Practical Course on Molecular Dynamics and Trajectory Analysis

Episode 4: Advanced simulations and free energy

Jordi Villà i Freixa

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

jordi.villa@uvic.cat

MD Course and Trajectory Analysis
Concepcion, January 2026

Contents

1. Episode 4: Advanced simulations and free energy
   - Objectives and roadmap
   - Thermodynamic control
   - Pressure and barostats
   - Constraints and integration
   - Complex systems
   - Enhanced sampling
   - Metadynamics
   - FEP and TI
   - Thermodynamic cycles
   - Non-equilibrium
   - Official advanced sampling
   - Convergence and statistics
   - Summary

Notation used throughout the notes

- Phase-space point: $z = (q, p) \in \mathbb{R}^{3N} \times \mathbb{R}^{3N}$, with positions q and momenta p.
- Potential energy (force field): $U(q)$; forces $\mathbf{F}(q) = -\nabla_q U(q)$.
- Hamiltonian: $H(q, p) = K(p) + U(q)$ with $K(p) = \sum_i \frac{\|p_i\|^2}{2m_i}$.
- Temperature T, Boltzmann constant k_B, inverse temperature $\beta = (k_\mathrm{B} T)^{-1}$.
- Time step Δt; lag time for kinetic models τ.

Free energy: definition and estimators

For a reaction coordinate $\xi(q)$, the potential of mean force (PMF) is

$$F(s) = -k_{\mathrm{B}}\,T\,\ln p_\xi(s) + C, \qquad p_\xi(s) = \int \delta(\xi(q) - s)\,\pi(q)\,dq,$$

where $\pi(q) \propto e^{-\beta U(q)}$ is the configurational Boltzmann density. Common estimators: umbrella sampling + WHAM/MBAR; thermodynamic integration; free-energy perturbation (Zwanzig).

Suggested figure: free-energy perturbation (Zwanzig) / thermodynamic integration

Figure: Conceptual schematic for free-energy estimation (FEP/TI path in λ). Source: Wikipedia articles on free-energy perturbation / thermodynamic integration (use an image from those pages or linked Commons).

Episode objectives

- Apply advanced sampling and free energy techniques.
- Design robust protocols for complex systems.
- Evaluate convergence and reproducibility.

Advanced techniques map

- Highly stable thermostats and barostats.
- Multiscale integration and constraints.
- Enhanced sampling: umbrella, metadynamics, REMD.
- Free energies: FEP, TI, BAR/MBAR.

Control variables

$$\beta = \frac{1}{k_{\mathrm{B}} T}, \qquad \lambda \in [0, 1].$$

- The variable λ bridges the initial and final states.
- β defines the statistical weight in the ensemble.

Data requirements

- Trajectories long enough for each window.
- Independent sampling across windows or replicas.
- Energy reports and collective variables.

General workflow

1. Strict preparation and equilibration.
2. Define CVs and windows.
3. Production and convergence checks.
4. Post-process with robust estimators.

Advanced ensembles

- NVT for temperature control.
- NPT for realistic pressure and density.
- NPγT for surface tension in membranes.

Langevin equation

$$m\ddot{\mathbf{r}} = -\nabla U(\mathbf{r}) - \gamma m\dot{\mathbf{r}} + \sqrt{2\gamma m k_{\mathrm{B}} T}\,\mathbf{R}(t).$$

- Controls temperature via friction and noise.

Nosé-Hoover chain

$$\dot{\eta} = \frac{1}{Q}\left(\sum_i \frac{p_i^2}{m_i} - N_f\, k_\mathrm{B}\, T\right).$$

- Allows canonical sampling with less noise.

Energy equipartition

$$\langle K \rangle = \frac{N_f}{2} k_{\mathrm{B}} T.$$

- Check temperatures per subgroup (water, solute).

Temperature diagnostics

- Speed histogram vs Maxwell-Boltzmann.
- Drift monitor for non-equilibrated systems.
- Watch for jumps when parameters change.

Types of barostat

- Isotropic: box scales equally in all directions.
- Anisotropic: allows independent scaling.
- Monte Carlo barostat: discrete volume changes.

