

Curso práctico de Dinámica Molecular y Análisis de Trayectorias

OpenMM Application Layer y análisis de trayectorias

Jordi Villà i Freixa

Universitat de Vic - Universitat Central de Catalunya
Facultat de Ciències, Tecnologia i Enginyeries (FCTE)

jordi.villa@uvic.cat

Curso MD y Análisis de Trayectorias
Concepción, enero 2026

Índex

1 Episodio 1: Getting Started (entorno y datos)

- Panorama y escalas
- Variables y estados
- Entropía y energía libre
- Ensembles y partición
- Hamiltoniano y dinámica
- Fuerzas y gradientes
- Partición y probabilidades
- Distribución de Boltzmann
- Espacio de fases y ergodicidad
- Superficie de energía potencial
- Campos de fuerza
- Entorno y datos

2 Episodio 2: Running Simulations

- Discretización temporal
- Integradores clásicos
- Esquema leapfrog

Episodio 1: Getting Started (entorno y datos): Summary

1 Episodio 1: Getting Started (entorno y datos)

- Panorama y escalas
- Variables y estados
- Entropía y energía libre
- Ensembles y partición
- Hamiltoniano y dinámica
- Fuerzas y gradientes
- Partición y probabilidades
- Distribución de Boltzmann
- Espacio de fases y ergodicidad
- Superficie de energía potencial
- Campos de fuerza
- Entorno y datos

2 Episodio 2: Running Simulations

- Discretización temporal
- Integradores clásicos
- Escenarios leapfrog

Objetivos del episodio

- Conectar termodinámica, mecánica estadística y simulación.
- Definir variables, funciones de estado y observables.
- Establecer el flujo de datos y el entorno de trabajo.

Escalas de tiempo y longitud

$$\Delta t_{\text{vibraciones}} \sim 10^{-15} \text{ s},$$

$$\Delta t_{\text{conformaciones}} \sim 10^{-9} - 10^{-3} \text{ s},$$

$$L_{\text{átomos}} \sim 10^{-10} \text{ m}, \quad L_{\text{proteínas}} \sim 10^{-8} \text{ m}.$$

- La simulación atómica resuelve escalas ultrarrápidas.
- La estadística conecta microestados con promedios macroscópicos.

Modelo microscópico

Estado = $(\mathbf{r}_1, \dots, \mathbf{r}_N, \mathbf{p}_1, \dots, \mathbf{p}_N)$,

$$\mathbf{p}_i = m_i \dot{\mathbf{r}}_i.$$

- El sistema está definido por $6N$ coordenadas en el espacio de fases.
- La energía potencial codifica la interacción entre partículas.

Del micro al macro

$$\langle A \rangle = \int A(\mathbf{r}, \mathbf{p}) \rho(\mathbf{r}, \mathbf{p}) d\mathbf{r} d\mathbf{p},$$

$$\bar{A} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T A(t) dt.$$

- La ergodicidad conecta promedios temporales y de conjunto.

Pipeline de simulación

- ① Preparación del sistema (topología, coordenadas, parámetros).
- ② Definición del Hamiltoniano y condiciones de contorno.
- ③ Integración temporal y generación de trayectoria.
- ④ Análisis estadístico de observables.

Variables de estado

$$P, V, T, U, H = U + PV, G = H - TS, A = U - TS.$$

- Funciones de estado: dependen solo del estado macroscópico.
- Funciones de trayectoria: trabajo W y calor Q .

Primera ley y trabajo

$$dU = \delta Q + \delta W,$$

$$\delta W = -P dV + \sum_i f_i dx_i.$$

- La energía interna cambia por intercambio de calor o trabajo.
- En simulación, δW proviene de fuerzas sobre partículas.

Segunda ley y entropía

$$dS \geq \frac{\delta Q}{T},$$
$$S = k_B \ln \Omega.$$

- Ω es el número de microestados compatibles con el macroestado.

Potenciales termodinámicos

$$G = U + PV - TS,$$

$$A = U - TS.$$

- G minimiza a T, P constantes; A minimiza a T, V constantes.

Relaciones diferenciales

$$dG = -S \, dT + V \, dP + \sum_i \mu_i \, dN_i,$$

$$dA = -S \, dT - P \, dV + \sum_i \mu_i \, dN_i.$$

- Las derivadas parciales dan observables medibles.

Microestados y macroestados

$$S = k_B \ln \Omega, \quad \Omega = \sum_{\text{micro}} 1.$$

- Incrementos de entropía favorecen estados con más configuraciones.

Energía libre y equilibrio

$$\Delta G = \Delta H - T\Delta S,$$

$$\Delta A = \Delta U - T\Delta S.$$

- En equilibrio, $\Delta G = 0$ para procesos a T, P fijos.

Potencial químico

$$\mu_i = \left(\frac{\partial G}{\partial N_i} \right)_{T,P,N_{j \neq i}} .$$

- Controla intercambios de materia y equilibrio de fases.

Relación con probabilidades

$$P(\text{estado}) \propto e^{-\beta G}, \quad \beta = 1/(k_B T).$$

- La energía libre gobierna el peso estadístico de macroestados.

Ejemplo: equilibrio $A \rightleftharpoons B$

$$K = e^{-\beta \Delta G},$$
$$\Delta G = -k_B T \ln K.$$

- La estabilidad relativa se expresa en términos de ΔG .

Ensembles clásicos

- Microcanónico (NVE): N, V, E fijos.
- Canónico (NVT): N, V, T fijos.
- Isotermo-isobárico (NPT): N, P, T fijos.

Función de partición canónica

$$Z = \frac{1}{h^{3N} N!} \int e^{-\beta H(\mathbf{r}, \mathbf{p})} d\mathbf{r} d\mathbf{p}.$$

- Central para derivar energías libres y promedios.

Observables en NVT

$$\langle A \rangle = \frac{1}{Z} \int A e^{-\beta H} d\mathbf{r} d\mathbf{p}.$$

- En simulación, se estiman con promedios temporales.

Ensemble NPT

$$\Delta = \int dV e^{-\beta PV} Z(N, V, T),$$
$$G = -k_B T \ln \Delta.$$

- Incluye fluctuaciones de volumen controladas por el barostato.

Fluctuaciones termodinámicas

$$C_V = \frac{\langle E^2 \rangle - \langle E \rangle^2}{k_B T^2},$$

$$\kappa_T = \frac{\langle V^2 \rangle - \langle V \rangle^2}{k_B T \langle V \rangle}.$$

- Las fluctuaciones conectan estadística y respuesta macroscópica.

Hamiltoniano clásico

$$H(\mathbf{r}, \mathbf{p}) = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r}).$$

- Se separa en energía cinética y potencial.

Ecuaciones de Hamilton

$$\dot{\mathbf{r}}_i = \frac{\partial H}{\partial \mathbf{p}_i},$$
$$\dot{\mathbf{p}}_i = -\frac{\partial H}{\partial \mathbf{r}_i}.$$

- Equivalentes a la segunda ley de Newton.

Ecuaciones de Newton

$$m_i \ddot{\mathbf{r}}_i = -\nabla_i U(\mathbf{r}).$$

- La fuerza es el gradiente negativo del potencial.

Conservación de energía

$$\frac{dH}{dt} = 0 \quad (\text{NVE ideal}).$$

- La integración numérica introduce errores controlables.

Flujo en el espacio de fases

$$\frac{d\rho}{dt} = \{\rho, H\} = 0 \quad (\text{ecuación de Liouville}).$$

- La densidad de probabilidad se conserva en el flujo hamiltoniano.

Fuerzas a partir del potencial

$$\mathbf{F}_i = -\nabla_i U(\mathbf{r}).$$

- En MD, el coste dominante es evaluar U y \mathbf{F}_i .

Potenciales de pares

$$U = \sum_{i < j} u(r_{ij}), \quad r_{ij} = \|\mathbf{r}_i - \mathbf{r}_j\|.$$

- Simplifican la energía total con interacciones binarias.

Ejemplo: Lennard-Jones

$$u_{\text{LJ}}(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right].$$

- Repulsión a corta distancia, atracción a media distancia.

Energía electrostática

$$u_C(r) = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}.$$

- Dominante en sistemas biomoleculares.

Gradientes y Hessiano

$$\nabla U \equiv \left(\frac{\partial U}{\partial r_1}, \dots, \frac{\partial U}{\partial r_{3N}} \right),$$
$$\text{Hess} = \nabla \nabla U.$$

- El Hessiano describe curvaturas locales (modos normales).

Separación cinética/potencial

$$Z = Z_{\text{kin}} Z_{\text{conf}},$$

$$Z_{\text{kin}} = \prod_i \left(\frac{2\pi m_i}{\beta h^2} \right)^{3/2}.$$

- Z_{conf} depende solo de $U(\mathbf{r})$.

Distribución de configuraciones

$$P(\mathbf{r}) = \frac{e^{-\beta U(\mathbf{r})}}{Z_{\text{conf}}}.$$

- Base del muestreo Monte Carlo y MD con termostato.

Energía libre configuracional

$$A = -k_B T \ln Z_{\text{conf}} + \text{const.}$$

- Relaciona el muestreo con ΔA entre estados.

Distribución de velocidades

$$P(\mathbf{v}) \propto e^{-\beta \sum_i \frac{1}{2} m_i v_i^2}.$$

- Permite inicializar velocidades a temperatura T .

Teorema de equipartición

$$\langle K \rangle = \frac{3N}{2} k_B T.$$

- Cada grado de libertad cuadrático aporta $\frac{1}{2} k_B T$.

Forma general

$$P(E) = \frac{1}{Z} e^{-\beta E}.$$

- El peso exponencial penaliza estados de alta energía.

Distribución de Maxwell-Boltzmann

$$f(v) = 4\pi \left(\frac{m}{2\pi k_B T} \right)^{3/2} v^2 e^{-\frac{mv^2}{2k_B T}}.$$

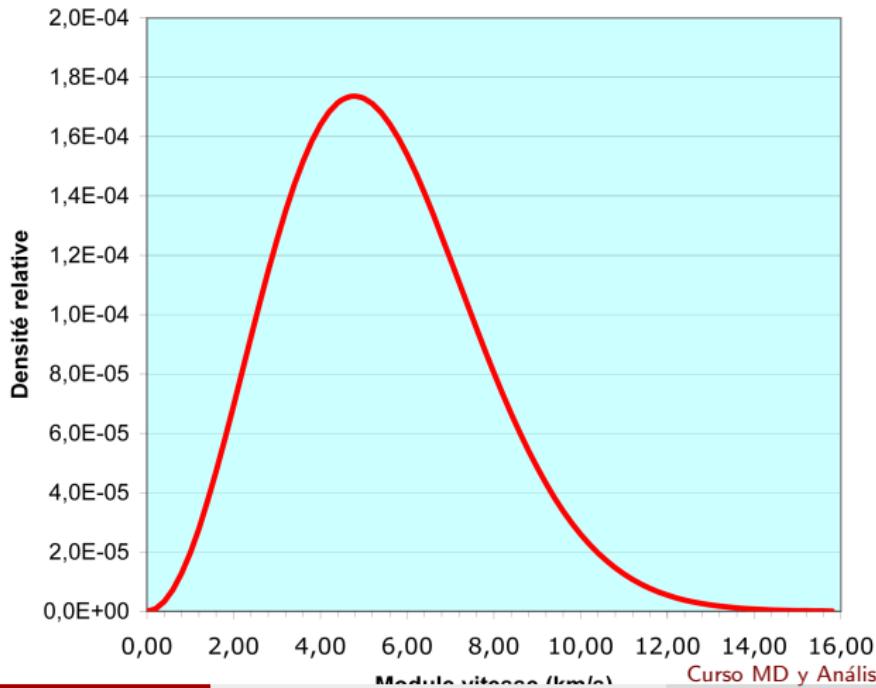
- Caracteriza la distribución de velocidades en equilibrio.

Interpretación estadística

- La mayoría de partículas se agrupan alrededor de v_m .
- El ensanchamiento crece con T .

Ejemplo gráfico

Distribution des vitesses des molécules d'hélium à 5500 K (Distribution de Maxwell-Boltzmann)



Aplicación en simulación

- Inicializar velocidades según $f(v)$ evita sesgos térmicos.
- Permite comprobar que el termostato reproduce el equilibrio.

Espacio de fases

$$\Gamma = (\mathbf{r}, \mathbf{p}) \in \mathbb{R}^{6N}.$$

- Cada punto representa un microestado completo.

Ecuación de Liouville

$$\frac{\partial \rho}{\partial t} + \{\rho, H\} = 0.$$

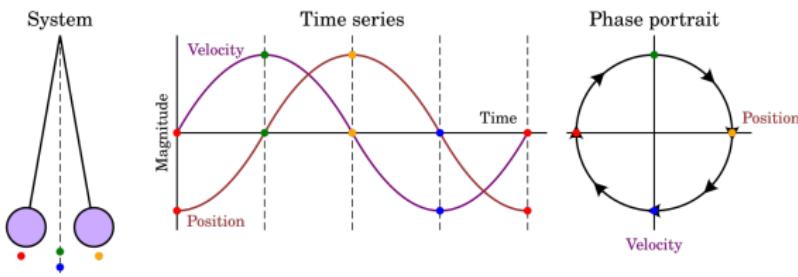
- El volumen en Γ se conserva (teorema de Liouville).

