QSAR PRELIMINAR DE TOXICIDAD MEDIANTE UN MODELO DE APRENDIZAJE SUPERVISADO EN *Drosophila melanogaster* (DROSOPHILIDAE)

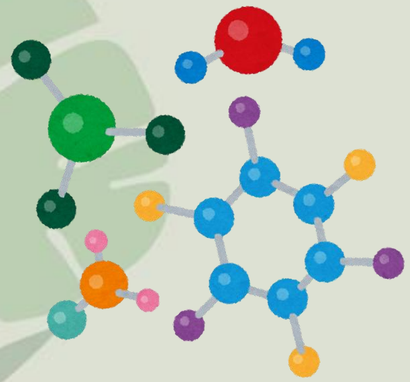
Xavier Clemente García Cevallos,
William Orlando Castillo-Ordóñez

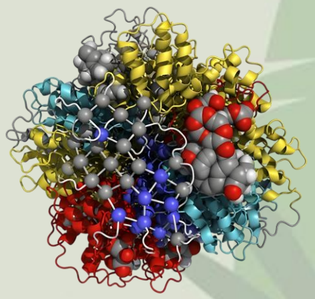
email: xgarcia@unicauca.edu.co
email: bioden@unicauca.edu.co (Semillero)

Grupo de Investigación en Microscopía y Análisis de Imágenes (GIMAI).
Semillero de Investigación en Biología del Desarrollo y Plasticidad Neural

Departamento de Biología
Universidad del Cauca
2025

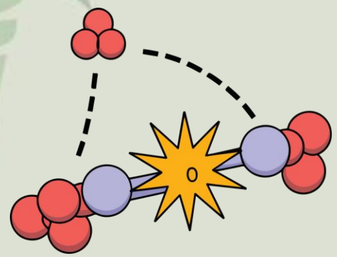
Universidad del Cauca

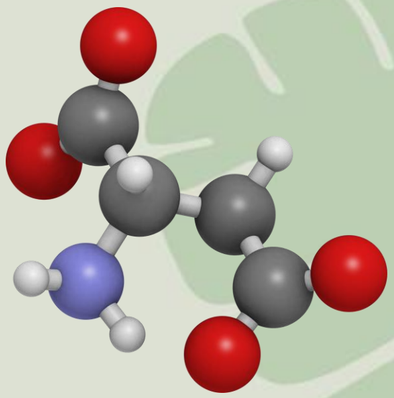




Introducción

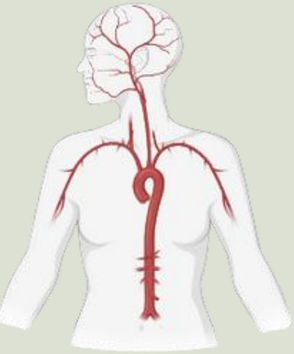
Se escribieron aproximadamente unas 500 líneas de código





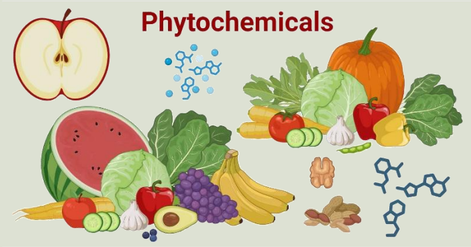
¿Qué es la toxicología?

Ciencia que estudia y describe los efectos nocivos de las sustancias químicas, físicas o biológicas en los seres vivos.



Olker, J. H et al., (2022).

Toxicología



¿Cómo afecta la exposición de diferentes compuestos a un organismo vivo?

¿Dónde podrían afectar?

¿Qué preguntas deberíamos hacernos?

¿Cómo son los mecanismos subyacentes?

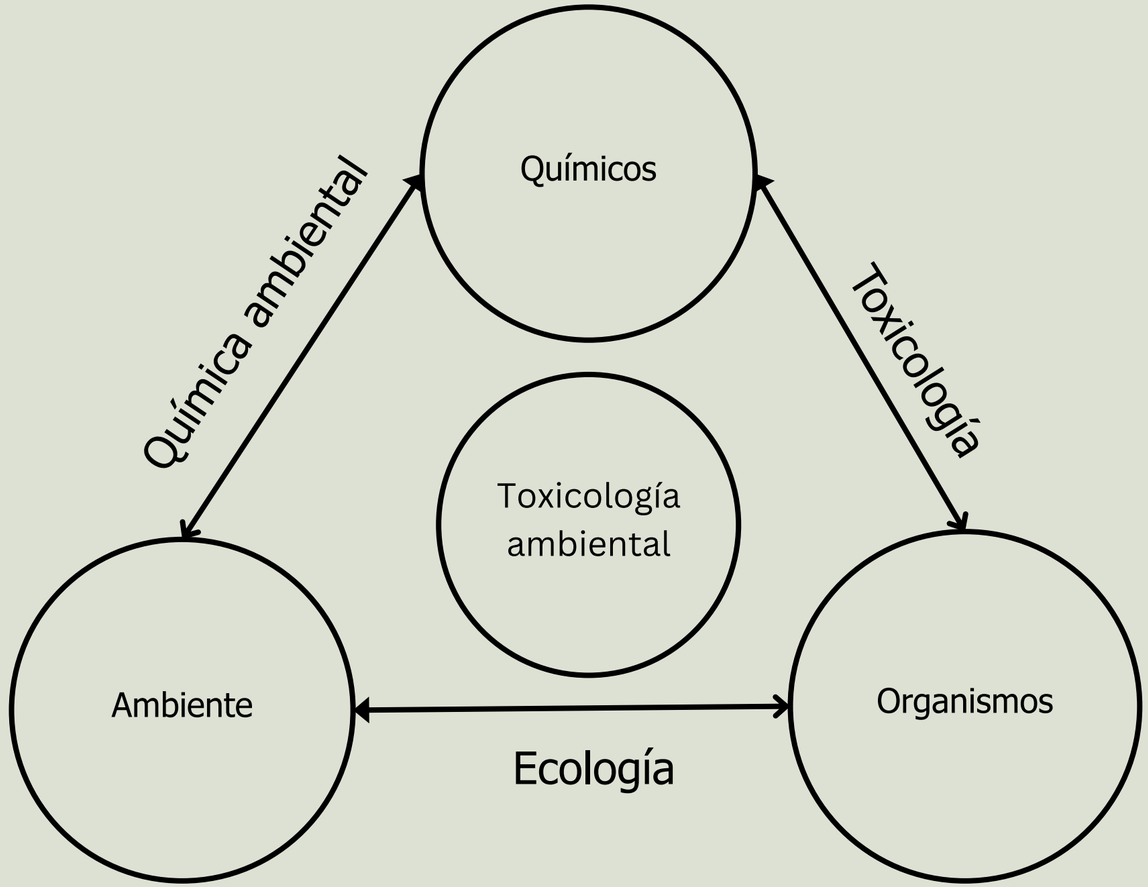
¿Soluciones?

