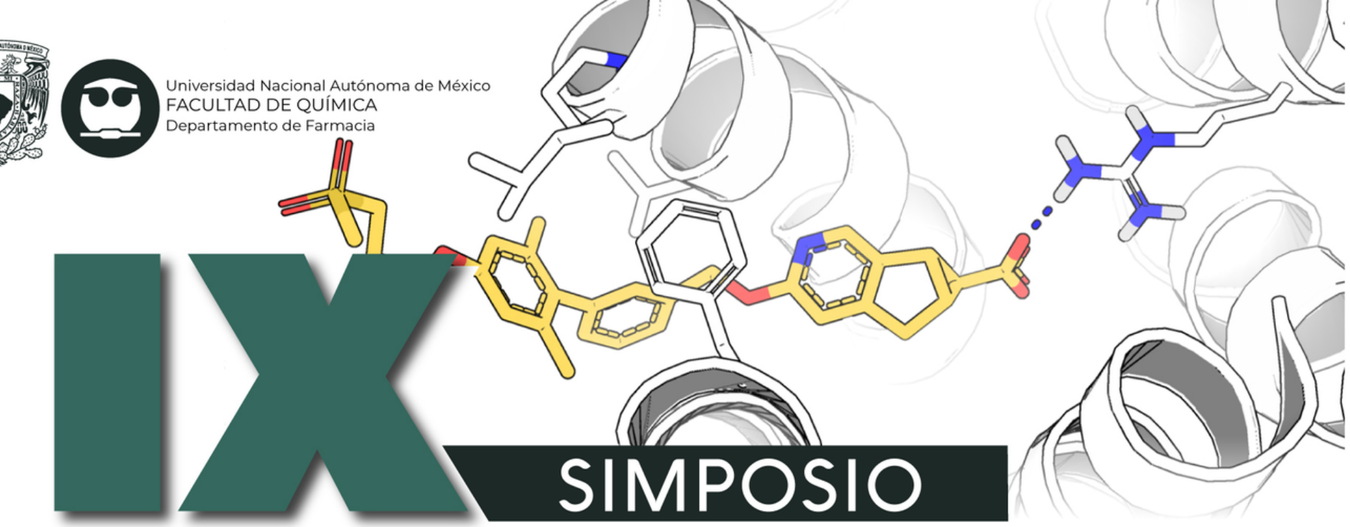
Universidad Nacional Autónoma de México
FACULTAD DE QUÍMICA
Departamento de Farmacia



IX SIMPOSIO

Tendencias actuales en la búsqueda y desarrollo de fármacos

Talleres pre-simposio

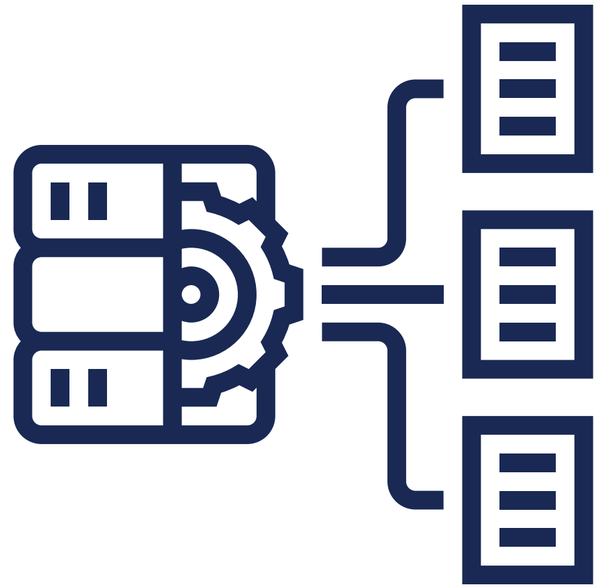
Herramientas quimioinformáticas para el diseño de fármacos

12 y 13 de junio de 2023

ANÁLISIS EXPLORATORIO DE BASES DE DATOS

CONTENIDO

- Datos a información
- Descriptores y núcleos base
- Análisis exploratorio de datos
- Análisis de datos moleculares
- Aplicación
- Herramientas quimioinformáticas



