

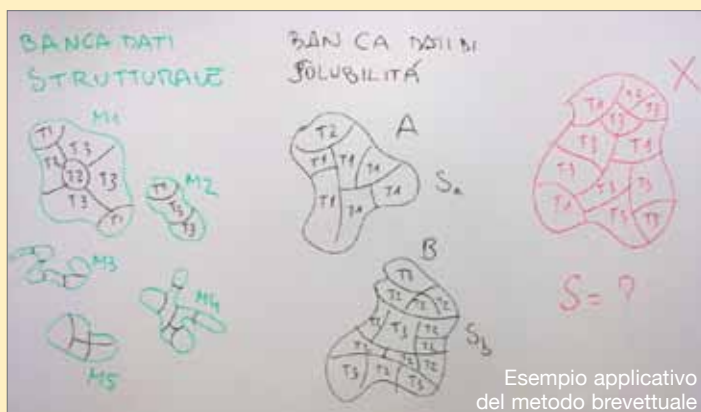
## Nuovo brevetto CRS4



# *Metodo per determinare valori di proprietà molecolari*

Alla fine di un percorso di lavoro di sei anni, tre ricercatori del CRS4, M. Valentini, E. Pieroni e D. Hofmann, hanno recentemente depositato la domanda di brevetto italiano dal titolo **“Metodo per determinare valori di proprietà molecolari”**. L'invenzione consente di predire tramite calcolatore le proprietà chimico-fisiche di interesse, quali ad esempio densità, solubilità, coefficiente di ripartizione acqua-ottanolo, per molecole non ancora sintetizzate o non ancora analizzate in laboratorio.

*La progettazione di un nuovo composto chimico per una specifica applicazione richiede normalmente personale altamente qualificato per il design iniziale, l'analisi dei risultati e grosse infrastrutture sperimentali, con conseguenti tempi lunghi e costi elevati. L'intelligenza artificiale e l'elaborazione delle banche dati esistenti attraverso l'utilizzo di potenti calcolatori paralleli (di cui il CRS4 è dotato) può consentire di abbreviare questo percorso consentendo di identificare da subito i composti potenzialmente più interessanti.*



Normalmente le esigenze industriali nel settore della progettazione di molecole (molecular design) sono di due tipi: progettare un nuovo composto con specifici requisiti funzionali oppure cercare possibili nuovi ambiti applicativi (targets) per un composto noto o nuovo e per tutte le sue possibili nuove varianti.

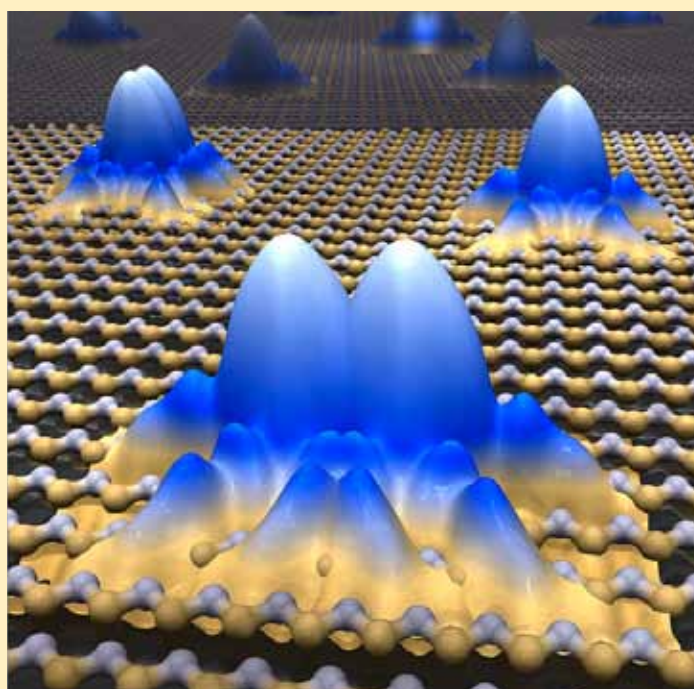
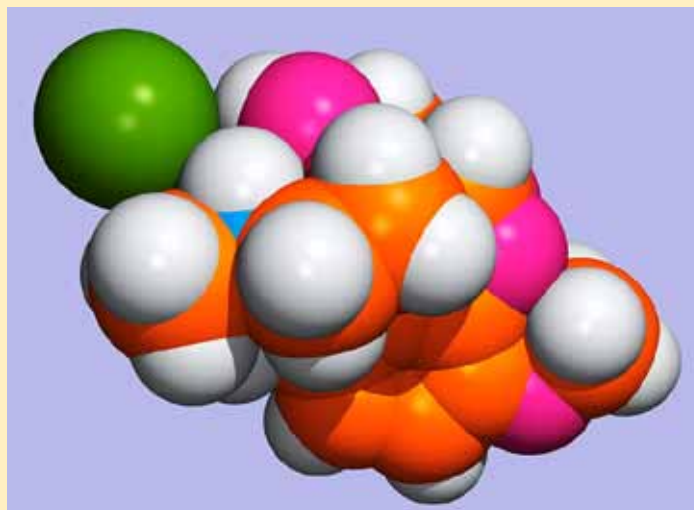
Nel primo caso occorre sintetizzare in quantità adeguata, con costi elevati, un elevato numero di nuove molecole, talvolta concepite a partire da strutture di riferimento (leads), e di effettuare su ciascuna di esse una serie di misure sperimentali, con l'obiettivo finale di pervenire a un numero limitato di composti interessanti (hits). Nel secondo caso occorre sintetizzare un buon numero di nuove molecole sulle quali effettuare un'estesa batteria di test differenti tesi a verificarne il possibile uso in un diverso ambito applicativo. Con un'enfasi diversa rispetto alle molecole o ai test, i due problemi sono molto simili e richiedono il supporto di personale specializzato, con una forte competenza nel settore, che identifichi le possibili strutture interessanti, esegua gli esperimenti con protocolli ben definiti tramite costose attrezzature a elevata processazione (high throughput screening) e analizzi i risultati.

