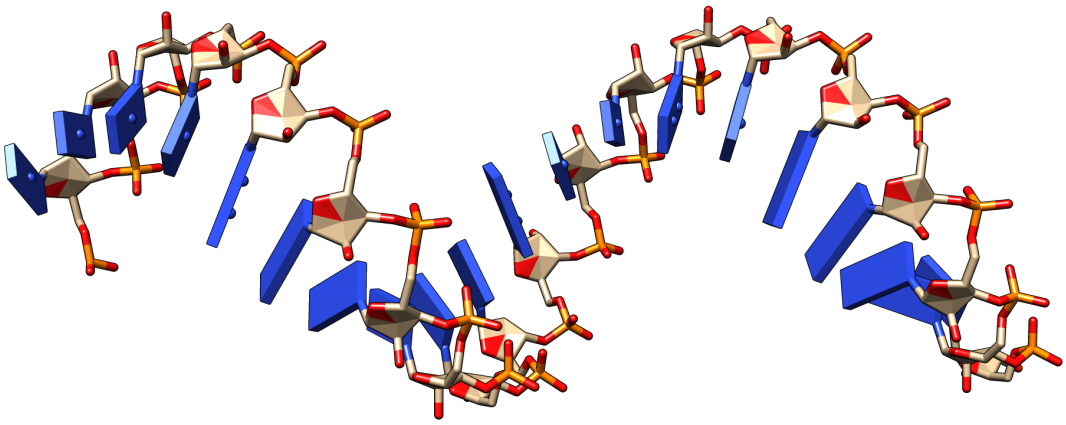
Exploración *in silico*

Química Digital para Democratizar la Búsqueda de Fármacos

Dr. Jesús Alvarado-Huayhuaz



2026

Índice general

Capítulo 1

Química Digital

La química es una ciencia maravillosa. En cada lugar de la naturaleza podemos encontrar un basto ambiente para explorar y aprender de la mano de las teorías químicas. El aprendizaje clásicamente proviene de la experimentación y la construcción del conocimiento empírico, sin embargo, la ciencia ha avanzado hacia nuevos paradigmas, pasando por la ciencia teórica (mecánica, termodinámica, leyes de Maxwel, etcétera), la ciencia computacional o de las simulaciones (teoría del funcional de la densidad, dinámica molecular, etcétera) y la ciencia de los datos, y hago énfasis de la gran cantidad de datos (inteligencia artificial, minería de datos, reconocimiento de patrones, etcétera) [7].

La química digital es el campo que integra la ciencia química con tecnologías digitales, computación avanzada, ciencia de datos, inteligencia artificial, automatización y laboratorios conectados, para modelar, predecir, controlar y acelerar descubrimientos y procesos químicos. En vez de depender sólo de experimentos manuales, la química digital combina simulación teórica, aprendizaje automático, flujos de trabajo automatizados y gestión de datos para diseñar las moléculas, optimizar reacciones y se gestionar resultados experimentales [2].

La química computacional y la quimioinformática se han consolidado como pilares fundamentales de la investigación química moderna, permitiendo la predicción de propiedades moleculares, el diseño racional de fármacos y materiales, y la interpretación de fenómenos fisicoquímicos complejos. Sin embargo, el acceso efectivo a estas herramientas no está distribuido de manera equitativa a nivel global. Diversos estudios han señalado que las brechas en infraestructura computacional, licencias de software propietario y formación especializada limitan la participación de instituciones de países en desarrollo en investigación computacional de frontera [5]. Uno de los principales factores que contribuyen a estas brechas es la dependencia histórica de software propietario de alto costo y de infraestructura de cómputo de alto desempeño. Aunque los centros de investigación en países industrializados cuentan con acceso a clústeres y aceleradores especializados, muchas universidades carecen de los recursos financieros necesarios para sostener estas inversiones. En este contexto, el desarrollo y adopción de software libre y de código abierto ha sido identificado como una estrategia clave para democratizar la química computacional [5, 6].

Las brechas de accesibilidad no son exclusivamente tecnológicas, sino también pedagógicas. La formación en química computacional y en inteligencia artificial requiere competencias interdisciplinarias que integran química, matemáticas, estadística y ciencias de la computación. La literatura en educación química destaca que la falta de programas formativos estructurados y de capacitación docente limita la incorporación efectiva de estas metodologías en el aula [1, 3].