Objetivos

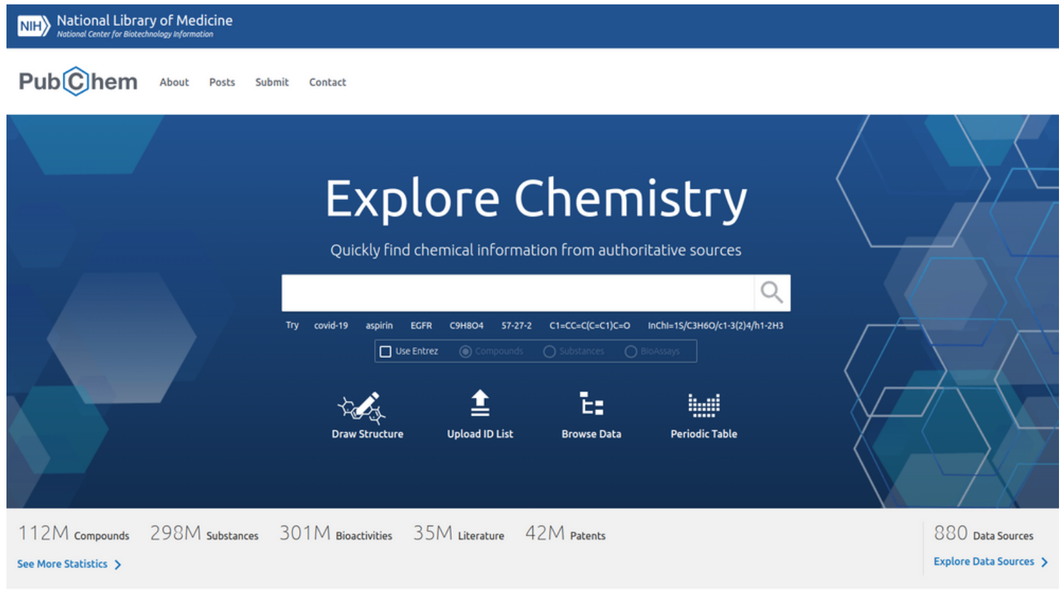
1) Introducir a la visualización y al análisis de datos químicos.

2) Graficar histogramas, *boxplots* y *violinplots* para analizar propiedades fisicoquímicas de importancia farmacéutica.

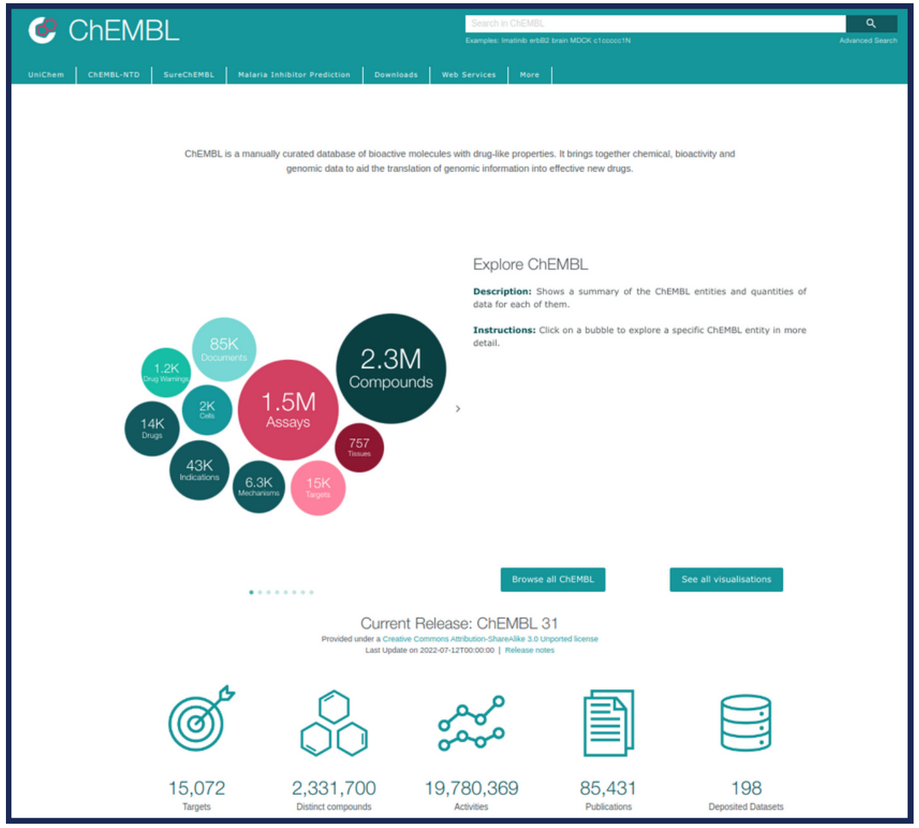
3) Identificar posibles correlaciones entre variables.

Base de datos

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



https://pubchem.ncbi.nlm.nih.gov/



https://www.ebi.ac.uk/chembl/

De los datos al conocimiento

Actividad biológica

- Ensayos *in vitro*, *in vivo*.
- Cribado de alto rendimiento (HTS).

Datos

Cálculo de descriptores moleculares

- Propiedades.
- Huellas digitales moleculares.

Información

Relaciones estructura-actividad

Organización y contextualización de los datos..

