



Simulazioni di Dinamica Molecolare con GROMACS: l'accelerazione delle GPU

GROMACS è tra i software più diffusi per i calcoli di Dinamica Molecolare. Calcoli che l'uso del Cloud GPU accelera, agevolando la ricerca biomolecolare



03

Introduzione alla Dinamica Molecolare

04

Sviluppo e applicazioni della Dinamica Molecolare
Progressi e diffusione della DM nella ricerca scientifica

05

Il ruolo delle GPU per accelerare le simulazioni
GROMACS, GPU e Cloud Computing

08

GROMACS nel mondo della ricerca biomedica:
alcune applicazioni d'uso

10

Conclusioni

11

Bibliografia



Introduzione alla Dinamica Molecolare

Nel 1648 Isaac Newton formulò le famose tre leggi che portano il suo nome, dalle quali ha avuto inizio la fisica-matematica classica [1].

- **Lex. I.** *Corpus omne perseverare in statu suo quiescendi vel movendi uniformiter in directum, nili quatenus a viribus impressis cogitur statum illum mutare.*
- **Lex. II.** *Mutationem motus proportionalem esse vi motrici impressae, et fieri secundum lineam rectam qua vis illa imprimitur.*
- **Lex. III.** *Actioni contrariam semper et aequalem esse reactionem: sive corporum duorum actiones in se mutuo semper esse aequales et in partes contrarias dirigi.*

Isaac Newton. *Philosophiae Naturalis Principia Mathematica.* London, 1686.

A 375 anni dalla pubblicazione di quello che è ritenuto uno dei più grandi capolavori scientifici, il *Principia Mathematica*, le stesse leggi enunciate per descrivere il moto delle stelle e dei pianeti trovano ancora applicazione per semplificare la descrizione del mondo atomico tramite la Dinamica Molecolare [2].





Sviluppo e applicazioni della Dinamica Molecolare

La Dinamica Molecolare (DM) è un metodo computazionale per simulare e analizzare il movimento degli atomi e delle molecole, permettendo in tal modo la caratterizzazione l'“evoluzione” dinamica del sistema in esame.

Le prime simulazioni di DM furono effettuate da Alder e Wainwright [3] negli anni '50, agli albori dell'era informatica, per studiare fluidi semplici mediante modelli che rappresentano gli atomi come dischi e sfere rigide. Il sistema simulato era costituito da 32 atomi appena.

La prima **simulazione di un sistema biologico**, in particolare di un enzima, risale agli ultimi anni '70 e fu pubblicata da McCammon et al. sulla famosa rivista Nature [4].

Lo studio delle basi teoriche che hanno permesso tali risultati è valso il Nobel per la Chimica nel 2013 a M. Karplus, M. Levitt e A. Warshel [5,6], i quali possono essere considerati i padri delle moderne tecniche di DM. Oggi, a distanza di 70 anni dalla storica pubblicazione di Alder, è possibile **simulare sistemi rappresentati da diverse centinaia di migliaia di atomi**, ottenendo dati nell'arco di giorni o settimane, grazie all'utilizzo di hardware e software altamente specializzati. Molto recentemente alcuni gruppi di ricerca sono riusciti a simulare intere particelle virali [7,8] o addirittura intere cellule viventi [9], via via più complesse [10].



Progressi e diffusione della DM nella ricerca scientifica

La DM è tra le tecniche computazionali che richiedono più tempo, perché dal carattere molto iterativo e incentrato sul calcolo. Poiché i calcoli vengono eseguiti miliardi di volte, ciò significa che sono **coinvolte milioni di righe di codice**.

Negli ultimi decenni l'utilizzo della DM è divenuto sempre più accessibile e diffuso nell'ambito della ricerca scientifica, fino a divenire il **metodo di elezione per lo studio teorico del dettaglio atomico delle biomolecole**.

