

PREDICCIÓN DE BLANCOS MOLECULARES COMBINANDO DOCKING E INTELIGENCIA ARTIFICIAL

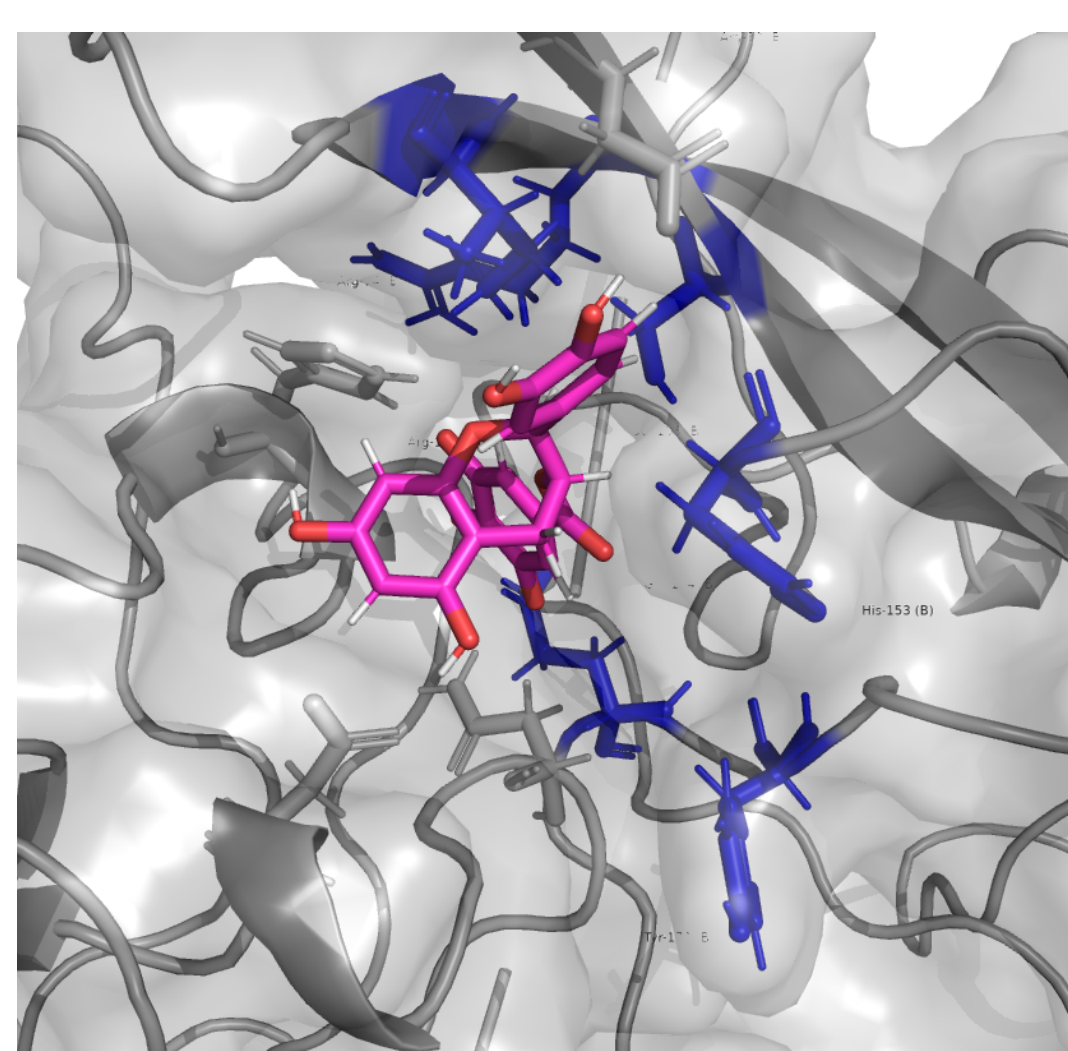
Fernández, Julián F.[1,2] y Lavecchia, Martín J. [3]

Introducción

La predicción de blancos moleculares es una rama de la química computacional que cuenta con poco desarrollo en la actualidad. Las herramientas disponibles dan resultados confusos y carentes de métricas, que imposibilitan su introducción en un proyecto de investigación.