Conocimiento

Interpretación y generalización de los modelos

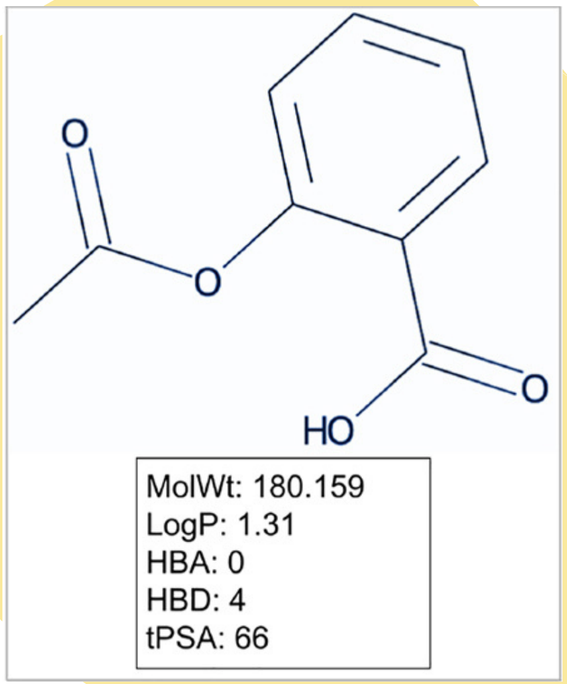
Aplicaciones quimioinformáticas

- Diseño de nuevas moléculas con propiedades de interés.
- Fármacos.
- Materiales, polímeros.
- Moléculas para la industria alimentaria, fragrancias, etc.

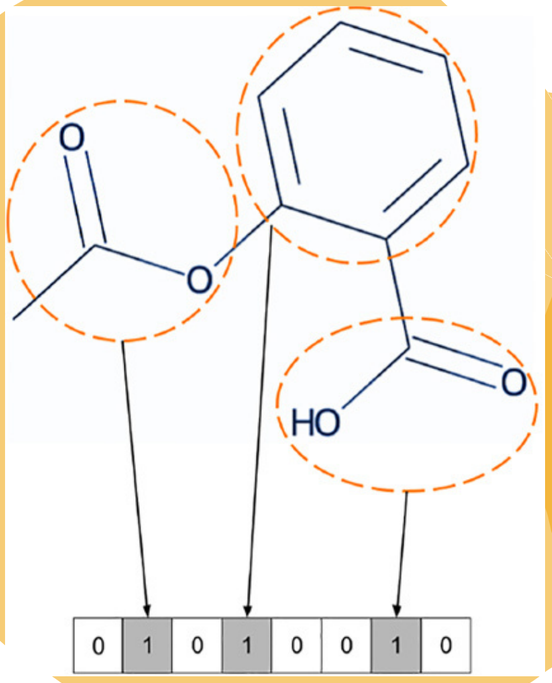
Descriptores

Características y propiedades de las moléculas.