Ergodicidad

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T A(t) dt = \langle A \rangle_{\text{ensemble}}.$$

- Supuesto clave para sustituir promedios de conjunto por tiempo.

Trayectorias en Γ



Fuente: Wikimedia Commons (CC BY-SA 4.0). [9]

Interpretación física

- Órbitas cerradas: movimiento periódico.
- Órbitas abiertas: difusión en regiones de alta energía.

Definición

$$U(\mathbf{r}) : \mathbb{R}^{3N} \rightarrow \mathbb{R}.$$

- Describe el paisaje energético donde se mueve el sistema.

Coordenadas de reacción

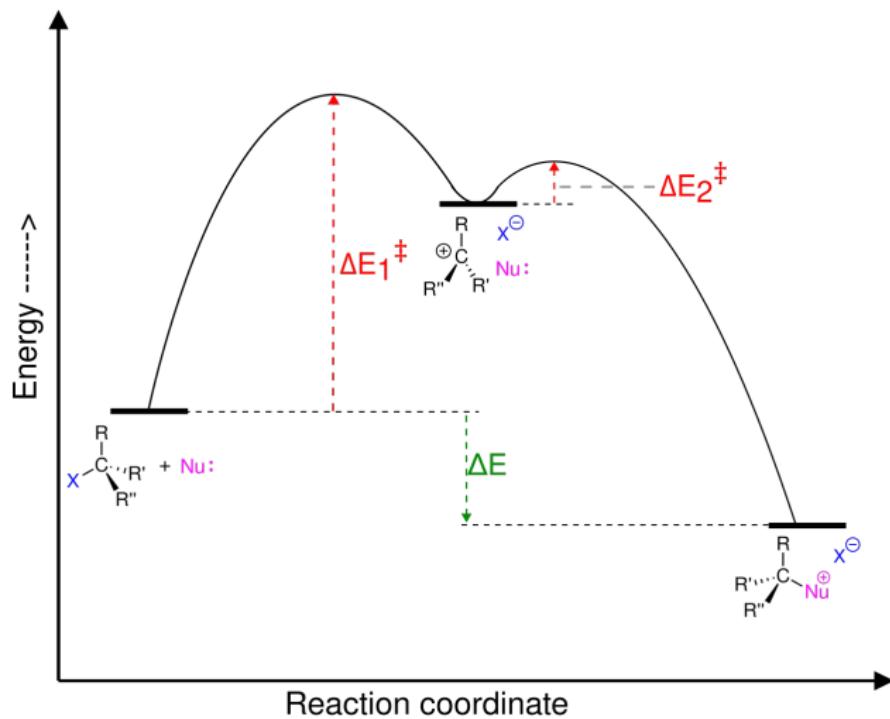
$$\xi = \xi(\mathbf{r}), \quad F(\xi) = -k_B T \ln P(\xi).$$

- Proyección que resume la dinámica en una coordenada clave.

Barreras y estados metaestables

- Mínimos locales: conformaciones metaestables.
- Barreras: controlan las escalas de transición.

Ejemplo de superficie



Fuente: Wikimedia Commons (CC BY-SA 4.0). [15]

Dinámica sobre la superficie

- Integración clásica produce trayectorias en $U(\mathbf{r})$.
- Los termostatos aseguran muestreo canónico.

Descomposición del potencial

$$U = U_{\text{bond}} + U_{\text{angle}} + U_{\text{dihedral}} + U_{\text{nonbond}}.$$

- Separación entre términos enlazados y no enlazados.

Términos enlazados

$$U_{\text{bond}} = \sum_b k_b (r_b - r_b^0)^2,$$

$$U_{\text{angle}} = \sum_a k_a (\theta_a - \theta_a^0)^2.$$

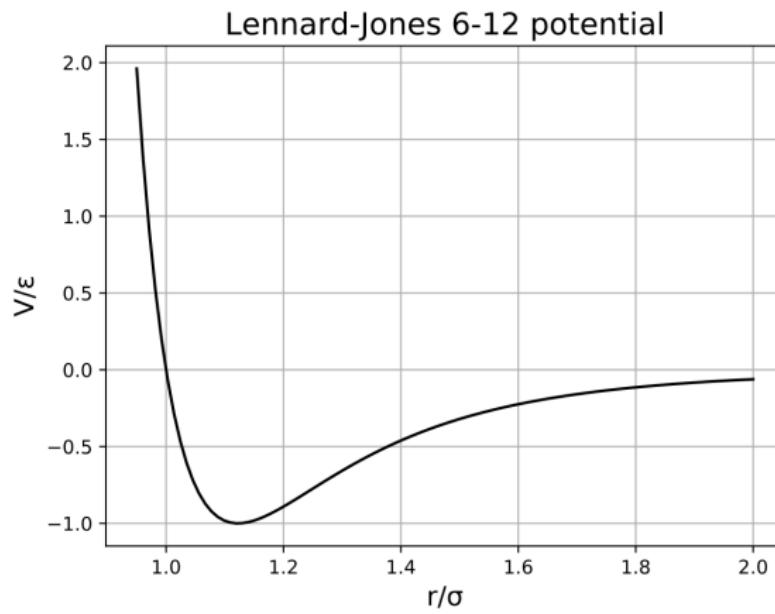
- Aproximación armónica alrededor de geometrías de equilibrio.

Dihedros y rotación

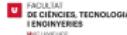
$$U_{\text{dihedral}} = \sum_d \frac{V_d}{2} [1 + \cos(n_d \phi_d - \gamma_d)].$$

- Controlan barreras rotacionales y preferencias conformacionales.

Ejemplo LJ



Fuente: Wikimedia Commons (CC BY-SA 4.0). [2]



Parámetros y validación

- Los campos de fuerza se ajustan a datos cuánticos y experimentales.
- La validación incluye densidades, energías y conformaciones.

Directorio de trabajo

- Variable de entorno COURSE_DIR.
- Estructura: data/ (entradas) y results/ (salidas).

Sistemas del curso

- Sistema simple: alanina.
- Sistema complejo: proteína + ligando.

Flujo de datos

- ① Descarga de datos base.
- ② Preparación de topología y coordenadas.
- ③ Validación del entorno de cálculo.

Reproducibilidad

- Controlar versiones de librerías y semillas aleatorias.
- Guardar archivos de entrada y scripts.

Resumen del episodio

- La base termodinámica se conecta con el muestreo estadístico.
- El potencial define fuerzas y probabilidades.
- El entorno de trabajo garantiza reproducibilidad.

Episodio 2: Running Simulations: Summary

1 Episodio 1: Getting Started (entorno y datos)

- Panorama y escalas
- Variables y estados
- Entropía y energía libre
- Ensembles y partición
- Hamiltoniano y dinámica
- Fuerzas y gradientes
- Partición y probabilidades
- Distribución de Boltzmann
- Espacio de fases y ergodicidad
- Superficie de energía potencial
- Campos de fuerza
- Entorno y datos

2 Episodio 2: Running Simulations

- Discretización temporal
- Integradores clásicos
- Esquema leapfrog

Ecuaciones continuas y discretas

$$\dot{\mathbf{r}}(t) = \mathbf{v}(t), \quad \dot{\mathbf{v}}(t) = \frac{\mathbf{F}(t)}{m}.$$

- La simulación sustituye derivadas por incrementos finitos.

Paso de tiempo

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \mathcal{O}(\Delta t^2).$$

- Δt debe resolver las vibraciones más rápidas.

Error local y global

$$\begin{aligned}\text{error local} &= \mathcal{O}(\Delta t^{p+1}), \\ \text{error global} &= \mathcal{O}(\Delta t^p).\end{aligned}$$

- Los integradores de orden p controlan la acumulación de error.

Escala de estabilidad

$$\Delta t \lesssim \frac{1}{10 \omega_{\max}}.$$

- ω_{\max} proviene de los modos vibracionales más rápidos.

Criterio práctico

- Sin restricciones: $\Delta t \approx 1$ fs.
- Con SHAKE/RATTLE: $\Delta t \approx 2$ fs.

Euler explícito

$$\mathbf{v}_{t+\Delta t} = \mathbf{v}_t + \frac{\mathbf{F}_t}{m} \Delta t,$$
$$\mathbf{r}_{t+\Delta t} = \mathbf{r}_t + \mathbf{v}_t \Delta t.$$

- Simple pero inestable para dinámica molecular.

Verlet clásico

$$\mathbf{r}_{t+\Delta t} = 2\mathbf{r}_t - \mathbf{r}_{t-\Delta t} + \frac{\mathbf{F}_t}{m}\Delta t^2.$$

- Simétrico y reversible en el tiempo.

Velocity Verlet

$$\begin{aligned}\mathbf{r}_{t+\Delta t} &= \mathbf{r}_t + \mathbf{v}_t \Delta t + \frac{\mathbf{F}_t}{2m} \Delta t^2, \\ \mathbf{v}_{t+\Delta t} &= \mathbf{v}_t + \frac{\mathbf{F}_t + \mathbf{F}_{t+\Delta t}}{2m} \Delta t.\end{aligned}$$

- Combina estabilidad y acceso directo a velocidades.

Simetría y estabilidad

- Integradores simécticos conservan volumen en el espacio de fases.
- Evitan deriva energética en NVE.

Elección práctica

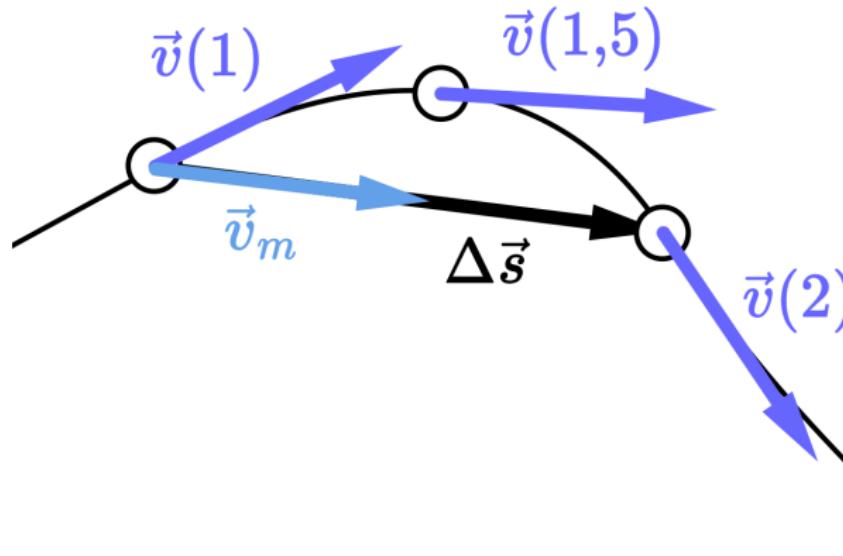
- OpenMM usa integradores tipo Verlet o Langevin.
- La elección depende del ensemble deseado.

Leapfrog

$$\mathbf{v}_{t+\Delta t/2} = \mathbf{v}_{t-\Delta t/2} + \frac{\mathbf{F}_t}{m} \Delta t,$$
$$\mathbf{r}_{t+\Delta t} = \mathbf{r}_t + \mathbf{v}_{t+\Delta t/2} \Delta t.$$

- Velocidades y posiciones “se alternan” en el tiempo.

Interpretación gráfica



Fuente: Wikimedia Commons (CC BY-SA 4.0). [11]

Ventajas

- Simétrico, reversible y de bajo coste.
- Conserva mejor la energía que Euler.

Relación con Verlet

- Leapfrog y velocity Verlet son algebraicamente equivalentes.
- Diferencias en el almacenamiento de velocidades.

Elección de integrador

- NVE: Verlet/Leapfrog.
- NVT: Langevin o Nosé-Hoover.

Energía total

$$E(t) = K(t) + U(t).$$

- En NVE ideal, $E(t)$ es constante.

Deriva energética

- Errores numéricos producen deriva secular.
- Se controla reduciendo Δt .

Temperatura instantánea

$$T = \frac{2\langle K \rangle}{3Nk_B}$$

- En NVE, T fluctúa alrededor de un valor medio.

Comprobaciones rápidas

- Estabilidad de $E(t)$ y distribución de velocidades.
- Comparar $\langle K \rangle$ con el valor teórico.

Tiempo de equilibración

- Un pre-equilibrado estabiliza la energía antes de producir datos.

Objetivo del termostato

- Imponer una distribución canónica a temperatura T .
- Extraer o inyectar energía de forma controlada.

Dinámica de Langevin

$$\begin{aligned}m\ddot{\mathbf{r}} &= -\nabla U - \gamma m\dot{\mathbf{r}} + \mathbf{R}(t), \\ \langle \mathbf{R}(t)\mathbf{R}(t') \rangle &= 2\gamma m k_B T \delta(t - t').\end{aligned}$$

- Termostato estocástico con fricción y ruido.