Volume fluctuations

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

- Compare with experimental values if available.

Compressibility and stability

- Excessive values signal poor convergence.
- Adjust barostat frequency and damping.

Surface tension

$$\gamma = \frac{L_z}{2}\left(P_{zz} - \frac{P_{xx} + P_{yy}}{2}\right).$$

- Relevant for membranes and lipid bilayers.

Pressure control

- Log volume and density per window.
- Avoid overly aggressive couplings.
- Stabilize before long sampling runs.

Geometric constraints

- SHAKE/LINCS fix fast bonds.
- Allow larger integration steps.

Verlet integrator

$$\mathbf{r}_{t+\Delta t} = \mathbf{r}_t + \mathbf{v}_t \Delta t + \frac{1}{2}\mathbf{a}_t \Delta t^2.$$

- Conserves energy in short integrations.

Scale separation (RESPA)

$$U = U_{\text{fast}} + U_{\text{slow}}.$$

- Integrate slow forces with larger steps.

Mass repartitioning

- HMR: increases hydrogen mass for stability.
- Enables Δt of 4–5 fs in well-constrained systems.

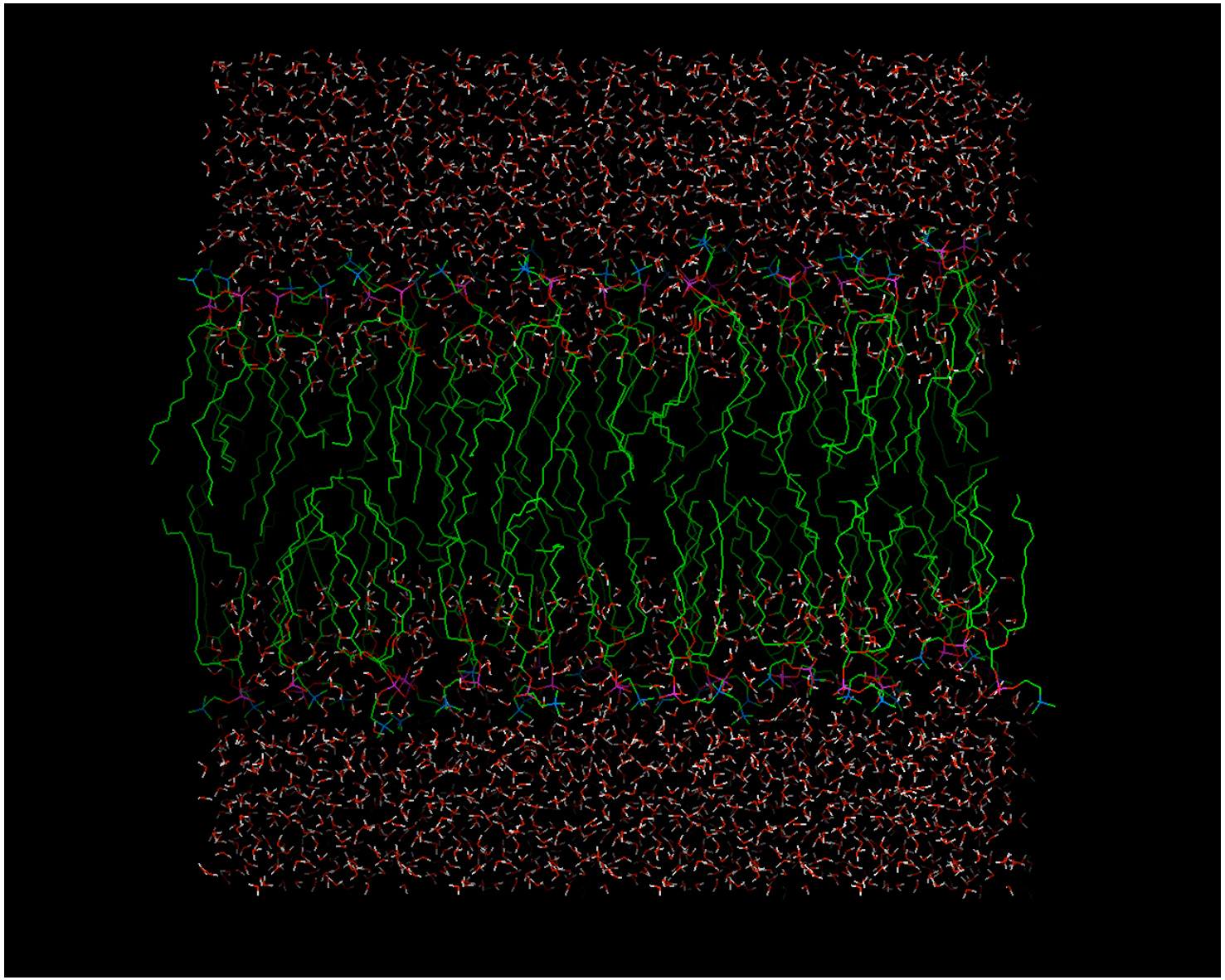
Numerical stability

- Monitor total energy drift.
- Adjust Δt according to stiffness and temperature.

Membranes and anisotropic systems

- Independent pressure scaling per axis.
- Control thickness and area per lipid.

Lipid bilayer



Explicit vs implicit solvation

$$\Delta G_{\text{solv}} = \Delta G_{\text{elec}} + \Delta G_{\text{np}}.$$

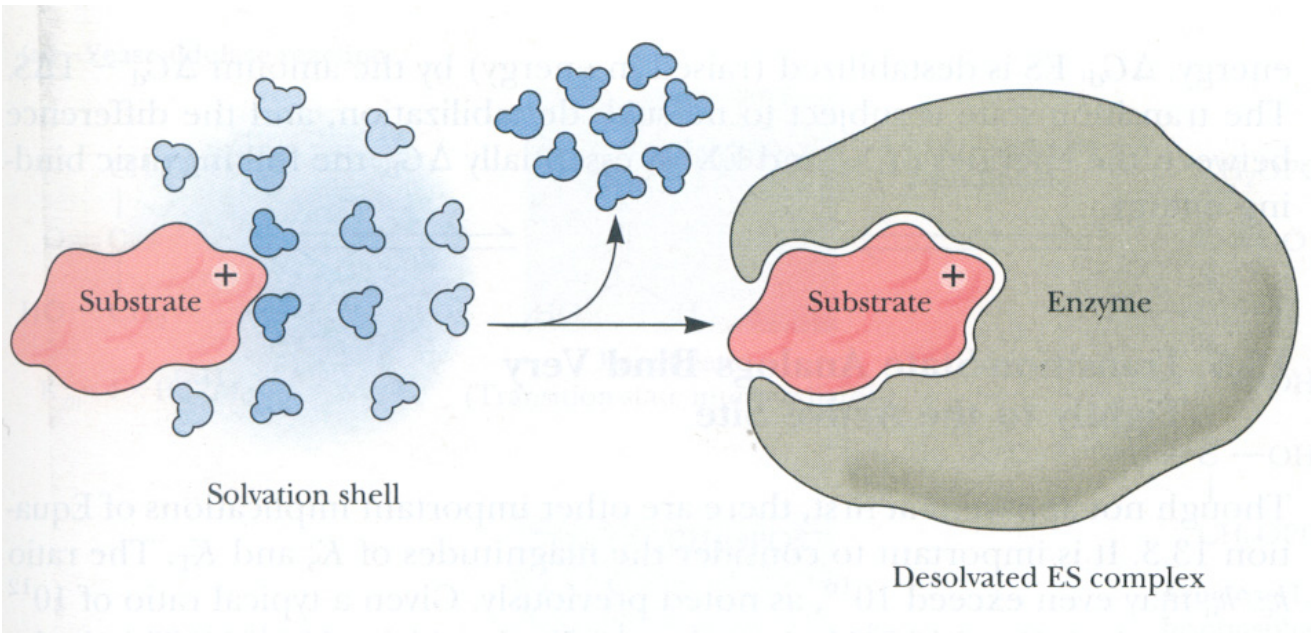
- Implicit models reduce computational cost.