La irrupción de la inteligencia artificial, particularmente del aprendizaje automático y los modelos generativos, ofrece oportunidades sin precedentes para reducir barreras de

entrada mediante interfaces más intuitivas y flujos de trabajo automatizados. No obstante, también plantea nuevos riesgos de exclusión asociados al acceso desigual a datos de calidad, a recursos computacionales para entrenamiento de modelos y a la alfabetización en IA. Estudios recientes subrayan la necesidad de integrar principios de equidad, transparencia y sostenibilidad en el diseño de herramientas de IA para la química [4, 1].

En el marco de este curso, reconocer estas brechas resulta fundamental para adoptar un enfoque crítico e inclusivo. Se priorizará el uso de herramientas abiertas, reproducibles y accesibles, así como el desarrollo de competencias que permitan a los estudiantes evaluar de manera informada tanto el potencial como las limitaciones de la química computacional y la inteligencia artificial en distintos contextos.

La **Exploración *in silico*** constituye un punto de partida fundamental para comprender cómo la química computacional y la inteligencia artificial permiten estudiar sistemas químicos sin recurrir, al menos en una primera etapa, a experimentación directa en el laboratorio.

El término *"in silico"* se utiliza por analogía con las expresiones *in vivo* e *in vitro*. Mientras que *in vivo* hace referencia a estudios realizados en organismos vivos y *in vitro* a experimentos llevados a cabo en sistemas controlados fuera de ellos, *in silico* alude a experimentos, simulaciones y análisis realizados mediante computadoras. El nombre proviene del **silicio**, elemento base de los semiconductores utilizados en los microprocesadores, y refleja el uso de modelos matemáticos, algoritmos y datos digitales para explorar propiedades moleculares, reactividad, interacciones y comportamiento dinámico de sistemas químicos.

Una pieza central de la exploración *in silico* es la representación molecular, ya que la forma en que se visualiza una molécula influye directamente en su interpretación química. Por ejemplo, consideremos a la molécula "3-hydroxypropanal", de acuerdo con su nombre IUPAC, entre las representaciones más comunes se encuentra la bidimensional, con los hidrógenos apolares implícitos (Figura 1.1)

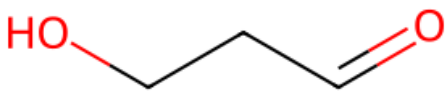


Figura 1.1: Representación bidimensional de 3-hydroxypropanal.

O el modelo **stick**, en el cual los enlaces se muestran como líneas o cilindros delgados y los átomos suelen representarse de forma implícita o con pequeños nodos.

Este tipo de representación es especialmente útil para analizar conectividad, geometría y conformación molecular. El modelo **ball-and-stick** incorpora esferas para los átomos y cilindros para los enlaces, facilitando la identificación de elementos químicos, ángulos de enlace y estereoquímica. Por ejemplo, en la Figura 1.3, las esferas etiquetadas como 1 y 5 corresponden a átomos de oxígeno, 2, 3 y 4, con átomos de carbono y todas las demás

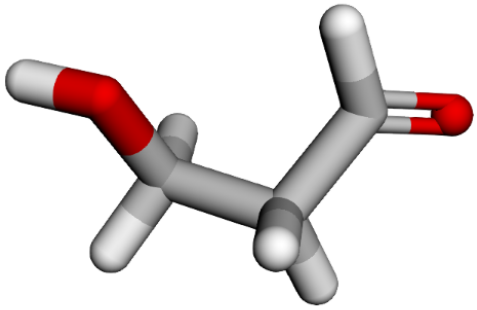


Figura 1.2: Representación tridimensional para 3-hydroxypropanal en modo de varillas o sticks.

son átomos de hidrógeno.

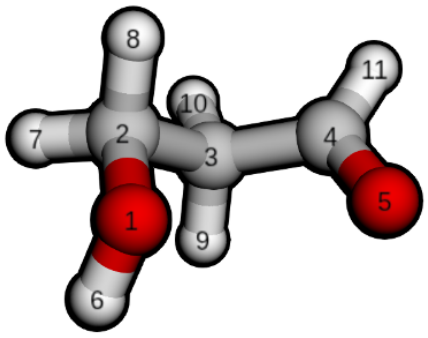


Figura 1.3: Representación 3D para 3-hydroxypropanal en modo de esferas y varillas.

Por otro lado, el modelo cartoon es ampliamente utilizado en bioquímica y biología estructural, particularmente para proteínas y ácidos nucleicos, donde se resaltan elementos de estructura secundaria como hélices alfa, láminas beta y la organización global del biomacromolécula, más que la posición exacta de cada átomo. Como es el caso de la molécula de la portada.