¿Aproximaciones?

¿Puede afectar el desarrollo?
¿Hay inhibición enzimática?
¿Cambios en el comportamiento?

- Efectos reproductivos y neurotóxicos
- Daño mitocondrial
- Rompimiento del ADN
- Rutas moleculares alteradas

Demir, E., & Demir, F. T. (2022)



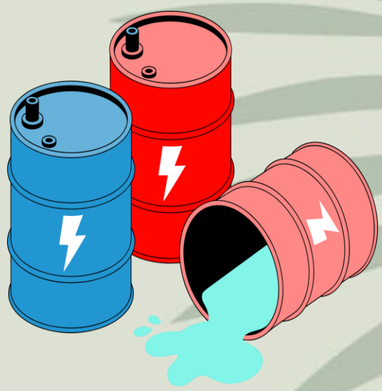


Figura 1. Fundamentos de la toxicología y ecotoxicología

Vieira, L. R., & Morgado, F. (2020)

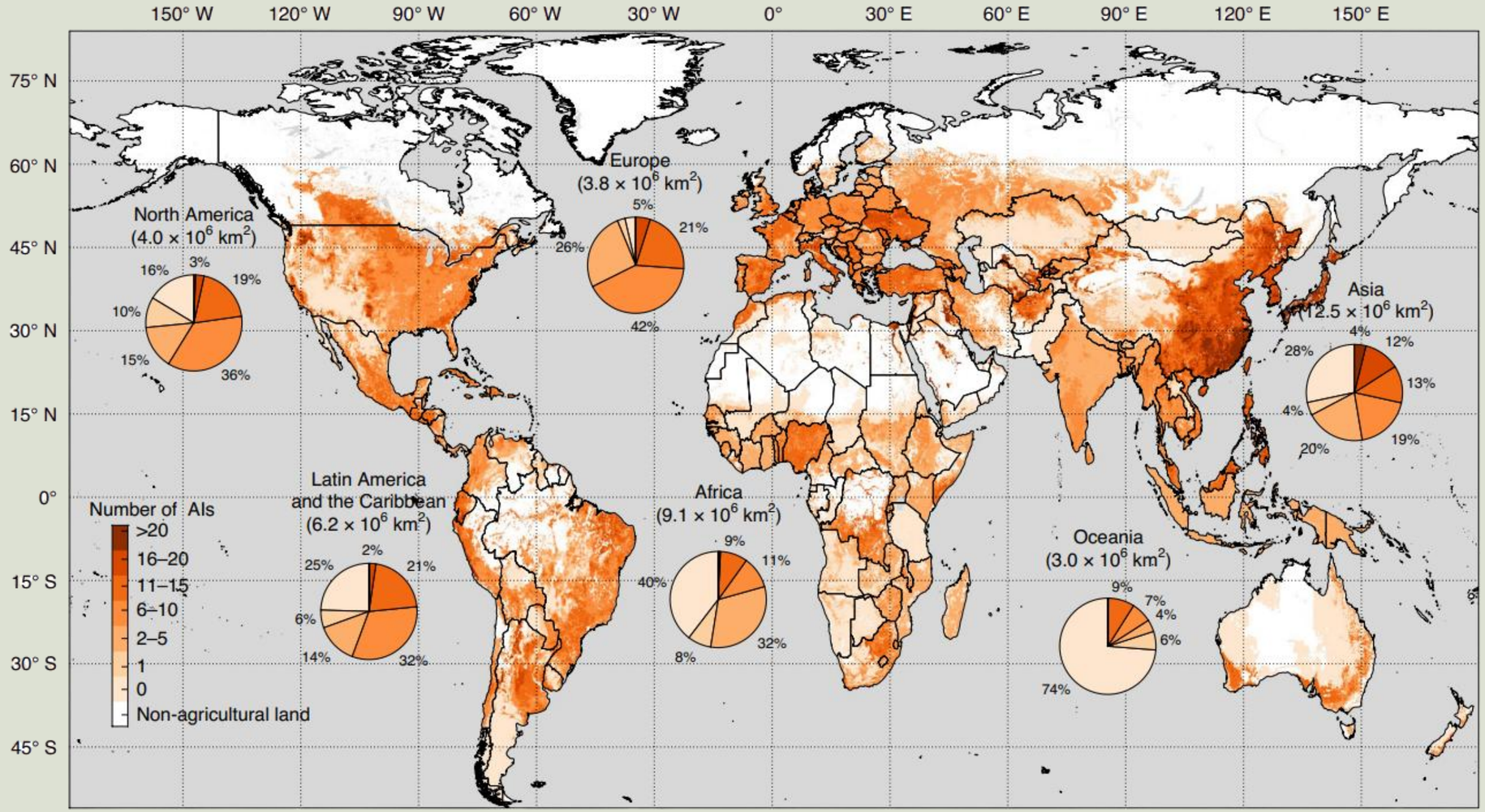
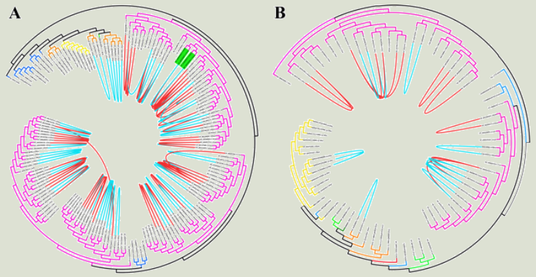
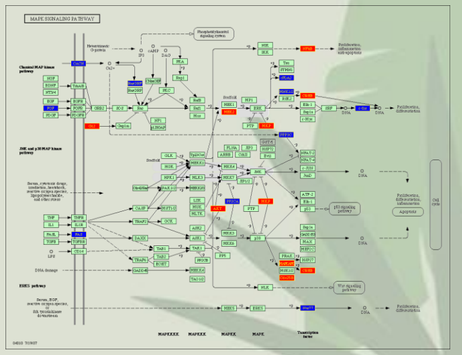
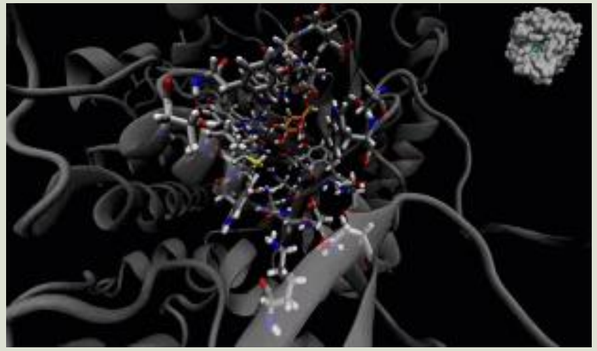