Ionic screening

$$\kappa^{-1} = \sqrt{\frac{\varepsilon k_{\mathrm{B}} T}{2 N_A e^2 I}}.$$

- The ionic strength I controls the screening.

Desolvation energy



Barriers and rare events

- Transitions with high $\Delta G^{\ddagger}$ are rare.
- A bias potential accelerates sampling.

Window sampling (umbrella)

$$U_{\text{sesgo}}(\xi) = \frac{k}{2}(\xi - \xi_0)^2.$$

- Controls sampling in windows of the CV ξ.

PMF from distributions

$$F(\xi) = -k_{\mathrm{B}}\, T \ln P(\xi) + C.$$

- WHAM combines histograms from multiple windows.

Replica exchange

$$P_{\text{acc}} = \min\left(1, \exp\left[(\beta_i - \beta_j)(U_j - U_i)\right]\right).$$

- Temperature swaps to overcome barriers.

Adaptive sampling

- Select new seeds based on uncertainty.
- Combine short simulations with iterative analysis.

Collective variables

- Reduced coordinates that describe the transition.
- Must separate metastable states.

Bias via Gaussian hills

$$V(s,t) = \sum_i w \exp\left[-\frac{(s-s_i)^2}{2\sigma^2}\right].$$

Well-tempered metadynamics

$$w(t) = w_0 \exp\left(-\frac{V(s,t)}{k_\mathrm{B} T \, \Delta T}\right).$$

- Smooths the growth of the bias over time.

Free energy reconstruction

$$F(s) \approx -\frac{T + \Delta T}{\Delta T} V(s, t \to \infty).$$

Common pitfalls

- Poorly chosen CVs create artifacts.
- Deposition interval too short.
- Lack of sampling in key regions.

