



a cura di Silvia Cauteruccio e Monica Civera

Dipartimento di Chimica

Università di Milano

silvia.cauteruccio@unimi.it

monica.civera@unimi.it

Un'architettura *open source* per predire l'affinità di legame

A sei mesi dall'uscita di Boltz-1 (pre-print: [10.1101/2024.11.19.624167](https://arxiv.org/abs/2011.1101)), un *tool open source* capace di predire la struttura 3D di biocomplessi con un'accuratezza paragonabile a quella di AlphaFold3, lo stesso gruppo di ricercatori del MIT, in collaborazione con Recursion (una compagnia TechBio dotata del supercomputer BioHive-2), presenta Boltz-2, (<https://github.com/jwohlwend/boltz>) capace di predire anche l'affinità di legame. Il *tool*, rilasciato sotto licenza *open* del MIT, include il codice sorgente, i pesi dei modelli, dataset e benchmark, con l'obiettivo di promuovere la collaborazione globale, accelerare le scoperte scientifiche e fornire una piattaforma solida e condivisa. Rispetto alla prima versione, Boltz-2 calcola l'affinità di legame per piccole molecole avvicinandosi all'accuratezza di metodi computazionali più sofisticati ed impegnativi (come FEP), pur mantenendo una velocità di calcolo circa 1.000 volte superiore. Questo risultato è stato realizzato grazie agli sforzi applicati nella parte di *data curation* utilizzata per il *training* della rete. Le grandi quantità di dati di affinità disponibili nei database pubblici sono spesso caratterizzate da un'elevata eterogeneità e dalla presenza di rumore. Per migliorare le prestazioni del modello, gli autori hanno selezionato esclusivamente saggi di alta qualità provenienti da fonti affidabili come PubChem e ChEMBL, standardizzando i dati e riducendo il rischio di *overfitting* tramite l'inclusione di *decoys*, ovvero molecole inattive utilizzate come controlli negativi. Come AlphaFold3, anche in Boltz la struttura delle proteine viene dedotta a partire dalla sequenza di amminoacidi, sfruttando l'allineamento multiplo (MSA) addestrato su diversi complessi presenti nel PDB. I ligandi sono rappresentati tramite stringhe SMILES, quindi da una codifica 2D. Per la

predizione delle coordinate 3D dei complessi ligando-proteina, anche Boltz utilizza un modello di diffusione basato su trasformatori multi-risoluzione per il *denoising* delle coordinate atomiche, ma, rispetto ad AlphaFold3, introduce alcuni vincoli di distanza ed allena il sistema includendo dati strutturali derivati da simulazioni di dinamica molecolare. Per identificare la tasca di legame durante la fase di training, si utilizzano i dati strutturali, tuttavia, i residui etichettati come parte del sito attivo vengono selezionati in modo casuale tra quelli situati entro 6 Å dagli atomi del ligando, e questa informazione viene utilizzata solo nel 30% dei cicli di training, con l'obiettivo di migliorare la capacità di generalizzazione del modello.

Nella competizione CASP16 (*Critical Assessment of Structure Prediction*) dedicata alla predizione delle affinità di legame, Boltz-2 ha superato tutte le altre soluzioni computazionali. Inoltre, ha ottenuto risultati migliori rispetto a molti software di docking e algoritmi di *machine learning* nella classificazione di molecole in composti attivi e *decoys* (Fig. 1)

Reazioni di *cross-coupling* catalizzate da "ppm" di palladio

I (pre)catalizzatori utilizzati nelle reazioni di *cross coupling* contengono generalmente metalli preziosi e, tra questi, il palladio fornisce ancora oggi sistemi catalitici di gran lunga superiori in termini di efficienza e selettività rispetto ai catalizzatori

