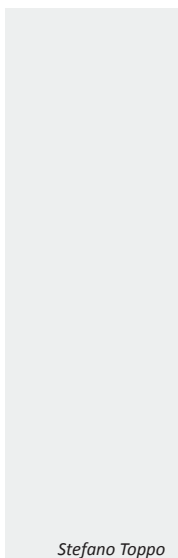


di Stefano Toppo\*

Il termine bioinformatica fu coniato sul finire degli anni Settanta dalla biologa teoretica olandese Paulien Hogeweg riferendosi agli studi di simulazione *in silico*, ovvero tramite computer, dei processi di crescita cellulare. Da semplice termine a definitiva consacrazione come disciplina scientifica il passo fu breve, ma fino ai primi anni Novanta la bioinformatica è coincisa, in realtà, con gli studi di filogenesi molecolare e le analisi di sequenze di DNA. Ai giorni nostri il suo campo d'azione si è esteso grazie all'avvento di nuove tecnologie che vanno sotto il nome generico di scienze "omiche", tra cui le più note sono la genomica, la trascrittomica e la proteomica, accomunate dalla produzione su



Stefano Toppo

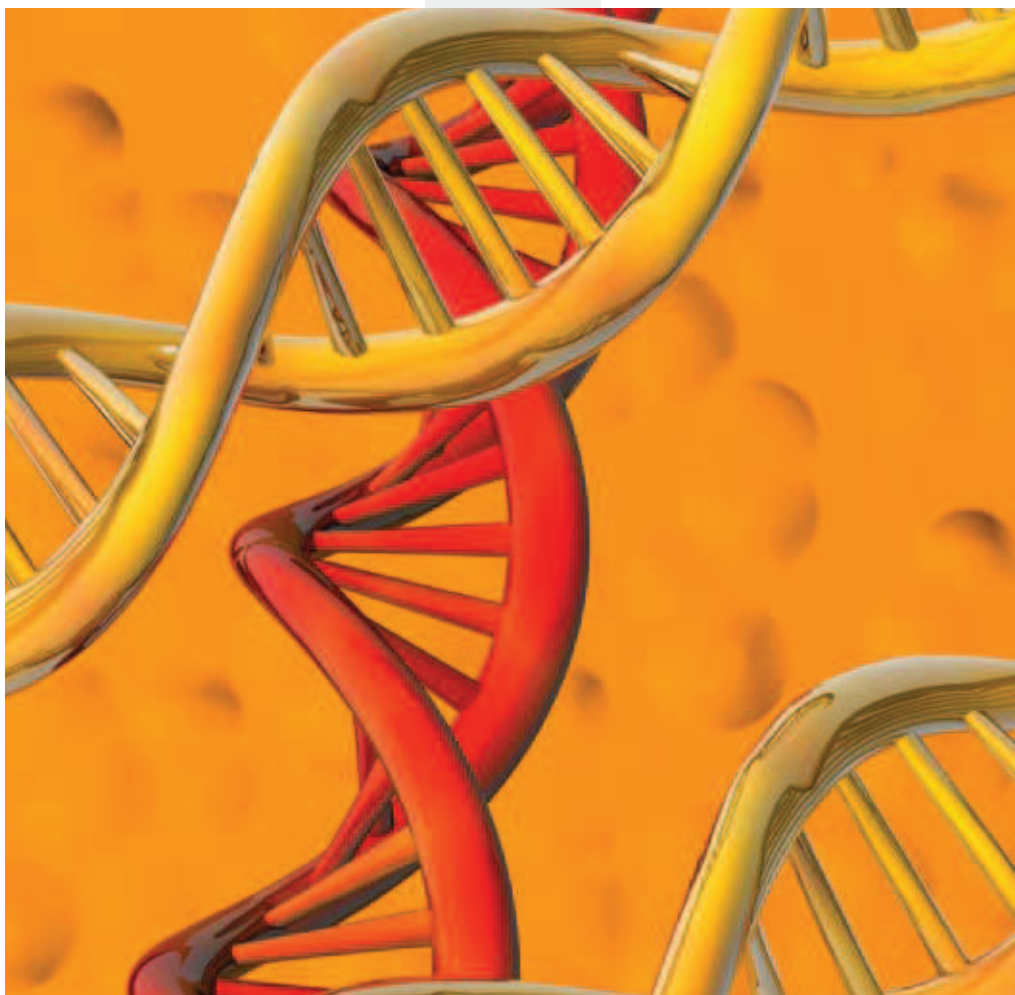


larga scala di dati grezzi che devono essere analizzati ed interpretati. I confini di questa disciplina si

sono, poi, ampliati a tal punto da diventare spesso confusi ed oggetto di discussione, ma nella più ampia accezione del termine, la bioinformatica abbraccia anche altri settori come la metabolomica, la *systems biology*, la biostatistica, l'*imaging*, la farmaco-genomica o *drug design*, la chimica computazionale, la bioinformatica strutturale e funzionale. Le competenze necessarie sono, quindi, trasversali ed il background culturale del bioinformatico spazia dalla matematica all'informatica, alla fisica, alla chimica, alla biologia e alla medicina.

