



LE PROMESSE DELL'INTELLIGENZA ARTIFICIALE NEL *DRUG DISCOVERY*

I metodi di computer-aided drug design (CADD) sono ormai indispensabili nel processo di drug discovery e hanno favorito lo sviluppo di numerosi farmaci approvati. L'integrazione dell'intelligenza artificiale nei metodi CADD, resa possibile dalla crescente disponibilità di dati e dal miglioramento dei modelli, apre nuove prospettive nella previsione delle strutture dei target terapeutici, delle proprietà molecolari e nella generazione di nuove molecole.



Fig. 1 - La copertina del numero di ottobre 1981 di *Fortune* dedicata al CADD

L'intelligenza artificiale (IA) ha trovato numerose applicazioni nel settore farmaceutico, tra cui la previsione delle reazioni avverse ai farmaci, il supporto alle decisioni cliniche, l'analisi di mercato e l'elaborazione di strategie di posizionamento dei prodotti farmaceutici. È soprattutto nella progettazione di nuovi farmaci che è possibile percepire il profondo cambiamento che gli approcci IA, *machine learning* (ML) e la recente estensione della sua *toolbox*, il *deep learning* (DL), possono apportare. L'integrazione dei dati genomici con altre fonti cliniche e omiche, supportata dall'automazione tramite l'IA, consente infatti di ottenere informazioni più accurate e di correlare le varianti genetiche con gli obiettivi clinici. Questo approccio ha rivoluzionato il *drug discovery*, spostandolo dai tradizionali *target* convenzionali a una strategia personalizzata, basata sul profilo genomico del paziente. In modo inevitabile, l'IA è stata integrata anche nella tecnologia CADD (*Computer-Aided Drug Design*), i cui progressi hanno contribuito in modo significativo ad accelerare il processo di scoperta di nuovi farmaci, rendendo possibile selezionare e sintetizzare rapidamente e a costi ridotti ampie librerie di composti [1] (Fig. 1).



Le potenzialità di queste tecniche computazionali emergono negli studi condotti nel 2002 dagli scienziati di Pharmacia (ora Pfizer) che utilizzando lo *screening* virtuale (VS) di librerie di composti chimici per scoprire nuovi inibitori della tirosina fosfatasi-1B, un enzima associato al diabete, identificavano 365 composti, di cui 127 con una efficace capacità di inibizione, ottenendo un tasso di successo superiore al 34%. Al contrario, l'*high throughput screening* tradizionale di 400 mila composti riusciva a identificare soltanto 85 molecole promettenti, con un tasso di successo decisamente inferiore e pari allo 0,021% [2]. Sunitinib, Venetoclax sono solo alcuni tra gli oltre 70 farmaci attualmente in commercio, a partire dal Captopril (1981), che devono la loro scoperta all'utilizzo di avanzate tecniche computazionali e soprattutto alla loro efficace integrazione. Il successo di questi studi dimostra che le metodologie *in silico* possono essere determinanti non solo nell'identificare le più promettenti molecole candidate a diventare farmaci, ma anche nel guidarne l'intero processo di sviluppo fino alla generazione dei prodotti farmaceutici destinati al mercato [3].

I dati presenti in letteratura evidenziano un aumento costante e progressivo nel numero di farmaci scoperti grazie all'impiego delle tecniche CADD. Questo incremento può essere attribuito a diversi fattori, tra cui i progressi raggiunti nella modellistica molecolare e nella chimica computazionale, che hanno reso queste tecniche sempre più precise ed efficaci. Il dato può anche trovare spiegazione nella crescente diffusione della ricerca multidisciplinare, che facilita sicuramente una maggiore integrazione tra diverse discipline scientifiche, favorendo l'adozione di approcci innovativi e collaborativi nella scoperta di nuovi farmaci. Il Premio Nobel per la Chimica del 2013, assegnato a Martin Karplus, Michael Levitt e Arieh Warshel, ha evidenziato i progressi rivoluzionari raggiunti nella comprensione dei processi chimico-fisici e biologici a livello atomico. Il loro lavoro pionieristico ha dimostrato il potenziale delle tecniche computazionali nell'integrare la meccanica quantistica con la fisica classica, permettendo di simulare e prevedere le interazioni molecolari complesse con estrema precisione e aprendo la strada ad applicazioni pratiche, come ad esempio il *drug design*. Con una quantità sempre crescente di dati sull'attività biologica e sulle proprietà dei compo-