Nel 1997, sommando tutti gli articoli che si avvalgono della DM pubblicati sulle riviste scientifiche più autorevoli, si contavano meno di 175 manoscritti. Dopo venti anni, nel 2017, questo numero ha superato i 1000 [11]. Senza dubbio tale progresso è stato fortemente spinto dallo sviluppo di nuovi hardware (grazie all'uso di unità di processazione grafica, GPU), software e algoritmi (tecniche di programmazione in parallelo e codici di DM più efficienti), ma anche nella progettazione di modelli fisici sempre più accurati (che nel linguaggio tecnico sono detti 'force-fields').





Il ruolo delle GPU per accelerare le simulazioni

Le GPU sono circuiti elettronici altamente specializzati, progettati per accelerare la grafica computerizzata e il processamento dell'immagine. Oltre allo scopo iniziale per cui questi circuiti furono progettati, ci si è resi conto che la loro particolare **struttura "parallela"** permetteva di eseguire calcoli non-grafici per risolvere problemi computazionali suddivisi in attività di piccole dimensioni, semplici e indipendenti che possono essere eseguite contemporaneamente (spesso con poca o nessuna comunicazione tra di loro) proprio come nel caso delle simulazioni di DM. Di conseguenza, l'avvento del **GPU Computing ha apportato una vera e propria rivoluzione nell'ambito della DM**, rendendo accessibili e relativamente economiche le alte prestazioni che in passato erano riservate ai supercomputer.

Ad esempio, un computer fornito di GPU riesce a eseguire mediamente simulazioni fino a dieci volte più velocemente di uno stesso computer fornito solo di CPU. Inoltre, per ottenere le stesse prestazioni, lo stesso computer con CPU dovrebbe essere cinque volte più grande rispetto a un computer con GPU: di conseguenza, utilizzando GPU anche il consumo energetico si dimezza. Ad oggi, **la maggior parte dei software per calcoli di DM sono compatibili per GPU** [12].

Tra questi, **GROMACS** rappresenta indubbiamente una delle scelte più comuni ed efficienti.

La più recente versione di GROMACS (2023), permette di ridurre l'overhead tra GPU e CPU, incrementando notevolmente le prestazioni del software di simulazione biomolecolare [13]

