

CORAL-DOCK

Evaluación científica multicohorte de recuperación de pose en Astex y LIT-PCBA

Roberto Ibanez

Francisco Arancibia

ISR-BIO

28 de julio de 2026

Resumen

Se evaluó CORAL-DOCK frente a AutoDock Vina 1.2.7 mediante redocking en dos cohortes estructurales independientes: 85 complejos Astex y 127 pares proteína–ligando analizables de un total de 129 estructuras LIT-PCBA (127/129; 98,4 %).

CORAL-DOCK recuperó una pose top-1 con $\text{RMSD} \leq 2 \text{ \AA}$ en 60/85 complejos Astex (70,6 %), frente a 56/85 (65,9 %) para Python Vina. En los 127 casos analizables de LIT-PCBA, obtuvo 87/127 (68,5 %) frente a 83/127 (65,4 %). Las diferencias pareadas no fueron estadísticamente concluyentes (McNemar exacto: $p = 0,219$ y $p = 0,424$). La mediana RMSD top-1 fue menor para CORAL-DOCK en Astex (1,139 frente a 1,230 Å) y LIT-PCBA (1,194 frente a 1,282 Å), sin diferencia global significativa. La cobertura top-9 favoreció a Python Vina en Astex (88,2 % frente a 84,7 %) y a CORAL-DOCK en LIT-PCBA (88,2 % frente a 79,5 %). Las afinidades fueron altamente concordantes en ambas cohortes. Los resultados respaldan una exactitud de redocking comparable y una señal descriptiva favorable para CORAL-DOCK en top-1, no superioridad universal.

1. Introducción y objetivos

La utilidad de un motor de docking depende de su capacidad para generar conformaciones cercanas a la pose cristalográfica y situarlas entre las soluciones mejor puntuadas. Astex ofrece una cohorte diversa de validación estructural; LIT-PCBA fue diseñado para cribado virtual y aporta una cohorte adicional de complejos cristalográficos. En este trabajo ambas fuentes se analizaron por separado para responder cuatro preguntas:

1. ¿Con qué frecuencia CORAL-DOCK recupera una pose cercana a la cristalográfica?
2. ¿La pose correcta aparece en primer lugar o entre candidatos secundarios?
3. ¿La señal observada en Astex se reproduce en los complejos estructurales LIT-PCBA?
4. ¿Cómo varía el resultado entre dianas y estratos de complejidad molecular?

2. Materiales y métodos

2.1. Cohortes estructurales

Astex aportó 85 pares de proteínas PDB y ligandos cristalográficos SDF. Ambos motores produjeron resultados para los 85 casos. La cohorte cubrió ligandos de 10 a 41 átomos pesados y de 0 a 16 torsiones activas.

En LIT-PCBA se analizaron 127 pares cristalográficos distribuidos entre 15 dianas, a partir de 129 estructuras disponibles. Los archivos de activos e inactivos LIT-PCBA no se utilizaron; por tanto, este estudio evalúa redocking y no enriquecimiento de cribado virtual. La nota explicativa al final del documento detalla los dos pares que no pudieron incorporarse al análisis.

2.2. Preparación y protocolo de docking

Se conservaron únicamente los registros ATOM del receptor, excluyendo ligandos, cofactores, iones y aguas. Se agregaron hidrógenos y se generaron archivos PDBQT con Open Babel 3.1.0. El centro de búsqueda correspondió al centro geométrico del ligando cristalográfico y la caja fue de $24 \times 24 \times 24$ Å. Se usó la semilla 181129 y se solicitaron hasta nueve poses. CORAL-DOCK empleó 8000 carriles de búsqueda y Python Vina exhaustiveness=8. Los receptores permanecieron rígidos.

2.3. Métricas y análisis estadístico

La métrica primaria fue éxito top-1, definido como RMSD de átomos pesados ≤ 2 Å respecto de la pose cristalográfica. Se calculó además recuperación acumulada top-3, top-5 y top-9, RMSD mediano y medio, cuartiles, concordancia de afinidad y resultados estratificados por complejidad.