Figura 2. Uso de pesticidas a nivel global. Tomado Tang et al. (2021)

¿Y la bioinformática?







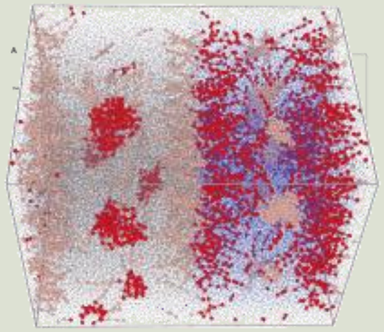
Bases de datos

- Bases de datos de toxicidad
- Estructura química de datos
- Actividad biológica
- Rutas bioquímicas
- Metabolómica
- Toxicogenómica

Análisis de datos minimizados

- Modelos de relación de datos
- Inferencia toxicológica
- Estandarización de datos
- Toxicoinformática

Importancia de las herramientas *in silico*



Modelado QSAR

- Descriptores moleculares
- Propiedades fisicoquímicas
- Relación estadística
- Métrica de similiridad molecular

Quimioinformática

- Simulación de mecánica cuántica (QM)
- Simulación de mecánica molecular (MM)
- Bibliotecas de estandarización química
- Fingerprints químicos
- Diversidad química
- Incorporación de datos

de Lacam, E. G. C., & Chipot, C. J. (2023).

Relación actividad-estructura cuantitativa (QSAR)

Compilación de datos

Datos libres

Literatura primaria

Compilar

Filtración por importancia

Selección

Procesamiento de datos

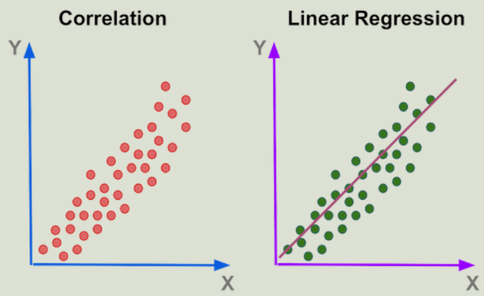
Curación y procesamiento

Cálculo de descriptores

Selección de características

Balance de datos

Corte de datos



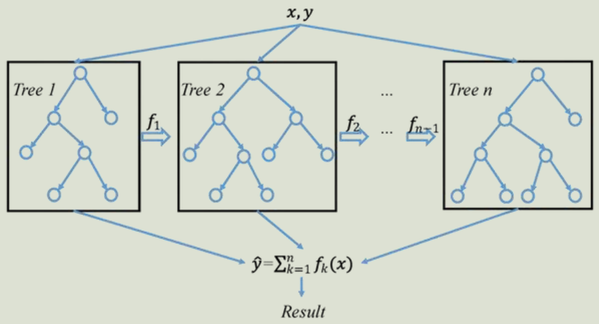
Construcción del modelo

Algoritmo de aprendizaje

RF, XGBoost, SVM, Regresión lineal

Construcción del QSAR

Estadística para un modelo robusto



Interpretación

Validación

Puntuación con modelos previos

Identificar características importantes

Interpretación

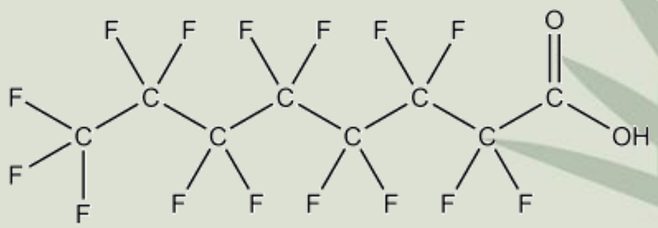
SIB

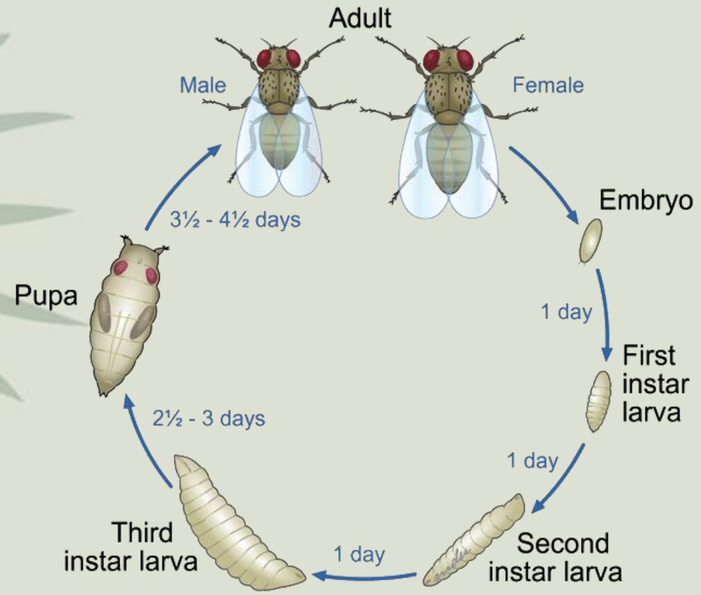
PROTOX
TOX.CHARITE.DE

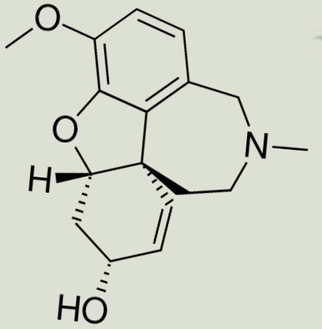
Kapsiani, S., & Howlin, B. J. (2021).