Andersen

- Reasigna velocidades aleatoriamente con frecuencia ν .
- Bueno para muestreo, menos realista dinámicamente.

Nosé-Hoover

$$\dot{\mathbf{p}}_i = \mathbf{F}_i - \xi \mathbf{p}_i, \quad \dot{\xi} = \frac{1}{Q} \left(\sum_i \frac{\mathbf{p}_i^2}{m_i} - 3Nk_B T \right).$$

- Termostato determinista con variable extendida.

Elección del termostato

- Langevin: robusto y estable.
- Nosé-Hoover: dinámica más realista si está bien parametrizado.

Presión microscópica

$$P = \frac{Nk_B T}{V} + \frac{1}{3V} \sum_{i < j} \mathbf{r}_{ij} \cdot \mathbf{F}_{ij}.$$

- La presión depende de energía cinética y fuerzas internas.

Barostato de Berendsen

$$\frac{dV}{dt} = \frac{1}{\tau_P} (P_0 - P) V.$$

- Rápido para equilibrar, no reproduce fluctuaciones exactas.

Parrinello-Rahman

- Escala la celda de simulación con variables dinámicas.
- Permite cambios anisotrópicos del volumen.

NPT

- Ensemble realista para condiciones experimentales.
- Combina termostato y barostato.

Elección práctica

- Equilibración: Berendsen + Langevin.
- Producción: barostato más riguroso (Monte Carlo barostat).

Motivación

- Eliminar vibraciones rápidas permite aumentar Δt .

Restricciones holónomas

$$g_k(\mathbf{r}) = 0, \quad k = 1, \dots, M.$$

- Fijan distancias o àngulos internos.

SHAKE/RATTLE

- SHAKE corrige posiciones; RATTLE corrige posiciones y velocidades.

Impacto numérico

- Aumenta estabilidad y reduce coste de simulación.
- Puede alterar modos de alta frecuencia.

Recomendación

- Usar restricciones en enlaces con H para biomoléculas.

Motivación de PBC

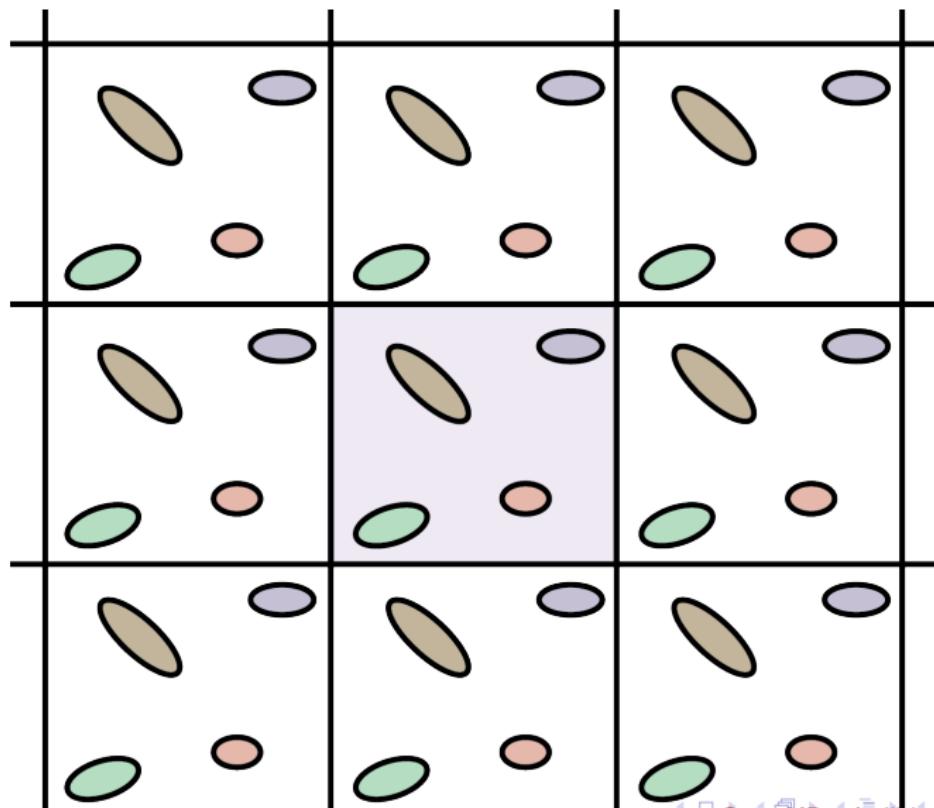
- Evita efectos de superficie en sistemas finitos.
- Emula un sistema infinito por replicación de la celda.

Celda periódica

$$\mathbf{r} \rightarrow \mathbf{r} + n_x \mathbf{a} + n_y \mathbf{b} + n_z \mathbf{c}.$$

- **a, b, c** definen la caja de simulación.

Ejemplo 2D



Implicaciones en dinámica

- Las partículas que salen reingresan por la cara opuesta.
- Se preserva la densidad en el volumen simulado.

Cajas no ortogonales

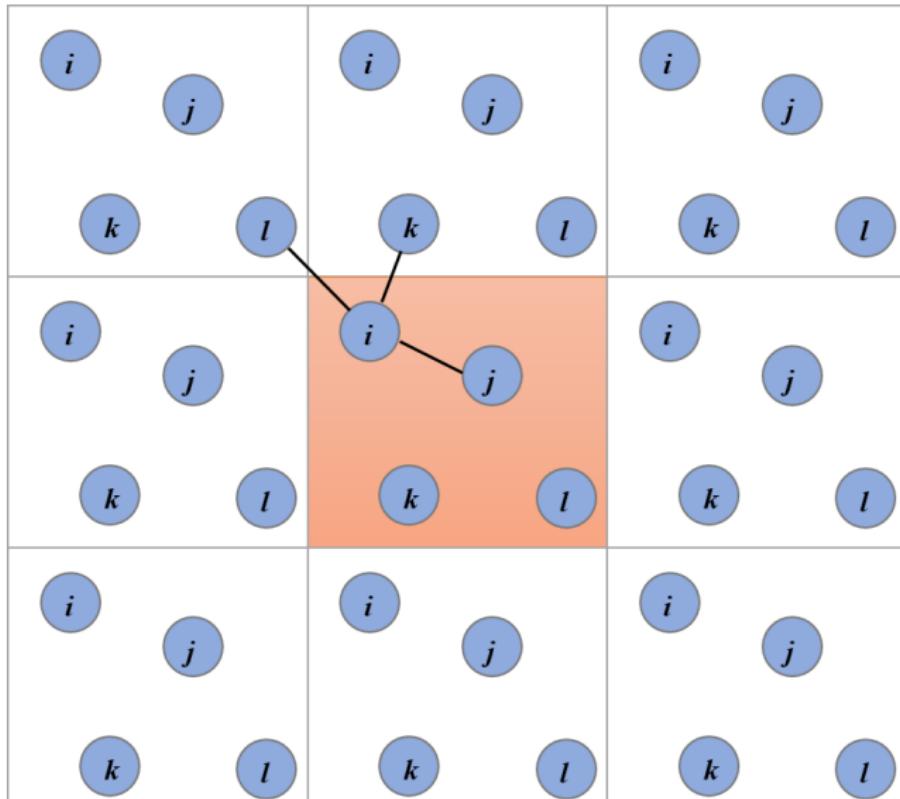
- Celdas triclínicas para cristales o membranas inclinadas.

Convención de imagen mínima

$$r_{ij} = \min_{\mathbf{n}} \|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}\|.$$

- Usa la imagen periódica más cercana para calcular interacciones.

Ejemplo visual



Radio de corte

- La imagen mínima requiere $r_c < L/2$ para evitar dobles conteos.

Impacto en energía

- Errores si el corte es grande o la densidad baja.

Buenas prácticas

- Ajustar r_c y tamaño de caja según la densidad del sistema.

Interacciones de corto alcance

- LJ se trunca a r_c con o sin función de suavizado.

Función switch

$$u_{\text{switch}}(r) = s(r) u(r), \quad s(r) \in [0, 1].$$

- Evita discontinuidades en fuerzas.

Corrección de cola

$$U_{\text{tail}} \approx 2\pi\rho \int_{r_c}^{\infty} r^2 u(r) dr.$$

- Corrige energía y presión al truncar LJ.

Escogiendo r_c

- Mayor r_c mejora precisión, pero aumenta coste.

Implicaciones en estabilidad

- Un corte abrupto puede introducir ruido en la dinámica.

Problema de Coulomb

$$u_C(r) \propto \frac{1}{r} \quad (\text{largo alcance}).$$

- El truncamiento directo produce grandes errores.

Ewald

$$U = U_{\text{real}} + U_{\text{rec}} + U_{\text{self}}.$$

- Divide Coulomb en suma real y recíproca.

PME

- Particle Mesh Ewald usa FFT para acelerar el cálculo.

Precisión

- Controla el error con parámetros de malla y àlfa de Ewald.

Elección práctica

- Para biomoléculas en solvente, PME es el estándar.

Trayectorias

- Guardar coordenadas y velocidades en intervalos regulares.
- Balance entre almacenamiento y resolución temporal.

Observables

$$\langle A \rangle \approx \frac{1}{M} \sum_{k=1}^M A_k.$$

- La autocorrelación determina el número efectivo de muestras.

Semillas aleatorias

- Documentar seeds para reproducir termostatos estocásticos.

Checkpoints

- Guardar estados para reiniciar simulaciones largas.

Resumen del episodio

- La integración numérica controla estabilidad y fidelidad.
- Termostatos y barostatos definen el ensemble.
- PBC y electrostática de largo alcance son críticos en biomoléculas.

Episodio 3: Model Building & Editing: Summary

1 Episodio 1: Getting Started (entorno y datos)

- Panorama y escalas
- Variables y estados
- Entropía y energía libre
- Ensembles y partición
- Hamiltoniano y dinámica
- Fuerzas y gradientes
- Partición y probabilidades
- Distribución de Boltzmann
- Espacio de fases y ergodicidad
- Superficie de energía potencial
- Campos de fuerza
- Entorno y datos

2 Episodio 2: Running Simulations

- Discretización temporal
- Integradores clásicos
- Esquema leapfrog

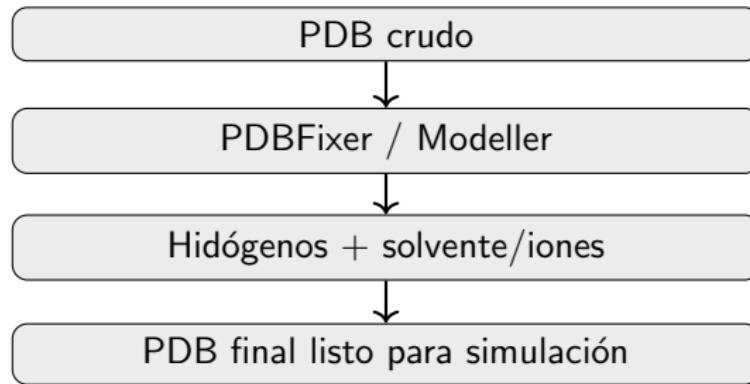
Objetivos de la sesión

- Preparar sistemas listos para simular.
- Aplicar operaciones de Modeller.
- Guardar estructuras reproducibles.

Modeller: operaciones clave

- Añadir hidrógenos y solvente.
- Añadir membranas o partículas extra.
- Eliminar agua para solvente implícito.

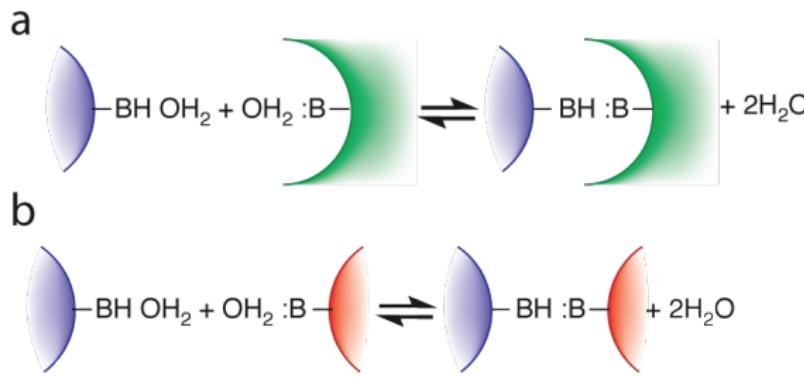
Model building (pipeline visual)



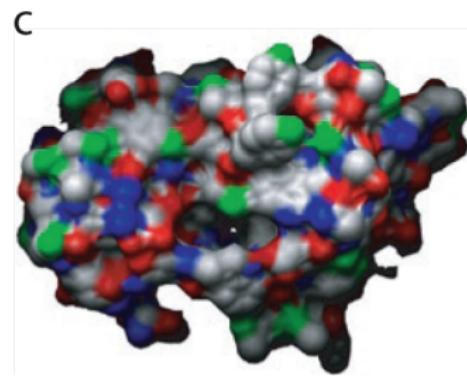
Guardar resultados

- Minimizar y escribir PDB final.
- Evitar repetir operaciones en cada simulación.
- Controlar versiones y trazabilidad.