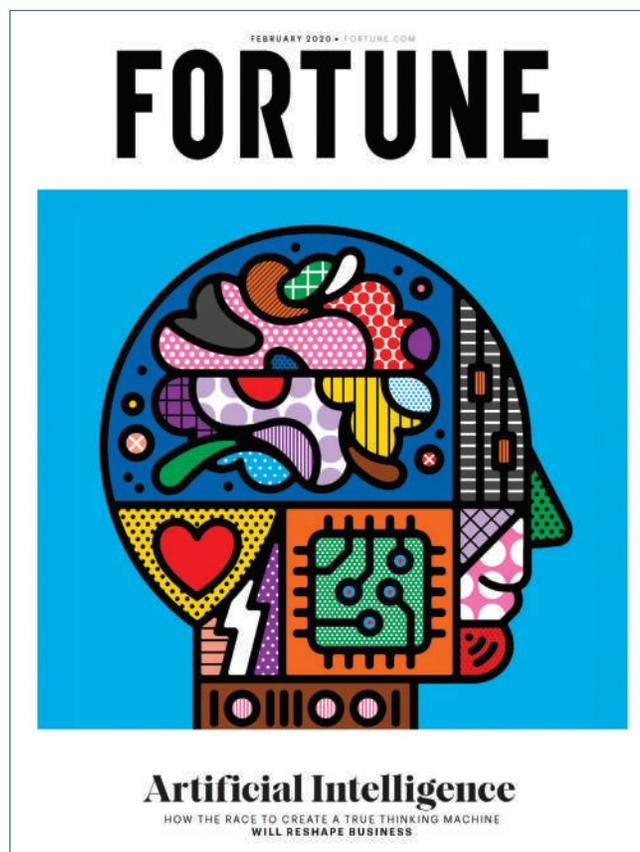


Fig. 2 - La copertina del numero di febbraio 2020 di *Fortune* dedicata all'IA

sti chimici, i metodi di IA si sono presentati quindi come una opportunità senza precedenti. Numerose aziende farmaceutiche hanno sviluppato modelli di *business* basati sulla combinazione di approcci di modellistica molecolare *physics based* con il *deep learning* e, sebbene sia ancora presto per aspettarsi farmaci approvati col supporto dell'IA, queste nuove metodologie stanno già producendo un numero crescente di composti candidati farmaci, promettendo tempi "*target-to-lead*" di soli 1-2 mesi e "*target-to-clinic*" inferiori a un anno (Fig. 2).

Non è un caso quindi che, a dieci anni circa dal Nobel del 2013, il Premio Nobel per la Chimica sia stato assegnato a John Jumper e Demis Hassabis, che hanno guidato lo sviluppo di AlphaFold, e a David Baker dell'Università di Washington per i suoi successi nella progettazione computazionale delle proteine. In particolare, AlphaFold, lo strumento di IA sviluppato da Google DeepMind, ha segnato un momento di svolta nella modellistica molecolare e nel *drug design* riuscendo a prevedere, dalle