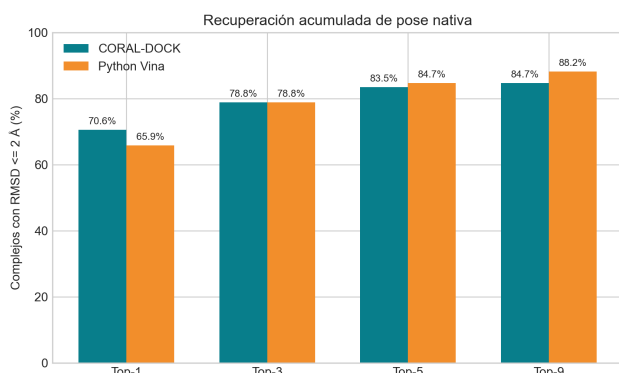
El RMSD se calculó emparejando átomos por orden en el mismo marco del receptor, sin corrección por simetría. Para proporciones se informaron intervalos Wilson al 95 %. La comparación pareada de éxito top-1 utilizó McNemar exacto; la distribución pareada de RMSD, Wilcoxon; y la diferencia de medianas, bootstrap con 20.000 remuestreos. Las cohortes no se fusionaron en una única tasa. El análisis LIT-PCBA no corrigió la agrupación de complejos dentro de una misma diana.

3. Resultados

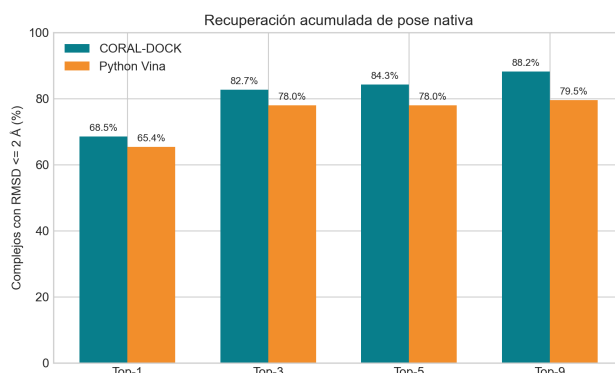
3.1. Recuperación de pose nativa

Cuadro 1: Recuperación acumulada de una pose con $\text{RMSD} \leq 2$ Å.

Cohorte	Nivel	CORAL-DOCK	Python Vina	Diferencia
Astex	Top-1	60/85 (70,6 %)	56/85 (65,9 %)	+4,7 pp
Astex	Top-3	67/85 (78,8 %)	67/85 (78,8 %)	0,0 pp
Astex	Top-5	71/85 (83,5 %)	72/85 (84,7 %)	-1,2 pp
Astex	Top-9	72/85 (84,7 %)	75/85 (88,2 %)	-3,5 pp
LIT-PCBA	Top-1	87/127 (68,5 %)	83/127 (65,4 %)	+3,1 pp
LIT-PCBA	Top-3	105/127 (82,7 %)	99/127 (78,0 %)	+4,7 pp
LIT-PCBA	Top-5	107/127 (84,3 %)	99/127 (78,0 %)	+6,3 pp
LIT-PCBA	Top-9	112/127 (88,2 %)	101/127 (79,5 %)	+8,7 pp



Astex



LIT-PCBA, 127 casos analizables

Figura 1: Recuperación acumulada de pose nativa por cohorte.

La dirección top-1 fue consistente y favoreció descriptivamente a CORAL-DOCK. La cobertura top-k fue dependiente del conjunto: Python Vina rescató más poses secundarias en Astex, mientras CORAL-DOCK amplió su diferencia hasta top-9 en LIT-PCBA.

3.2. Comparación pareada top-1

Cuadro 2: Contingencia pareada del éxito top-1.

Cohorte	Ambos	Sólo CORAL	Sólo Vina	Fallo ambos	McNemar
Astex	55	5	1	24	$p = 0,219$
LIT-PCBA	78	9	5	35	$p = 0,424$

Ninguna comparación alcanzó significación estadística. En LIT-PCBA, los nueve éxitos exclusivos de CORAL-DOCK incluyeron casos de ESR1, GBA, MAPK1, PKM2, PPARG y TP53; los cinco exclusivos de Vina correspondieron a ADRB2, KAT2A, PKM2, PPARG y TP53. La dirección repetida de la diferencia constituye una señal descriptiva y no evidencia suficiente de superioridad.

