

# Reporte de Análisis del Proyecto: Generación de Moléculas Similares a Inhibidores de la Telomerasa Usando WGAN-GP con R-GCN

**Autor:** Brandon Villada

---

## 1. Introducción

El presente documento resume el trabajo realizado en la generación de estructuras moleculares mediante redes generativas adversarias (WGAN-GP) aplicadas a representaciones tipo grafo de moléculas. El proyecto tiene como objetivo general explorar el uso de inteligencia artificial para la generación de compuestos bioactivos, con un interés particular en estructuras similares a Telomerase-IN-2, un inhibidor de telomerasa con propiedades anticancerígenas y antienviejamiento.

---

## 2. Dataset y Preparación

Inicialmente se consideró el uso del dataset **QM9**, ampliamente utilizado para tareas de predicción de propiedades moleculares. Sin embargo, se identificó una limitación clave: **las moléculas contenidas en QM9 tienen pesos moleculares mucho menores (< 200 Da)** y no representan adecuadamente la complejidad de fármacos reales como Telomerase-IN-2. Por esta razón, se decidió cambiar al dataset "250k\_rndm\_zinc\_drugs\_clean\_3.csv" de **ZINC**, el cual ofrece moléculas más grandes y realistas.

Se filtraron 175,396 moléculas con peso molecular entre 300 y 600 Da para representar estructuras comparables a Telomerase-IN-2 (MW  $\approx$  480 Da). Dejando 122034.

Las moléculas se convirtieron de SMILES a grafos mediante RDKit, generando:

- Tensores de adyacencia (tipo de enlace)
- Tensores de características atómicas (one-hot)

### Modificaciones al código:

- Se amplió el diccionario de átomos para incluir elementos como C, N, O, F.
  - Se incrementó `NUM_ATOMS` de 9 (valor usado para QM9) a 80 para permitir representar estructuras más grandes.
  - Se ajustaron las dimensiones `ATOM_DIM` y `BOND_DIM` para reflejar el nuevo espacio de átomos y enlaces.
  - Se incorporó manejo robusto de errores en la conversión `SMILES -> grafo` para evitar descartes silenciosos.
-

### 3. Arquitectura del Modelo

Se desarrolló una arquitectura basada en WGAN-GP:

- **Generador:** recibe un vector latente de dimensión 64 y produce grafos moleculares.
- **Discriminador:** evalúa la validez de grafos reales vs. generados.

Se implementó WGAN con penalización de gradiente (WGAN-GP), mejorando la estabilidad mediante:

- `gp_weight = 1.0`
- Clipping de logits del discriminador
- Control de norm de gradiente (`clip_by_norm`)

Entrenamiento realizado en GPU T4 por 100 épocas, usando `batch_size = 16`.

---

### 4. Resultados del Entrenamiento

- El generador mostró pérdida (`loss_gen`) oscilando entre -3 y +2.5, indicando aprendizaje progresivo.
- El discriminador mostró estabilidad general con picos ocasionales ( $> 100$ ), esperados en GANs.

Tras 20 épocas:

- Solo se generó 1 molécula válida.
- No se observó correlación evidente con Telomerase-IN-2.
- Muchas salidas fueron `None` debido a problemas en sanitización o conectividad del grafo.

Tras 100 épocas:

- Solo se generaron 2 moléculas válidas.
  - No se observó correlación evidente con Telomerase-IN-2.
  - Muchas salidas fueron `None` debido a problemas en sanitización o conectividad del grafo.
- 

### 5. Análisis Crítico

A pesar de un entrenamiento estable, el modelo no logró generalizar hacia estructuras significativas. Se identificaron las siguientes limitantes:

- **Dataset genérico:** ZINC contiene muchas moléculas disimilares al target deseado.

- **Falta de guía:** No se incluye ninguna función de optimización hacia Telomerase-IN-2.
  - **Ausencia de reconstrucción:** Al no usar VAE o autoencoder, el modelo no tiene presión por conservar información estructural.
- 

## 6. Recomendaciones y Sigüientes Pasos

1. **Reentrenamiento dirigido:**
  - Buscar análogos estructurales de Telomerase-IN-2 en ZINC o ChEMBL.
  - Filtrar por similitud Tanimoto  $> 0.7$  y motivos estructurales comunes.
2. **Modelos alternativos:**
  - Usar un **GraphVAE** o **JTVAE** con función de reconstrucción.
  - Incorporar refuerzo por propiedades deseadas (QED, docking, Lipinski).
3. **Evaluación automática:**
  - Medir validez, unicidad y similitud estructural por cada bloque de épocas