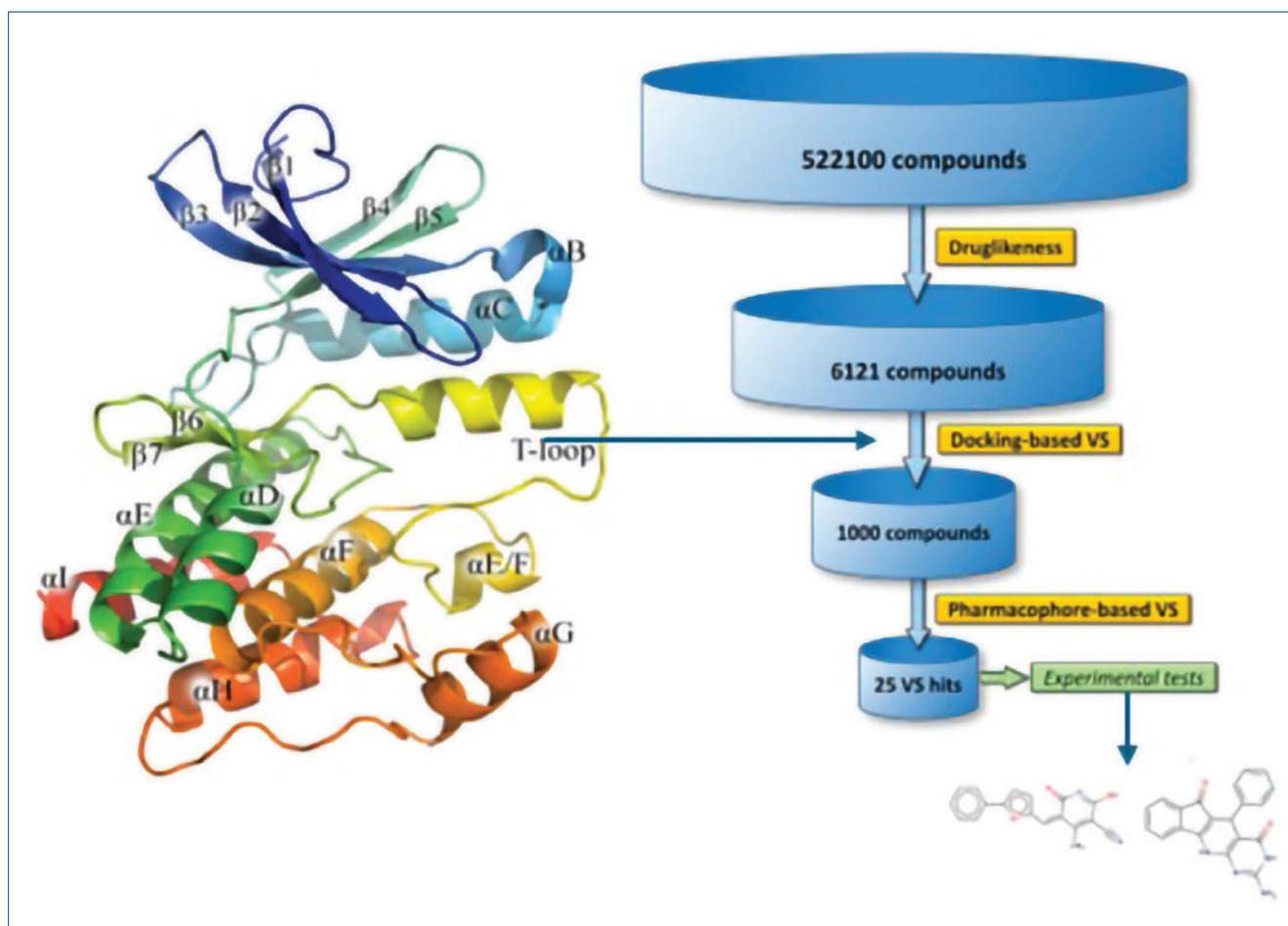


Fig. 3 - Strategia di VS per l'identificazione di molecole attive contro i tumori solidi (brevetto WO2019193494A1), adattata da [7]

informazioni di sequenza, la struttura tridimensionale delle proteine [4], e recentemente anche dei complessi ligando-recettore [5], con un'accuratezza estremamente elevata, paragonabile a quella ottenuta sperimentalmente attraverso processi che richiedono da mesi ad anni.

L'identificazione del target e la conoscenza della sua struttura tridimensionale rappresentano i primi e fondamentali passaggi nella progettazione e nello sviluppo di farmaci efficaci. Questo processo deriva dalla combinazione di approcci diversi provenienti da biologia strutturale, biologia molecolare, genomica e proteomica, la cui integrazione consente sia di comprendere i meccanismi molecolari alla base della patogenesi, sia di analizzare i *pathways* correlati, aspetti determinanti per lo sviluppo di nuovi approcci terapeutici. Le stesse tecniche CADD *ligand-based*, quali ad esempio la relazione quantitativa struttura-attività (QSAR) e gli studi farmacoforici, quando impiegate in combinazione con gli approcci *structure-based* si sono rivelate particolarmente efficaci nello scoprire nuovi candidati clinici [6, 7] (Fig. 3).