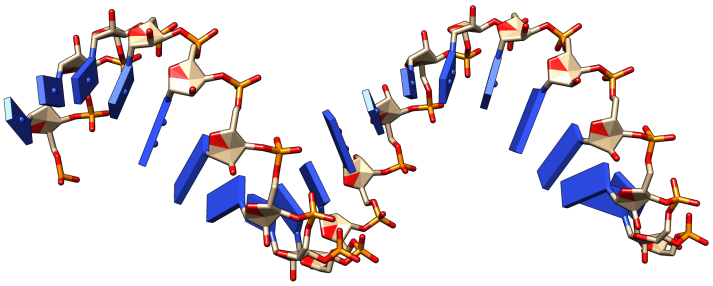


Figura 1.4: Representación cartoon de fragmento de ADN.

Para trabajar con estas representaciones, existen diversos programas de visualización molecular en tres dimensiones que se han consolidado como estándares de facto en investigación y docencia. PyMOL es uno de los más utilizados para la visualización de proteínas, complejos proteína-ligando y estructuras cristalográficas, destacando por su calidad gráfica y flexibilidad. VMD (Visual Molecular Dynamics) es ampliamente empleado

en simulaciones de dinámica molecular, permitiendo el análisis de trayectorias y grandes sistemas biomoleculares. UCSF Chimera y su versión más reciente, ChimeraX, combinan visualización avanzada con herramientas de análisis estructural. Para moléculas pequeñas y quimioinformática, Avogadro y Jmol ofrecen entornos accesibles, multiplataforma y adecuados para enseñanza, mientras que programas como GaussView actúan como interfaces gráficas para paquetes de cálculo cuántico.

La exploración *in silico* se apoya, además, en el uso de formatos digitales de archivos moleculares, los cuales definen cómo se almacena y transmite la información estructural. Entre los formatos más básicos se encuentra el XYZ, que contiene el número de átomos seguido de una lista con el símbolo químico y las coordenadas cartesianas de cada átomo. Este formato es simple y ampliamente utilizado en química computacional, especialmente como salida o entrada en cálculos cuánticos. Sin embargo, el formato XYZ no incluye información explícita sobre enlaces, cargas formales ni tipos de átomos, lo que limita su uso para ciertos análisis.

```
11
Generated by RDKit
C -0.8002 0.3024 0.6968
O -2.0905 -0.0785 0.2395
C 0.2632 -0.2620 -0.2229
C 1.6139 0.2333 0.2027
O 2.5087 -0.5094 0.5938
H -0.7608 1.3966 0.7161
H -0.6708 -0.0508 1.7255
H -2.1696 -1.0433 0.3428
H 0.2477 -1.3570 -0.2060
H 0.0877 0.0450 -1.2593
H 1.7708 1.3238 0.1377
```

Figura 1.5: 3-hydroxypropanal escrita en formato XYZ.

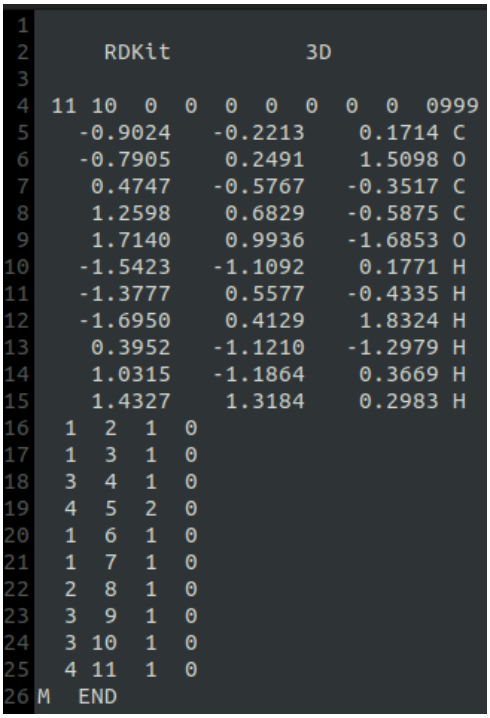
El formato MOL, desarrollado originalmente dentro del ecosistema MDL, incorpora información más rica sobre la molécula, incluyendo conectividad, tipos de enlace y, en algunos casos, propiedades adicionales.

