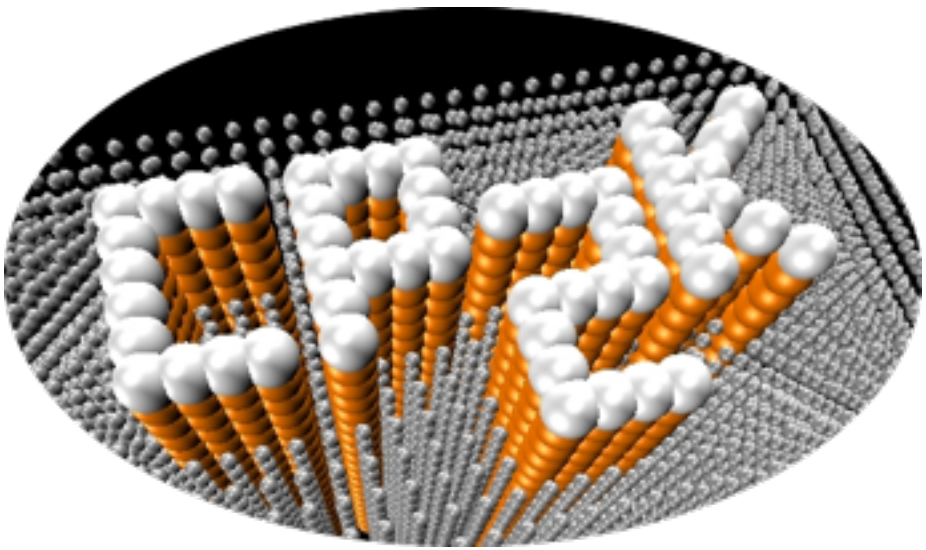
3rd CP2K tutorial: Enabling the Power of Imagination in MD Simulations
June 17-21 2013, Zürich

Sampling Free Energy Surfaces by MD

Marcella Iannuzzi

Department of Chemistry, University of Zurich



http://www.cp2k.org

OUTLINE

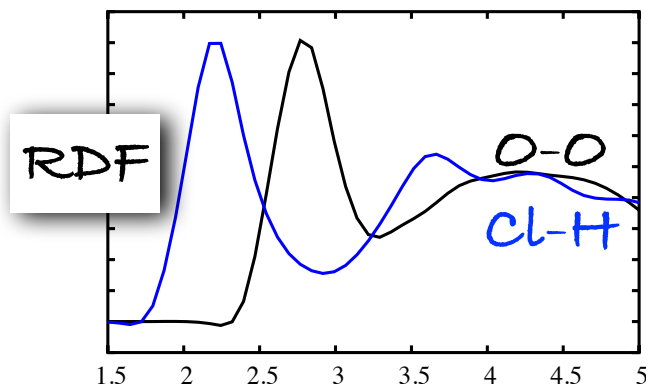
- Free energy in statistical mechanics

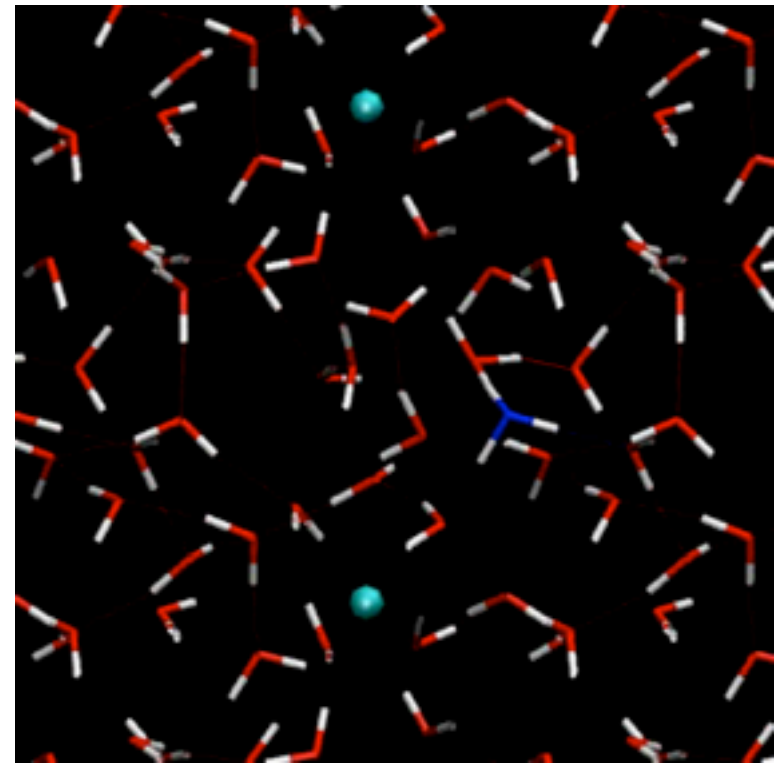
- Free energy difference by improving the sampling along the evolution order parameters