In assenza di strutture sperimentali ottenute da cristallografia, Cryo-EM o NMR, si è fatto ampio ricorso alle tecniche computazionali per generare modelli 3D di proteine e complessi molecolari e, attraverso simulazioni di dinamica molecolare, ottenere una visione dettagliata a livello atomico dei processi dinamici che regolano il riconoscimento molecolare e le perturbazioni dei *pathways* coinvolti indotte dalle piccole molecole. Recentemente agli approcci computazionali tradizionali di *modeling* proteico per omologia e *de novo modeling* si sono quindi affiancati quelli basati sull'apprendimento automatico, grazie alla crescita esponenziale dei dati, all'aumento della potenza di calcolo (es. GPU, TPU) e all'ottimizzazione continua degli algoritmi quali le reti neurali convoluzionali (cNN) e le reti generative avversarie (GAN) (Fig. 4).

Oltre ad AlphaFold, attualmente esistono molti metodi di modellazione proteica basati sul DL, tra cui RoseTTAFold [9] ed ESMFold [10]. Quest'ultimo offre un ampio *database* di previsioni strutturali delle proteine, che include 617 milioni di strutture proteiche metagenomiche. Rispetto al *modeling* per

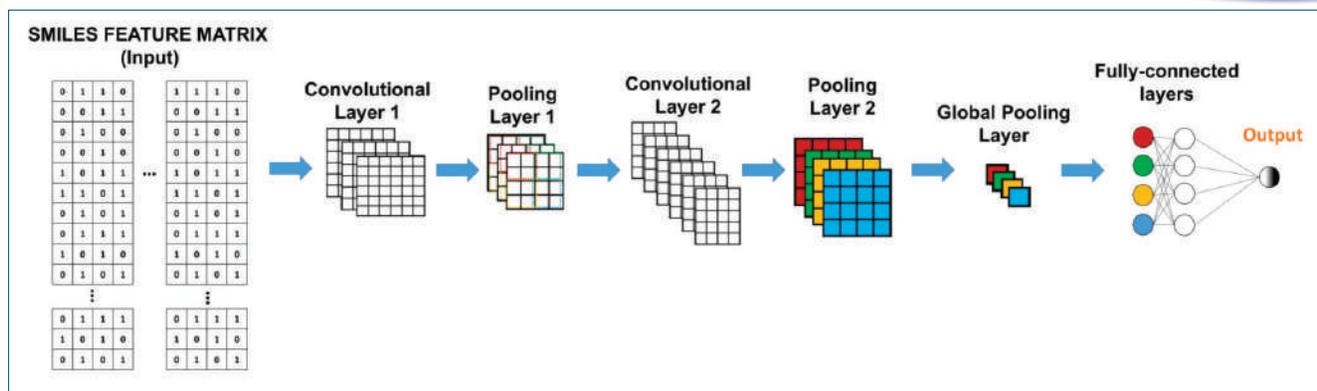


Fig. 4 - Schema della CNN utilizzata per generare il modello QSAR utile a identificare inibitori dell'interazione proteina-proteina [8]

omologia e al *de novo modeling* i metodi basati sul DL sono guidati dai dati e non presentano quindi le limitazioni del *modeling* proteico basato sulla conoscenza di strutture template, e dello stesso *de novo modeling* che è ancora applicabile solo a proteine di piccole dimensioni a causa dell'enorme numero di possibili stati conformazionali.

