

# Predicción del coeficiente de partición a través del aprendizaje automático

Dan Santivañez, Rafael Pulido

AI QUIMIA

*aiquimia.oficial.info@gmail.com*

June 26, 2021

## Presentación

Somos una comunidad multidisciplinaria de investigación, desarrollo y divulgación de proyectos de Inteligencia Artificial y Ciencia de datos aplicados a la Química.

# Dan Santivañez Gutarra



Bachiller en química de la Universidad Peruana Cayetano Heredia (UPCH).  
Estudiante de maestría de Computer Science en la Universidad de Ingeniería y Tecnología (UTEC). Miembro del grupo de investigación URPUNANO liderado por la Dra. María Quintana.  
Investigador en el campo de inteligencia artificial aplicado al desarrollo de celdas solares. Cofundador de AI Química y actual director ejecutivo.

Correo: [dan.santivaney@utec.edu.pe](mailto:dan.santivaney@utec.edu.pe)

# Rafael Pulido del Pino



Bachiller en química de la Universidad Peruana Cayetano Heredia (UPCH). Actualmente trabaja como asistente de investigación y desarrollo en RENASA. Pasante de investigación en la Universidad de Córdoba (2019) en el laboratorio del Dr. Alejandro Fracaroli. Investigador en el campo de materiales nanoestructurados. Cofundador de AI Química y actual director del área de investigación.

Correo: [rafael.pulido@upch.pe](mailto:rafael.pulido@upch.pe)

# Contenido de la presentación

- 1 Introducción
  - Concepto
  - Medición
  - Aplicaciones
- 2 Enfoque dirigido por datos
  - Factores fisicoquímicos
  - Descriptores moleculares
  - Relaciones cuantitativas estructura-propiedad (QSPR)
  - Aprendizaje automático
- 3 Bibliografía
- 4 Contacto

# Definición

## Coeficiente de Partición - $K_D$ .

Parámetro que describe la distribución de un soluto en un sistema bifásico de solventes.

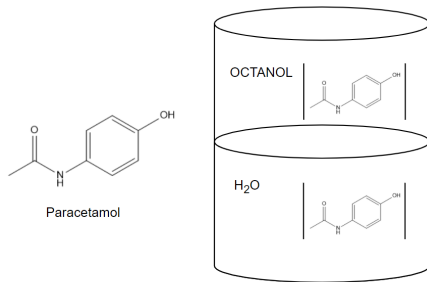


Figure: Distribución del paracetamol en un sistema octanol-agua

# ¿Qué permite medir?

Coeficiente de Partición -  $K_D$ .

Permite medir indirectamente la hidrofobicidad de una molécula

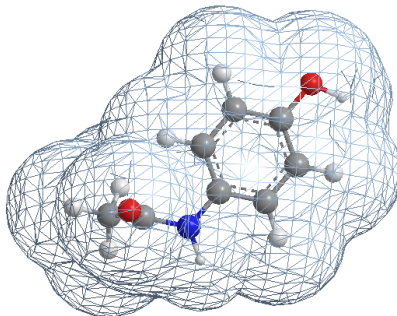


Figure: Vista 3D de la superficie de la molécula del paracetamol

# Estimación teórica

El **Coefficiente de Partición ( $\text{Log}K_D$ )** como constante en un sistema octanol-agua se puede estimar de la siguiente forma:

$$\text{Log}K_D = \text{Log}\left(\frac{[D]_{\text{oct}}}{[D]_{\text{agua}}}\right) = \text{Log}[D]_{\text{oct}} - \text{Log}[D]_{\text{agua}}$$

Donde  $[D]_{\text{oct}}$  y  $[D]_{\text{agua}}$  son las concentraciones del soluto en el medio octanol y en el medio de acuoso respectivamente.



# Técnicas Experimentales

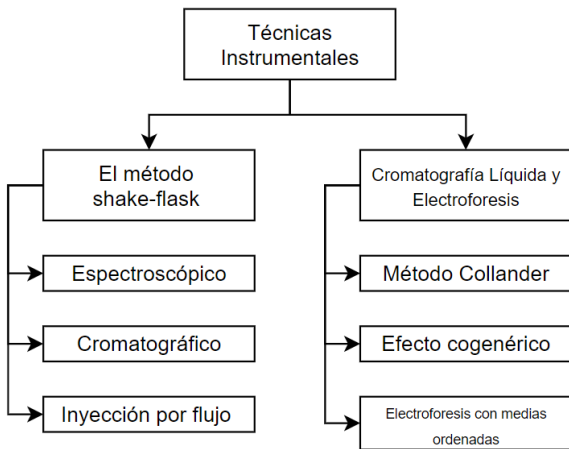


Figure: Técnicas experimentales utilizadas para determinar  $K_D$

# ¿Dónde se puede aplicar?

## Academia

Estudios cinéticos de compuestos orgánicos e inorgánicos complejos.