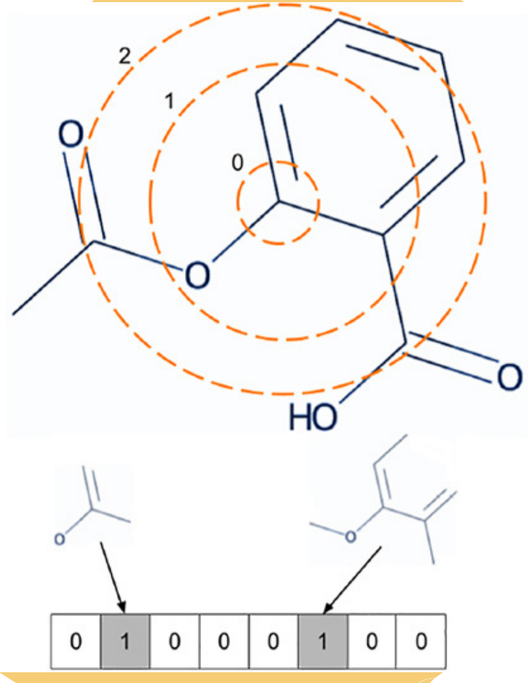
Propiedades



Huellas digitales



Diccionario



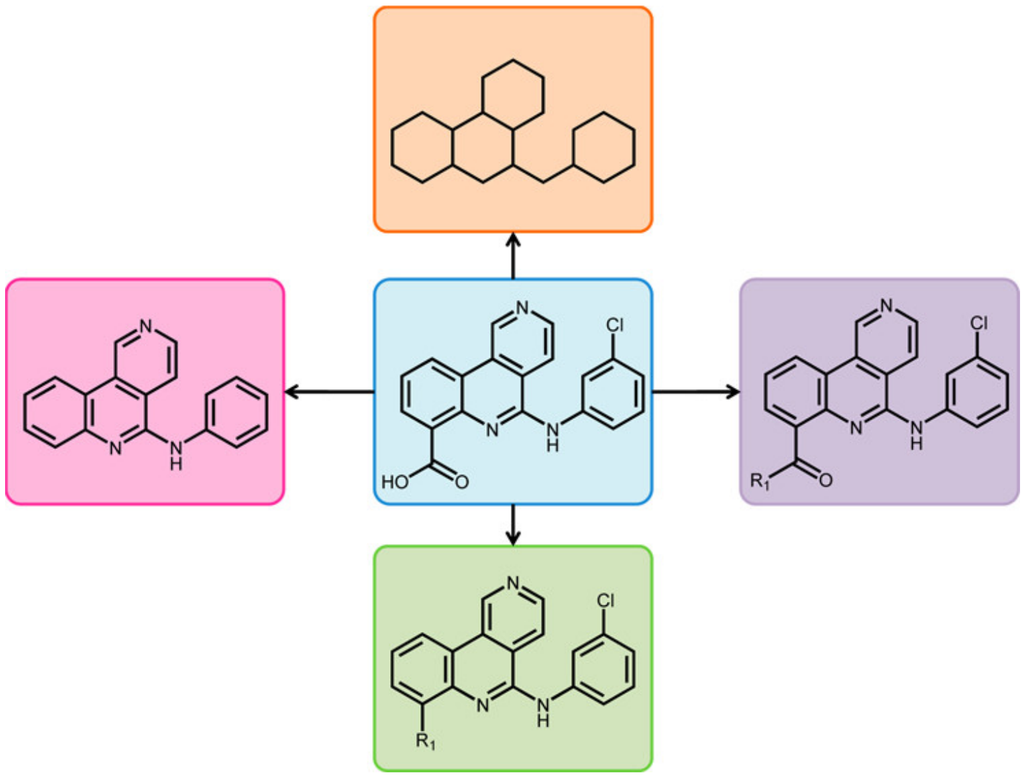
Circular

J. Integr. Bioinform. 2022, 19, 20220006.

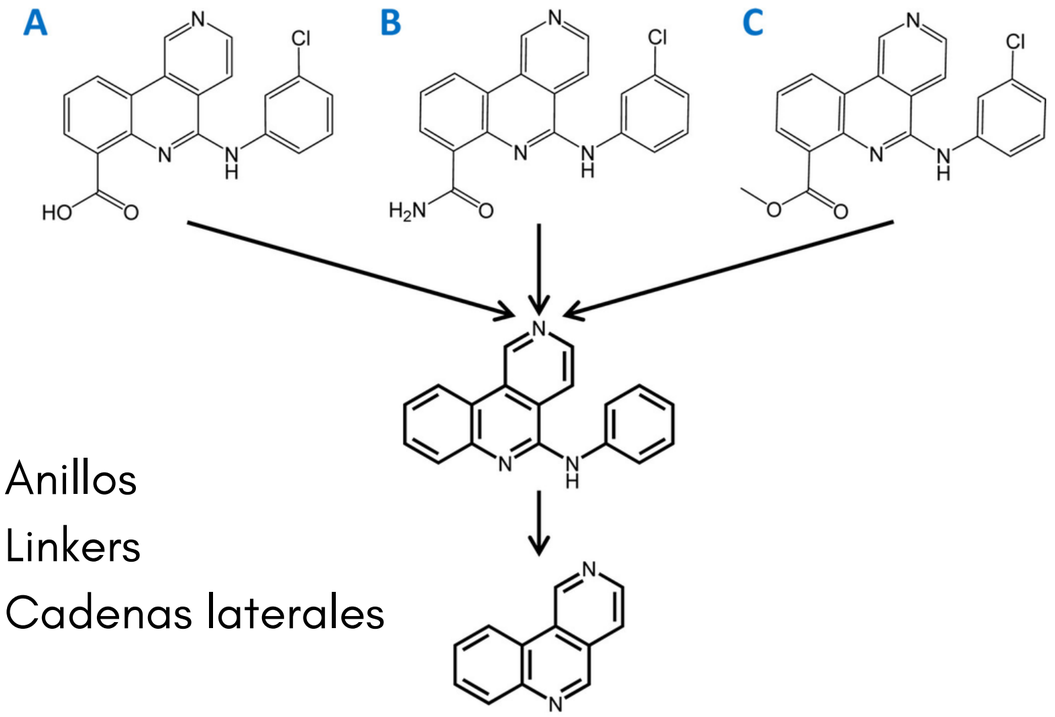
Núcleos base (Scaffolds)

Andamios o *frameworks*.

Distintas definiciones.