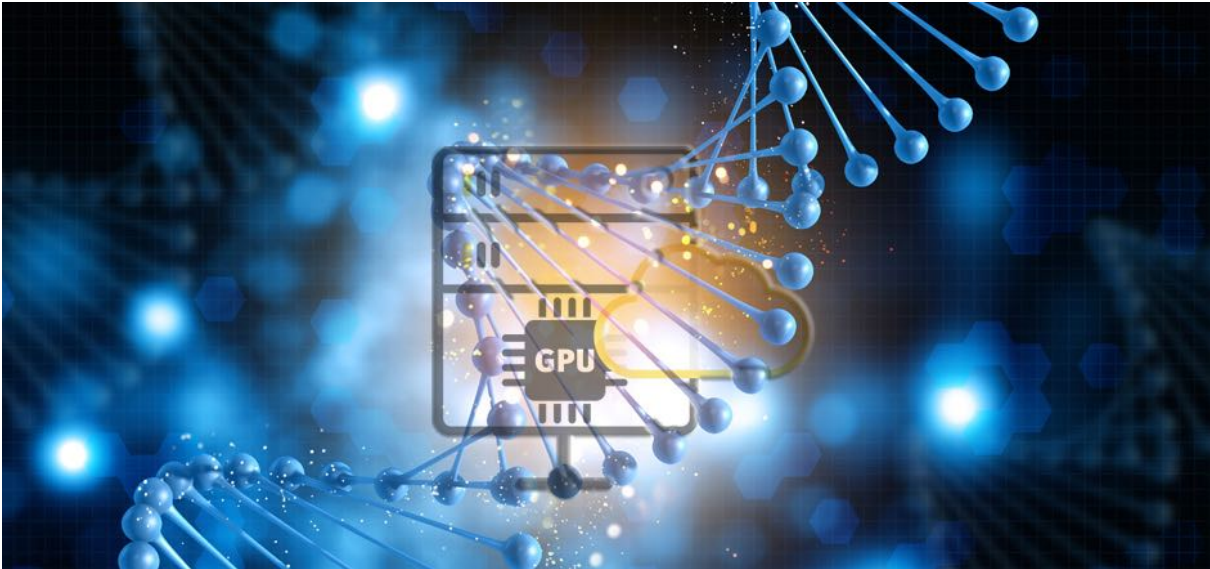
GROMACS, GPU e Cloud Computing

"GROMACS è una delle applicazioni di dinamica molecolare open source più utilizzate al mondo ed è facile capire perché. Le simulazioni che possiamo condurre con questa applicazione ci permettono di capire meglio cose piccole, come le proteine del nostro corpo, e grandi come le galassie dell'universo. In particolare, il nostro lavoro con GROMACS - sviluppato e ottimizzato con oneAPI - consente a Intel di contribuire a progressi significativi nella scoperta di farmaci e di espandere lo sviluppo aperto di GROMACS su diverse architetture di calcolo. E tutto questo collaborando con la comunità open source che apprezziamo molto".

– **Roland Schulz**, ingegnere del software parallelo di Intel

GROMACS (*GRO*ningen *MACH*ine for *SIM*ulations) è un **software open-source progettato per simulazioni atomistiche ad alta prestazione** di proteine, lipidi, carboidrati, acidi nucleici e particelle. Sviluppato inizialmente dall'Università di Groninga, il progetto è supportato da altre università e dalla comunità europea attraverso il programma Horizon2020. Viene usato principalmente per simulazioni di DM di sistemi biologici (nonostante possa essere impiegato per simulazioni di altre tipologie di sistemi) e offre un set completo di varie tipologie di calcolo, preparazione degli input e analisi degli output.





GROMACS supporta tutti gli algoritmi più comuni che ci si aspetta da un moderno software di DM, ma comprende alcune caratteristiche innovative e peculiari che lo distinguono dalla concorrenza [14]:

- GROMACS raggiunge prestazioni estremamente alte rispetto agli altri software disponibili. È in grado di sfruttare al massimo il potenziale delle moderne CPU e GPU grazie ad algoritmi altamente ottimizzati.
- GROMACS utilizza simultaneamente le CPU e GPU disponibili nel sistema, fornendo opzioni in grado di bilanciare il carico tra diverse risorse, sia in modo statico che dinamico.
- GROMACS è *user-friendly*. I file relativi a topologie e parametri sono scritti in forma testuale. Il software esegue molti controlli per la coerenza interna dei dati ed emette chiari messaggi quando si verifica un errore.
- GROMACS non richiede scripting: tutti i moduli di GROMACS utilizzano una semplice interfaccia da linea di comando per gestire input e output. Attualmente gli sviluppatori di GROMACS stanno lavorando per implementare una Python API (“interfaccia di programmazione dell’applicazione”) che permette lo scripting della preparazione, esecuzione e analisi delle simulazioni.
- Sia i file di input che di output possono essere letti da qualsiasi versione di GROMACS, essendo indipendenti dalla ‘endianness’ dell’hardware.
- GROMACS scrive gli output utilizzando una compressione dati di tipo ‘lossy’, utile per archiviare grandi quantitativi di dati in modo ottimale. L’accuratezza della compressione può essere selezionata dall’utente.
- GROMACS comprende una varietà di tools per analizzare gli output. I formati di output di GROMACS sono supportati dalla maggior parte dei software di analisi e visualizzazione.
- GROMACS può eseguire calcoli in parallelo, sia utilizzando protocolli standard di comunicazione MPI o tramite la libreria “Thread-MPI” sviluppata da GROMACS per workstations a singolo nodo.





La chiave per ottimizzare la performance delle simulazioni di DM risiede nella comprensione dell'**utilizzo dell'hardware da parte degli algoritmi di DM**. In particolare, GROMACS combina diverse tecniche di parallelizzazione, tra cui MPI e OpenMP, GPU "offloading" e calcoli separabili delle interazioni elettrostatiche a lungo raggio.