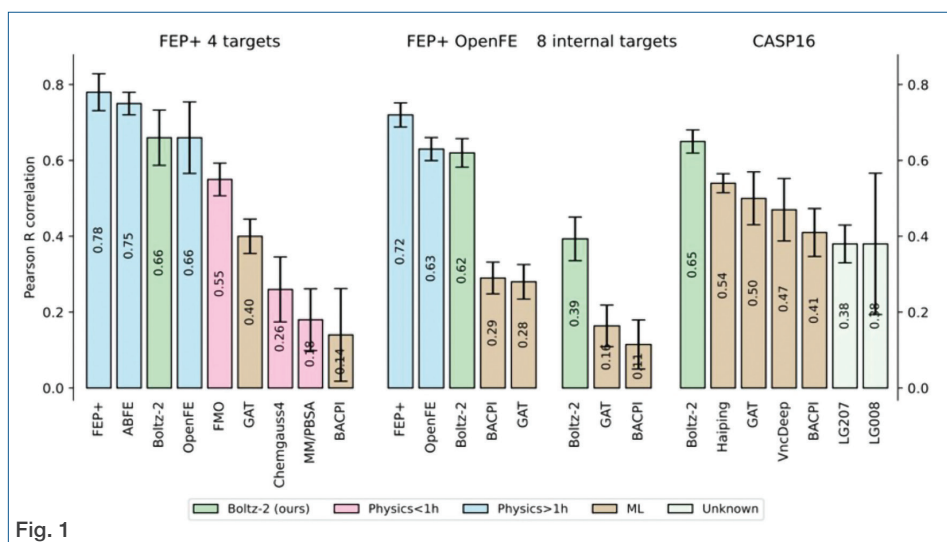


Fig. 1



a base di metalli più abbondanti e a basso costo (ferro, manganese, rame o cobalto), sebbene questo divario si stia riducendo. Sembrerebbe quindi che del palladio non possiamo farne a meno per ora, anche se il suo costo inizia a rappresentare un ostacolo non indifferente, oltre al fatto che catalizzatori al palladio si ritrovano in passaggi chiave di sintesi brevettate e attualmente in uso, per i quali è spesso difficile sviluppare delle valide alternative meno costose.

Accanto, quindi, all'intenso studio attualmente in atto per mettere a punto processi catalitici con metalli più sostenibili, un approccio alternativo alle reazioni catalitiche su cui si è focalizzata la ricerca negli ultimi anni consiste nell'utilizzare concentrazioni di palladio nell'ordine delle parti per milione (ppm), spesso associando alle concentrazioni in ppm di Pd il concetto di sostenibilità. Fairlamb nella sua review del 2022 [J.S. Fairlamb *et al.*, *Org. Process Res. Dev.*, 2022, **26**, 2240] riporta un interessante approfondimento per valutare se il palladio "in ppm" sia un descrittore praticamente utile delle quantità di catalizzatore nelle reazioni e come questo possa essere usato come parametro di sostenibilità. La sostenibilità di un processo infatti dipenderà in larga misura dalla scala operativa, e l'uso del palladio a "livello di ppm" (in isolamento) può non descrivere necessariamente un processo sostenibile. A livello di ricerca accademica, le operazioni su piccola scala utilizzeranno livelli relativamente bassi di palladio (indipendentemente dall'utilizzo della terminologia del palladio a "livello di ppm") e continueranno a essere sostenibili, in particolare se il palladio di scarto può essere raccolto e rigenerato, mentre i processi catalitici che necessitano di palladio su larga scala potrebbero non essere sostenibili a meno che il suo impiego non sia in linea con la domanda e l'offerta di produzione. In questa review viene inoltre sottolinea-

ta la necessità di uniformare i calcoli sulla quantità di palladio in ppm in una reazione, considerando tutti i componenti, ovvero tenendo presente che i livelli di ppm dipendono in larga misura dal numero di altre specie utilizzate in ciascuna reazione e dal volume di solvente. Il volume di solvente gioca un ruolo importante nel determinare il valore in ppm complessivo in una reazione, poiché spesso è superiore al 90% della miscela. D'altra parte, comprendere la quantità di Pd o di qualsiasi specie (pre)catalitica presente nelle reazioni può aiutare a capire il livello di Pd catalitico attivo e quali livelli di (pre)catalizzatore sono effettivamente necessari. La letteratura degli ultimi mesi riporta, ad esempio, una reazione altamente efficiente e regioselectiva di arilazione di alcoli insaturi ad opera di ioduri arilici per la preparazione di composti carbonilici diversamente funzionalizzati [H. Ge *et al.*, *Chem. Sci.*, 2025, DOI: [10.1039/d5sc02745d](https://doi.org/10.1039/d5sc02745d)], utilizzando $\text{Pd}(\text{OAc})_2$ (50 ppm) e formiato di sodio in toluene a 125 °C (Fig. 2a). L'impiego di mezzi micellari in acqua è un altro metodo ben noto per condurre reazioni catalitiche con metalli di transizione con elevata resa e selettività, consentendo spesso il riciclo del catalizzatore. Recentemente [J. Zhang *et al.*, *Org. Lett.*, 2025, **27**, 4343] è stata riportata la sintesi di catalizzatori polimerici micellari contenenti palladio (Fig. 2b), i quali sono in grado di promuovere efficientemente reazioni di Suzuki-Miyaura tra bromuri (etero)arilici ed acidi arilboronici in acqua. In questo caso sono state ottenute rese molto elevate nell'arco delle due ore utilizzando un carico catalitico relativamente basso (100 ppm).