Bemis-Murcko

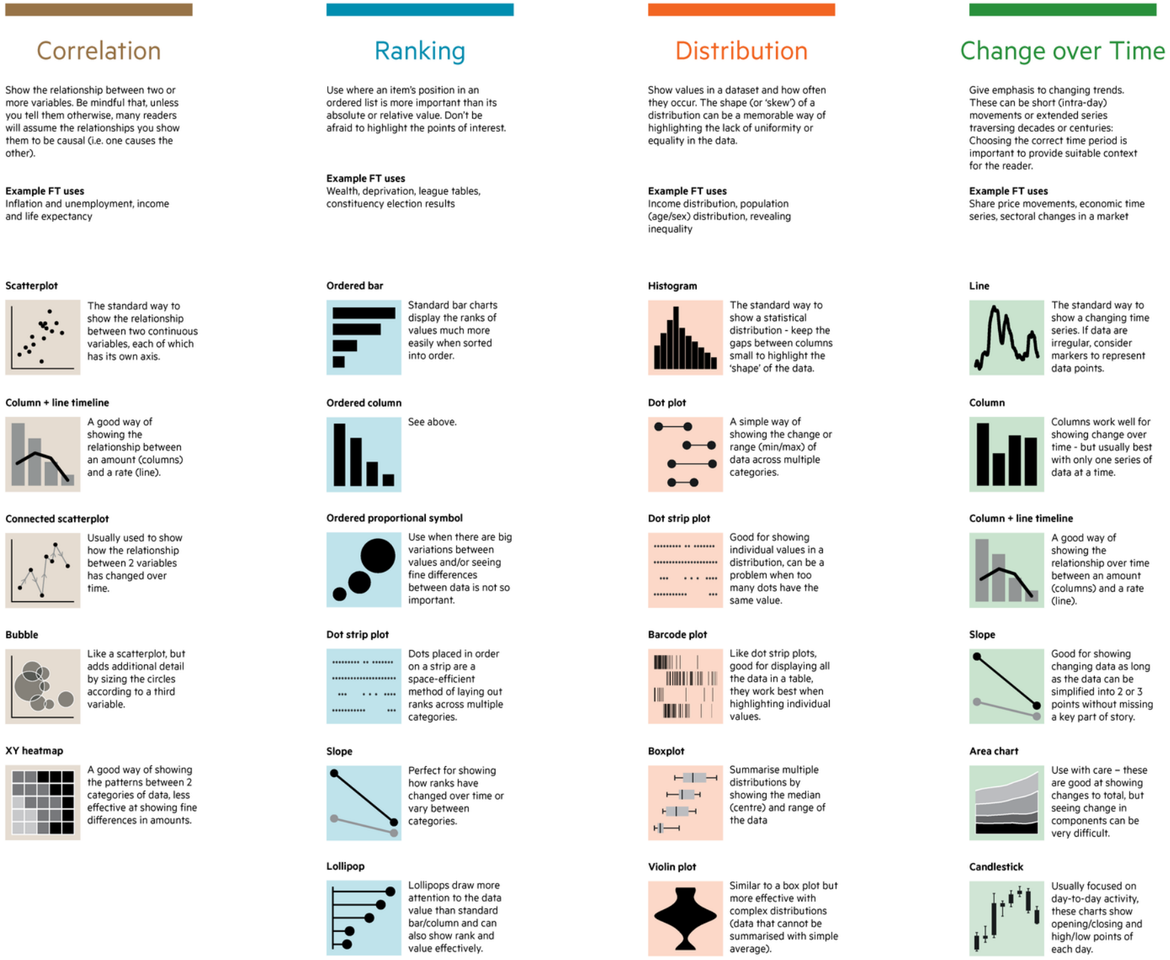


- Anillos
- Linkers
- Cadenas laterales

J. Med. Chem. 2016, 59, 9, 4062-4076

Análisis exploratorio de datos

- *Exploratory Data Analysis (EDA)*
- Investigar conjuntos de datos y resumir sus características principales, empleando métodos gráficos.
- Descubrir patrones o tendencias, valores atípicos, anomalías o verificar suposiciones.



Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e. one causes the other).

Example FT uses
Inflation and unemployment, income and life expectancy

Scatterplot
The standard way to show the relationship between two continuous variables, each of which has its own axis.

Column + line timeline
A good way of showing the relationship between an amount (columns) and a rate (line).

Connected scatterplot
Usually used to show how the relationship between 2 variables has changed over time.

Bubble
Like a scatterplot, but adds additional detail by sizing the circles according to a third variable.

XY heatmap
A good way of showing the patterns between 2 categories of data, less effective at showing fine differences in amounts.

Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

Example FT uses
Wealth, deprivation, league tables, constituency election results

Ordered bar
Standard bar charts display the ranks of values much more easily when sorted into order.

Ordered column
See above.

Ordered proportional symbol
Use when there are big variations between values and/or seeing fine differences between data is not so important.

Dot strip plot
Dots placed in order on a strip are a space-efficient method of laying out ranks across multiple categories.

Slope
Perfect for showing how ranks have changed over time or vary between categories.