En el presente trabajo se desarrolló una nueva metodología que combina *docking* molecular con inteligencia artificial para elaborar predicciones de actividad. La misma fue evaluada en primer lugar sobre un blanco modelo y luego extendida a un conjunto de 20 blancos moleculares.

¿Cuál es la idea?



Descomponer las interacciones ligando - proteína generadas a partir de las poses de *docking* de compuestos activos e inactivos, con el fin de utilizar esa información para entrenar un modelo que luego permita elaborar predicciones.

Metodología

Docking: DOCK 6.9, Vina, FRED
Descomposición de interacciones: DOCK 6.9
Datos de actividad: ChEMBL
Machine Learning: Sci-kit learn.
Todos los modelos fueron entrenados utilizando validación cruzada y los hiperparámetros fueron seleccionados tomando un consenso entre distintas métricas, como se puede ver en [este link](#).

Resultados

Prueba piloto sobre la Caseína Quinasa 2 (CK2)

La prueba preliminar de la metodología fue realizada sobre la quinasa CK2. En primer lugar, se *dockearon* los compuestos activos e inactivos obtenidos de ChEMBL y luego se obtuvo la "huella" de interacción de cada uno con el sitio activo. Esta información fue utilizada para entrenar distintos modelos, con diferentes algoritmos de decisión (regresión lineal, *random forests* y *support vector machines* (SVM)), siendo el último el de mejor resultado.

El modelo generado mostró una mejora significativa respecto al no entrenado (Figura 1, columna azul), demostrando además ser una técnica que puede utilizarse para "filtrar" los falsos positivos generados durante el *docking* (Figura 2).