Binding of cetuximab



[14]



Electrostatics principles

fons

Electrostatics principles

fons

- Charge interactions obey Coulomb's law

Electrostatics principles

fons

- Charge interactions obey Coulomb's law
- When more than two charges interact, the energies are sums of Coulombic interactions

Electrostatics principles

fons

- Charge interactions obey Coulomb's law
- When more than two charges interact, the energies are sums of Coulombic interactions
- Electric field, Gauss's law and electrostatic potentials allow us to do such calculations

Coulomb's law I



$$U(r) = \frac{1}{4\pi\epsilon_0\epsilon_r} \frac{q_1 q_2}{r} \quad (1)$$

- Clearly long ranged (r^{-1}) with respect to dispersion forces (r^{-6}) and strong
- Polarizable media (charges redistribute in response to an electric field) shield charges strongly (large ϵ_r)
- As ϵ_r depends on temperature, Eq. 1 resembles, actually, a free energy expression.
- Polarizability arises from: permanent dipoles, atomic polarizabilities, hydrogen bonds

Coulomb's law II

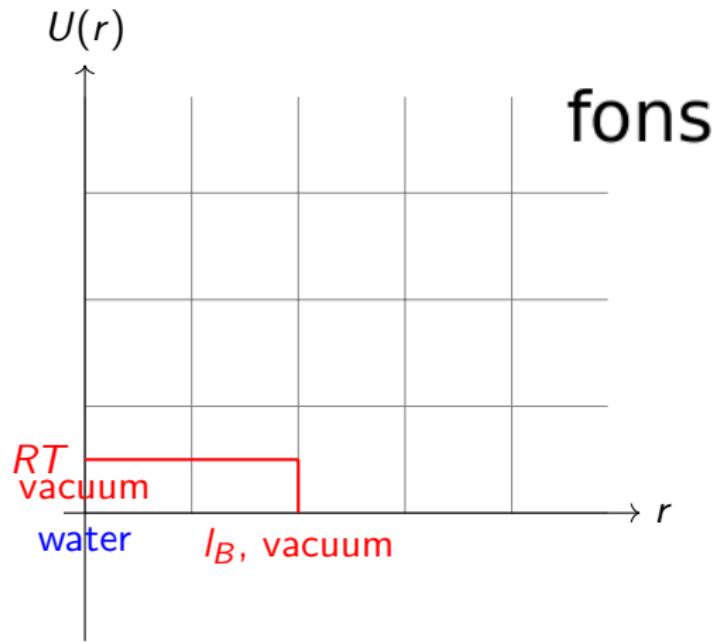
fons

- The Bjerrum length I_B describes the charge separation at which the Coulomb energy $U(r)$ equals the thermal energy RT [5]. For example, for $q_1 = q_2 = e$:

$$I_B = \frac{1}{4\pi\epsilon_0\epsilon_r} \frac{e^2 N}{RT}$$

At vacuum this occurs around 560 Å (in water this needs to be divided by $\epsilon_r = 80$). At bigger distances, the interactions are weaker than thermal energy RT , and particles are governed by Brownian motion.

Charges interact weaker in water



Electrostatic force and field

$$\mathbf{F} = -\nabla U = \frac{1}{4\pi\epsilon_0\epsilon_r} \frac{q_1 q_2}{r^2} \frac{\mathbf{r}}{r}$$

which, for a unit charge, becomes the electrostatic field if dependent on just one particle with charge q :

$$\mathbf{E}(\mathbf{r}) = \frac{\mathbf{F}(\mathbf{r})}{q_{\text{test}}} = \frac{1}{4\pi\epsilon_0\epsilon_r} \frac{q}{r^2} \frac{\mathbf{r}}{r}$$

For more complex settings of charges, one may use Gauss' law, which equals the flux of the electrostatic field through any bounding surface to the sum of all charges enclosed:

$$\phi = \int_{\text{surface}} \epsilon \mathbf{E} \cdot d\mathbf{s} = \frac{1}{\epsilon_0} \sum_{i=1}^n q_i \quad (2)$$

Electrostatic potentials

Electrostatic field describes forces (vectorial). Electrostatic potential describes energies (scalar). We start by:

$$\delta w = -\mathbf{F}d\mathbf{l} = -q\mathbf{E}d\mathbf{l}$$

which allows us to define the work done *against* the field when moving a charge q between two points as:

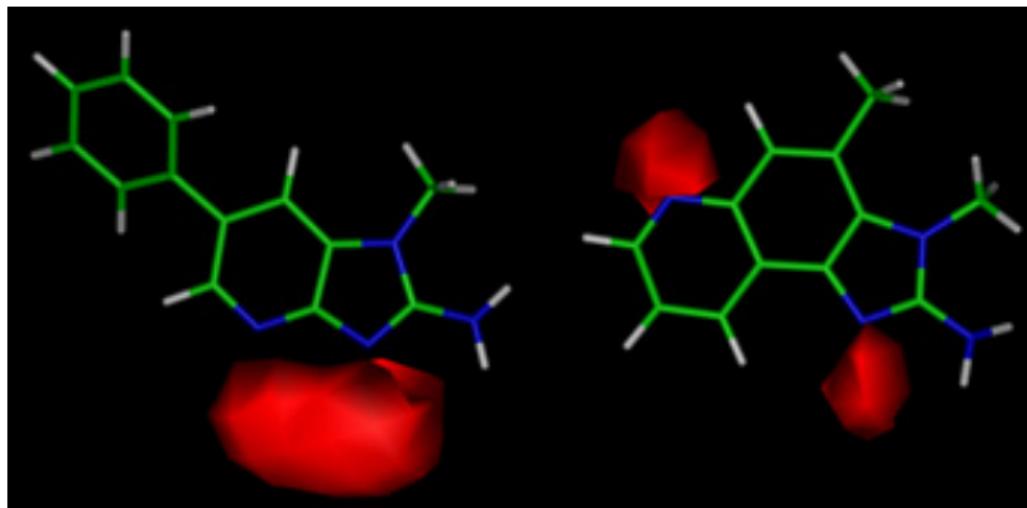
$$w_{A \rightarrow B} = -q \int_A^B \mathbf{E}d\mathbf{l}$$

and the corresponding difference in electrostatic potentials as:

$$\psi_B - \psi_A = \frac{w_{AB}}{q_{\text{test}}} = - \int_A^B \mathbf{E}d\mathbf{l}$$

(equivalent to $\mathbf{E} = -\nabla\psi$). Thus, the electrostatic potential around a point charge, $\psi_{\text{test}} = \frac{q_{\text{fixed}}}{4\pi\varepsilon_0\varepsilon_r r}$, and that produced by a charge density, $\psi_{\text{test}} = \frac{1}{4\pi\varepsilon_0\varepsilon_r} \int_V \frac{\rho_{\text{fixed}}}{r_{12}} dV$ (r_{12} , distances between all the charges and the test charge).

Electrostatic potential surfaces



The work along equipotential curves is zero. In addition, electrostatic interactions are conservative forces.

Mixing thermodynamics with electrostatics

We shall investigate two situations:

- Moving charges from one point to another within a fixed electrostatic field (Session V on ion channels)
- Computing free energies for creating the electrostatic fields: "charging up" an assembly of originally uncharged particles:

$$\Delta G_{el} = w_{el} = \frac{1}{2} \sum_i q_i \psi_i = \frac{1}{8\pi\epsilon_0\epsilon_r} \sum_i \sum_{j \neq i} \frac{q_i q_j}{r_{ij}}$$

Ion solvation: Born Energy I

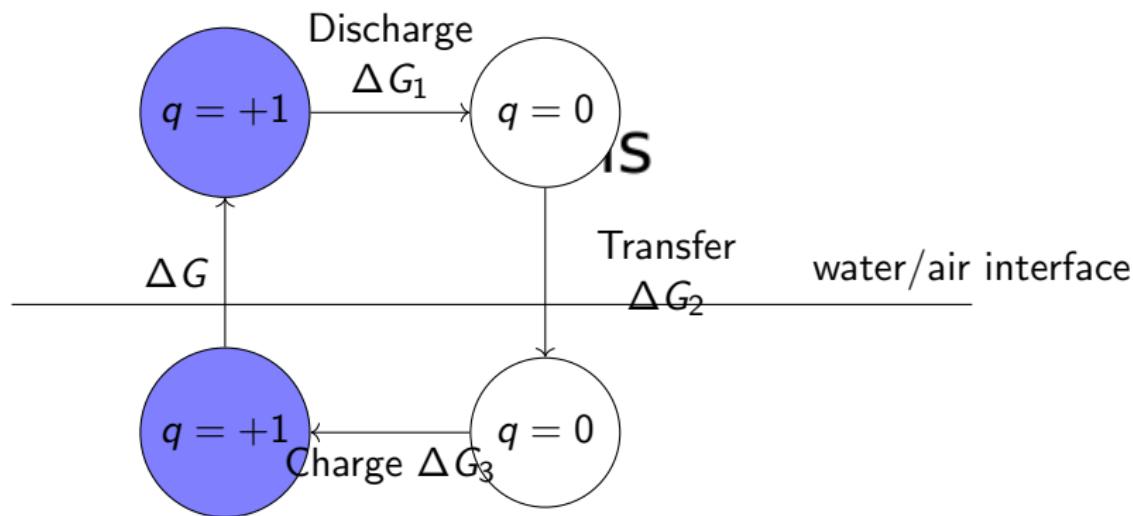
Charging up a continuous distribution resembles the above equation, but using a continuous description of the charge:

$$\Delta G_{el} = \frac{1}{2} \int_V \rho \psi_V dV$$

or, for a charged sphere of radius a :

$$\Delta G_{el} = \frac{1}{2} \int_S \sigma \psi_S dS = \frac{1}{2} \left(\frac{q}{4\pi a^2} \right) \left(\frac{1}{4\pi \epsilon_0 \epsilon_r} \frac{q}{a} \right) 4\pi a^2 = \frac{1}{4\pi \epsilon_0 \epsilon_r} \frac{q^2}{2a}$$

Ion solvation: Born Energy II



$$\Delta G_{el} = \Delta G_1 + \underbrace{\Delta G_2}_{\approx 0} + \Delta G_3 = \frac{q^2}{8\pi\epsilon_0 a} \left(\frac{1}{\epsilon_0} - \frac{1}{\epsilon_w} \right)$$

Episodio 4: Advanced Simulation Examples: Summary

1 Episodio 1: Getting Started (entorno y datos)

- Panorama y escalas
- Variables y estados
- Entropía y energía libre
- Ensembles y partición
- Hamiltoniano y dinámica
- Fuerzas y gradientes
- Partición y probabilidades
- Distribución de Boltzmann
- Espacio de fases y ergodicidad
- Superficie de energía potencial
- Campos de fuerza
- Entorno y datos

fons

2 Episodio 2: Running Simulations

- Discretización temporal
- Integradores clásicos
- Esquema leapfrog

Objetivos de la sesión

fons

- Aplicar técnicas avanzadas sobre sistemas pequeños.
- Extender los scripts con fuerzas y reportes personalizados.
- Conocer el ecosistema OpenMM.

Ejemplos avanzados de la guía

fons

- Simulated annealing (control de temperatura).
- Fuerza externa en contenedor esférico.
- Reportes personalizados de fuerzas.
- Cálculo de energías en conjuntos de estructuras.

Técnicas avanzadas (mapa visual)

fons

Control

- Annealing
- Barostatos
- Thermostatos

Exploración

- Bias/forces externas
- Sampling acelerado
- Análisis rápido de energías

MD vs Docking (testimoni)

fons

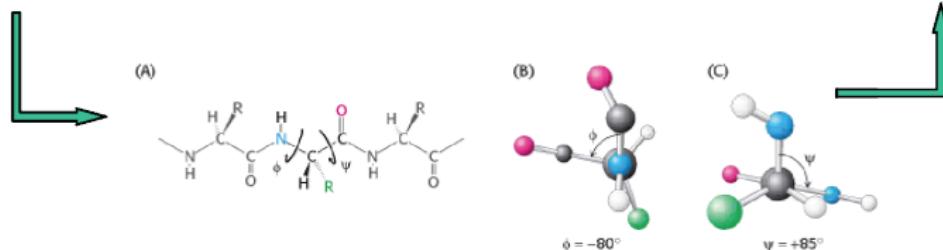
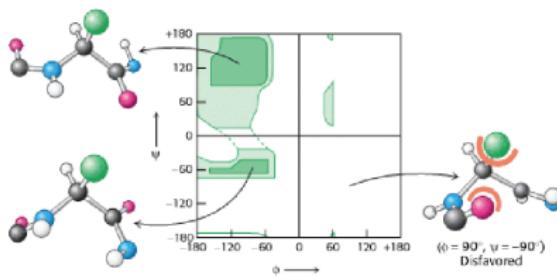
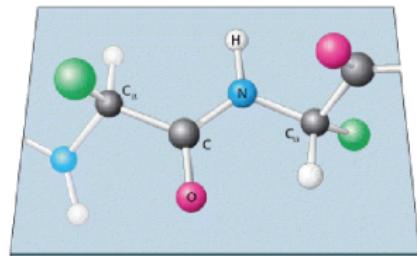
- Dinàmica molecular: trajectòries i mostreig continu del paisatge energètic.
- Docking: orientacions discretes i scoring per estimar interaccions.
- El docking aporta hipòtesis ràpides; la MD valida estabilitat i dinàmica.