Lollipop
Lollipops draw more attention to the data value than standard bar/column and can also show rank and value effectively.

Distribution

Show values in a dataset and how often they occur. The shape (or 'skew') of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.

Example FT uses
Income distribution, population (age/sex) distribution, revealing inequality

Histogram
The standard way to show a statistical distribution - keep the gaps between columns small to highlight the 'shape' of the data.

Dot plot
A simple way of showing the change or range (min/max) of data across multiple categories.

Dot strip plot
Good for showing individual values in a distribution, can be a problem when too many dots have the same value.

Barcode plot
Like dot strip plots, good for displaying all the data in a table, they work best when highlighting individual values.

Boxplot
Summarise multiple distributions by showing the median (centre) and range of the data

Violin plot
Similar to a box plot but more effective with complex distributions (data that cannot be summarised with simple average).

Change over Time

Give emphasis to changing trends. These can be short (intra-day) movements or extended series traversing decades or centuries: Choosing the correct time period is important to provide suitable context for the reader.

Example FT uses
Share price movements, economic time series, sectoral changes in a market

Line
The standard way to show a changing time series. If data are irregular, consider markers to represent data points.

Column
Columns work well for showing change over time - but usually best with only one series of data at a time.

Column + line timeline
A good way of showing the relationship over time between an amount (columns) and a rate (line).

Slope
Good for showing changing data as long as the data can be simplified into 2 or 3 points without missing a key part of story.

Area chart
Use with care – these are good at showing changes to total, but seeing change in components can be very difficult.

Candlestick
Usually focused on day-to-day activity, these charts show opening/closing and high/low points of each day.

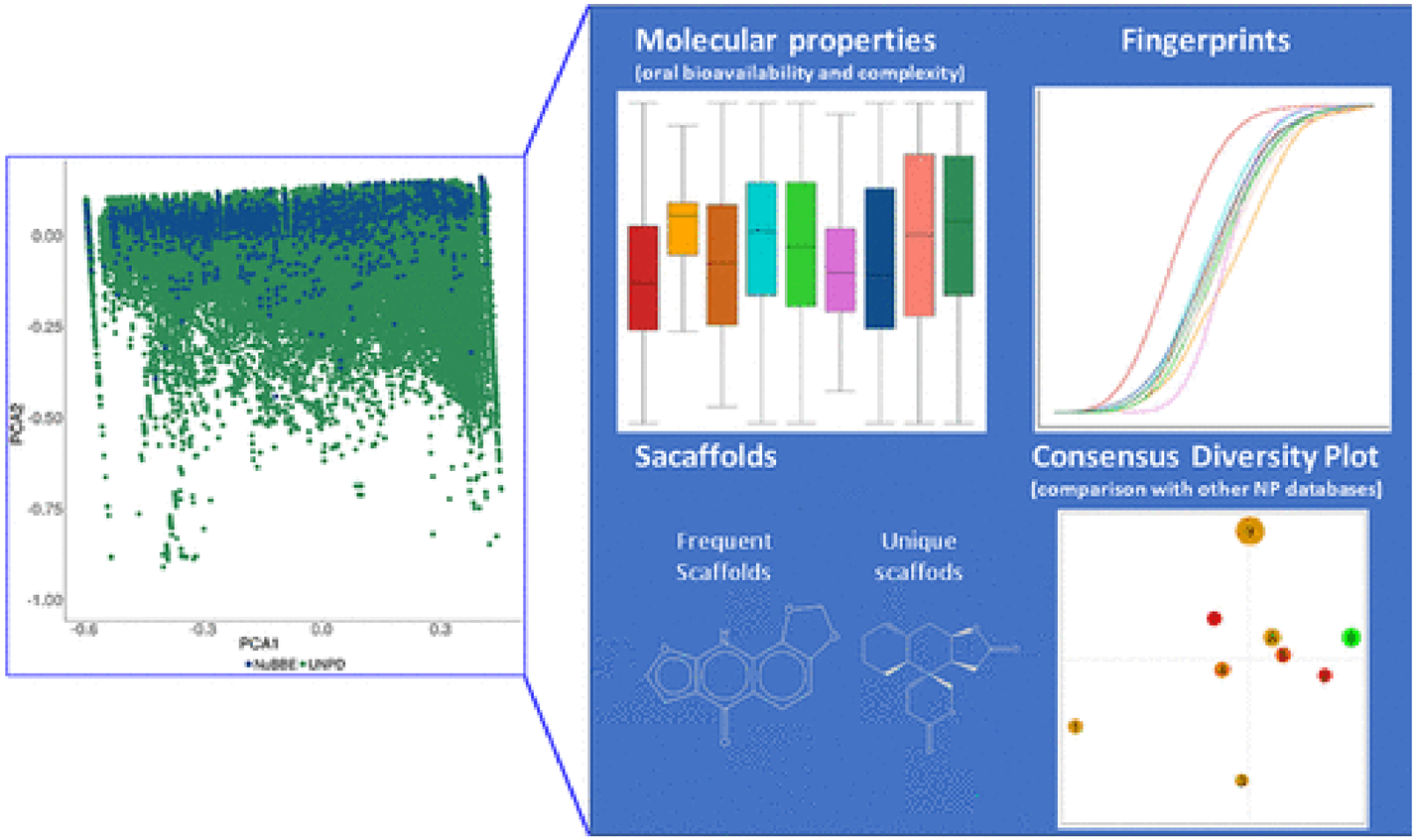
https://www.ibm.com/mx-es/topics/exploratory-data-analysis
https://tinyurl.com/26v26265