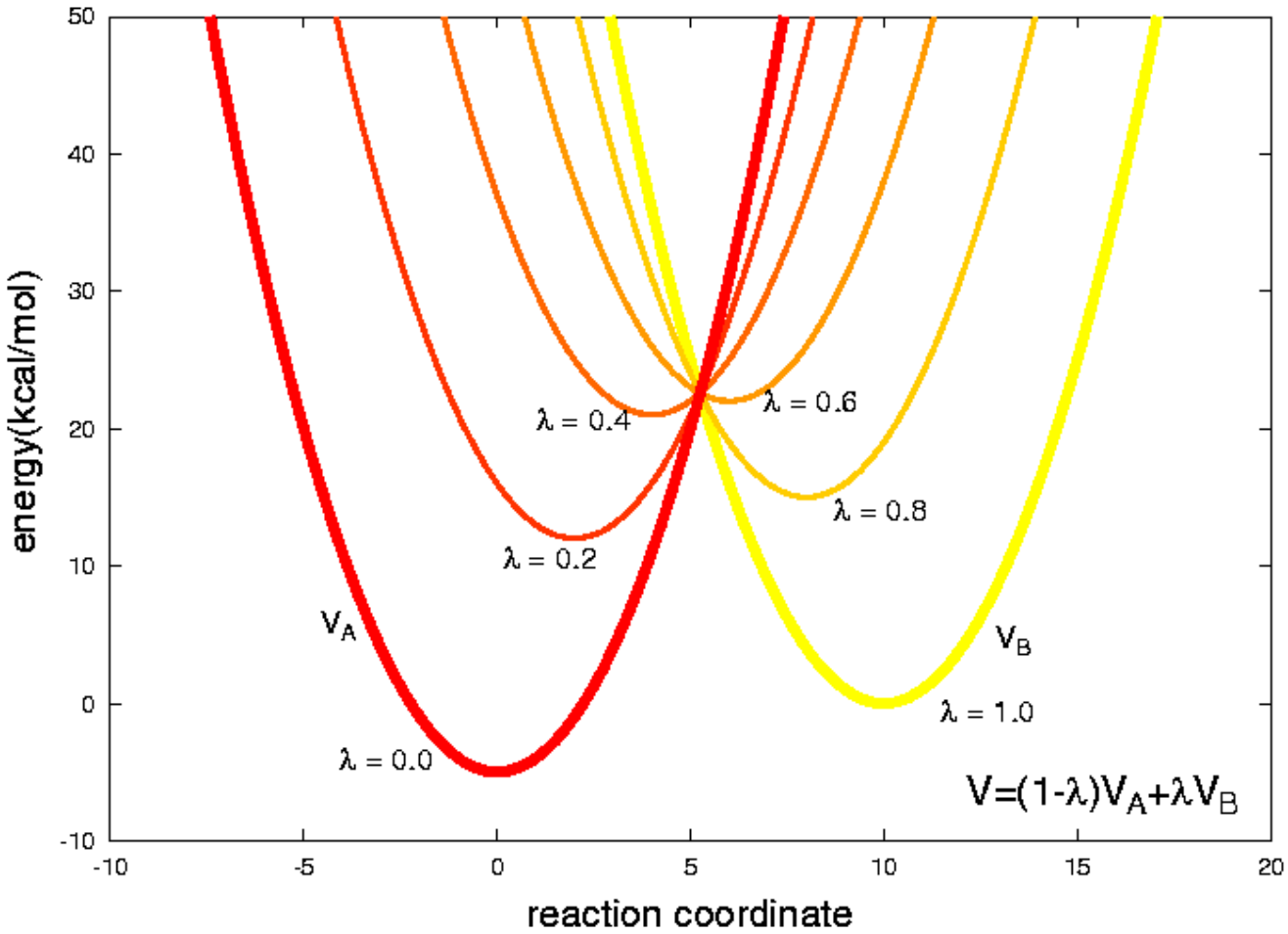
Free energy perturbation

$$\Delta G = -k_{\mathrm{B}} T \ln \left\langle \exp \left[-\beta \Delta U \right] \right\rangle_0 .$$

Thermodynamic integration

$$\Delta G = \int_0^1 \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda.$$

FEP scheme


energy(kcal/mol)
reaction coordinate
50
40
30
20
10
0
-10
-10
-5
0
5
10
15
20
λ = 0.0
λ = 0.2
λ = 0.4
λ = 0.6
λ = 0.8
λ = 1.0
V_A
V_B
V=(1-λ)V_A+λV_B

Soft-core potentials

- Prevent singularities when turning off interactions.
- Ensure overlap of neighboring distributions.