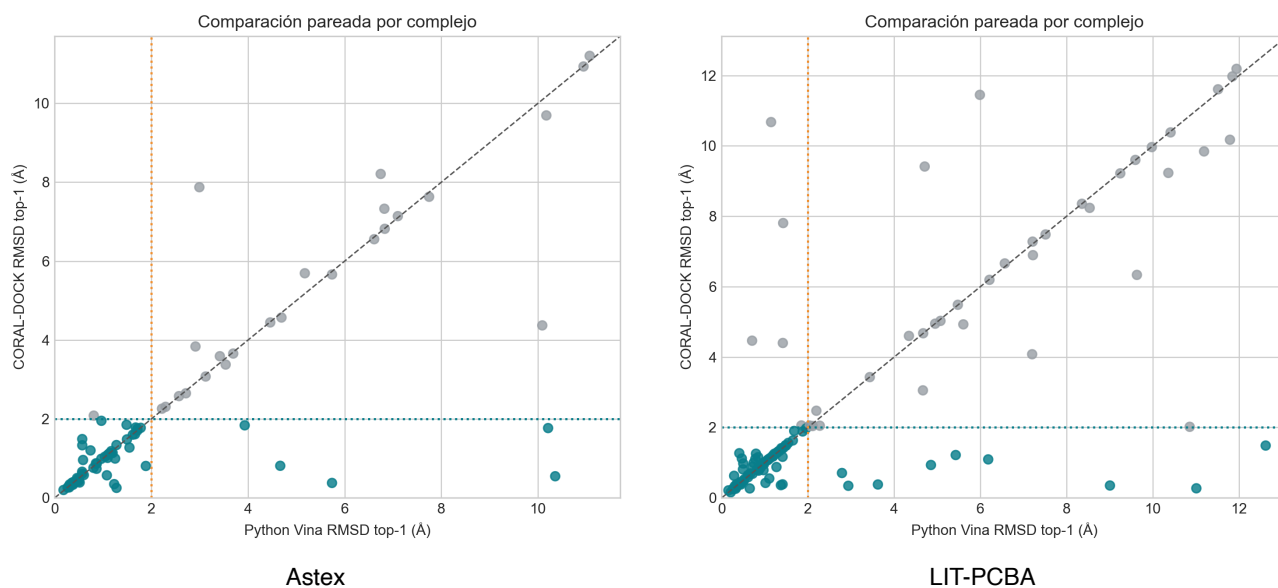


Figura 2: Comparación pareada del RMSD top-1 por complejo.

3.3. Distribución de RMSD

Cuadro 3: Estadísticos de RMSD top-1.

Cohorte	Estadístico	CORAL-DOCK	Python Vina
Astex	Mediana	1,139 Å	1,230 Å
Astex	Media	2,243 Å	2,561 Å
Astex	Primer–tercer cuartil	0,502–2,578 Å	0,569–3,417 Å
Astex	Menor RMSD pareado	50/85	35/85
LIT-PCBA	Mediana	1,194 Å	1,282 Å
LIT-PCBA	Media	2,680 Å	2,998 Å
LIT-PCBA	Primer–tercer cuartil	0,664–3,241 Å	0,727–4,690 Å
LIT-PCBA	Menor RMSD pareado	69/127	58/127

La mediana pareada CORAL-DOCK menos Python Vina fue $-0,006$ Å en Astex (IC95 % bootstrap: $-0,373$ a $+0,203$; Wilcoxon $p = 0,219$) y $-0,003$ Å en LIT-PCBA (IC95 %: $-0,264$ a $+0,011$; $p = 0,140$). Los intervalos incluyeron cero.

3.4. Cobertura conformacional y ranking

En Astex hubo 12 rescates adicionales para CORAL-DOCK y 19 para Python Vina entre top-1 y top-9. Siete casos fallaron con ambos motores aun en top-9. En LIT-PCBA hubo 25 rescates adicionales para CORAL-DOCK y 18 para Vina; once casos fallaron con ambos. Estos fallos compartidos son prioritarios para revisar protonación, tautomería, aguas funcionales, cofactores, flexibilidad del receptor y suficiencia de búsqueda.

3.5. Concordancia de afinidad

Cuadro 4: Concordancia de afinidad de la pose top-1.