Es un formato muy utilizado en quimioinformática y bases de datos químicas, ya que permite reconstruir la estructura molecular de manera inequívoca. Cuando se requiere almacenar múltiples moléculas en un solo archivo, el formato SDF (Structure Data File) se convierte en una extensión natural del MOL, ya que permite incluir varias estructuras junto con campos de datos asociados, como propiedades fisicoquímicas, identificadores o resultados experimentales y computacionales. Por esta razón, SDF es ampliamente empleado en bibliotecas virtuales, cribado molecular y aprendizaje automático aplicado a química.

En el ámbito de la bioquímica estructural, el formato PDB (Protein Data Bank) ocupa un lugar central. Este formato fue diseñado para describir estructuras tridimensionales de macromoléculas biológicas, como proteínas y ácidos nucleicos, e incluye información detallada sobre coordenadas atómicas, tipos de residuos, cadenas, ocupación, factores de temperatura y, en algunos casos, conectividad. Aunque presenta ciertas limitaciones históricas, como el manejo de sistemas muy grandes o de nuevos tipos de átomos, el formato PDB sigue siendo esencial para el estudio de interacciones biomoleculares, docking y simulaciones de dinámica molecular.

Existen otros formatos complementarios, como CIF para cristalografía, SMILES e InChI para representaciones lineales y compactas de moléculas, y formatos específicos

de programas de simulación. Por ejemplo, para 3-hydroxypropanal el código SMILES es Ç(O)CC=O". En conjunto, el conocimiento de estos formatos y de sus alcances es un componente clave de la alfabetización digital en química, ya que permite al estudiante moverse con solvencia entre visualización, modelado, simulación e inteligencia artificial dentro del ecosistema de la química computacional moderna.



```
1
2      RDKit          3D
3
4  11 10  0  0  0  0  0  0  0  0999
5    -0.9024   -0.2213    0.1714 C
6    -0.7905    0.2491    1.5098 O
7     0.4747   -0.5767   -0.3517 C
8     1.2598    0.6829   -0.5875 C
9     1.7140    0.9936   -1.6853 O
10   -1.5423   -1.1092    0.1771 H
11   -1.3777    0.5577   -0.4335 H
12   -1.6950    0.4129    1.8324 H
13    0.3952   -1.1210   -1.2979 H
14    1.0315   -1.1864    0.3669 H
15    1.4327    1.3184    0.2983 H
16  1  2  1  0
17  1  3  1  0
18  3  4  1  0
19  4  5  2  0
20  1  6  1  0
21  1  7  1  0
22  2  8  1  0
23  3  9  1  0
24  3 10  1  0
25  4 11  1  0
26 M  END
```

Figura 1.6: 3-hydroxypropanal escrita en formato MOL.

Parte 2

Sesión práctica

A continuación exploraremos tres programas que nos servirán como punto de partida:
- 1. quimicaorganica.streamlit.org
- 2. PyMOL
- 3. Python

Bibliografía

[1] Sandra Berber, Mathea Bruckner, Nikolai Maurer, and Johannes Huwer.
Artificial intelligence in chemistry research - implications for teaching and learning.
Journal of Chemical Education, 102(4):1445–1456, 2025.

[2] Stefan Brase.
Digital chemistry: navigating the confluence of computation and experimentation–definition, status quo, and future perspective.
Digital Discovery, 3(10):1923–1932, 2024.

[3] Yael Feldman-Maggor, Ron Blonder, and Giora Alexandron.
Perspectives of generative ai in chemistry education within the tpack framework.
Journal of Science Education and Technology, 34(1):1–12, 2025.

[4] Erik Hermann, Gunter Hermann, and Jean-Christophe Tremblay.
Ethical artificial intelligence in chemical research and development: a dual advantage for sustainability.
Science and Engineering Ethics, 27(4):45, 2021.

[5] Susi Lehtola and Antti J Karttunen.
Free and open source software for computational chemistry education.
Wiley Interdisciplinary Reviews: Computational Molecular Science, 12(5):e1610, 2022.

[6] Johannes Pernaa, Aleksi Takala, Veysel Ciftci, Jose Hernandez-Ramos, Lizethly Caceres-Jensen, and Jorge Rodriguez-Becerra.
Open-source software development in cheminformatics: A qualitative analysis of rationales.
Applied Sciences, 13(17):9516, 2023.

[7] Gabriel R Schleder, Antonio CM Padilha, Carlos Mera Acosta, Marcio Costa, and Adalberto Fazzio.
From dft to machine learning: recent approaches to materials science a review.
Journal of Physics: Materials, 2(3):032001, 2019.