Ecosistema OpenMM

fons

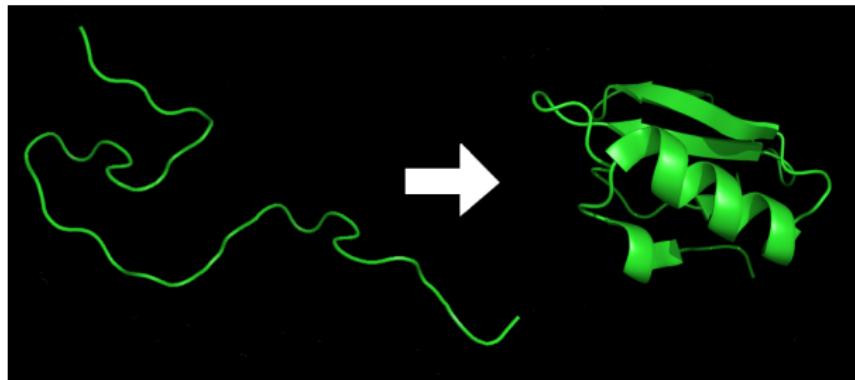
- OpenMM-Tools, OpenMM-ML, OpenMM-Torch.
- OpenMMForceFields + OpenFF Toolkit.
- Integración con PLUMED y flujos de sampling.

Levinthal's paradox

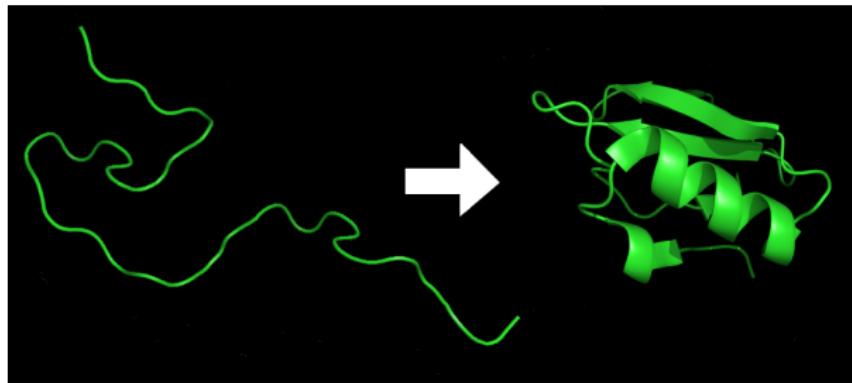
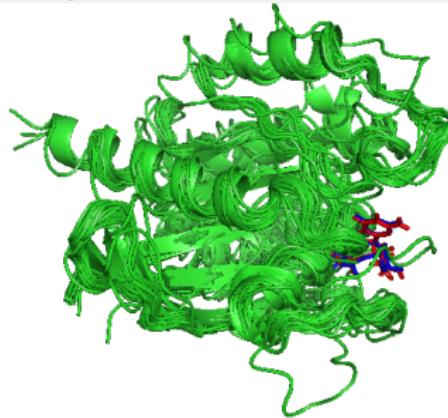


Folding

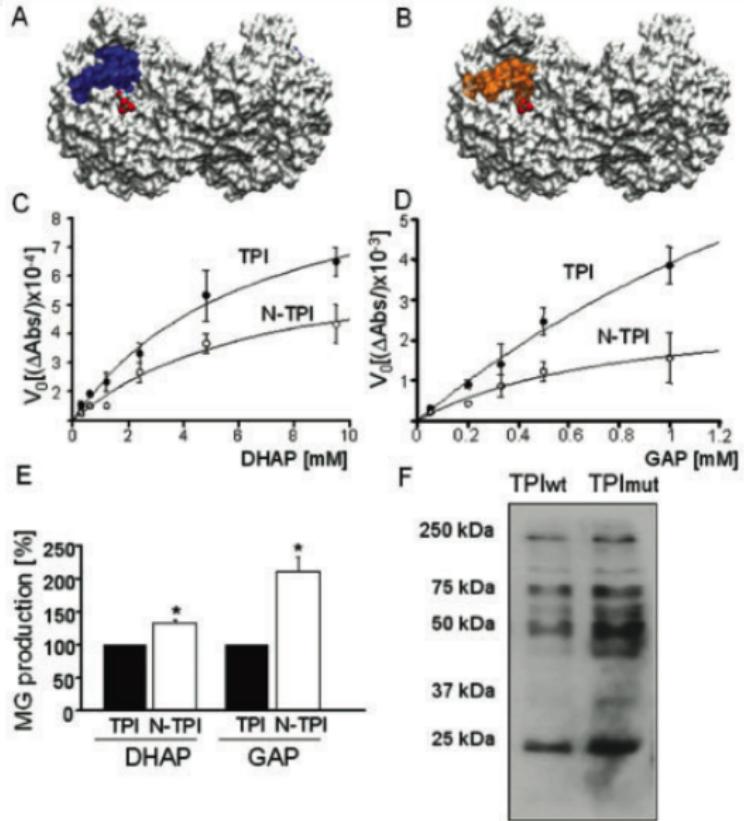
fons



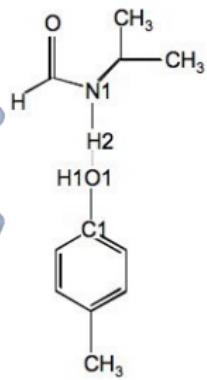
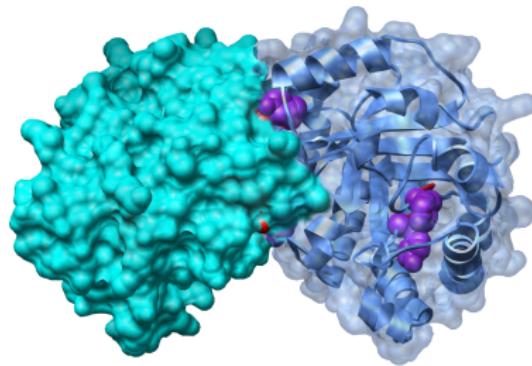
Folding and simple conformational changes



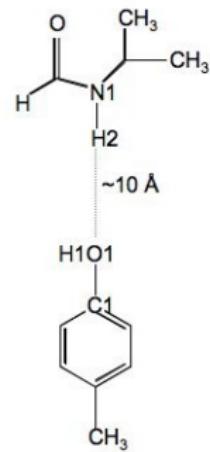
Subtile effects on conformational changes



Subtile effects on conformational changes

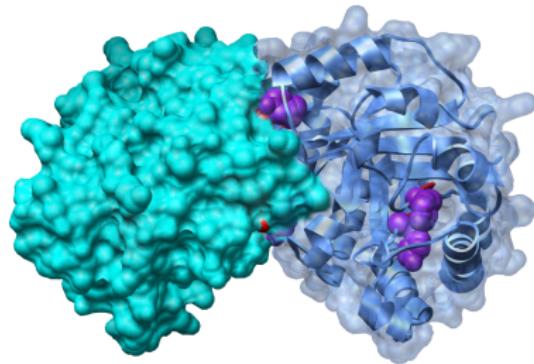


closed state

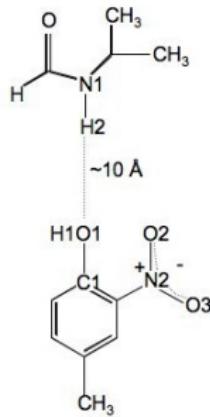
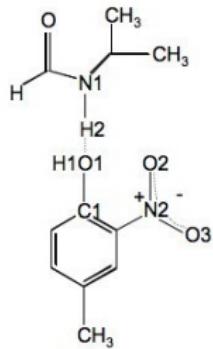


open state

Subtile effects on conformational changes

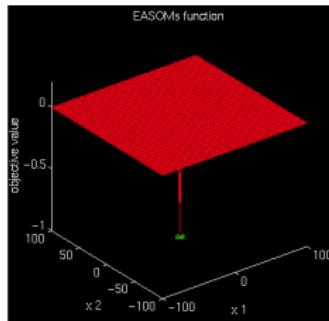


fonsc

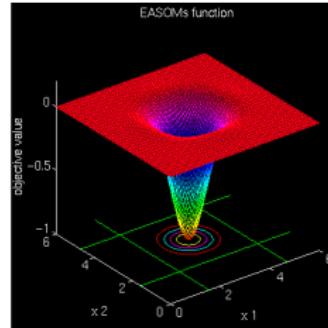


Miscione, Giupponi, Villà-Freixa, unsubmitted

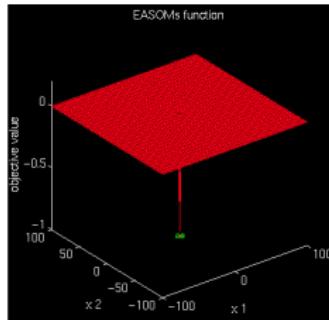
A simple abstract example



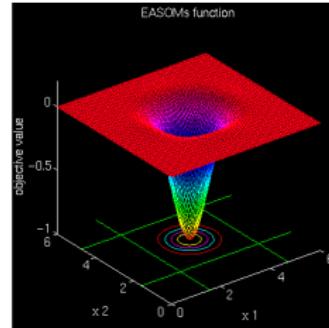
fons



A simple abstract example



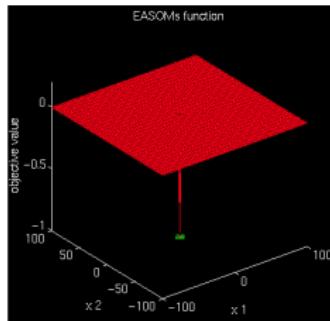
fons



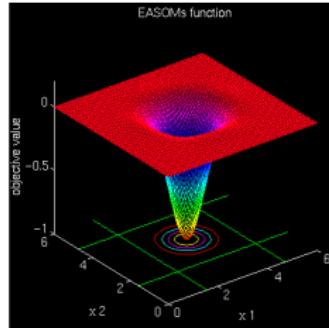
so...

...in what case do we reach the minimum before?

A simple abstract example



fons

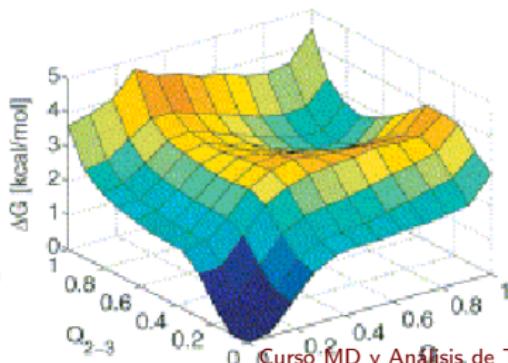
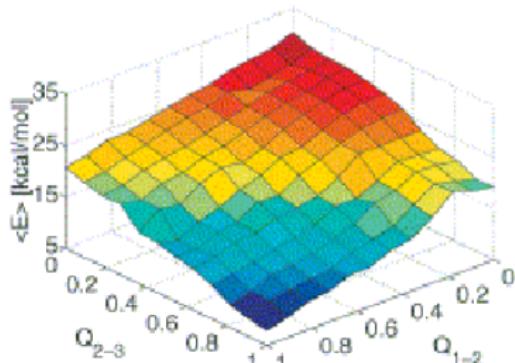
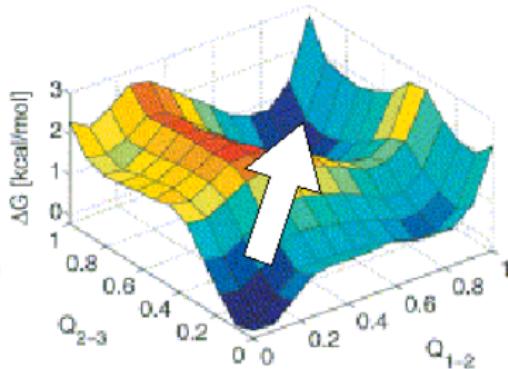
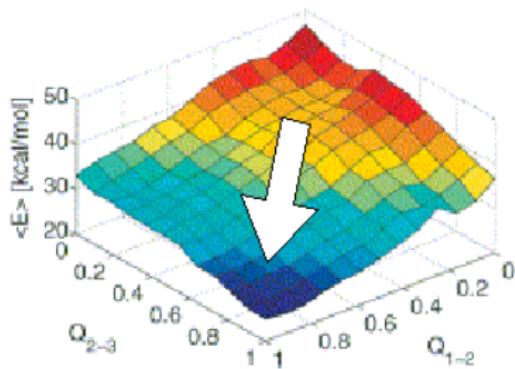


so...