Métrica	Astex	LIT-PCBA
Pearson r	0,990	0,988
Spearman ρ	0,978	0,977
Delta medio CORAL-DOCK – Vina	+0,044	-0,059
Error absoluto medio (kcal/mol)	0,147	0,186
RMSE (kcal/mol)	0,276	0,478
Diferencia absoluta $\leq 0,5$ kcal/mol	81/85	118/127

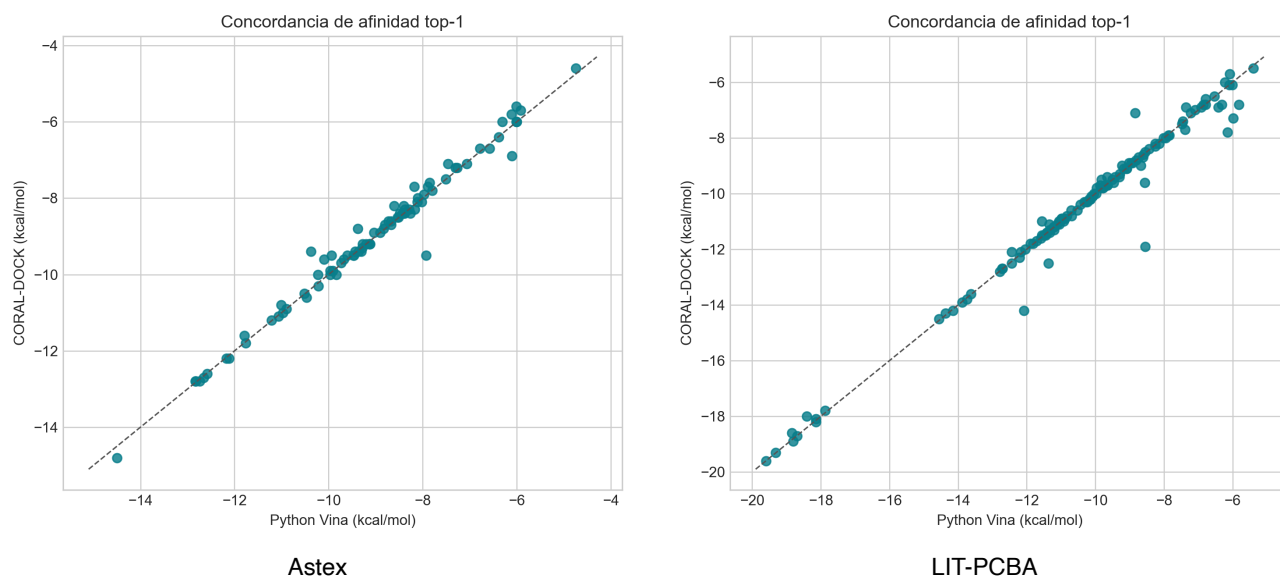


Figura 3: Concordancia de afinidad entre CORAL-DOCK y Python Vina.

La concordancia energética fue alta en ambas cohortes. Debido a que los sistemas derivan de la familia Vina, esto demuestra consistencia numérica entre implementaciones, no exactitud termodinámica frente a afinidades experimentales.

3.6. Heterogeneidad por diana en LIT-PCBA

Cuadro 5: Éxito top-1 y top-9 por diana LIT-PCBA.

Diana	<i>n</i>	C1	V1	C9	V9	Diana	<i>n</i>	C1	V1	C9	V9
ADRB2	8	75,0	87,5	100,0	100,0	MAPK1	14	71,4	64,3	100,0	78,6
ALDH1	8	50,0	50,0	50,0	75,0	MTORC1	11	100,0	100,0	100,0	100,0
ESR1_ago	15	86,7	86,7	100,0	100,0	OPRK1	1	100,0	100,0	100,0	100,0
ESR1_ant	15	80,0	73,3	86,7	73,3	PKM2	9	55,6	44,4	100,0	88,9
FEN1	1	0,0	0,0	100,0	0,0	PPARG	15	53,3	46,7	66,7	60,0
GBA	6	66,7	33,3	100,0	66,7	TP53	6	50,0	50,0	100,0	50,0
IDH1	13	53,8	53,8	76,9	76,9	VDR	2	100,0	100,0	100,0	100,0
KAT2A	3	33,3	66,7	66,7	66,7						

C1 y V1 indican top-1 para CORAL-DOCK y Vina; C9 y V9, top-9. La heterogeneidad entre dianas fue mayor que la diferencia agregada entre motores. Los porcentajes con *n* pequeño son inestables y no deben generalizarse por blanco.

3.7. Influencia de la complejidad molecular

Cuadro 6: Éxito top-1 estratificado por complejidad del ligando.