Una peculiarità di GROMACS consiste nell'eseguire diversi tipi di calcoli, molto dispendiosi a livello computazionale, sulla GPU. Nel dettaglio, uno dei più grandi benefici a livello di performance è dato dal calcolo delle interazioni di non legame a corto raggio (i.e.: Coulomb e van der Waals) sulla GPU. La possibilità di eseguire questo tipo di calcoli in parallelo (in cui ogni MPI rank sfrutta una GPU per il calcolo delle interazioni a corto raggio) ottimizza ulteriormente la performance. Al contrario, il calcolo delle interazioni a lungo raggio non poteva essere distribuito su GPU multiple, fino ai primi mesi del 2023. A seguito del rilascio della versione GROMACS 2023, è stato possibile migliorare significativamente la performance delle simulazioni di DM, rendendo il sistema ancora più scalabile di quanto fosse in precedenza [13,14].

GROMACS viene usato nel settore sanitario da organizzazioni biotecnologiche, università e centri di ricerca, istruzione, organizzazioni farmaceutiche, ospedali e cliniche.

Comunque, nonostante i notevoli e costanti progressi tecnologici nell'ambito dell'hardware e del software che abbiamo appena sottolineato, **eseguire simulazioni di DM rimane un compito molto dispendioso dal punto di vista computazionale**.

Ma dove trovano i ricercatori il tempo di calcolo necessario? Tra le soluzioni consolidate si annoverano centri di calcolo di università e istituti di ricerca, centri nazionali di supercomputing, clusters locali, e workstation a singolo-nodo, ognuno con vantaggi e svantaggi riguardo la facilità di accesso alle risorse, la loro disponibilità e il relativo costo.

Recentemente, una valida soluzione a tali problematiche è rappresentata dal cloud computing, che permette di gestire al meglio le risorse per il calcolo: cioè in questo modo l'utente ha accesso a sistemi HPC (*High Performance Computing*) che possono essere facilmente adattati e rimodulati in funzione di particolari esigenze in un dato momento. Di conseguenza, nel cloud computing le risorse possono essere finemente regolate con lo scopo di ottimizzare i costi o minimizzare i tempi di calcolo.

Diversi studi hanno dimostrato come la performance di GROMACS e il rapporto P/C (performance/costo, ovvero la quantità di dati prodotti per € investiti) siano altamente dipendenti dalla tipologia di hardware a disposizione [16]. A tal proposito, singoli nodi di calcolo equipaggiati con multiple GPUs forniscono performance migliori. D'altra parte, le GPUs più performanti ad oggi hanno un prezzo che si aggira intorno a diverse (talvolta decine di) migliaia euro e, tra tutte le componenti hardware, sono quelle più soggette a fluttuazioni economiche.

Di nuovo, **il cloud computing si rivela un'ottima scelta che permette di massimizzare simultaneamente performance e rapporto P/C**. Tale approccio fornisce all'utilizzatore una maggiore semplicità di utilizzo del sistema e una drastica riduzione delle tempistiche necessarie per installare e testare software – operazioni necessarie e preliminari ai calcoli di DM – che possono essere lunghe anche per gli utenti più esperti.