- Enhanced exploration of the configuration space and disclosure of mechanisms of transformation

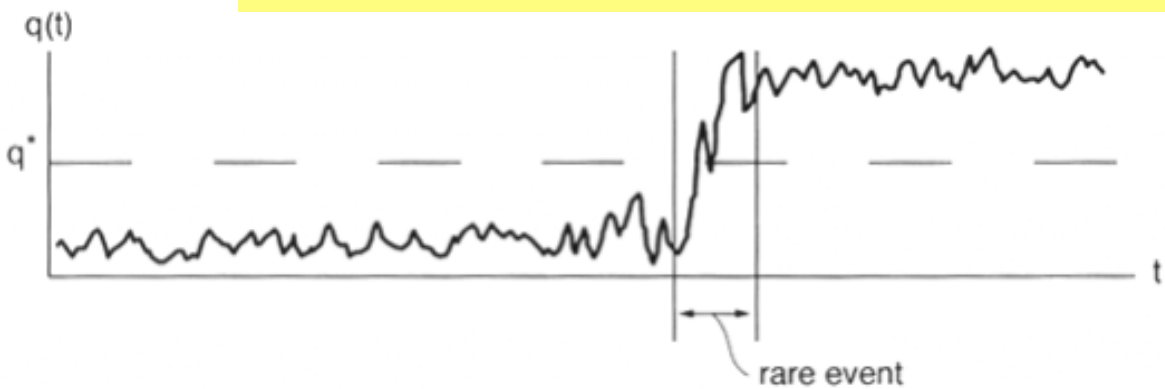
Complex Processes by MD

- Choose a suitable model of the system
- Determine the thermodynamic conditions ⇒ Ensemble averages
- Equilibrium sampling of physical quantities





Predictive power frustrated by sampling fast degrees of freedom with time-steps from < 0.1 fs (CPMD) up to 1 fs (MM)

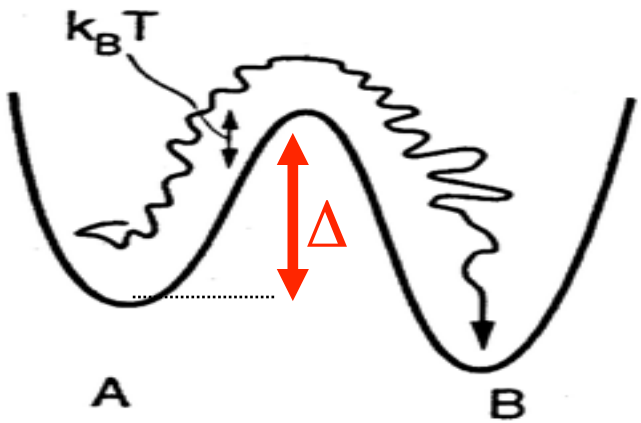


~ few ns

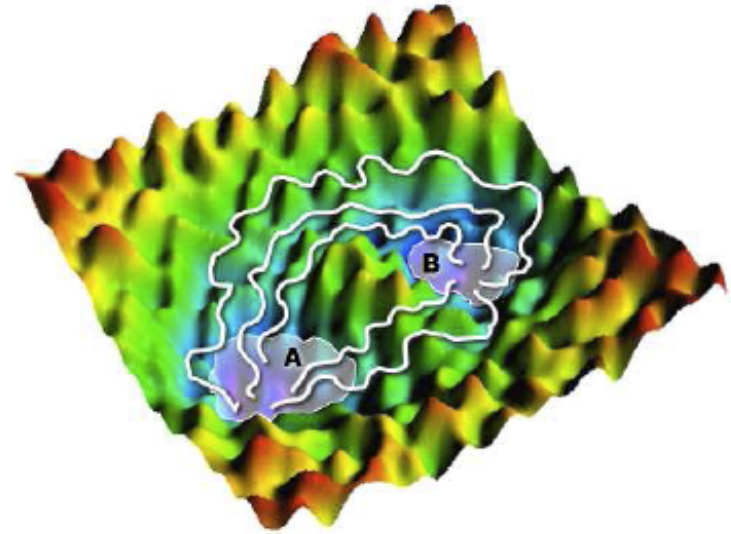
Rare Events

Phase Transitions, Conformational Rearrangements,
Chemical Reactions, Nucleation, Diffusion, Growth, etc.