Cohorte	Estrato	<i>n</i>	CORAL-DOCK	Python Vina
Astex	0–5 torsiones	33	69,7 %	63,6 %
Astex	6–10 torsiones	35	80,0 %	77,1 %
Astex	> 10 torsiones	17	52,9 %	47,1 %
LIT-PCBA	0–5 torsiones	41	61,0 %	56,1 %
LIT-PCBA	6–10 torsiones	51	76,5 %	68,6 %
LIT-PCBA	> 10 torsiones	35	65,7 %	71,4 %
Astex	≤ 20 átomos pesados	34	61,8 %	52,9 %
Astex	21–35 átomos pesados	48	75,0 %	75,0 %
Astex	> 35 átomos pesados	3	100,0 %	66,7 %
LIT-PCBA	≤ 20 átomos pesados	19	42,1 %	31,6 %
LIT-PCBA	21–35 átomos pesados	86	73,3 %	69,8 %
LIT-PCBA	> 35 átomos pesados	22	72,7 %	77,3 %

Astex mostró menor recuperación con más de 10 torsiones. LIT-PCBA no presentó una degradación monotónica con flexibilidad o tamaño. Los estratos combinan dianas distintas y no permiten atribuir causalidad a una sola propiedad molecular.

4. Discusión

CORAL-DOCK mostró una señal top-1 favorable y reproducible en dirección: +4,7 puntos porcentuales en Astex y +3,1 en LIT-PCBA. También obtuvo una mediana RMSD menor en ambas cohortes. Sin embargo, las pruebas pareadas no fueron concluyentes y los intervalos de la diferencia de medianas incluyeron cero. La interpretación respaldada es comparabilidad bajo este protocolo, no superioridad universal.

Top-1 representa la utilidad de aceptar automáticamente la primera solución; top-5/top-9, la posibilidad de rescatar una conformación mediante inspección o re-ranking. La inversión del patrón top-9 entre Astex y LIT-PCBA demuestra que ranking y cobertura deben informarse por separado y en más de una cohorte.

LIT-PCBA fue concebido como benchmark de cribado virtual. Este estudio empleó sólo complejos cristalográficos para redocking: no utilizó sus activos e inactivos y no calculó AUROC, PR-AUC, BEDROC ni enriquecimiento temprano. No se puede inferir capacidad de discriminación activo–inactivo a partir de estos resultados.

Las limitaciones del protocolo incluyen una sola semilla, RMSD sin corrección por simetría, receptor rígido, preparación con Open Babel, exclusión de aguas/cofactores y esfuerzos de búsqueda no matemáticamente equivalentes. La agrupación por diana en LIT-PCBA tampoco fue modelada.

5. Conclusiones

1. El éxito top-1 favoreció descriptivamente a CORAL-DOCK en Astex (70,6 % frente a 65,9 %) y LIT-PCBA (68,5 % frente a 65,4 %).
2. Las diferencias top-1 no fueron estadísticamente concluyentes en ninguna cohorte.
3. La cobertura top-9 favoreció a Vina en Astex y a CORAL-DOCK en LIT-PCBA.
4. La afinidad fue numéricamente consistente entre implementaciones, pero no fue validada contra datos experimentales.
5. Los datos respaldan exactitud de redocking comparable y una señal top-1 favorable para CORAL-DOCK bajo este protocolo.

Referencias

1. Hartshorn MJ, Verdonk ML, Chessari G, et al. Diverse, high-quality test set for the validation of protein-ligand docking performance. *J Med Chem*. 2007;50:726–741. doi:10.1021/jm061277y.
2. Tran-Nguyen VK, Jacquemard C, Rognan D. LIT-PCBA: An unbiased data set for machine learning and virtual screening. *J Chem Inf Model*. 2020;60:4263–4273. doi:10.1021/acs.jcim.0c00155.
3. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*. 2010;31:455–461. doi:10.1002/jcc.21334.
4. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model*. 2021;61:3891–3898. doi:10.1021/acs.jcim.1c00203.
5. Buttenschoen M, Morris GM, Deane CM. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chem Sci*. 2024. doi:10.1039/D3SC04185A.

Nota explicativa sobre la cohorte LIT-PCBA

De las 129 estructuras LIT-PCBA disponibles, 127 pudieron incorporarse al análisis pareado. Los pares IDH1--4xs3 y MAPK1--5ax3 no pudieron analizarse debido a una incompatibilidad técnica durante la lectura de sus archivos de entrada PDBQT. En consecuencia, no se generó una comparación completa entre motores para esos dos pares. No corresponden a fallos de docking ni de recuperación de pose y no fueron incluidos en los denominadores de las métricas; todos los resultados LIT-PCBA del artículo se refieren a los 127 casos analizables.