Le nuove tecnologie analitiche in grado di generare grandi moli di dati hanno sicuramente accelerato lo sviluppo in campo bioinformatico, facendo aumentare le esigenze di grandi infrastrutture computazionali e rendendo necessario reinventare nuove soluzioni algoritmiche per l'analisi dei dati.

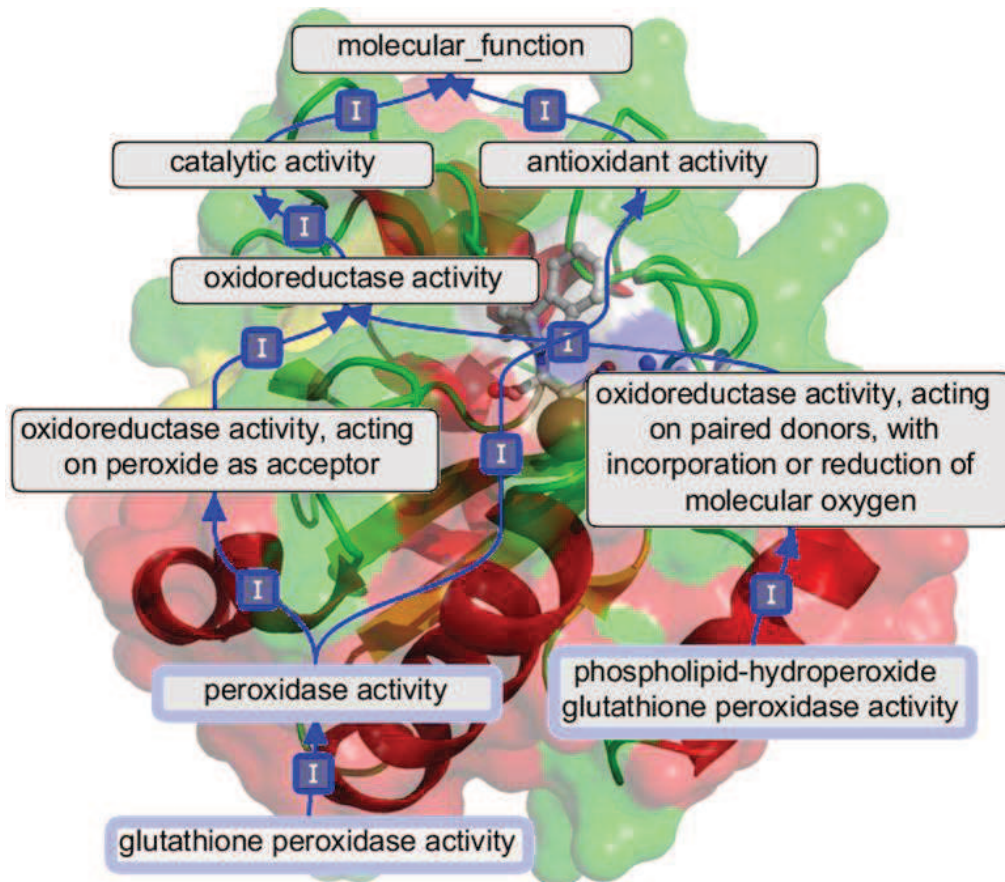
Il gruppo di bioinformatica del Dipartimento di Chimica Biologica sta portando avanti diversi progetti grazie anche alle collaborazioni con varie realtà presenti nell'Ateneo. Il nostro impegno principale è nel campo della predizione di strutture proteiche e nello studio dettagliato delle reazioni enzimatiche con tecniche di quantomeccanica e dinamica molecolare, in collaborazione con i Chimici Computazionali. Storicamente il nostro laboratorio si occupa di una categoria particolare di enzimi, le glutazione perossidasi, la cui peculiarità sta nel fatto di avere al sito attivo una selenocisteina, aminoacido raro contenente selenio. Dalle analisi comparate e di filogenesi effettuate *in silico* abbiamo scoperto un nuovo membro della famiglia che abbiamo chiamato GPX8, l'ultimo di una famiglia di otto membri. Abbiamo inoltre identificato nuovi aminoacidi chiave coinvolti nella reazione di ossidazione. Lo studio dettagliato del



sito catalitico con metodi computazionali sta inoltre rivelando sorprendenti nuovi meccanismi molecolari di azione fino ad ora rimasti vaghi ma che trovano un riscontro in dati sperimentali. Secondo alcune recenti evidenze alcuni membri di questa famiglia, in particolare la GPX4, sembrano essere coinvolti in vie di *signaling* dell'apoptosi, riportando alla ribalta questa classe di enzimi per il ruolo non trascurabile che potrebbero avere in terapie anti-tumorali.