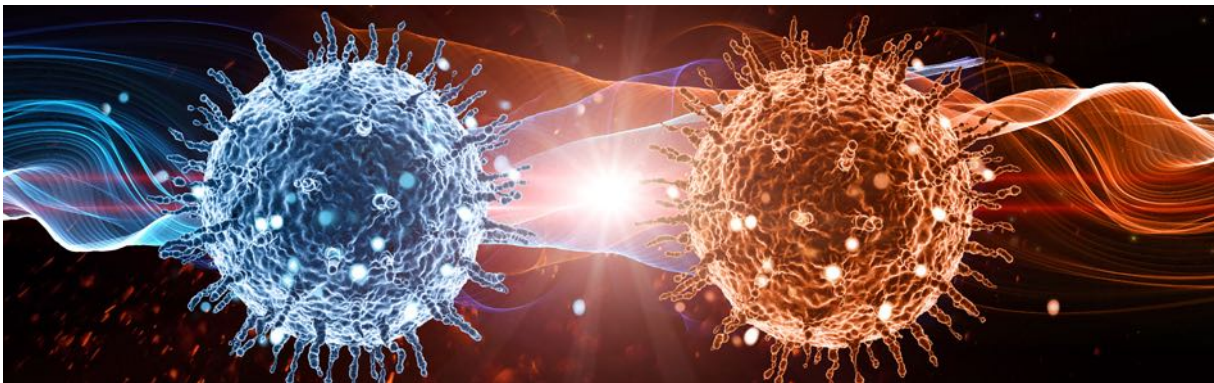


GROMACS nel mondo della ricerca biomedica: alcune applicazioni d'uso

Le simulazioni DM vengono costantemente applicate con successo nelle diverse fasi di ricerca e sviluppo di nuove molecole ad azione farmacologica, nei confronti delle più svariate patologie, per ottenere informazioni chiave sulla struttura tridimensionale delle biomolecole e la variazione della stessa nel tempo. Ciò consente di comprendere la relazione struttura-funzione che può guidare il processo di **scoperta e progettazione dei farmaci**.

Un esempio attuale e affascinante dell'utilizzo delle simulazioni di DM si contestualizza nella pandemia COVID-19, la quale non si combatte solo negli ospedali e nei laboratori, ma anche nei grandi centri di calcolo dove sono ospitati i potentissimi supercomputer.

Di recente è stato possibile, mediante DM, simulare e prevedere il comportamento delle particelle del coronavirus SARS-CoV-2 in innumerevoli contesti e condizioni. In uno degli studi più avanzati nel genere [7], il gruppo di ricerca guidato dalla Dr. R. Amaro, ha simulato in scala atomica una particella virale del coronavirus all'interno di una gocciolina di aerosol – un sistema composto da più di un miliardo di atomi – tracciando i loro movimenti con una precisione nell'ordine dei milionesimi di secondo. Ciò è stato reso possibile grazie a software come GROMACS, ed è stato fondamentale per studiare le interazioni e i possibili effetti sulla contagiosità.



Ancora più recentemente, lo stesso gruppo di ricerca ha portato avanti uno studio simile, ma questa volta indagando le proprietà strutturali e dinamiche delle glicoproteine del virus dell'influenza [8]. Le glicoproteine di superficie del virus sono responsabili del contagio e, mutando molto velocemente, rendono difficile sviluppare un vaccino universale. Questo studio ha permesso di comprendere dettagli precedentemente sconosciuti circa la dinamica di queste importanti glicoproteine e la loro interazione reciproca, suggerendo nuove strategie per lo sviluppo di vaccini e farmaci antivirali contro l'influenza.

Un altro esempio della sempre più stretta interconnessione tra le attività di laboratorio e quelle simulate in ambito biomedico viene da una ricerca scientifica pubblicata sulla rivista *Nature* a maggio scorso [15]. Tale ricerca ha permesso di comprendere la struttura tridimensionale di un recettore olfattivo umano, chiamato OR51E2. È un passo importante, che ci consente di capirne di più sul funzionamento di uno dei nostri sensi più complessi e meno conosciuti (l'olfatto umano coinvolge circa 400 recettori diversi!). Ciascuno delle centinaia di migliaia





di profumi che possiamo percepire è costituito da una miscela di diverse molecole odorose, ognuna delle quali, a sua volta, può essere rilevata da determinati tipi di recettori: si tratta, quindi, di un vero e proprio puzzle enorme, che il cervello deve risolvere ogni volta che il naso riconosce qualcosa nell'aria. L'ostacolo maggiore verso la comprensione di questo meccanismo a livello molecolare, però, è rappresentato dalla notevole difficoltà nell'ottenere in laboratorio informazioni sulla struttura dei recettori olfattivi, un'impresa considerata quasi impossibile. In questo studio gli autori sono riusciti a migliorare notevolmente la sensibilità degli strumenti utilizzati, cosa che ha permesso di lavorare anche con le quantità minime di recettori a disposizione.