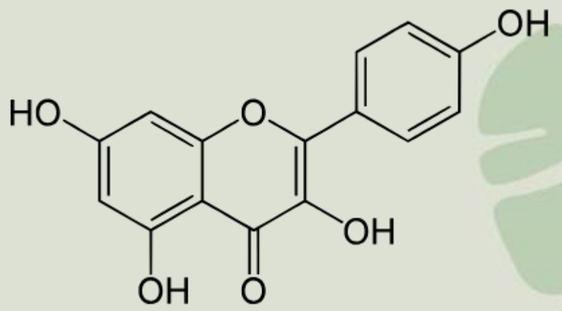
Drosophila melanogaster como biomodelo para el diseño de un QSAR

- **Bajo costo**
- **Alta disponibilidad de cepas transgénicas**
- **Homología con el ser humano del 70 %**
- **Rápido crecimiento**
- **Disponibilidad de protocolos**









- **Efectos:**
- **Neurotóxicos, reproductivos y locomotores**

Cambios:

- **Morfológicos e histológicos**

Alves, S. P. H. (2023)

¿Actualmente, qué limitaciones existen con los modelos QSAR?

Objetivos

Objetivo general:

- Diseñar un pipeline bioinformático para predecir parámetros toxicológicos en *Drosophila melanogaster.*

Objetivos específicos:

- Construir un modelo específico para $\mathrm{L}C_{50}, \mathrm{L}C_{90,}\, \mathrm{LD}_{50}, \mathrm{IC}_{50}$, actividad enzimática, inhibición enzimática y mortalidad.
- Evaluar robustez del modelo QSAR con los datos entrenados y predichos.

Materiales y métodos

Diseño de los modelos QSAR

Gramatica, P. (2020)

Se utilizaron 506 ensayos y se calcularon 3.300 registros a partir de 1.800 descriptores

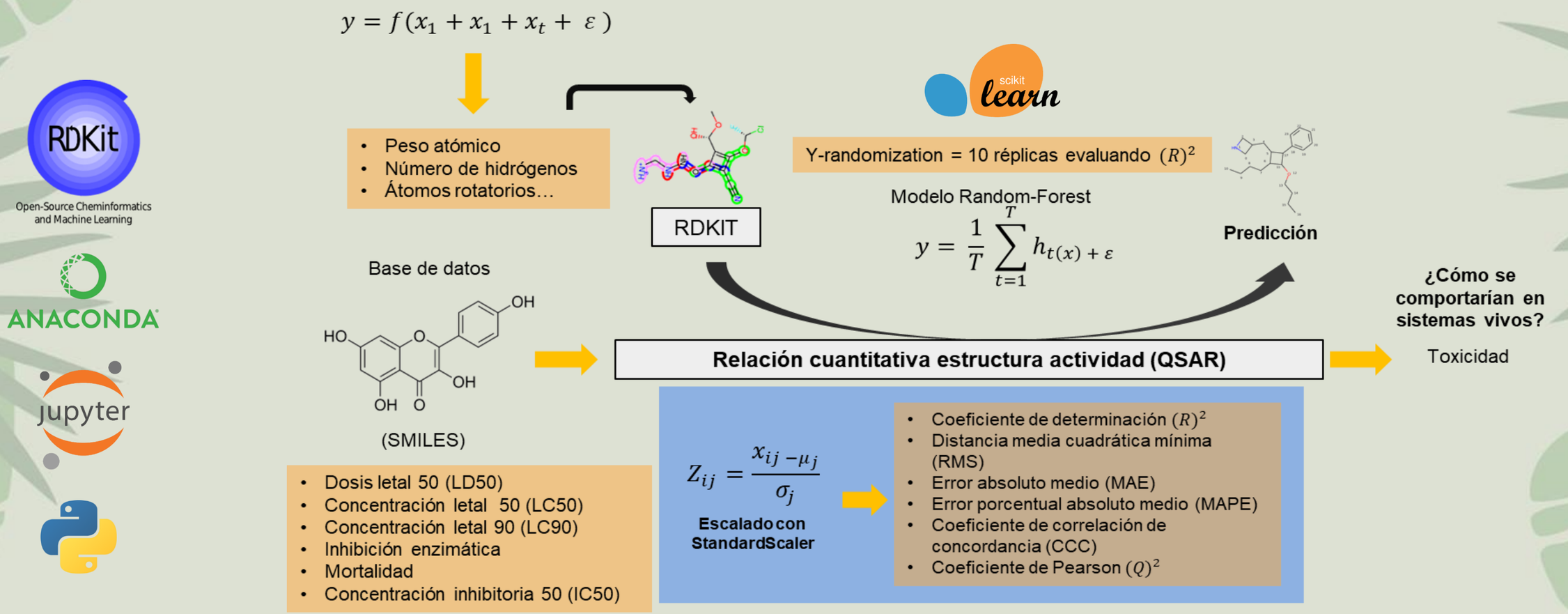


Figura 3. Metodología diseñada para el predictor de toxicidad.

Roy, K. (2017)

Eligiendo los mejores modelos

```
📈 Análisis exploratorio de datos...
Distribución de endpoints:
  mortality: 1515 muestras
  Activity: 676 muestras
  FC: 256 muestras
  LC50: 170 muestras
  LD50: 154 muestras
  Inhibition: 97 muestras
  AFI: 96 muestras
  EC50: 69 muestras
  LC90: 56 muestras
  RatioLC50: 33 muestras
  Ratio: 33 muestras
  Ratio CC50/IC50: 27 muestras
  IC50: 27 muestras
  CC50: 27 muestras
  Survival: 21 muestras
  GI: 20 muestras
  MTD: 18 muestras
  EC90: 5 muestras
```

Principios de la OECD

- Endpoint claro

Medidas de confiabilidad

- $(R)^2 => 0.7 - 0.9$ modelo bueno
- $(R)^2 => 0.6 - 0.7$ modelo acceptable
- $(R)^2 =< 0.5$ modelo rechazado
- $(Q)^2 =$ robustez interna

Y-randomization (Aleatorización de variable)

- El $(R)^2$ debe ser cercano a 0.