Nell'ambito della genomica, grazie all'impiego di nuove tecnologie di sequenziamento di DNA, siamo stati coinvolti nell'assemblaggio e nell'analisi del genoma di meningococco *Neisseria meningitidis*, responsabile di infezioni di meningite nel trevigiano. Dall'analisi sono emersi tratti peculiari e geni chiave probabilmente responsabili della virulenza dei ceppi isolati ed attualmente in corso di validazione sperimentale. Questo progetto ci ha portato allo sviluppo e alla realizzazione di diverse strategie per gestire ed assemblare i milioni di corte sequenze tipicamente prodotte dal sequenziatore. Una ricostruzione non banale, paragonabile ad un puzzle di milioni di pezzi di cui non si conosce l'immagine finale e che richiede molta potenza di calcolo e controlli stringenti.

Un altro filone di ricerca di grande attualità che il nostro gruppo sta portando avanti è la predizione di funzione ed annotazione di sequenze proteiche. È stato creato un nuovo algoritmo per "navigare" all'interno della rete di dati della *Gene Ontology*, in cui tutte le informazioni sulle funzioni proteiche e sui processi biologici sono ordinate e strutturate in una sorta di gerarchia. Grazie a processi di inferenza e ad altre proprietà, siamo in grado di predire l'ipotetica funzione di una proteina non ancora caratterizzata, associando ad essa un punteggio di significatività. L'obiettivo finale è quello di fornire valide ipotesi di lavoro nell'ottica di abbattere tempi e costi delle indagini di laboratorio. L'indubbio vantaggio del metodo è soprattutto anche quello di poter annotare velocemente ed automaticamente interi genomi contenenti decine di migliaia di proteine ed



ottenere un quadro complessivo delle loro funzioni. Particolarmente utile nel caso del sequenziamento di nuovi genomi batterici, ad esempio, di interesse microbiologico e medico per i quali esistono scarse informazioni sulle funzioni delle proteine.

Sempre nell'ambito della *Gene Ontology* è in corso l'ambizioso progetto, in collaborazione con il gruppo di Bioingegneria, di "ripulire" le banche dati pubbliche, contenenti milioni di proteine, dagli errori di annotazione che ammontano fino al 20% del totale e che si ripercuotono negativamente sulle indagini sperimentali di espressione genica, di reti di geni e di vie metaboliche.

Un notevole sforzo sarà inoltre richiesto al nascente centro di proteomica "Pietro d'Abano", dove tecniche di spettrometria di massa consentiranno l'identificazione di proteine e la caratterizzazione di interi proteomi. Come per la genomica, si prevede la produzione di enormi quantità di dati che dovranno essere gestiti e analizzati mediante procedure automatizzate: in questo caso le principali applicazioni saranno l'implementazione

Struttura della fosfolipid-idroperossido glutathione perossidasi e sua annotazione funzionale in *Gene Ontology*

“...risulta chiaro quanto la bioinformatica sia coinvolta, in modo sempre crescente, in un vasto spettro di settori di ricerca e di quanti vantaggi possa garantire in termini di riduzione dei tempi e dei costi delle attività di ricerca”

tazione di algoritmi di ricerca degli spettri di massa in banche dati specializzate, di quantificazione delle proteine nei campioni e di analisi di modificazioni post-traduzionali. Dalla panoramica dei progetti risulta chiaro quanto la bioinformatica sia coinvolta, in modo sempre crescente, in un vasto spettro di settori di ricerca e di quanti vantaggi possa garantire in termini di riduzione dei tempi e dei costi delle attività di ricerca e nella gestione e condivisione delle enormi moli di dati prodotti. Si renderà dunque necessario, nel breve e medio termine, potenziare la disponibilità dei servizi offerti dalla bioinformatica, sia in termini di addestramento e formazione di personale specializzato, sia in termini di ampliamento delle risorse hardware a disposizione. È una sfida importante per il futuro della Facoltà di Medicina, che potrà ritagliarsi un ruolo da protagonista nell'ambito della ricerca biomedica/informatica ed essere competitiva rispetto alle altre medical school in Italia e all'estero.

\* Stefano Toppo, docente di Biochimica, Dipartimento di Chimica biologica, Università di Padova