Activation Energies



Minimum energy pathways



Exploration of configurational space

- Complex and high dimensional configurational space
- Multitude of unknown intermediates and products
- Unforeseen processes, many irrelevant transition states
- Intrinsically multidimensional order parameter
- Entropic bottlenecks
- Diffusive trajectories

Hamiltonian MD

A system of N particles in a volume V is completely determined through its Hamiltonian

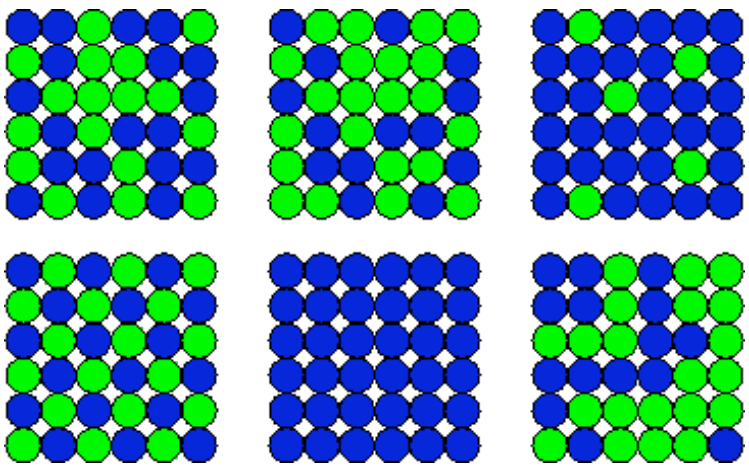
$$\mathcal{H}(\{\mathbf{R}_I\},\{\mathbf{P}_I\}) = \sum_I \frac{\mathbf{P}_I^2}{2M_I} + U(\mathbf{R}_I\})$$

NVE-P total energy and linear momentum are constant of motion

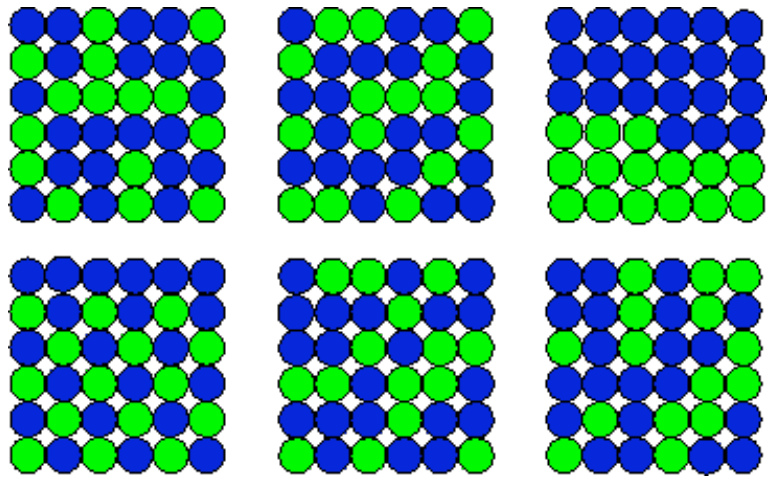
Choice of ensemble: portion of phase space, microstates
Difference in averages and fluctuations

Counters, blue on one side and green on the other, 6×6 checkerboard

Each pattern is a microstate



Subset belonging to macrostate "15 green"



Canonical Partition Function

The Laplace transform of the density of state

$$Q(N,V,T) = \int \exp(-\beta E)\Omega(N,V,E)dE$$

Probability of the macrostate at a given T

$$Q(N,V,T) = \frac{1}{N!h^{3N}} \int \exp\left[-\beta\mathcal{H}(\mathbf{r}^N,\mathbf{p}^N)\right] d\mathbf{r}^N d\mathbf{p}^N = \frac{1}{\Lambda(\beta)^{3N}N!} Z(N,V,T)$$

one dimensional integral over E replaced by configurational integral
analytic kinetic part is integrated out

$$Z(N,V,T) = \int e^{-\beta\mathcal{U}(\mathbf{r})} d\mathbf{r}$$

configurational partition function

Free Energy

Helmholtz free energy or thermodynamic potential

$$A = -\frac{1}{\beta} \ln Q(N, V, T)$$

Thermodynamics **Statistical Mechanics**

$$\Delta A = -\frac{1}{\beta} \ln \left(\frac{Z_1}{Z_0} \right)$$

entropic and enthalpic contributions

$$Q_0 \propto \int_{\Gamma_0} e^{-\beta \mathcal{H}(\mathbf{r},\mathbf{p})} d\mathbf{r} d\mathbf{p}$$