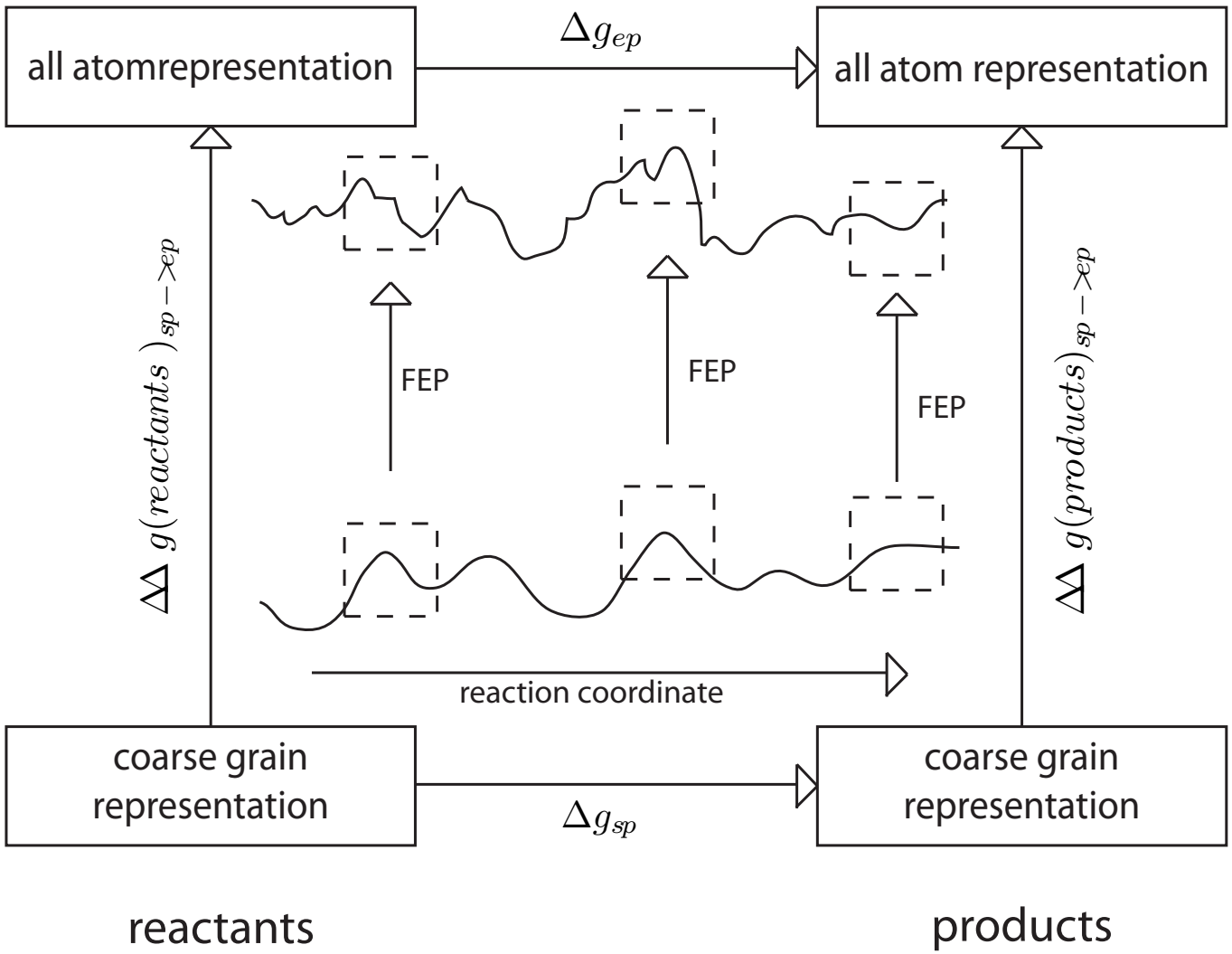
λ convergence

- Increase window density where changes are rapid.
- Monitor histograms of $\partial U/\partial\lambda$.

Double decoupling

- Turn off ligand interactions in solvent and complex.
- The difference yields the binding ΔG.

Thermodynamic cycle



Relation to affinity

$$\Delta G_{\text{bind}} = k_{\text{B}} T \ln K_d.$$

- Apply standard-state corrections when appropriate.

Cycle closure

$$\sum_{\text{ciclo}} \Delta G_i = 0.$$

- A deviation indicates systematic errors.

Error propagation

- Window errors add in quadrature.
- Report confidence intervals.