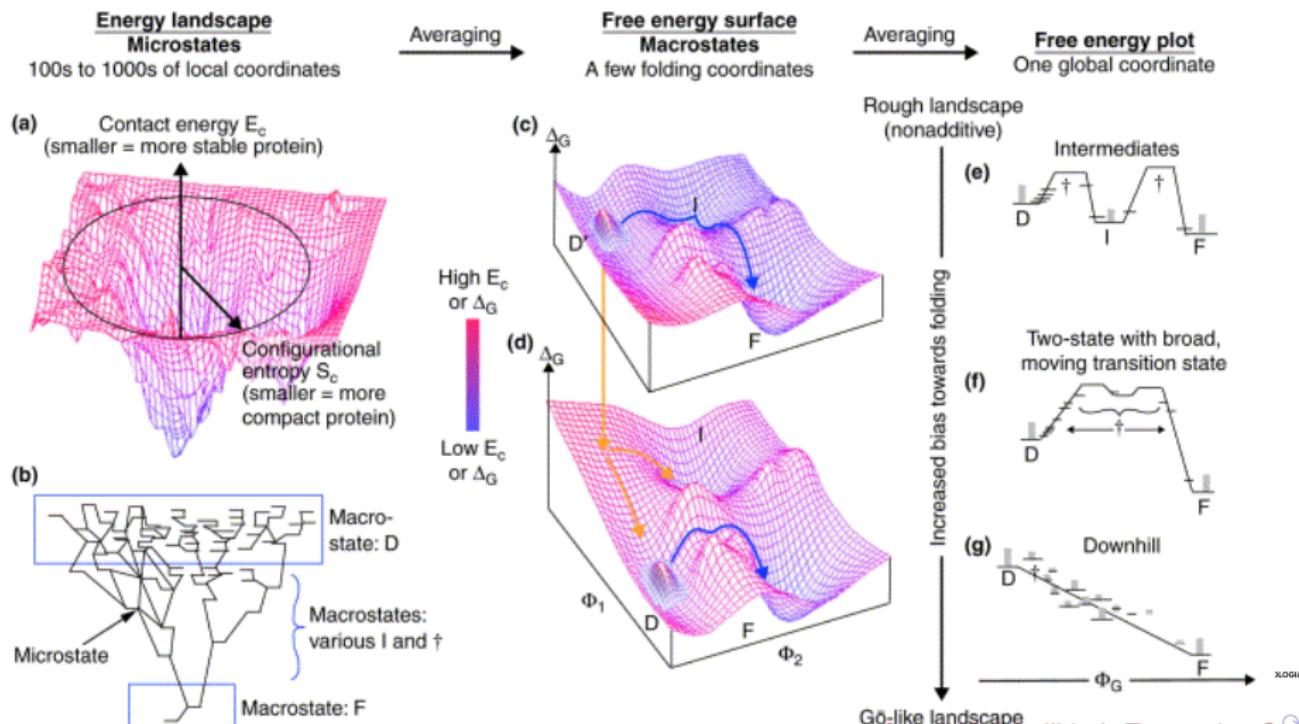
...in what case do we reach the minimum before?

$$\Delta G = \Delta H - T\Delta S$$

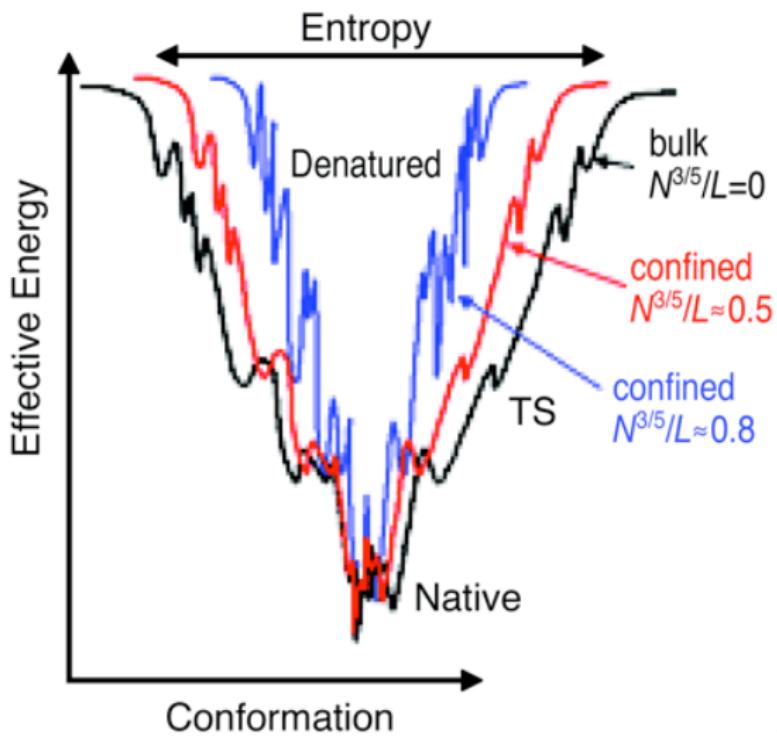
Energy and free energy can yield very different results → interpretations



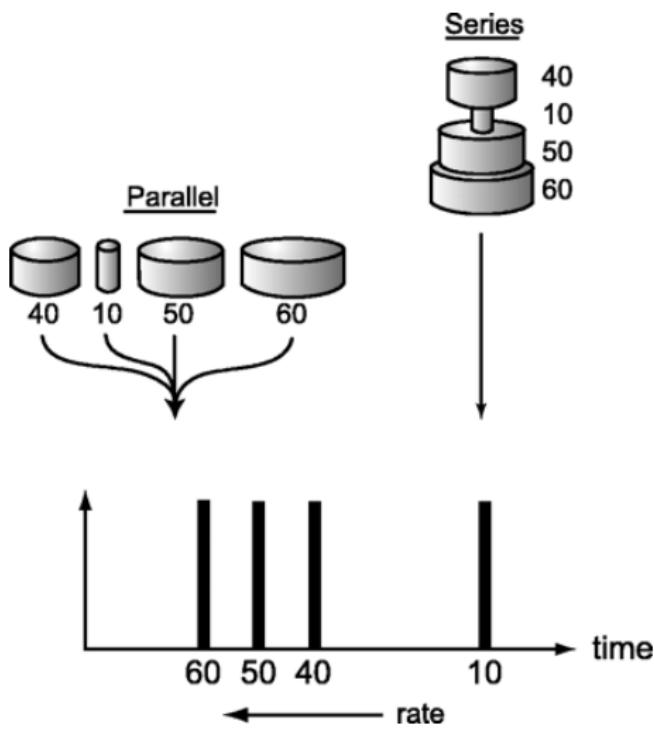
Energy and free energy can yield very different results → interpretations



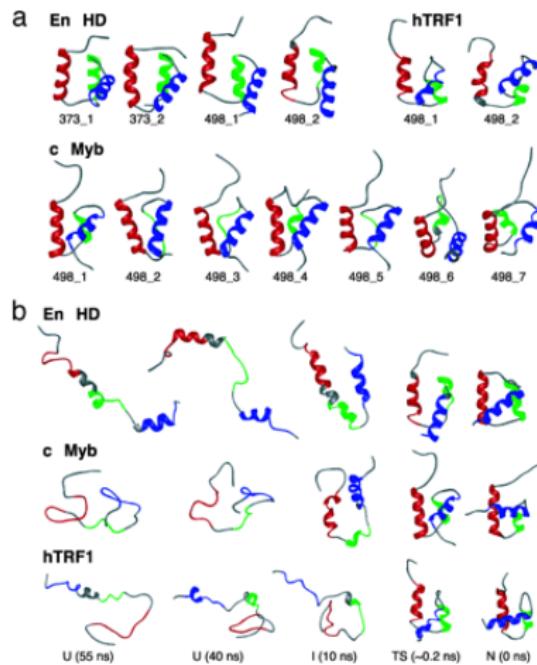
Energy and free energy can yield very different results → interpretations



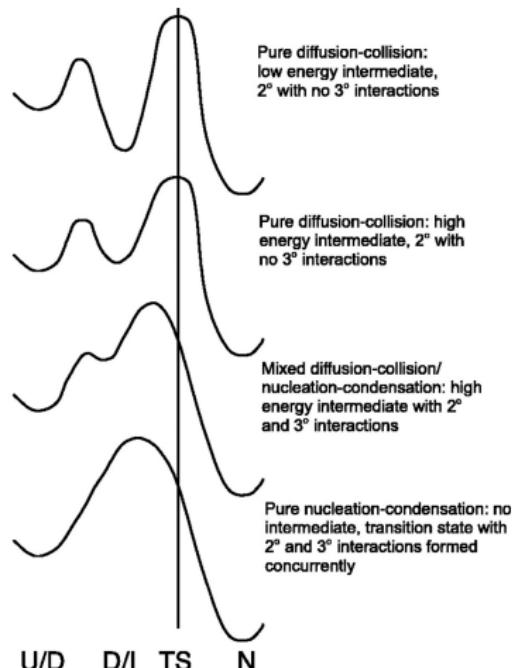
Different pathways to reach the same native structure



nucleation-condensation vs diffusion-collision



on



[6]

Investigating processes: free energy perturbations (FEP)

Absolute value of A (Helmholtz free energy, the “natural” quantity in the canonical, NVT , ensemble) is difficult to get, but its relative value is easier:

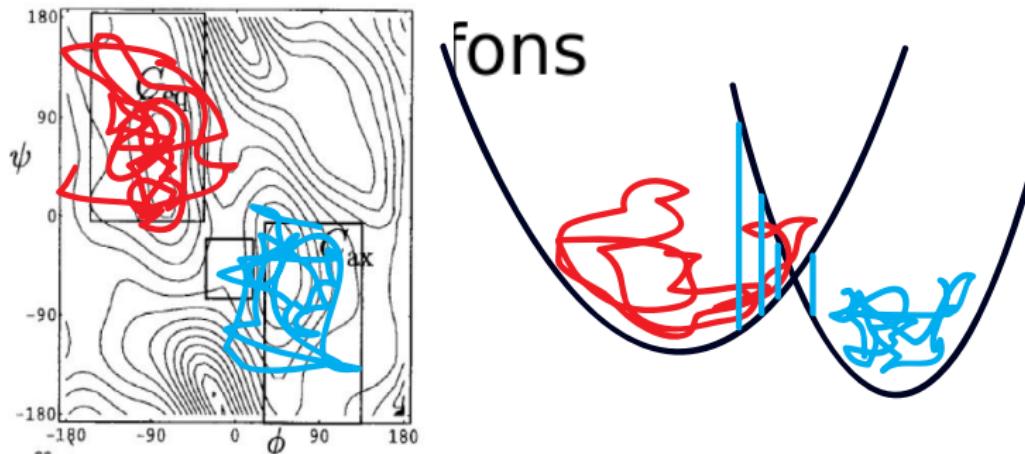
$$\Delta A = A_Y - A_X = -k_B T \ln \frac{Q_Y}{Q_X}$$

Zwanzig developed a better way to evaluate this quantity:

$$\Delta A = -k_B T \ln \langle \exp[(H_Y - H_X)/k_B T] \rangle_X$$

Most times X and Y do not overlap in phase space and thus the evaluation of the above average is difficult.

Free energy perturbations (FEP)



A is a state function, which implies we can use any path to go from one state to the other with identical result.

Free energy perturbations (FEP)

Thus, we can envisage an imaginary path that slowly leads from X to Y:

$$\begin{aligned}
 \Delta A &= A_Y - A_X \text{ fons} \\
 &= (A_Y - A_1) - (A_1 - A_X) \\
 &= -k_B T \ln \left[\frac{Q_Y}{Q_1} \frac{Q_1}{Q_X} \right] \\
 &= -k_B T \ln \langle \exp[(H_1 - H_X)/k_B T] \rangle_X \\
 &\quad - k_B T \ln \langle \exp[(H_Y - H_1)/k_B T] \rangle_1
 \end{aligned}$$

which can be extended to as many states as we need:

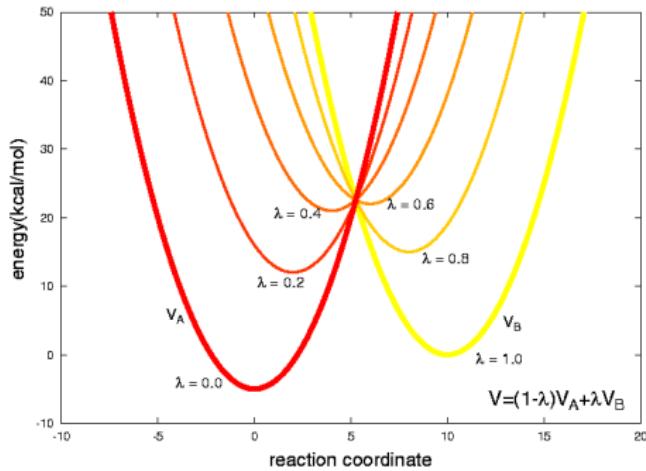
$$\Delta A = A_Y - A_X = -k_B T \ln \left[\frac{Q_Y}{Q_{Y-1}} \cdots \frac{Q_3}{Q_2} \frac{Q_2}{Q_1} \frac{Q_1}{Q_X} \right]$$

Free energy perturbations (FEP)

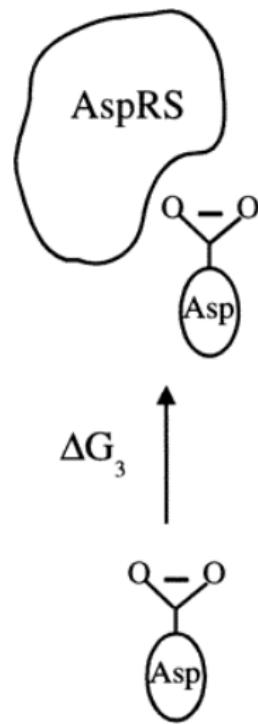
To efficiently implement a Free Energy Perturbation method, we need to move the systems from X to Y in small steps (even if they have no real physico-chemical meaning!) through a **mapping potential**:

$$E_m = \lambda_m E_Y + (1 - \lambda_m) E_X$$

where λ_m goes from 0 to 1.