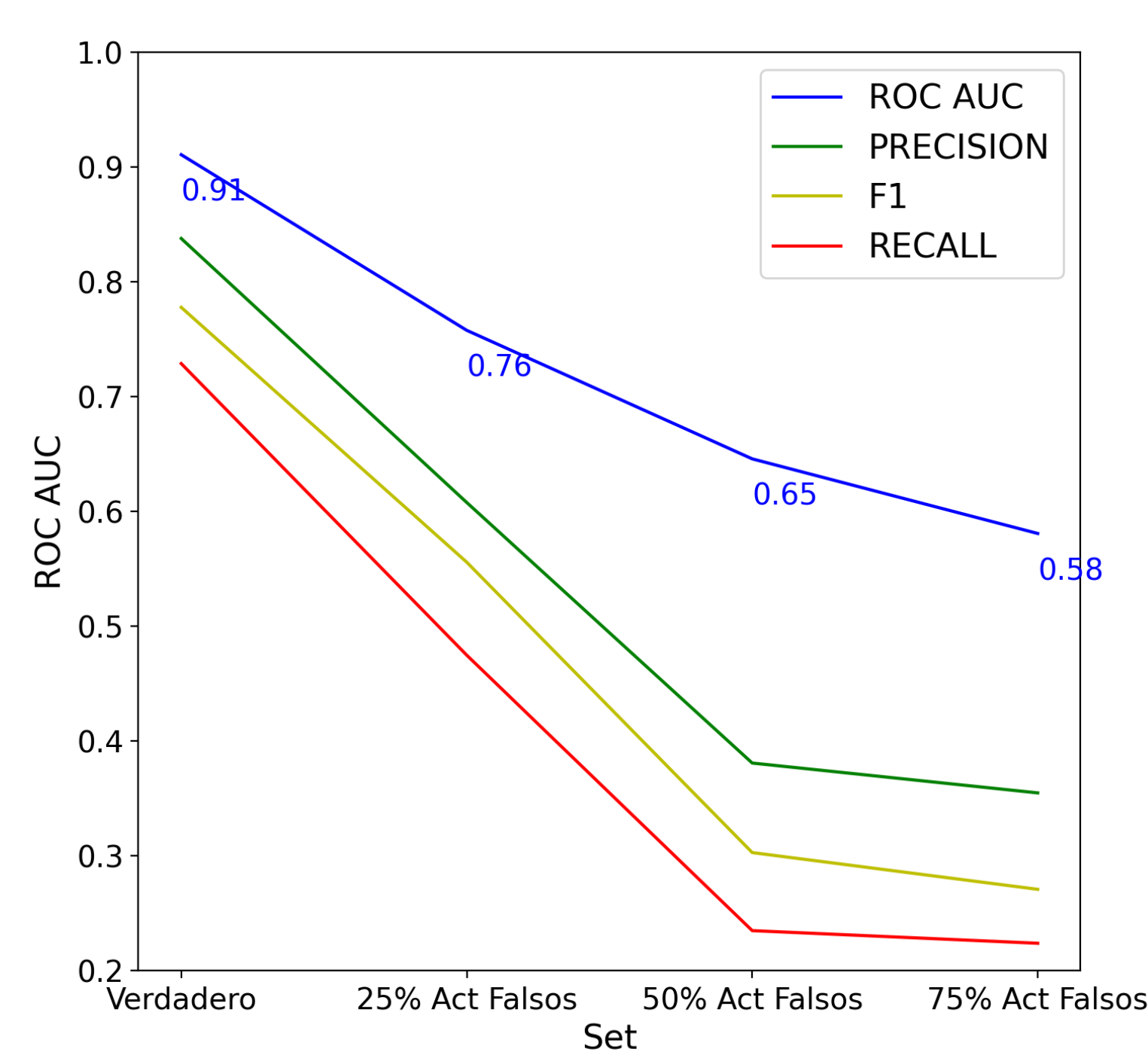


Figura 3

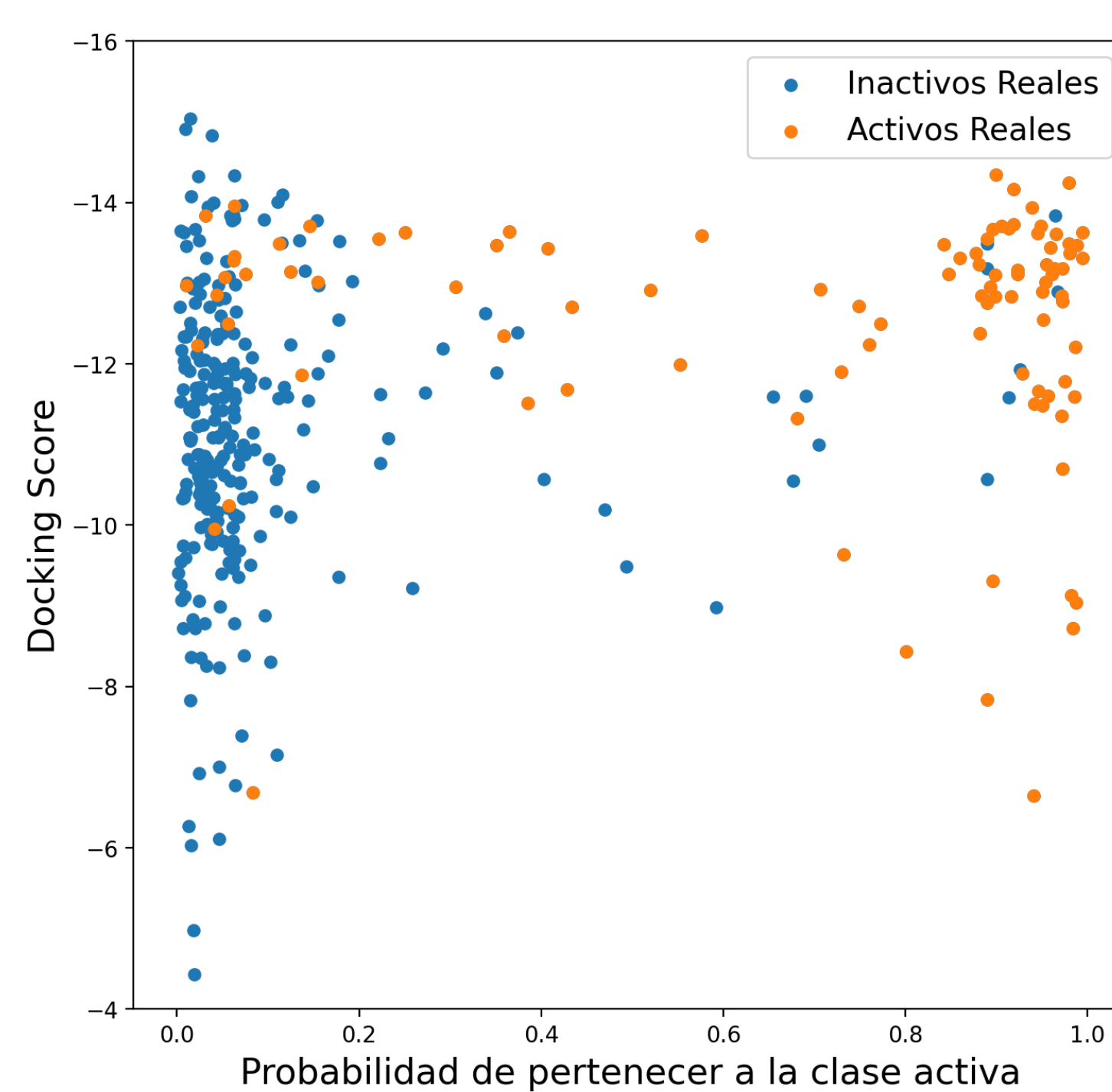


Figura 2

Ahora, ¿qué ocurre si al modelo le entrego información falsa? La Figura 3 muestra el resultado de sustituir los rótulos de distintas cantidades de ligandos activos por inactivos. Se puede observar que la performance del modelo en todas sus métricas empeora de manera gradual con la cantidad de falsos activos utilizada.