Macroscopic state 0 corresponds to a portion of the phase space : Γ_0

$$Q_0 \propto \int_{\Gamma} e^{-\beta \mathcal{H}_0(\mathbf{r},\mathbf{p})} d\mathbf{r} d\mathbf{p}$$

Macroscopic state 0 corresponds to H_0

$$Q_0 \propto \int_{\Gamma} e^{-\beta_0 \mathcal{H}(\mathbf{r},\mathbf{p})} d\mathbf{r} d\mathbf{p}$$

Macroscopic state 0 corresponds to a value of a macroscopic parameter, e.g T

Perturbation formalism

Reference (0) and target system (1) $\mathcal{H}_1(\mathbf{r},\mathbf{p}) = \mathcal{H}_0(\mathbf{r},\mathbf{p}) + \Delta\mathcal{H}(\mathbf{r},\mathbf{p})$

Probability of finding (0) in configuration (**r**,**p**) $\mathcal{P}_0(\mathbf{r},\mathbf{p}) = \dfrac{e^{-\beta\mathcal{H}_0(\mathbf{r},\mathbf{p})}}{\int e^{\mathcal{H}(\mathbf{r},\mathbf{p})} d\mathbf{r}d\mathbf{p}}$

Free energy difference

$$\Delta A = -\frac{1}{\beta}\ln\frac{\int e^{-\beta\mathcal{H}_1} d\mathbf{r}^N d\mathbf{p}^N}{\int e^{-\beta\mathcal{H}_0} d\mathbf{r}^N d\mathbf{p}^N} = -\frac{1}{\beta}\ln\frac{\int e^{-\beta\mathcal{H}_0} e^{-\beta\Delta\mathcal{H}} d\mathbf{r}^N d\mathbf{p}^N}{\int e^{-\beta\mathcal{H}_0} d\mathbf{r}^N d\mathbf{p}^N}$$

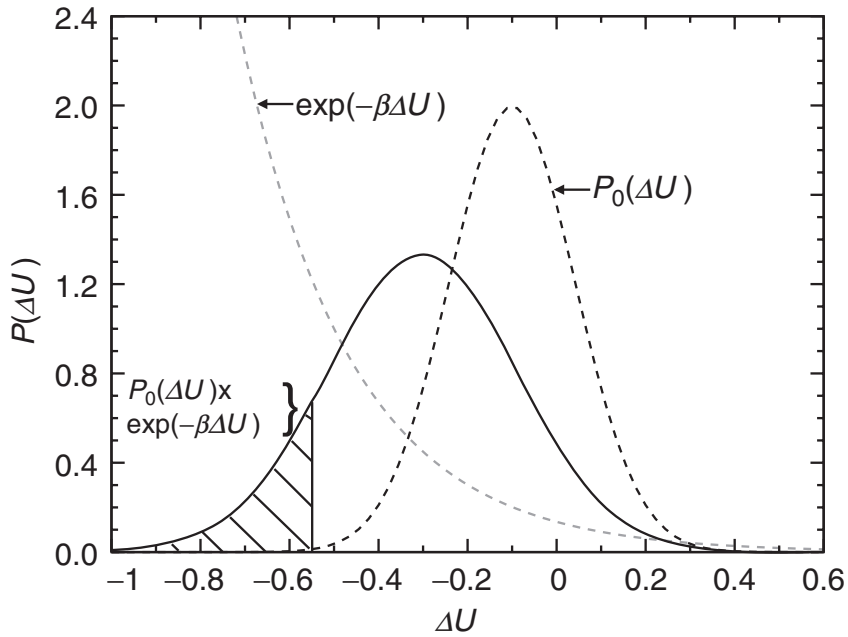
$$\Delta A = -\frac{1}{\beta}\ln\left\langle \exp\left[-\beta\Delta\mathcal{H}(\mathbf{r}^N,\mathbf{p}^N)\right]\right\rangle_0$$

Integrating out the analytic kinetic part

$$\Delta A_{0,1} = -\frac{1}{\beta}\ln\langle e^{-\beta\Delta U}\rangle_0 \qquad \langle\mathcal{F}(\mathbf{r},\mathbf{p})\rangle_1 = \frac{\langle\mathcal{F}e^{-\beta\Delta\mathcal{U}}\rangle_0}{\langle e^{-\beta\Delta\mathcal{U}}\rangle_0}$$

Limitations

$$\Delta A = -\frac{1}{\beta} \ln \int \exp\left[-\beta \Delta \mathcal{U}\right] \mathcal{P}_0(\Delta \mathcal{U}) d\Delta \mathcal{U}$$