Jarzynski equality

$$\Delta G = -k_{\mathrm{B}} T \ln \langle e^{-\beta W} \rangle .$$

Crooks relation

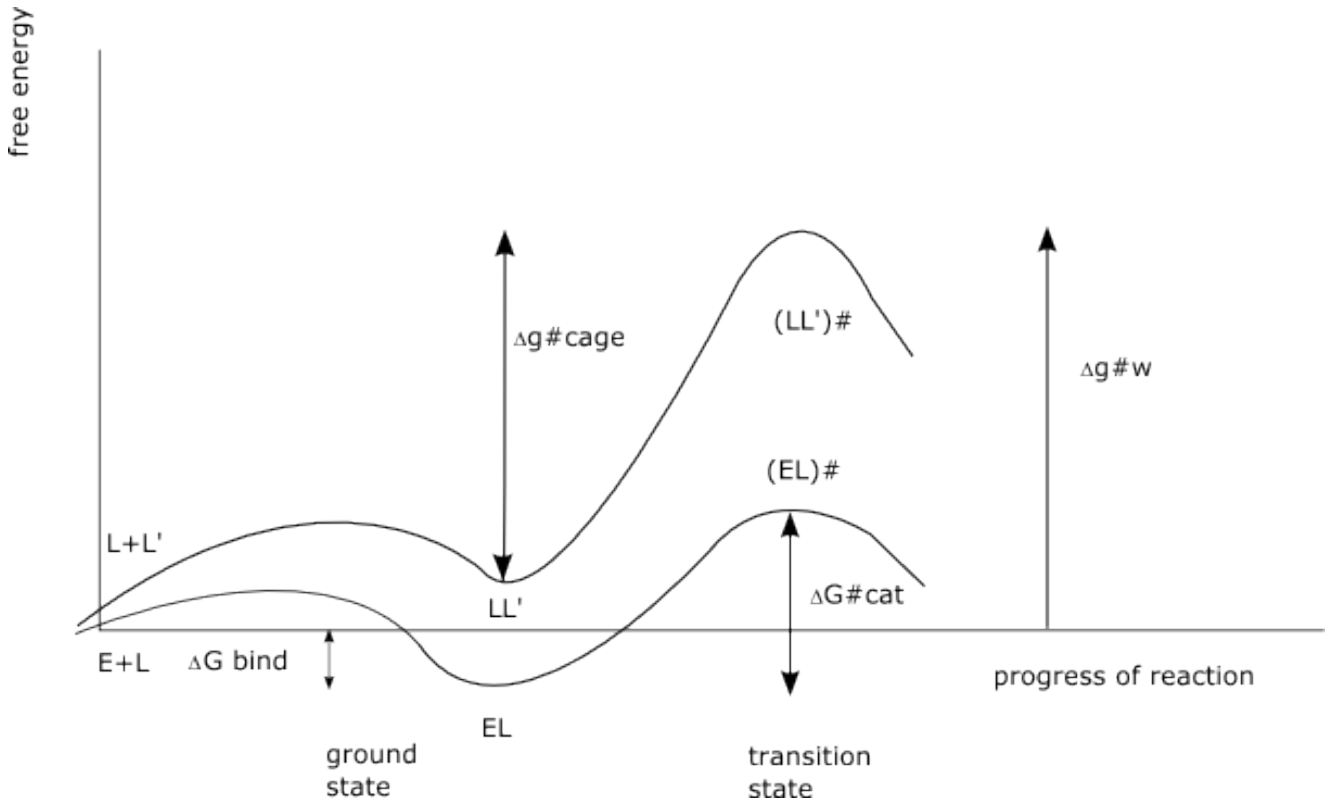
$$\frac{P_F(W)}{P_R(-W)} = e^{\beta(W-\Delta G)}.$$

Steered MD

$$W = \int_0^{\tau} \mathbf{F}_{\text{pull}}(t) \cdot d\mathbf{x}.$$

- Work depends on the pulling speed.

Reaction coordinate



Replica Exchange Solute Tempering (REST)

- REST uses ‘OpenMMTools‘ to scale energies on a subset of atoms (OpenMM Cookbook REST).
- Define modified ‘CustomBondForce‘, ‘CustomAngleForce‘, ‘CustomTorsionForce‘, and ‘NonbondedForce‘, and swap replicas with ‘ReplicaExchangeSampler‘.
- Useful for the protein–ligand complex where only the ligand is heated while the environment stays stable.

Guided umbrella sampling

- Apply a ‘HarmonicBias‘ on a collective variable $x(r)$ and collect histograms per window (OpenMM Cookbook Umbrella Sampling).
- From $\Delta F(x) = -k_{\mathrm{B}}\,T \log P(x)$, reconstruct the free energy profile and identify barriers.
- Combine with `argon-chemical-potential.py` to validate over simple LJ potentials before applying it to real proteins.

Pulling speed

- Slow pulls approximate quasi-equilibrium.
- Fast pulls require many repeats.

Block averages

- Split the time series into equal blocks.
- Estimate mean and error from independent blocks.