Ad esempio, progettare un farmaco significa avviare un percorso che parte da una prima concezione da parte di esperti, seguita da test preliminari di laboratorio su molti composti, da test su culture cellulari e poi su modelli animali di un numero ridotto di molecole, sino ad arrivare alla prova sui pazienti.

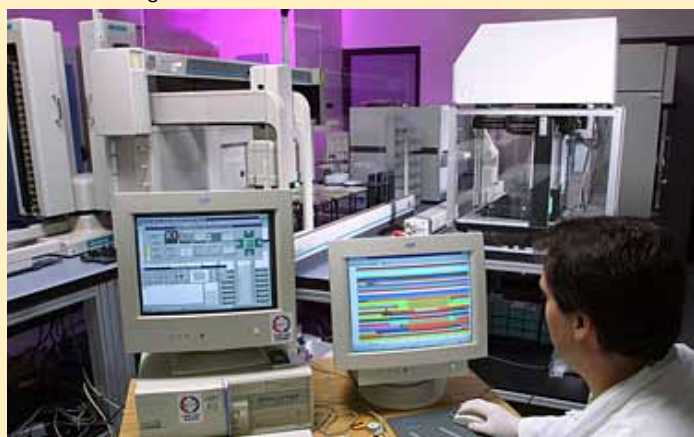
Questo complicato processo si traduce in tempi lunghi e costi elevati, con conseguente riduzione dei progetti nel settore e un accesso al mercato da parte di poche aziende dotate delle risorse economiche, umane e infrastrutturali sufficienti.

L'idea innovativa alla base del brevetto è invece quella di caratterizzare una molecola sconosciuta, ad esempio un farmaco, senza utilizzare alcuna esperienza o informazione a priori, ma interrogando direttamente la natura, setacciando tramite calcolatori paralleli le banche dati biologiche e chimiche realizzate dalla comunità scientifica e industriale negli ultimi vent'anni e che stanno avendo una crescita esponenziale.

Questo può essere utilizzato per progettare una nuova molecola oppure per scremare al computer (virtual screening) il ventaglio di potenziali composti candidati, per poi passare al bancone di laboratorio solo per la rosa ristretta delle molecole più promettenti. Focalizzandosi sulle applicazioni, quest'idea dovrebbe consentire, ad esempio, di 'progettare' un farmaco di massima efficacia (con elevata biodisponibilità negli organi obiettivo, alta selettività per le proteine obiettivo e ridotta tossicità), nel minor tempo possibile, riducendo i test di laboratorio e su animali e i costi a essi connessi, e nel complesso minimizzando gli effetti avversi e il rischio per i pazienti.