El método debe ser reproducible

Dominio de aplicabilidad

- ¿Para qué compuestos se puede usar?

OECD

OECD Principles for QSAR Validation" (2004).

Resultados

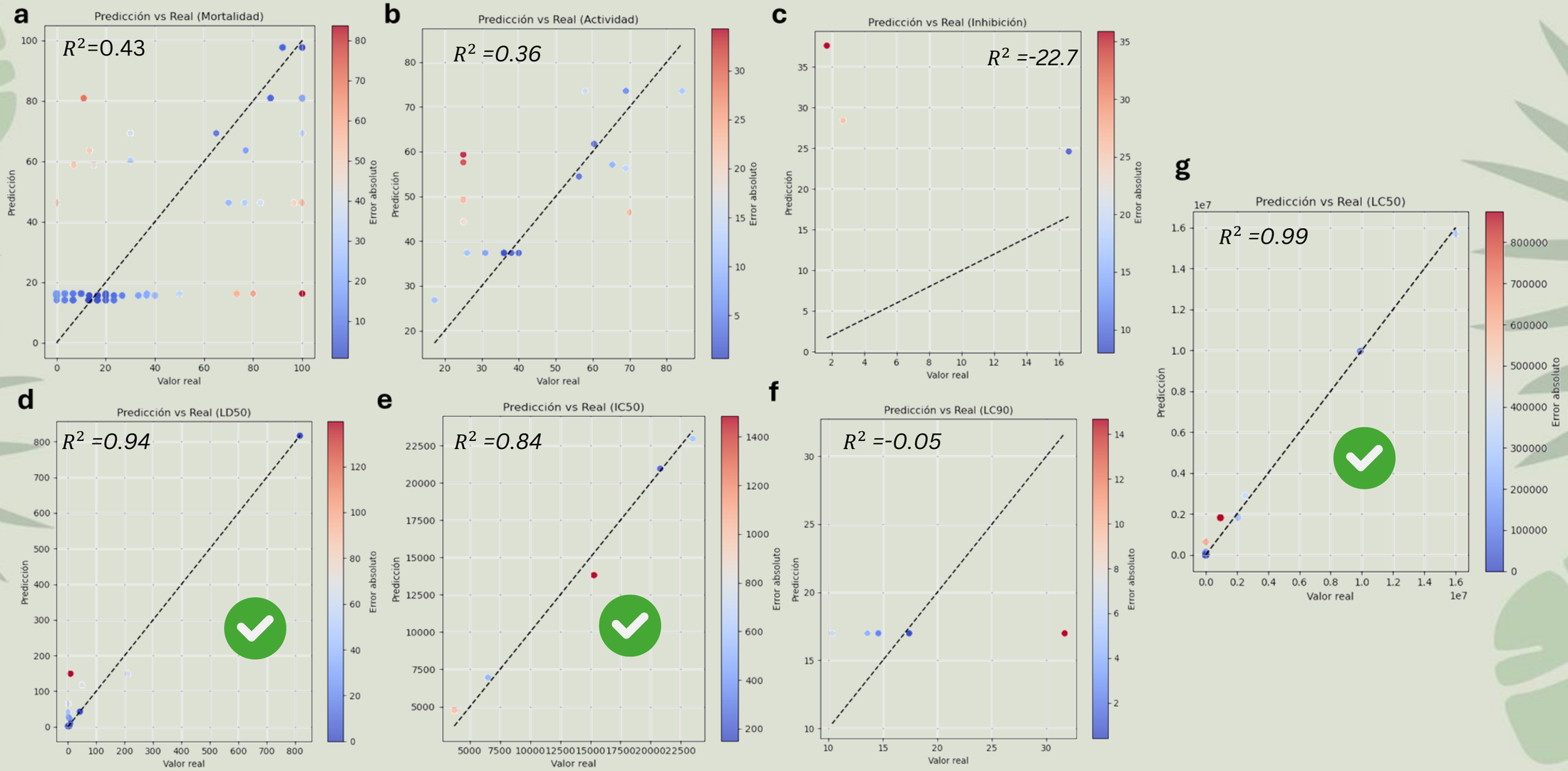


Figura 4. Los gráficos de dispersión muestran la relación entre los datos predichos y los valores reales soportados por la base de datos de entrenamiento.