Desempeño del método con distintos software de docking molecular

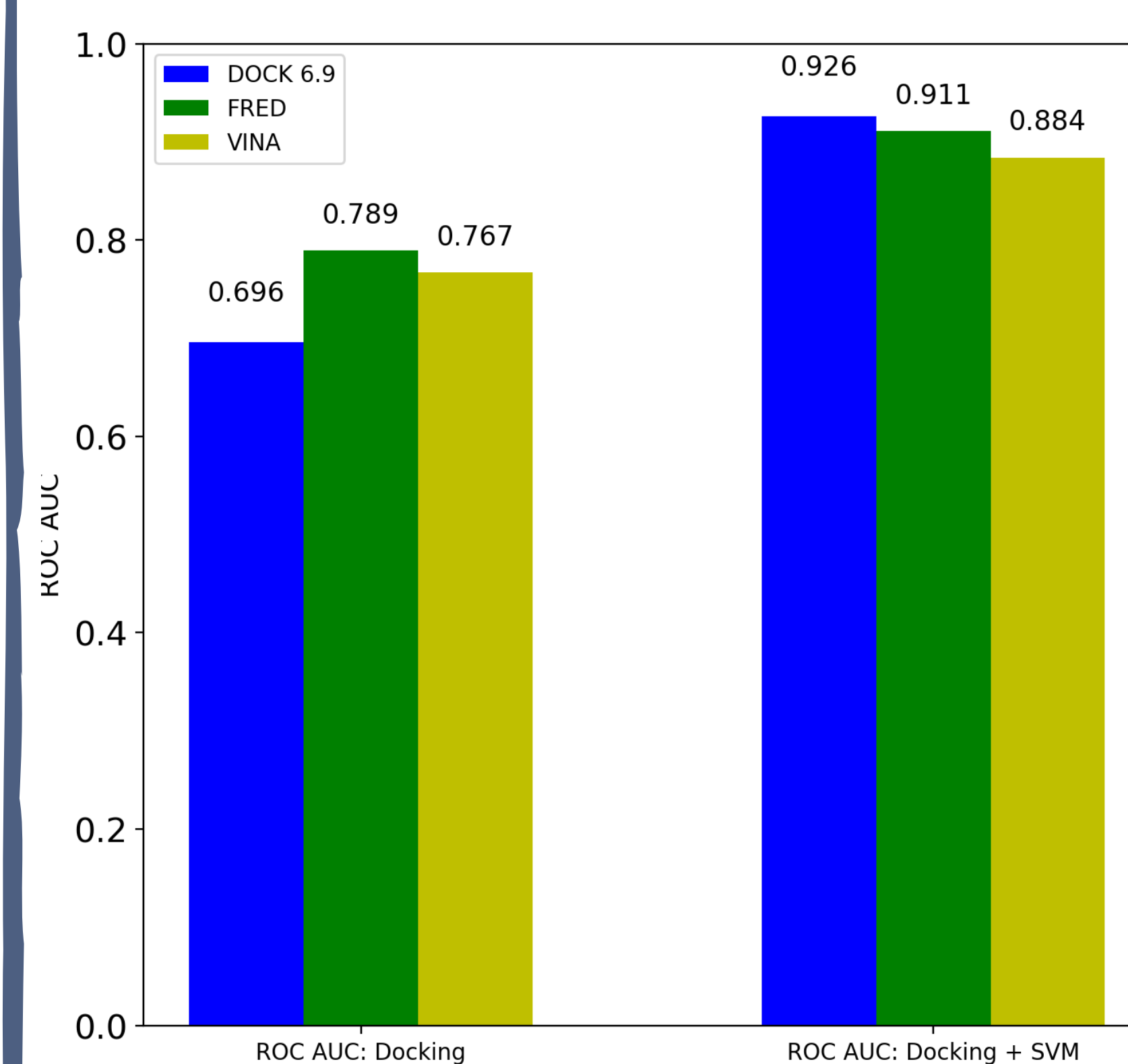


Figura 1

Otra aplicación interesante que tiene este enfoque es la posibilidad de acoplar la descomposición de residuos generada por DOCK 6.9 a cualquier conjunto de poses obtenido por otro programa de *docking*. Como se observa en la Figura 1, tanto para FRED como para Vina, la performance del modelo mejora significativamente de esta manera.

Extensión de la metodología a más blancos moleculares

Una vez realizada la prueba piloto sobre CK2, se procedió a expandir la metodología a un mayor número de blancos moleculares. Esto último se desarrolló con el objetivo de obtener una plataforma abierta que permita realizar *docking* inverso y elaborar predicciones de actividad.

Actualmente la misma cuenta con 20 blancos moleculares que fueron seleccionados tomando como base un enfoque similar realizado con GOLD y PADIF, incluidos en la suite de Schrödinger ([link](#)). En la Figura 4 se puede observar que el desempeño de la metodología planteada (DOCK 6.9 + SVM) es similar a la de referencia y muy superior al sólo uso de DOCK 6.9. El análisis sobre los blancos que mostraron peor desempeño (PI3KG y TPA) se puede atribuir a la gran cantidad de ligandos con macrociclos, que debieran ser tratados de manera independiente por el alto número de conformeros posible y la dificultad que esto provoca en la generación de poses.