Molecular design



Attrezzature High throughput screening (HTS)



Per le sue caratteristiche, lo stesso approccio può essere utilizzato in diversi ambiti: cosmetici, detersivi, materiali biodegradabili, alimenti nutraceutici, prodotti anallergici, additivi, antiparassitari e disinfestanti, nuovi materiali e prodotti del futuro. In pratica (si veda l'esempio nel Box), l'invenzione adotta delle tecniche tipiche dell'intelligenza artificiale e dell'analisi dei dati (data mining), che consentono automaticamente di predire le proprietà di interesse di molecole e composti non ancora sintetizzati né sperimentati, con un'accuratezza crescente con la dimensione e la qualità delle banche dati di riferimento.

Più in dettaglio, la procedura si basa su un'originale misura di somiglianza fra parti di molecole e prevede l'utilizzo di due banche dati (database):

- un database di strutture chimiche (structural database), sul quale viene applicata la misura di somiglianza, in modo da identificare le parti fondamentali e distinte che compongono ciascuna molecola (~106 strutture);
- una banca dati di molecole (~104-105), di cui siano note le proprietà chimiche di interesse (training set), tramite cui arriva ad associare a ciascuna delle parti fondamentali identificate in precedenza il contributo alla proprietà chimica di interesse.

Infine si definisce l'insieme delle molecole di interesse, non ancora sperimentate, e che possono essere centinaia di migliaia, e per ciascuna di esse si identificano le parti da cui è composta e quindi, sommando i rispettivi contributi, si arriva a predire la proprietà chimica di interesse. Al di là dei tecnicismi, il vantaggio competitivo offerto dall'invenzione è molteplice e spazia dai significativi risparmi in tempo e risorse ottenibili dalla preventiva selezione delle innumerevoli molecole potenzialmente utili, sino all'integrazione della fase di progettazione iniziale o di quella di ottimizzazione del prodotto finale.

**Come funziona il brevetto** - Immaginiamo di avere un database strutturale composto da cinque sole molecole M1, ..., M5 e di suddividere ciascuna molecola in tutte le parti possibili, ossia – a grandi linee – un singolo atomo, un atomo e il suo primo vicino, e così via. Attraverso la misura della somiglianza fra tutte le parti così ottenute da tutte e cinque le molecole, possiamo arrivare a distinguere solo quelle effettivamente diverse, che rappresentano i mattoni per ricostruire abbastanza fedelmente le cinque molecole originarie. Immaginiamo che queste parti siano solo tre: T1, T2 e T3.

Ad esempio, come si vede nella figura 1, immaginiamo che M1, M2, ..., M5 siano composte da varie proporzioni delle T1, T2 e T3. Consideriamo un training database composto da due molecole A e B di struttura nota e di solubilità nota Sa e Sb. Applicando la stessa misura di somiglianza, si ottenga ad esempio che A sia composta da sei parti T1 e una T2 e B sia composta da dieci parti T2 e quattro T3. In tal modo, una frazione della solubilità Sa di A potrà essere assegnata a T1, una frazione a T2, e in maniera simile per B. Si può così arrivare ad assegnare in maniera generale, per ogni molecola, alla parte T1 un contributo S1 alla solubilità, alla parte T2 un contributo S2 e infine alla parte T3 un contributo S3, tali da soddisfare in maniera ottimale il sistema  $S_a = 6 \cdot S_1 + S_2$  e  $S_b = 10 \cdot S_2 + 4 \cdot S_3$ . A questo punto si prenda una molecola X, di struttura nota e solubilità sconosciuta.

Procedendo come sinora descritto, si immagini trovare che X si costituisca da tre parti T1 e nove parti T3. In tal modo si arriva a predire la solubilità di X pari a  $3 \cdot S_1 + 9 \cdot S_3$ .



**Enrico Pieroni**, ricercatore del CRS4 laureato in fisica a Pisa, dopo un master in Bionformatica ha iniziato a occuparsi di biofisica, biologia dei sistemi, modellistica molecolare delle malattie autoimmuni e drug design.

**Maria Valentini**, ricercatore del CRS4 laureata in fisica a Bologna, si è occupata della modellistica di membrane per il trasporto di protoni e, dopo un master in Bioinformatica, ha cominciato a lavorare nell'ambito della genetica e del drug design.

**Detlef Hofmann**, ricercatore del CRS4 laureato in chimica teorica a Norimberga, ha contribuito sin dai primordi all'utilizzo dell'analisi dei database per la parametrizzazione di modelli chimici, con applicazioni in particolare alla struttura dei cristalli, dinamica molecolare e drug design.