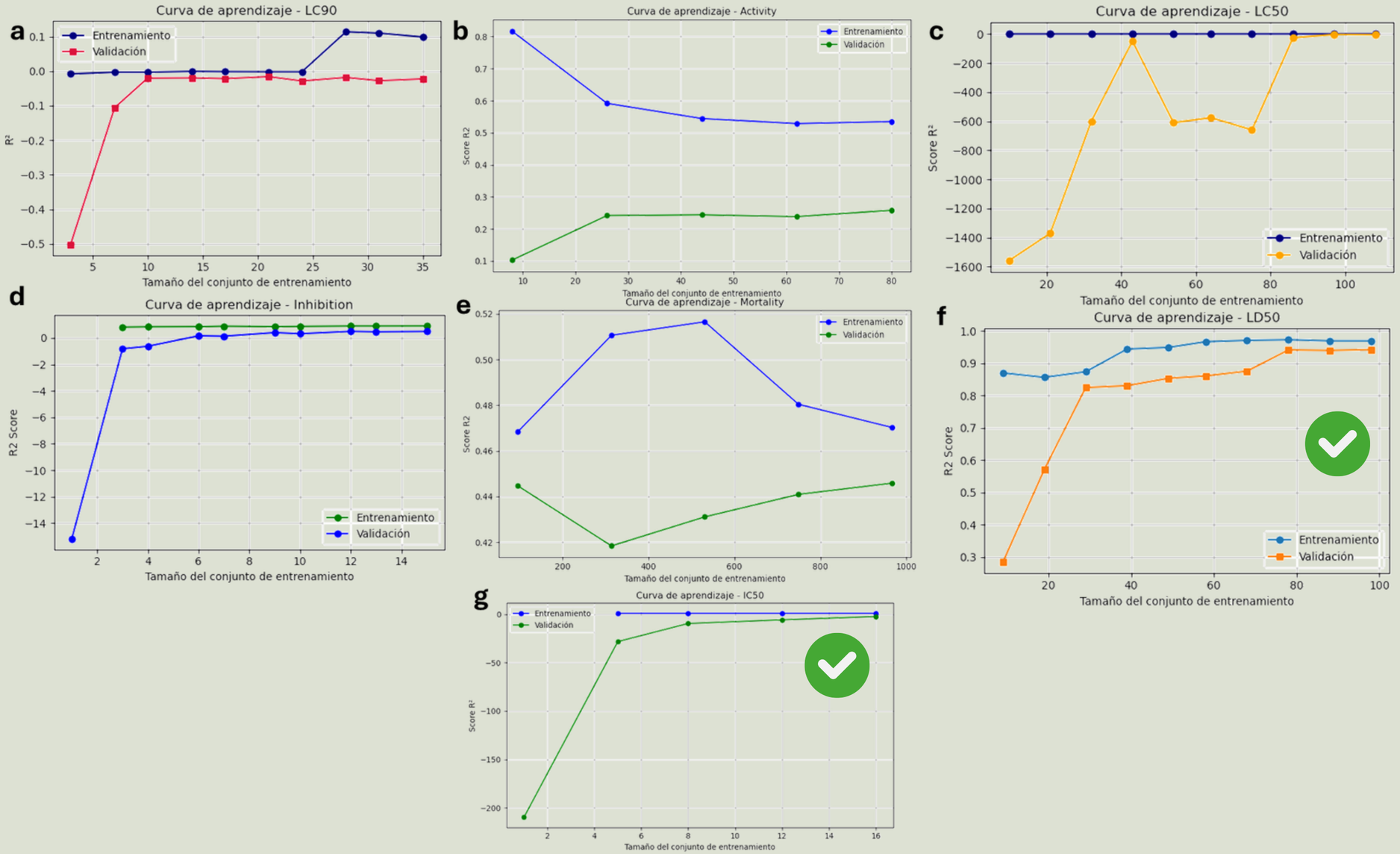


Figura 5. Resultados del modelo comparando el entrenamiento mediante Random-Forest y la validación con datos reales.

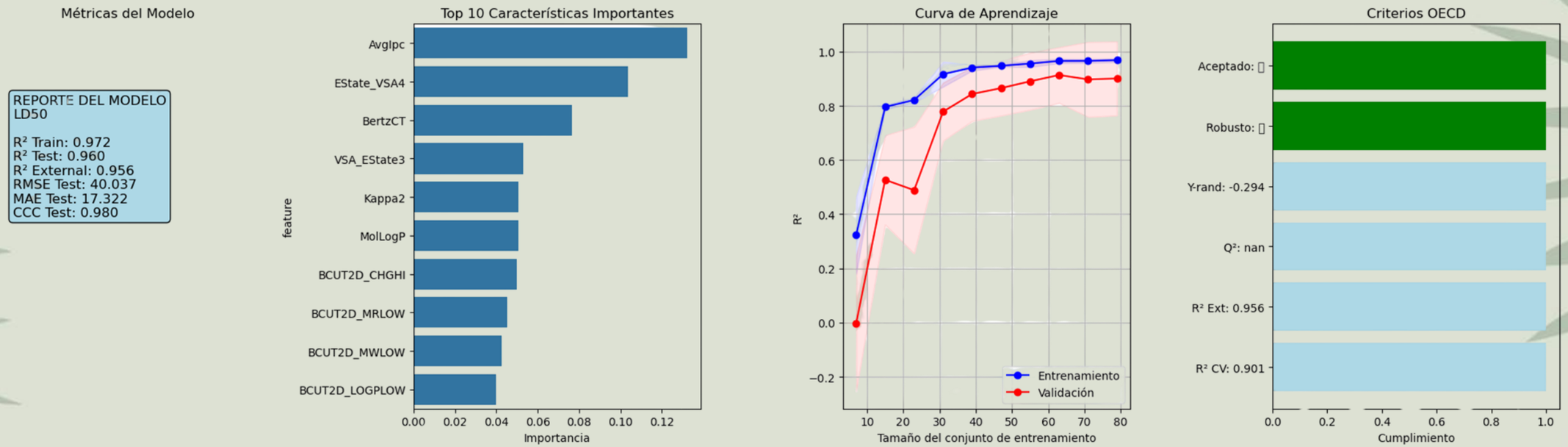


Figura 6. Modelo predictivo de toxicidad de LD50.

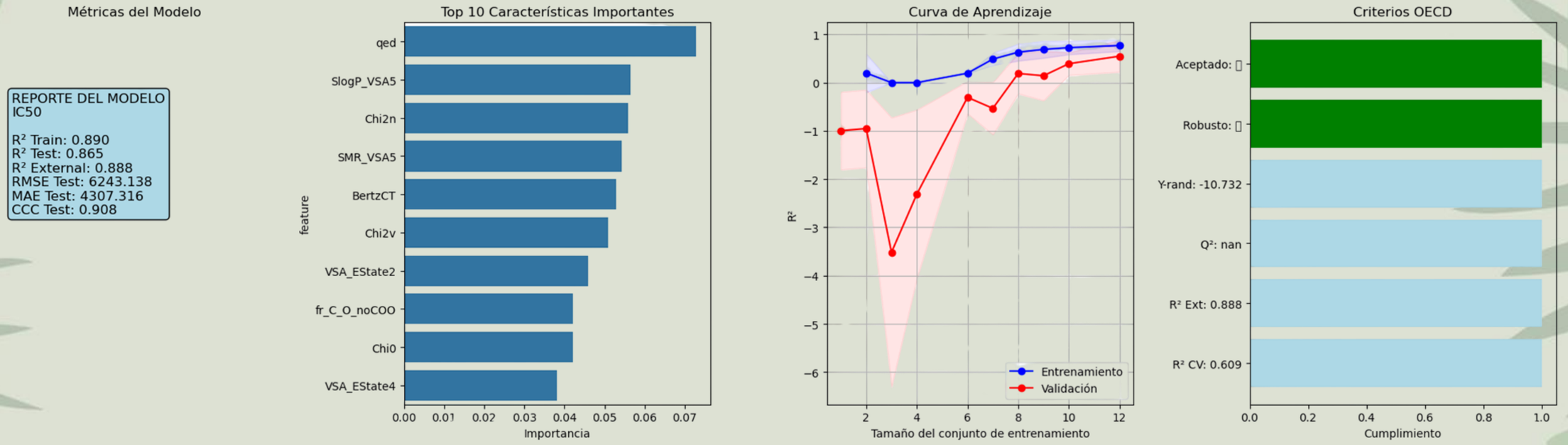


Figura 7. Modelo predictivo de toxicidad de IC50.