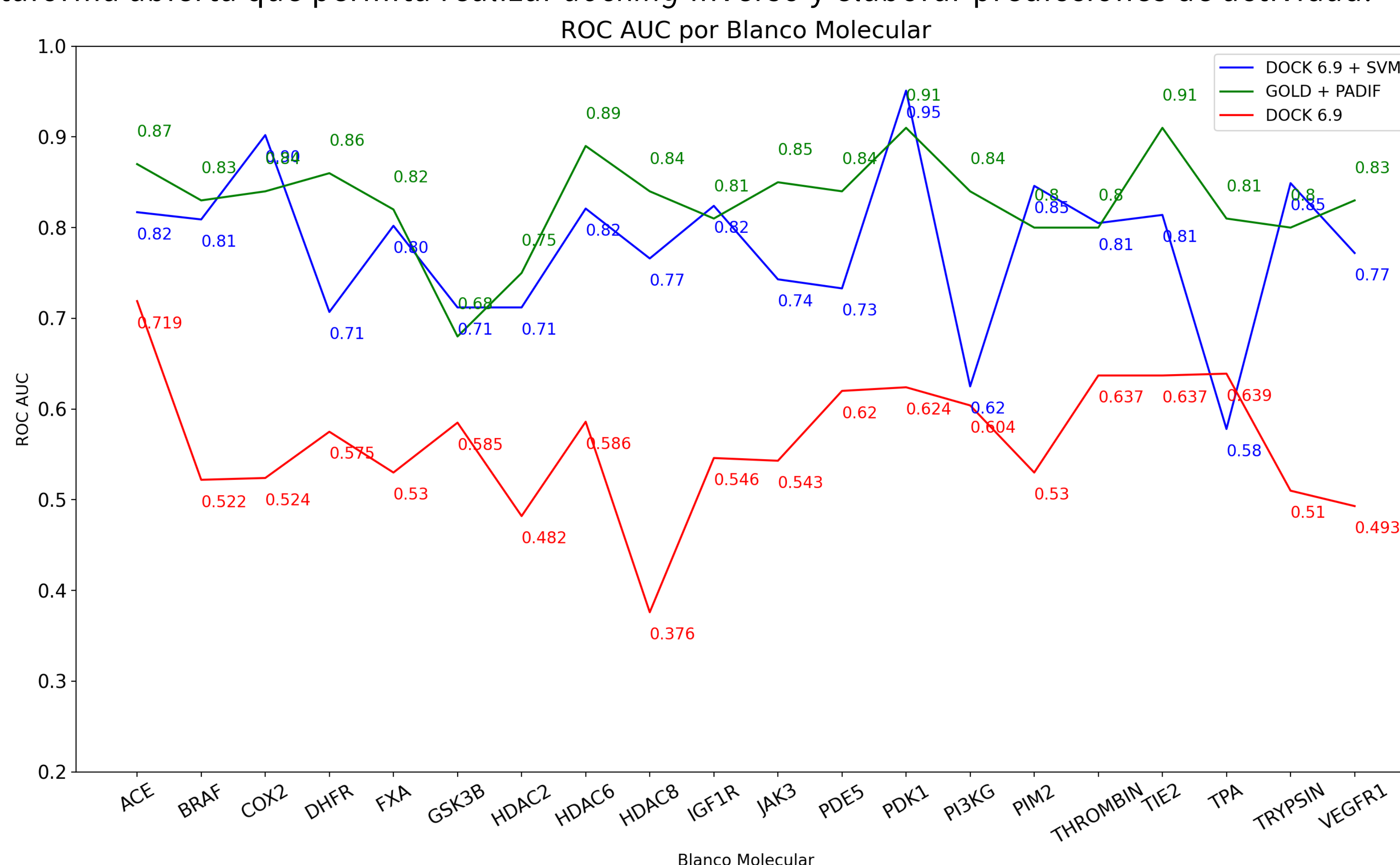
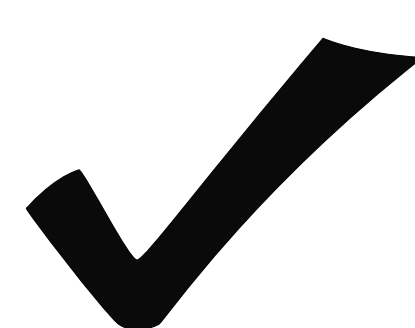


Figura 4

Conclusiones y Perspectivas



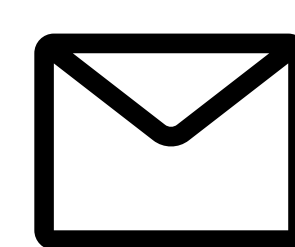
La metodología presentada en este trabajo permite desarrollar modelos de clasificación que pueden ser acoplados al *docking* molecular para mejorar su rendimiento predictivo, utilizando únicamente software gratuito.



SOON

Plataforma libre que permita *dockear* una dada molécula sobre múltiples blancos moleculares y elabore predicciones, conociendo para cada caso el desempeño del ligando en cada uno de los modelos.

[1] Departamento de Química Orgánica - FCEN - UBA
[2] UMYMFOR (UBA CONICET), Buenos Aires, Argentina
[3] CEQUINOR (UNLP CONICET), La Plata, Argentina



jffernandez@qo.fcen.uba.ar