"È la prima 'fotografia' di una molecola olfattiva che interagisce con uno dei nostri recettori olfattivi", spiega Aashish Manglik, uno degli autori.

In questo caso le simulazioni di DM sono state impiegate per **indagare dettagli su variazioni atomiche strutturali che gli esperimenti non possono ottenere**: si è riusciti a capire i cambiamenti nella struttura del recettore e le interazioni molecolari specifiche che si instaurano quando si lega la molecola odorosa, la quale avvia il processo di segnalazione alla base del senso dell'olfatto.

Nell'ambito della **chimica farmaceutica**, si fa spesso ricorso a un approccio chiamato "progettazione di farmaci mediante l'analisi dei frammenti" (*fragment-based drug design, FBDD*), il quale si è andato sempre più affermando ed è ormai riconosciuto come una valida alternativa alle procedure di HTS (*High Throughput Screening*). Esso presuppone di vagliare un numero iniziale molto limitato (dell'ordine di poche migliaia) di piccole molecole (frammenti), che vengono poi elaborate fino all'identificazione di nuovi composti guida da sviluppare in farmaci.

Attualmente l'**approccio FBDD** ha portato allo sviluppo di sei farmaci clinicamente approvati. Le piccole dimensioni dei frammenti chimici utilizzati, però, determinano un legame molto debole e transitorio nei confronti del bersaglio farmacologico. Ciò rende molto complicato determinare, tramite esperimenti classici, la struttura tridimensionale assunta dai frammenti a seguito del legame con la proteina target.



Conclusioni

Ancora una volta, le simulazioni di DM rappresentano un metodo che permette di superare il limite sopra descritto. Grazie al supercomputer *Anton 2*, è stato possibile eseguire simulazioni di DM nell'ordine dei millisecondi – tra le più lunghe mai eseguite – che hanno permesso di **determinare siti e modalità di legami di alcuni frammenti chimici** nei confronti della proteina PTP1b, un importante target farmaceutico notoriamente difficile da studiare [17]. Successivi esperimenti hanno dimostrato la validità delle predizioni effettuate tramite DM. Questo recente studio rappresenta la prima dimostrazione del fatto che le simulazioni di DM possono essere utilizzate in maniera *prospettiva* (e non retrospettiva; un uso già consolidato di questa tecnica computazionale) per identificare siti di legame non noti sulla superficie del bersaglio proteico. Come sottolineano gli autori stessi, un approccio del genere non è facilmente applicabile su larga scala (i.e.: screening di librerie di centinaia di frammenti), sebbene l'**utilizzo di hardware specializzato** possa renderlo possibile.

Nonostante ciò, grazie all'incremento della potenza di calcolo e allo sviluppo di modelli fisici sempre più accurati, l'utilizzo di simulazioni di DM su queste scale temporali sarà sempre più comune anche nell'ambito dell'identificazione di frammenti chimici di interesse farmacologico, affiancando le tecnologie già esistenti e applicate nelle prime fasi dello sviluppo dei farmaci.

In tale contesto, l'accelerazione fornita da [servizi di GPU Computing](#), ma anche la loro capacità di ottimizzare gli investimenti tecnologici aumentando il risultato, rappresenta una strategia fondamentale nell'accelerare i tempi della ricerca scientifica.