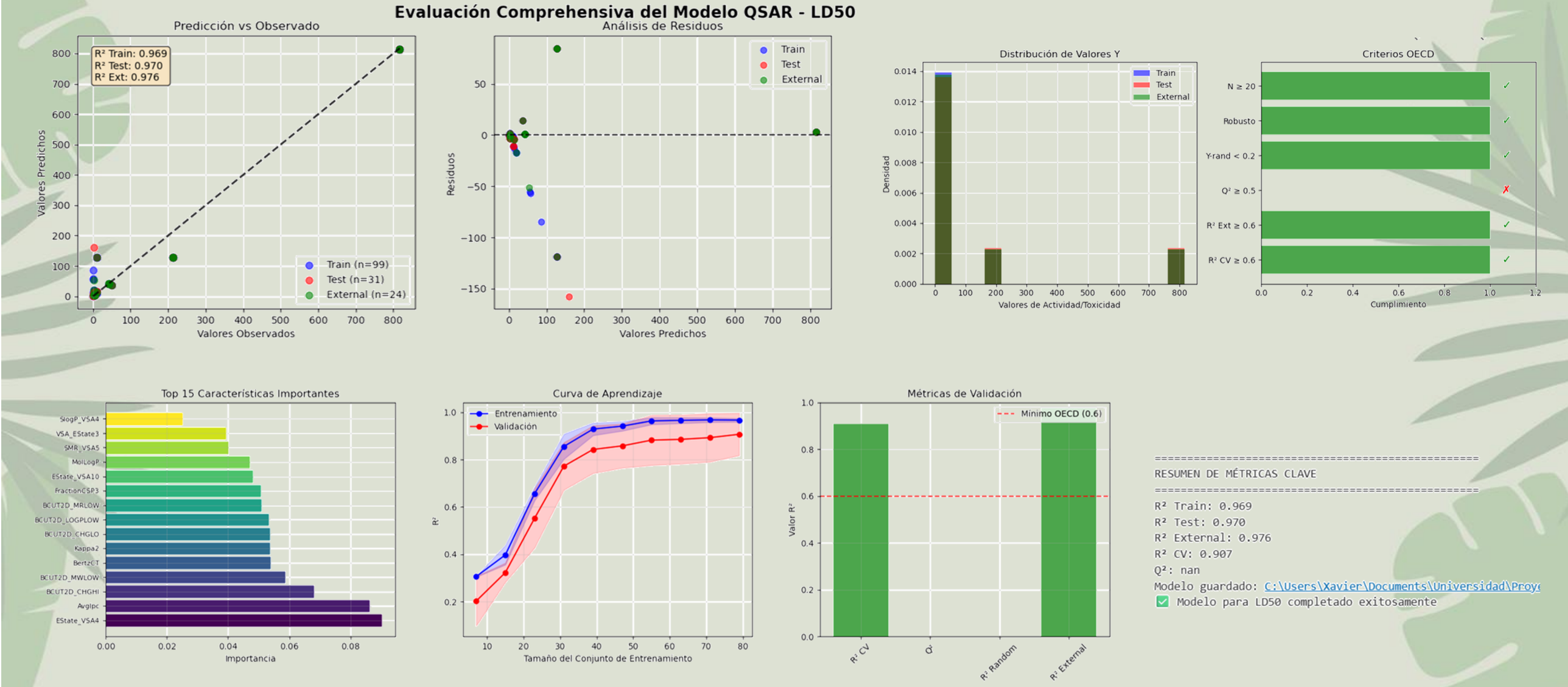
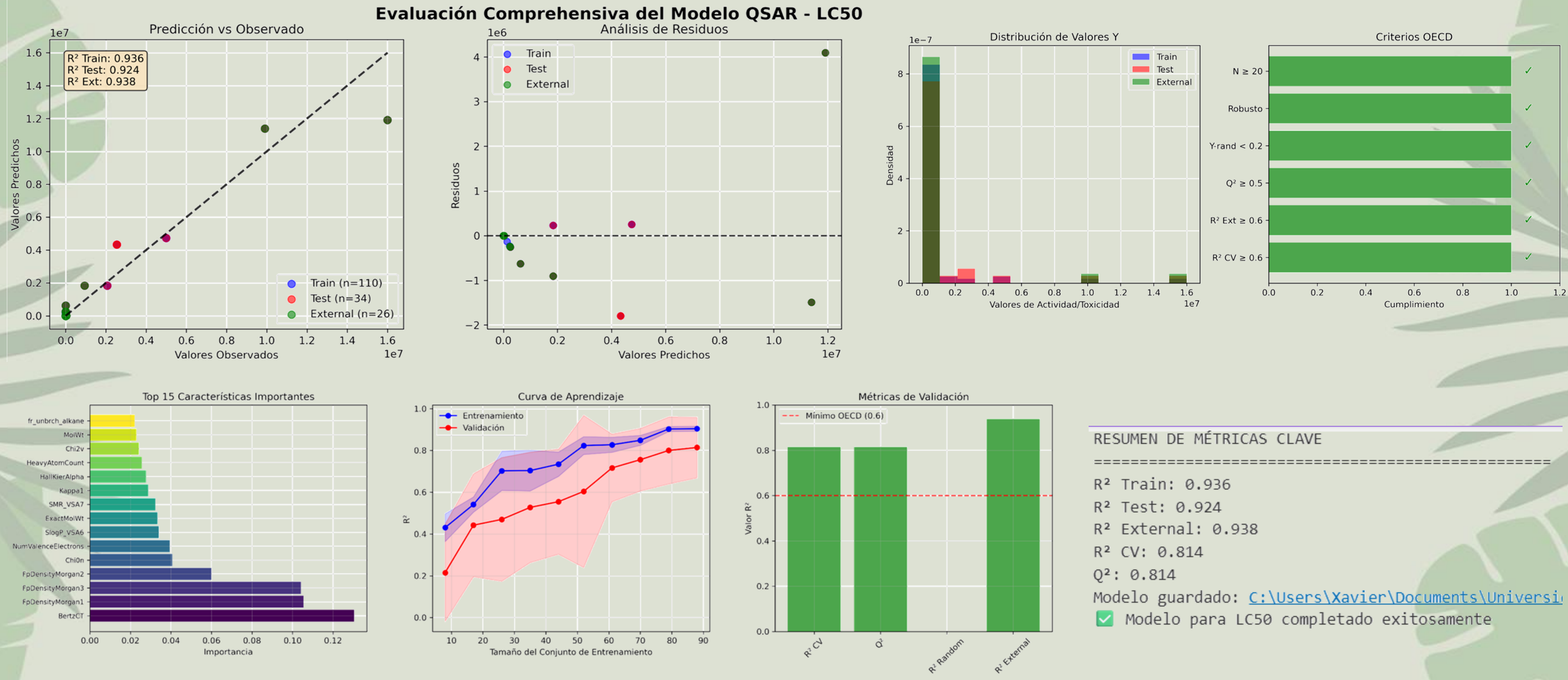