Non è soltanto nell'affrontare il "*protein folding problem*", tuttavia, che possiamo riconoscere le enormi potenzialità dell'IA nel *drug discovery*. Gli approcci di *deep learning* (DL) sono stati applicati alla predizione dell'interazione farmaco-recettore, l'IA generativa è stata impiegata per progettare nuove molecole, il riposizionamento dei farmaci (*drug repurposing* o *drug repositioning*) può essere guidato dall'IA. Una delle principali sfide degli approcci CADD è la crescente richiesta di risorse computazionali necessarie per eseguire il *ligand-based* (VS) di librerie chimiche in costante espansione grazie ai recenti progressi nella sintesi automatizzata e nella robotica. Tuttavia, i costi computazionali elevati rendono il *docking* convenzionale di librerie ultra-grandi spesso inaccessibile per la maggior parte della comunità scientifica. Per superare questo ostacolo, sono state sviluppate diverse soluzioni basate sull'apprendimento automatico che non richiedono grandi risorse computazionali. Ad esempio, *Deep Docking* (DD) esegue lo *screening* di miliardi di molecole utilizzando una rete neurale *deep* per escludere molecole per le quali si prevede un basso valore di *docking score*, seguendo una procedura di addestramento iterativa che riduce il numero di molecole necessarie per il *docking* a una piccola percentuale della libreria (solitamente, l'1%) [11]. In questo modo DD velocizza lo *screening* e allo stesso tempo, non introducendo una nuova funzione di *scoring*, non compromette l'accuratezza che viene garantita dall' algoritmo di *docking* utilizzato. Nel VS, che identifica i composti che si prevede si-

ano attivi contro specifici *target* sulla base della loro somiglianza con molecole attive già note, i modelli GCN (*Graph Convolutional Network*) hanno evidenziato ottime prestazioni, superiori a quelle tipicamente fornite dai classificatori ML classici, nel prevedere le proprietà molecolari. Modelli GCN sono stati impiegati per prevedere le proprietà di assorbimento, distribuzione, metabolismo, eliminazione e tossicità (ADMET) dei candidati farmaci, aspetti cruciali per garantirne l'efficacia e la sicurezza come agenti terapeutici [12].

Oltre a identificare da una libreria di composti quelle molecole che esibiscono l'attività biologica desiderata, l'applicazione dei metodi IA si estende alla progettazione *de novo* di molecole che soddisfino specifici criteri. Utilizzando modelli generativi l'IA può proporre, infatti, nuove strutture chimiche in virtù dell'attuale disponibilità di un enorme spazio chimico virtuale e di *database* sperimentali. Un esempio di tecnologia IA applicata in questo ambito sono le reti GAN, composte da due reti neurali, un generatore e un discriminatore, che vengono addestrate attraverso i dati del *training set* per creare nuove strutture chimiche, peptidi e proteine, oltre che per ottimizzare i composti *lead* per migliorarne potenza, selettività e proprietà farmacocinetiche [13, 14]. Nella fase *hit-to-lead* quindi, i modelli QSAR basati sull'IA, oltre ai più tradizionali approcci di FEP (*Free Energy Perturbation*), possono guidare l'ottimizzazione razionale dell'affinità del ligando per il *target*. L'IA facilita anche il *drug repurposing*, una strategia innovativa potenziata dalle tecniche CADD, che esplora nuovi usi terapeutici per farmaci già esistenti [15]. Questo permette di bypassare la fase preclinica e accedere direttamente alla fase II degli studi clinici, riducendo i tempi e i costi di sviluppo richiesti per i nuovi farmaci. L'IA, con la sua capacità di gestire e analizzare differenti tipologie di dati farmacologici e clinici su larga scala può identificare

rapidamente quei farmaci con un potenziale terapeutico per altre patologie come si è verificato ad esempio per il COVID-19 [16, 17] che necessitava l'identificazione di soluzioni terapeutiche in tempi rapidi, e per il cancro [18].

Il coinvolgimento dell'IA nel *drug discovery* ha quindi sicuramente trasformato in tempi rapidissimi l'industria farmaceutica, anche se la maggior parte degli sforzi si è concentrata prevalentemente su sviluppo e calibrazione di nuovi metodi computazionali, mentre le applicazioni che ne mostrano il reale impatto ad oggi sono poche. Una serie di candidati farmaci delle *biotech* InSilico Medicine, Exscientia, BenevolentAI, Recursion, sono attualmente in fase di sperimentazione clinica avanzata [19].

Tuttavia va ricordato che il primo farmaco generato dall'IA in grado di raggiungere la fase I degli studi clinici, DSP-1181, destinato ai pazienti con disturbo ossessivo-compulsivo, e che è stato sviluppato e ha passato i test preclinici in meno di un anno, è stato ritirato poiché non ha soddisfatto i criteri previsti. Al momento si può pertanto solo affermare che gli approcci IA hanno ridotto il tempo e i costi necessari per portare un composto nella fase di sperimentazione preclinica e oltre, mentre è probabile che la maggior parte dei candidati farmaci continuerà a fallire nelle fasi successive. I risultati eccezionali di AlphaFold dimostrano che l'efficacia di un sistema di IA dipende dalla qualità dei dati su cui è stato allenato che nel caso di AlphaFold è il Protein Data Bank, che attualmente contiene oltre 200 mila strutture proteiche.