Statistical inefficiency

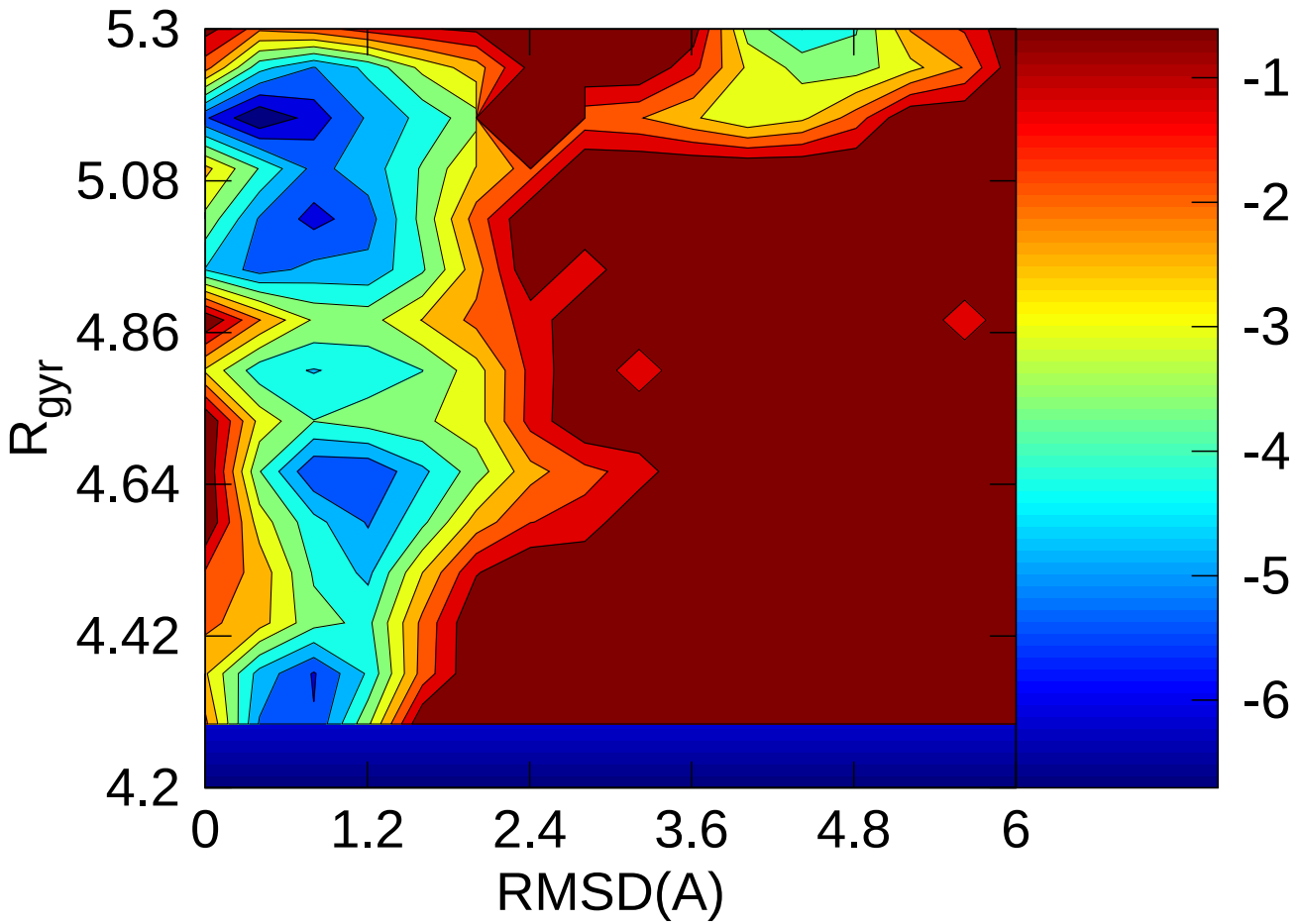
$$g = 1 + 2 \sum_{t=1}^{\infty} \rho(t).$$

- Reduces the effective number of samples.

BAR/MBAR

- BAR minimizes variance between two states.
- MBAR generalizes to multiple windows.

Free energy profile



Convergence checklist

- Overlapping histograms between windows.
- Stable errors over time.
- Repeats yielding consistent results.

Episode summary

- Reviewed advanced protocols and free energies.
- Emphasized convergence and statistical control.

Best practices

- Document versions, seeds, and parameters.
- Automate analyses and checks.

Common mistakes

- Insufficient windows during rapid changes.
- Lack of overlap between λ.

Reproducibility

- Save scripts and seeding.
- Report uncertainty and initial conditions.

Next step

- Dive into detailed trajectory analysis.

References I