Bibliografia

1. Isaac Newton. **Philosophiae Naturalis Principia Mathematica**. London, 1686.
2. Danilo Roccatano. **Introduzione alla Dinamica Molecolare**. Maggio 2015
3. Alder BJ, and Wainwright TE (1957). **Phase Transition for a Hard Sphere System**. *J Chem Phys* 27, 1208–1209.
4. McCammon JA, Gelin BR, and Karplus M (1977). **Dynamics of Folded Proteins**. *Nature* 267, 585–590.
5. Levitt M, and Lifson S (1969). **Refinement of Protein Conformations Using a Macromolecular Energy Minimization Procedure**. *Journal of Molecular Biology* 46, 269–279.
6. Lifson S, and Warshel A (1968). **Consistent Force Field for Calculations of Conformations Vibrational Spectra and Enthalpies of Cycloalkane and N-Alkane Molecules**. *J Chem Phys* 49, 5116–5129.
7. Jacob D. Durrant, Sarah E. Kochanek, Lorenzo Casalino, Pek U. leong, Abigail C. Dommer, and Rommie E. Amaro. **Mesoscale All-Atom Influenza Virus Simulations Suggest New Substrate Binding Mechanism**. *ACS Central Science* 2020 6 (2), 189-196.
8. Dommer A, Casalino L, Kearns F, Rosenfeld M, Wauer N, Ahn SH, Russo J, Oliveira S, Morris C, Bogetti A, Trifan A, Brace A, Sztain T, Clyde A, Ma H, Chennubhotla C, Lee H, Turilli M, Khalid S, Tamayo-Mendoza T, Welborn M, Christensen A, Smith DG, Qiao Z, Sirumalla SK, O'Connor M, Manby F, Anandkumar A, Hardy D, Phillips J, Stern A, Romero J, Clark D, Dorrell M, Maiden T, Huang L, McCalpin J, Woods C, Gray A, Williams M, Barker B, Rajapaksha H, Pitts R, Gibbs T, Stone J, Zuckerman DM, Mulholland AJ, Miller T 3rd, Jha S, Ramanathan A, Chong L, Amaro RE. **#COVIDisAirborne: AI-enabled multiscale computational microscopy of delta SARS-CoV-2 in a respiratory aerosol**. *Int J High Perform Comput Appl*. 2023 Jan;37(1):28-44.
9. Thornburg ZR, Bianchi DM, Brier TA, et al. **Fundamental behaviors emerge from simulations of a living minimal cell**. *Cell*. 2022 Jan;185(2):345-360.e28.
10. Stevens JA, Grünewald F, van Tilburg PAM, König M, Gilbert BR, Brier TA, Thornburg ZR, Luthey-Schulten Z and Marrink SJ (2023) **Molecular dynamics simulation of an entire cell**. *Front. Chem*. 11:1106495.
11. Hollingsworth SA, Dror RO. **Molecular Dynamics Simulation for All**. *Neuron*. 2018 Sep 19;99(6): 1129-1143.
12. <https://images.nvidia.com/content/tesla/pdf/Molecular-Dynamics-July-2017-MB-slides.pdf>
13. <https://developer.nvidia.com/blog/a-guide-to-cuda-graphs-in-gromacs-2023/>
14. GROMACS 2023 User's Manual (<https://doi.org/10.5281/zenodo.8134388>)
15. Billesbølle, C.B., de March, C.A., van der Velden, W.J.C. *et al.* **Structural basis of odorant recognition by a human odorant receptor**. *Nature* 615, 742–749 (2023).
16. Kutzner C, Páll S, Fechner M, Esztermann A, de Groot BL, Grubmüller H. Best bang for your buck: **GPU nodes for GROMACS biomolecular simulations**. *J Comput Chem*. 2015 Oct 5;36(26):1990-2008.
17. Jack B. Greisman, Lindsay Willmore, Christine Y. Yeh, Fabrizio Giordanetto, Sahar Shahamadtar, Hunter Nisonoff, Paul Maragakis, and David E. Shaw. **Discovery and Validation of the Binding Poses of Allosteric Fragment Hits to Protein Tyrosine Phosphatase 1b: From Molecular Dynamics Simulations to X-ray Crystallography**. *Journal of Chemical Information and Modeling* 2023 63 (9), 2644-2650.





seeweb

THINK CLOUD



Via Armando Vona 66
03100, Frosinone

Via Caldera, 21
Blue Building ala 1
20153 Milano



[T] +390775880041
[@] info@seeweb.it
[W] www.seeweb.it

cloud server gpu



Samuele Di Cristofano, PhD
Laboratorio di Bioinformatica, Prof. D. Raimondo.
Dip.to di Medicina Molecolare.
Facoltà di Farmacia e Medicina - Sapienza Università di Roma.