## Salud

Diseño y evaluación toxicológica de medicamentos (Ejm: Barrera ADME).

## Alimentos

Extracción y purificación de azúcares, proteínas y aceites.

## Industria

Optimización de productos y tratamientos de aguas.

# ¿De qué depende $K_D$ ?

## Termodinámica

$$\mu_{2A}^{\circ} - \mu_{1A}^{\circ} = R \times T \times \ln\left(\frac{x_{2A}}{x_{1A}}\right)$$

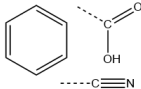
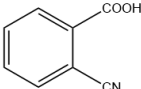
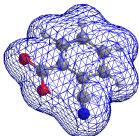
## Reacciones Químicas



$$D = \frac{K_D^{\circ} + K_D^{-}(K_a/[H^{+}])}{1 + (K_a/[H^{+}])}$$

# Influencia de la estructura química

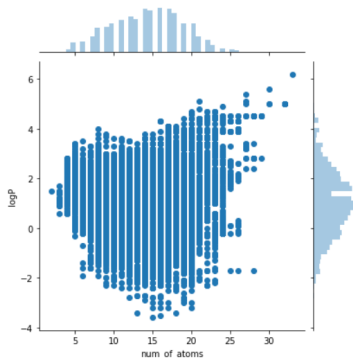
Estos factores dependen de la naturaleza de la molécula. Por la información estructural puede ayudar a estimar el  $K_D$ .

Tipos de descriptores moleculares															
0D	1D	2D	3D												
<table><tr><td>C:</td><td>8</td><td>96 g/mol</td></tr><tr><td>H:</td><td>5</td><td>5 g/mol</td></tr><tr><td>O:</td><td>2</td><td>32 g/mol</td></tr><tr><td>N:</td><td>1</td><td>14 g/mol</td></tr></table>	C:	8	96 g/mol	H:	5	5 g/mol	O:	2	32 g/mol	N:	1	14 g/mol			
C:	8	96 g/mol													
H:	5	5 g/mol													
O:	2	32 g/mol													
N:	1	14 g/mol													

**Table:** Tipo de información estructural de la molécula

# Relaciones cuantitativas estructura-propiedad (QSPR)

Se pueden construir modelos de predicción a partir de la relación entre la información estructural y la propiedad de interés.



**Figure:** Gráfico de dispersión de información 0D de un determinado conjunto de moléculas

# Introducción

## Definición

Campo de la inteligencia artificial que investiga el aprendizaje de las máquinas (algoritmos) direccionado por los datos. Confluyen disciplinas: Ciencias de la computación, Estadística y Matemática.

# Tipos

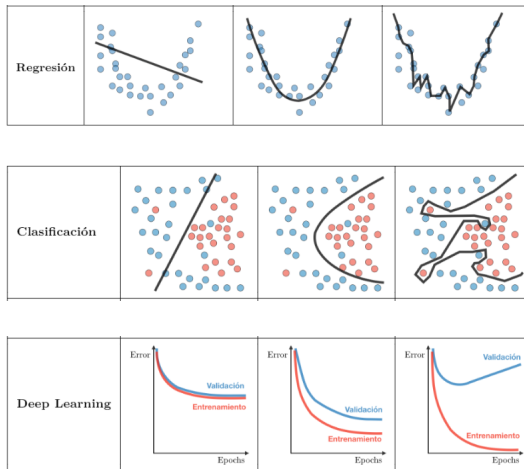


Figure: Modelos de aprendizaje automático

# Flujo de trabajo

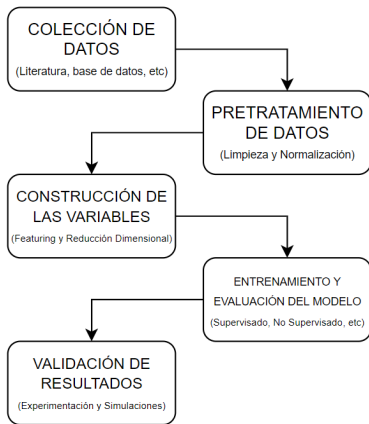


Figure: Flujo de trabajo de investigaciones de aprendizaje automático en química



# References



A. Berthod (2004)

Determination of liquid–liquid partition coefficients by separation methods

*Journal of Chromatography A* 1037 (2004) 3–14.



A. Amidi (2018)

Machine Learning cheatsheets for Stanford's CS 229

*github*

# Información de Contacto



- Correo: [aiquimia.oficial.info@gmail.com](mailto:aiquimia.oficial.info@gmail.com)
- Facebook: @Alquimia2021
- Instagram: @aiquimia