Figura 9. Modelo predictivo de toxicidad de LD50.




```
RESUMEN DE MÉTRICAS CLAVE
==================================================
R² Train: 0.936
R² Test: 0.924
R² External: 0.938
R² CV: 0.814
Q²: 0.814
Modelo guardado: C:\Users\Xavier\Documents\Universi
✅ Modelo para LC50 completado exitosamente
```

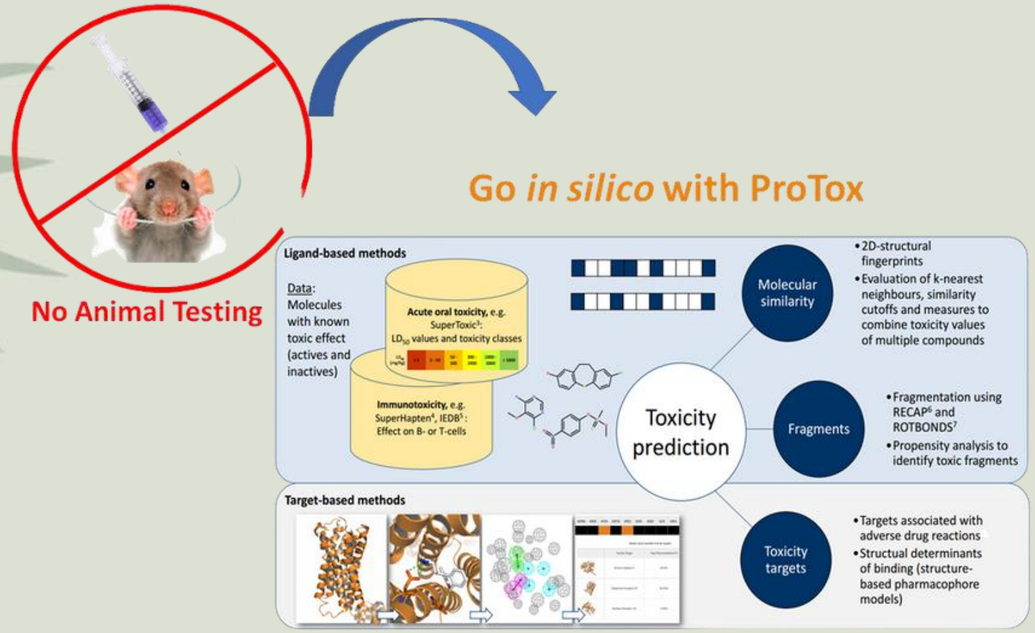
Figura 8. Modelo predictivo de toxicidad de LC50.

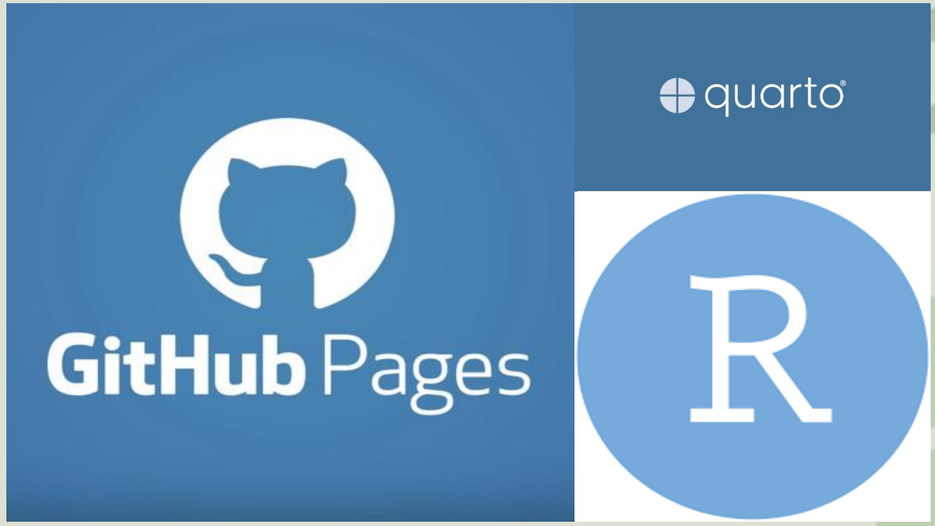
Conclusiones

- El modelo QSAR desarrollado demuestra un alto poder predictivo para endpoints toxicológicos en *Drosophila melanogaster*, especialmente en LD50 y LC50, cumpliendo los criterios de robustez y aceptación propuestos por la OECD.
- No todos los endpoints mostraron buena solidez, evidenciando la necesidad de ampliar las bases de datos existentes para el refinamiento del modelo.
- La validación interna y externa confirmó la capacidad de generalización de los modelos, con valores consistentes entre entrenamiento y validación, descartando sobreajuste.

Perspectivas futuras

- Desarrollo de un predictor de toxicidad open-source (Mantenido por la comunidad de interés) enfocado en farmacología y toxicidad por pesticidas sobre invertebrados.
- La base de datos deberá ser enriquecida por colaboradores del proyecto
- Publicar el modelo desarrollado en una revista para el conocimiento de la comunidad científica.





Agradecimientos

Referencias