Análisis exploratorio de bases de datos moleculares

Propiedades de relevancia farmacéutica
(Masa molar, logP, TPSA, aceptores y donadores de puente de hidrógeno)

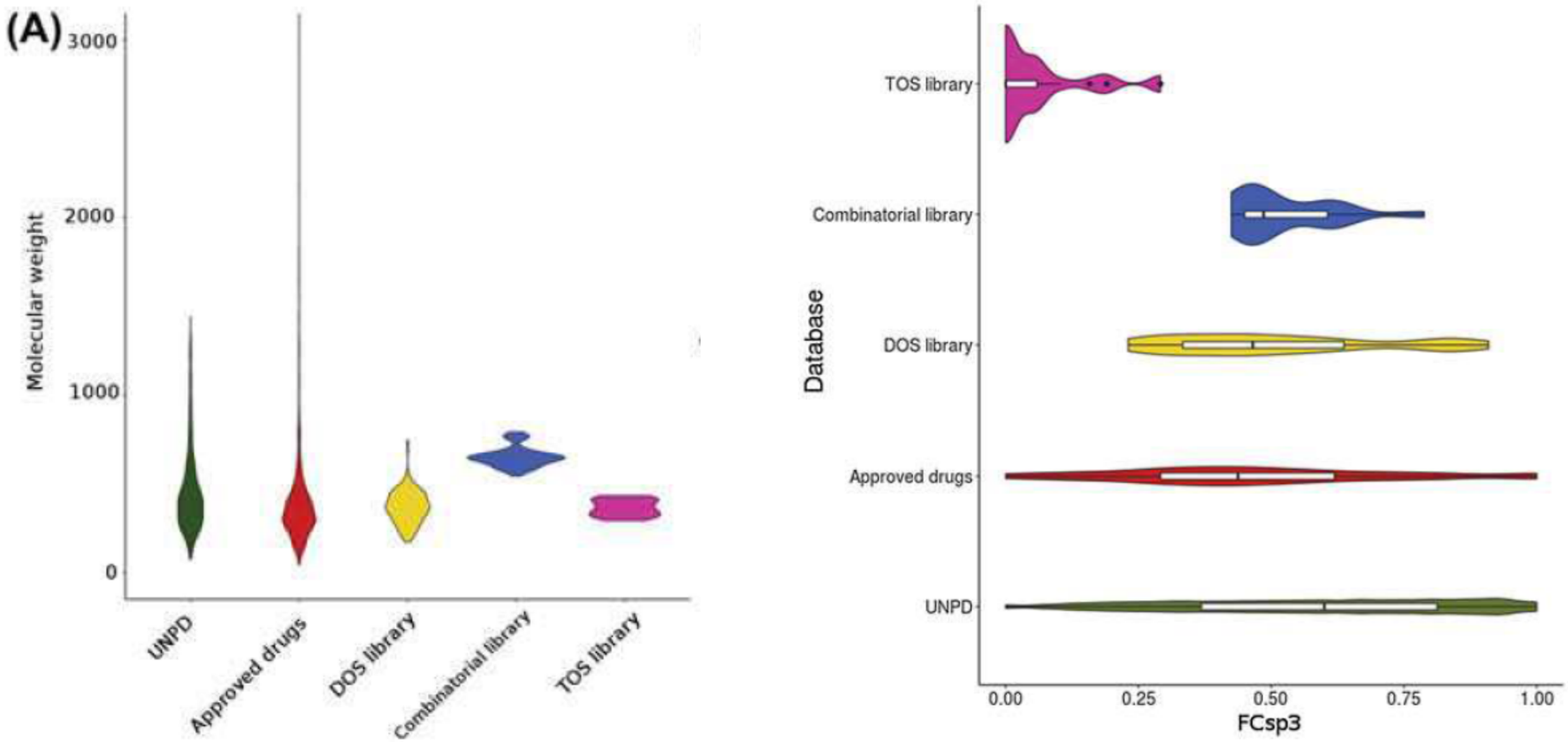
Diversidad de *scaffolds*
Curvas de Recuperación de Sistemas Cíclicos (CSR)

Espacio químico
(PCA, t-SNE, UMAP, TreeMap, etc.)