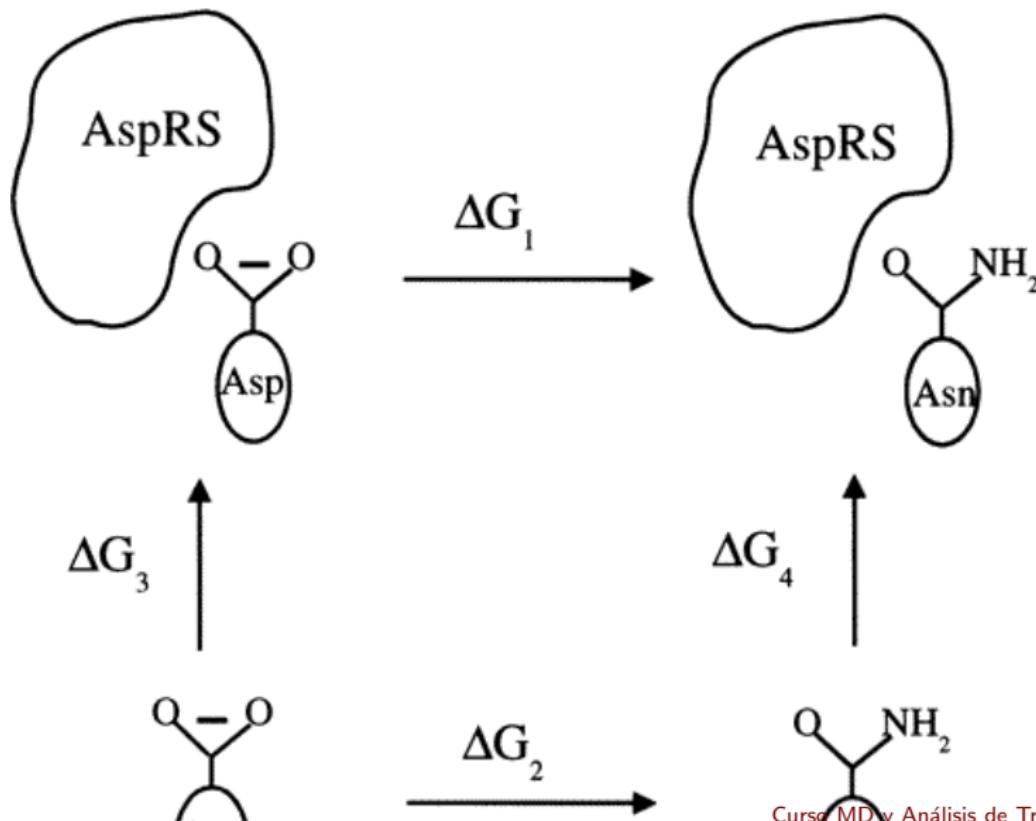


Alchemical transformations

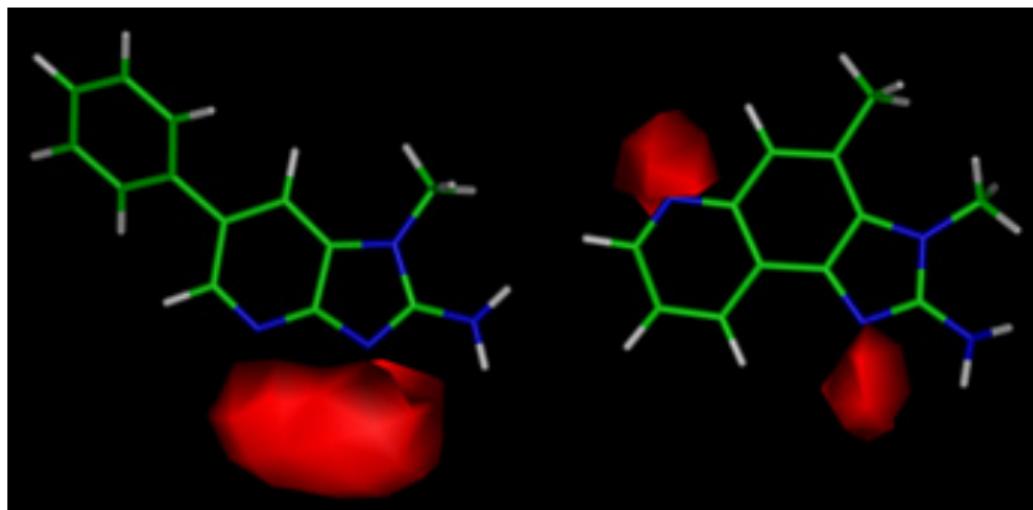


We can use FEP to do
alchemical
transformations

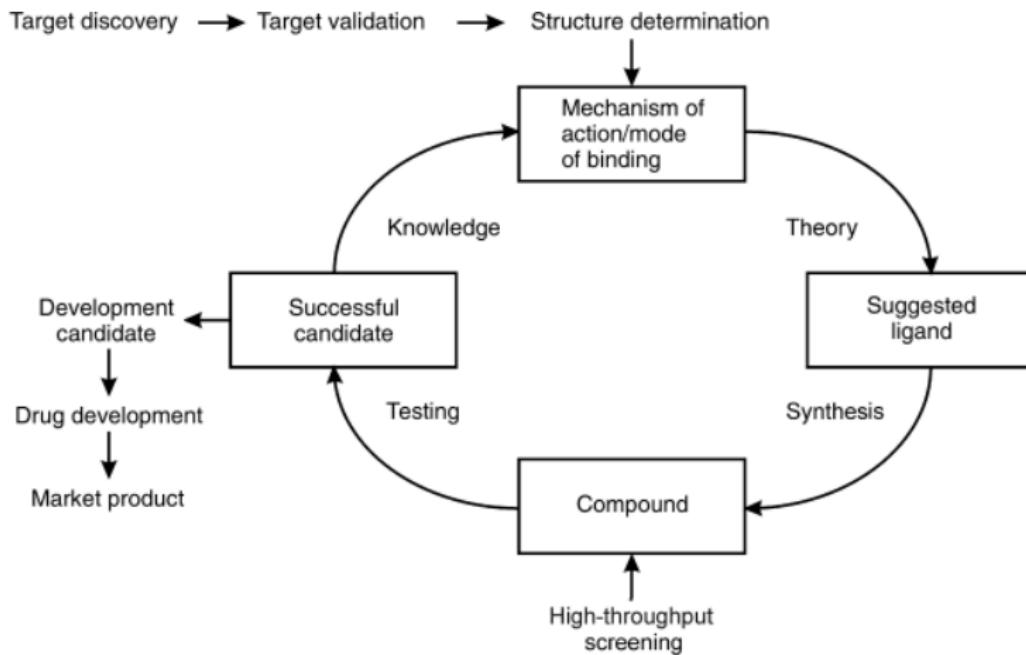
Alchemical transformations



Docking methods

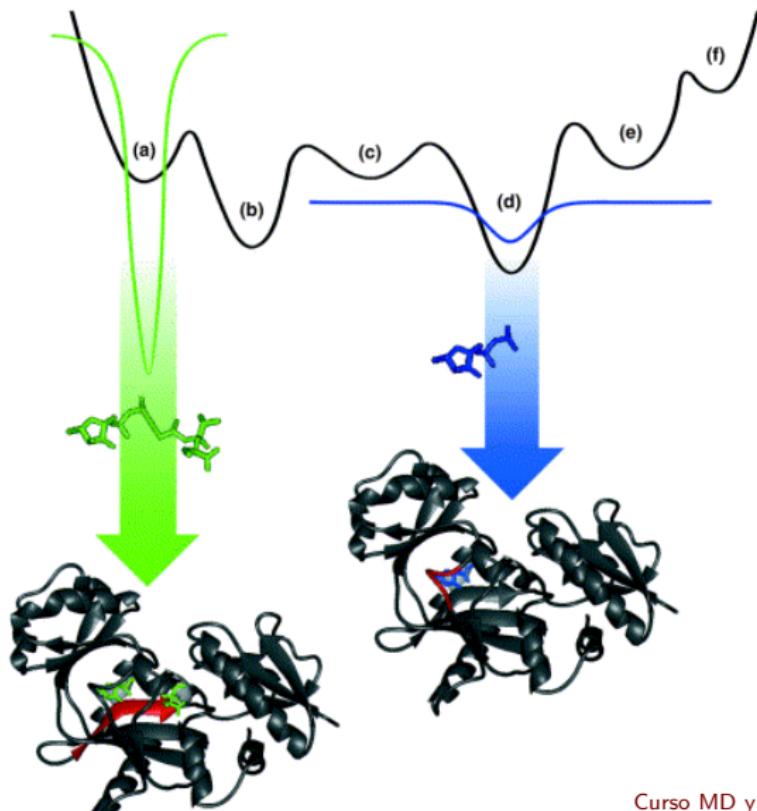


Drug discovery and drug design



[7]

Drug discovery and drug design



Energy (docking)

5100

J. Med. Chem. 1999, 42, 5100–5109

Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins

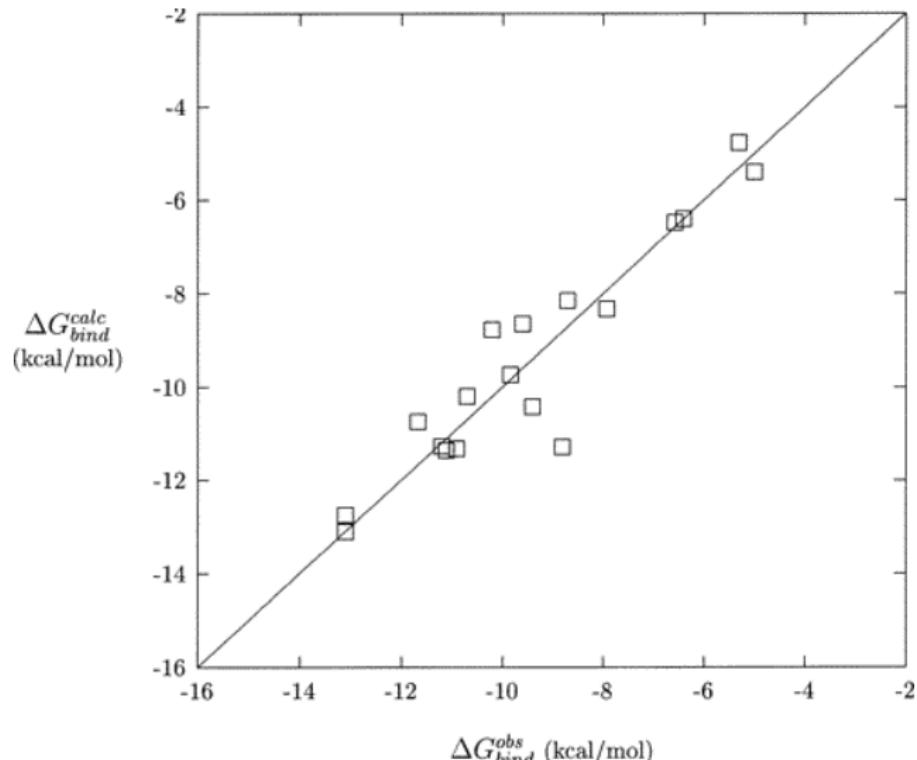
Paul S. Charifson,* Joseph J. Corkery, Mark A. Murcko, and W. Patrick Walters

Vertex Pharmaceuticals, 130 Waverly Street, Cambridge, Massachusetts 02139-4242

Received July 8, 1999

We present the results of an extensive computational study in which we show that combining scoring functions in an intersection-based consensus approach results in an enhancement in the ability to discriminate between active and inactive enzyme inhibitors. This is illustrated in the context of docking collections of three-dimensional structures into three different enzymes of pharmaceutical interest: p98 MAP kinase, inosine monophosphate dehydrogenase, and HIV protease. An analysis of two different docking methods and thirteen scoring functions provides insights into which functions perform well, both singly and in combination. Our data shows that consensus scoring further provides a dramatic reduction in the number of false positives identified by individual scoring functions, thus leading to a significant enhancement in hit-