Sviluppati oramai nuovi e validi metodi di *deep learning* è importante perciò, nel prossimo futuro, poter accedere anche a un numero maggiore di dati chimici accurati proprio per soddisfare tutte le esigenze del *drug discovery*. Di fronte a queste sfide, la comunità chimica si sta organizzando per rendere i dati più accessibili attraverso lo sviluppo di algoritmi che convertono i nomi chimici in strutture molecolari come, ad esempio, la soluzione *open source* OPSIN [20]. Sono state, inoltre, sviluppate piattaforme come l'*Open Reaction Database* che facilitano la condivisione e l'accesso ai dati delle reazioni chimiche (<https://open-reaction-database.org/>).

Quando anche nella chimica i sistemi di apprendimento automatico raggiungeranno il loro potenziale, l'efficienza della *pipeline* e la riduzione dei tempi e dei costi di sviluppo dei farmaci potrebbe rendere la democratizzazione del processo di *drug discovery* un obiettivo non più irraggiungibile.

BIBLIOGRAFIA

- [1] S.J. Macalino, V. Gosu *et al.*, *Arch. Pharm. Res.*, 2015, **38**, 1686.
- [2] T.N. Doman, S.L. McGovern *et al.*, *J. Med. Chem.*, 2002, **45**, 2213.
- [3] V.T. Sabe, T. Ntombela *et al.*, *Eur. J. Med. Chem.*, 2021, **224**, 113705.
- [4] J. Jumper, R. Evans *et al.*, *Nature*, 2021, **596**, 583.
- [5] J. Abramson, J. Adler *et al.*, *Nature*, 2024, **630**, 493.
- [6] A.K. Ghosh, Z.L. Dawson *et al.*, *Bioorg. Med. Chem.*, 2007, **15**, 7576.
- [7] M. De Donato, B. Righino *et al.*, *Sci. Rep.*, 2018, **8**, 16047.
- [8] D. Pirolli, B. Righino *et al.*, *Molecular informatics*, 2021, **40**, 2060080.
- [9] M. Baek, F. DiMaio *et al.*, *Science*, 2021, **373**, 871.
- [10] Z. Lin, H. Akin *et al.*, *Science*, 2023, **79**, 1123.
- [11] F. Gentile, J.C. Yaacoub *et al.*, *Nat. Protoc.*, 2022, **17**, 672.
- [12] E.N. Feinberg, E. Joshi *et al.*, *J. Med. Chem.*, 2020, **63**, 8835.
- [13] W.A. Warr, M.C. Nicklaus *et al.*, *J. Chem. Inf. Model.*, 2022, **62**, 2021.
- [14] A.E. Blanchard, C. Stanley *et al.*, *J. Cheminform.*, 2021, **23**, 14.
- [15] S. Pushpakom, F. Iorio *et al.*, *Nat. Rev. Drug Discov.*, 2019, **18**, 41.
- [16] S.A. Hooshmand, M. Zarei Ghobadi *et al.*, *Mol. Divers.*, 2021, **25**, 1717.
- [17] D. Pirolli, B. Righino *et al.*, *Sci Rep.*, 2023, **13**, 1494.
- [18] N.T. Issa, V. Stathias *et al.*, *Semin. Cancer Biol.*, 2021, **68**, 132.
- [19] C. Arnold, *Nature Medicine*, 2023, **29**, 1292.
- [20] D.M. Lowe, P.T. Corbett *et al.*, *J. Chem. Inf. Model.*, 2011, **51**, 739.

The Promises of Artificial Intelligence in Drug Discovery

Computer-aided drug design (CADD) is one of the pivotal approaches in the drug discovery process, contributing to the development of numerous approved drugs. Thanks to the increasing availability of data and continuous model optimization, integration with artificial intelligence opens up new perspectives in predicting therapeutically relevant target structures, molecular properties, and generating new molecules.