J. Chem. Inf. Model. 2019, 59, 1, 74-85

Análisis de diversidad química y complejidad molecular



Saldívar-González F.I. & Medina-Franco J.L.
Small molecule drug discovery. Elsevier, 2020. p. 83-102.

Análisis exploratorio de bases de datos moleculares

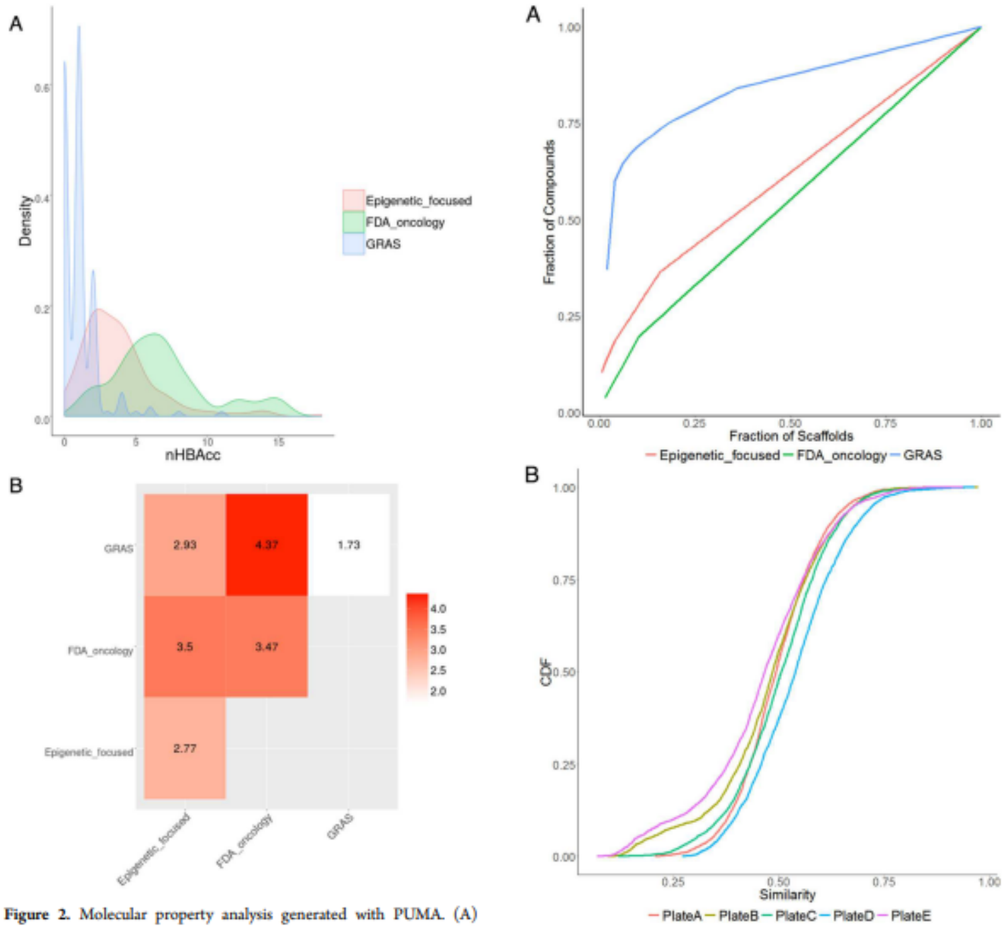


Figure 2. Molecular property analysis generated with PUMA. (A)

PUMA

DIFACQUIM

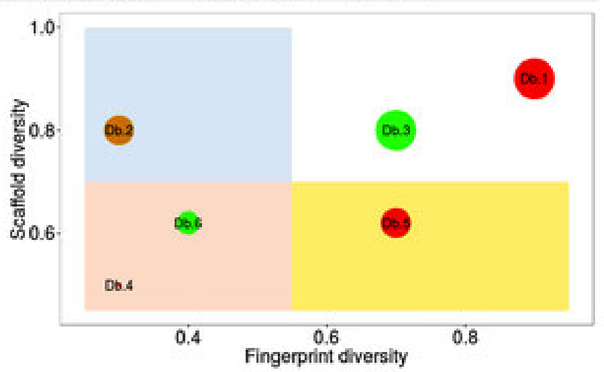
UNAM

Platform for Unified Molecular Analysis: PUMA

D-TOOLS

DIFACQUIM Tools for Chemoinformatics

D-Tools is an initiative that hosts free servers. Currently it has PUMA or Platform for Unified Molecular Analysis, Consensus Diversity Plots, and...



https://www.difacquim.com/d-tools/

¡Gracias por la atención!