Energy (docking) vs free energy (binding)



The LIE method [3]

Residue contributions to binding in Ras

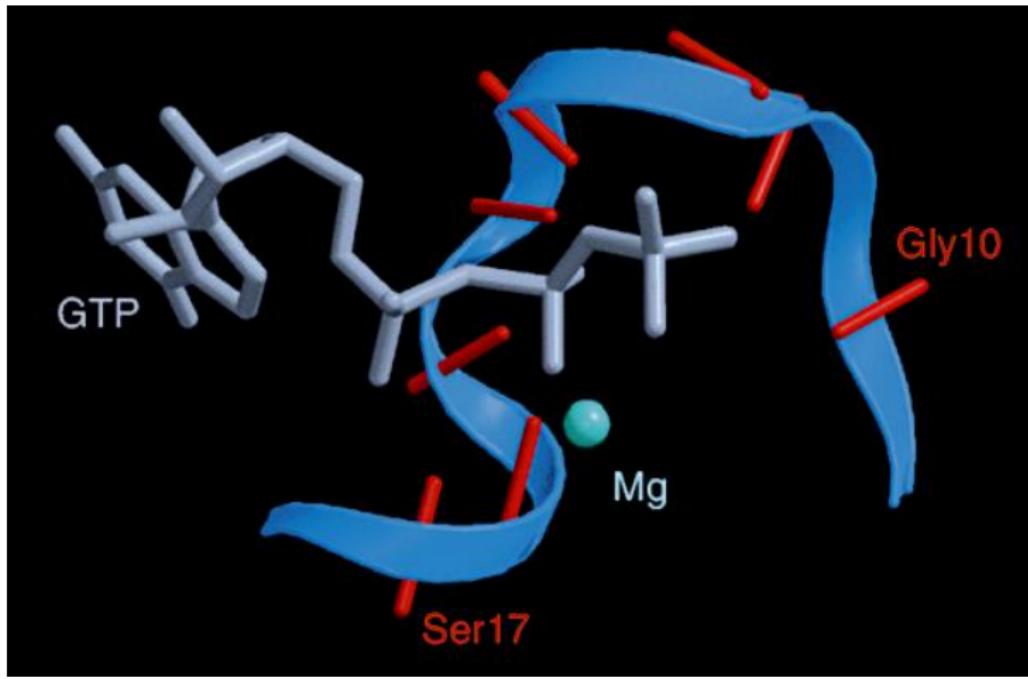
Table 4

Calculated and observed relative binding free energies of GTP and GDP in different p21^{ras} mutants.

Structure	PDLD/S estimate of relative binding energy (kcal mol ⁻¹) ^{*†} ΔG_3	Experimentally determined relative binding energies (kcal mol ⁻¹) [#]
(Gly15→Ala)p21-GDP	0.7	2.1 [24]
(Gly15→Ala)p21-GTP	3.2	>2.8 [24]
(Ser17→Asn)p21-GDP	0.9	0.3 [24]
(Ser17→Asn)p21-GTP	2.5	>2.8 [24]
(Ser17→Ala)p21-GDP	3.5	3.6 [20]
(Ser17→Ala)p21-GTP	6.0	5.6 [20]
(Gly60→Ala)p21-GTP	5.1	-0.2 [30]

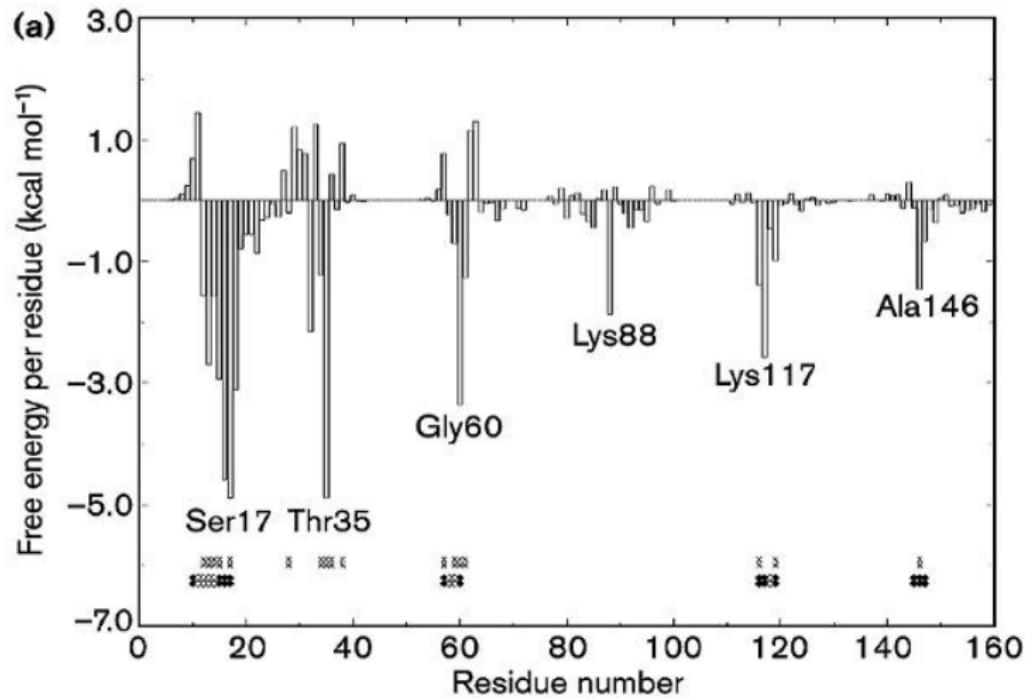
* ΔG_3 and the experimentally determined binding energies are given relative to the binding energies of the wild-type proteins in their GTP- and GDP-bound forms (which are both set to zero). †All calculations are done with $\epsilon_p = 4$. The calculated energies for ΔG_3 correspond to the equilibrium (k_3 and k_{-3}) of Figure 1. #The relevant reference is given in brackets.

Residue contributions to binding in Ras



[12]

Residue contributions to binding in Ras



[12]

Episodio 5: Análisis de trayectorias: Summary

1 Episodio 1: Getting Started (entorno y datos)

- Panorama y escalas
- Variables y estados
- Entropía y energía libre
- Ensembles y partición
- Hamiltoniano y dinámica
- Fuerzas y gradientes
- Partición y probabilidades
- Distribución de Boltzmann
- Espacio de fases y ergodicidad
- Superficie de energía potencial
- Campos de fuerza
- Entorno y datos

2 Episodio 2: Running Simulations

- Discretización temporal
- Integradores clásicos
- Escenarios leapfrog

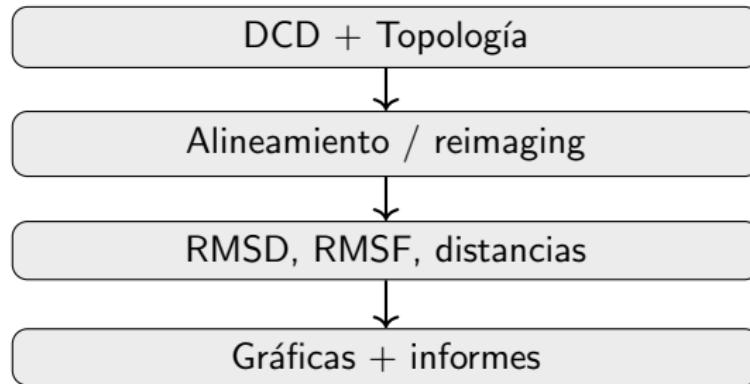
Objetivos de la sesión

- Analizar trayectorias simples y complejas.
- Medir RMSD, RMSF y métricas básicas.
- Generar figuras reproducibles.

Pipeline de análisis

- Lectura de DCD con MDTraj.
- Reimaging y superposición.
- Gráficas y exportación SVG/PNG.

De trayectorias a métricas



Episodio 6: MSM con PyEMMA: Summary

1 Episodio 1: Getting Started (entorno y datos)

- Panorama y escalas
- Variables y estados
- Entropía y energía libre
- Ensembles y partición
- Hamiltoniano y dinámica
- Fuerzas y gradientes
- Partición y probabilidades
- Distribución de Boltzmann
- Espacio de fases y ergodicidad
- Superficie de energía potencial
- Campos de fuerza
- Entorno y datos

2 Episodio 2: Running Simulations

- Discretización temporal
- Integradores clásicos
- Esquema leapfrog

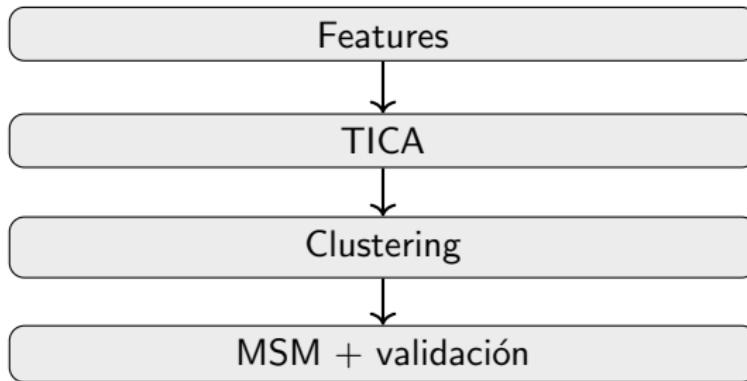
Objetivos de la sesión

- Construir MSM para alanina y complejo.
- Ejecutar featurización, TICA y clustering.
- Validar escalas de tiempo y estados metaestables.

Flujo de PyEMMA

- Features → TICA → clustering.
- Estimación MSM y validaciones ITS.
- PCCA y análisis de transiciones.

MSM (esquema visual)



Episodio 7: MSM con deeptime: Summary

1 Episodio 1: Getting Started (entorno y datos)

- Panorama y escalas
- Variables y estados
- Entropía y energía libre
- Ensembles y partición
- Hamiltoniano y dinámica
- Fuerzas y gradientes
- Partición y probabilidades
- Distribución de Boltzmann
- Espacio de fases y ergodicidad
- Superficie de energía potencial
- Campos de fuerza
- Entorno y datos

2 Episodio 2: Running Simulations

- Discretización temporal
- Integradores clásicos
- Esquema leapfrog

Objetivos de la sesión

- Repetir el flujo con deeptime.
- Comparar resultados con PyEMMA.
- Entender ventajas de un pipeline modular.

Comparación y cierre

- deptime ofrece una API moderna y ligera.
- Comparar escalas de tiempo y coherencia de estados.
- Preparar el camino para proyectos individuales.

Referencias I

- [1] Abusaleh AA. *Minimum Image Convention*. CC BY-SA 4.0. URL: https://commons.wikimedia.org/wiki/File:Minimum_Image_Convention.png (visited on 01/12/2026).
- [2] AdiPhysics. *Lennard Jones potential graph*. CC BY-SA 4.0. URL: https://commons.wikimedia.org/wiki/File:Lennard_Jones_potential_graph.svg (visited on 01/12/2026).
- [3] Aqvist. *Referencia pendiente*. Completar referencia completa. 2002.
- [4] Carlson. *Referencia pendiente*. Completar referencia completa. 2002.
- [5] Dill. *Referencia pendiente*. Completar referencia completa. 2000.
- [6] Gianni. *Referencia pendiente*. Completar referencia completa. 2003.
- [7] Gohlke. *Referencia pendiente*. Completar referencia completa. 2002.
- [8] Guix. *Referencia pendiente*. Completar referencia completa. 2009

Referencias II

- [9] Krishnavedala. *Pendulum phase portrait illustration*. CC BY-SA 4.0. URL: https://commons.wikimedia.org/wiki/File:Pendulum_phase_portrait_illustration.svg (visited on 01/12/2026).
- [10] Gilles MAIRET. *Distribution de Maxwell-Boltzmann. Distribution des vitesses de l'helium a 5500 K*. CC BY-SA 3.0. URL: https://commons.wikimedia.org/wiki/File:Distribution_de_Maxwell-Boltzmann._Distribution_des_vitesses_de_l%27helium_%C3%A0_5500_K.png (visited on 01/12/2026).
- [11] MikeRun. *Leapfrog-method-argument*. CC BY-SA 4.0. URL: <https://commons.wikimedia.org/wiki/File:Leapfrog-method-argument.svg> (visited on 01/12/2026).
- [12] Muegge. *Referencia pendiente*. Completar referencia completa.

Referencias III

- [13] Christopher Rowley. *Periodic Boundary Conditions in 2D*. CC BY-SA 4.0. URL: https://commons.wikimedia.org/wiki/File:Periodic_Boundary_Conditions_in_2D.png (visited on 01/12/2026).
- [14] Shoichet. *Referencia pendiente*. Completar referencia completa. 2007.
- [15] Unknown author. *Potential Energy Surface for a generic SN1 reaction*. CC BY-SA 4.0. Source: LibreTexts (Organic Chemistry, Morsch et al.) URL: https://commons.wikimedia.org/wiki/File:Potential_Energy_Surface_for_a_generic_SN1_reaction.svg (visited on 01/12/2026).