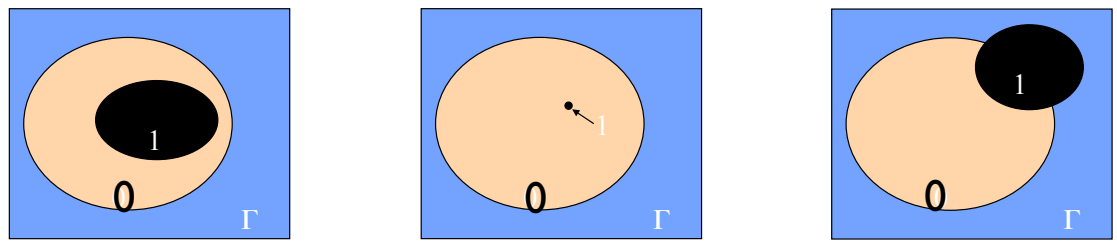
Shifted function

Low-ΔU tail is poorly sampled

low statistical accuracy

but important contribution to ΔA

Accuracy ⇒ target and reference systems are similar⇒ **overlapping regions**



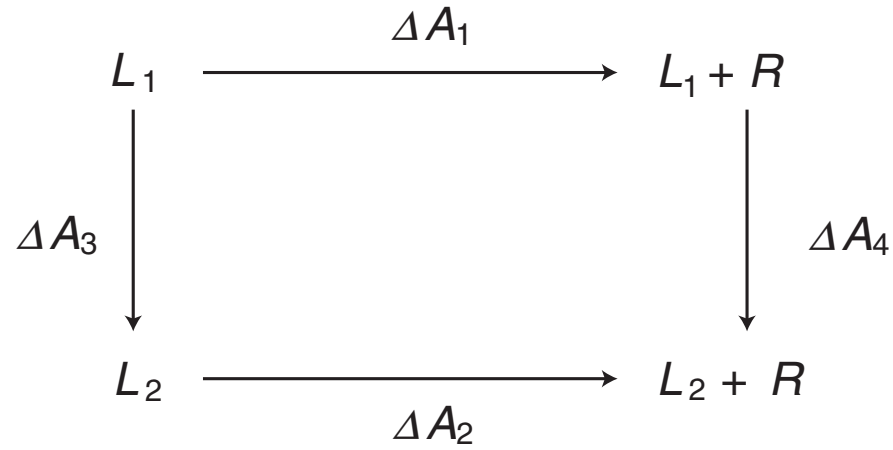
insufficient statistics or incomplete overlap ⇒ **enhanced sampling**

Alchemical Transformations

Protein ligand binding, host-guest chemistry, solvation properties, ...

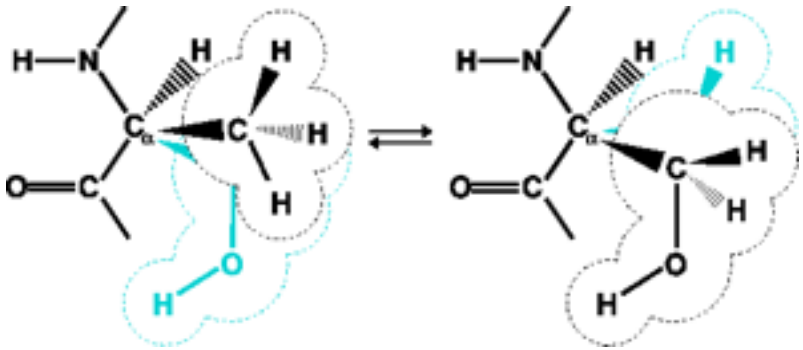
Transformation (0)→(1) as series of non-physical, intermediate states along a pathway characterized by the

"coupling parameter" λ



Free energy as continuous function of λ through H(λ)

$$\mathcal{H}(\lambda) = \mathcal{H}_{\text{env}} + \lambda\mathcal{H}_0 + (1-\lambda)\mathcal{H}_1$$



Point mutation of alanine into serine:
coexistence without seeing each other

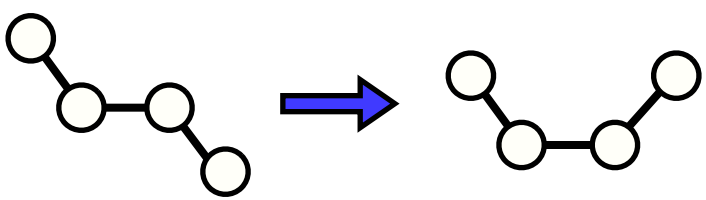
Interaction with side chain tuned through λ

A. E. Mark, *Encyclopaedia of computational chemistry*, 2, 1070 (1998)

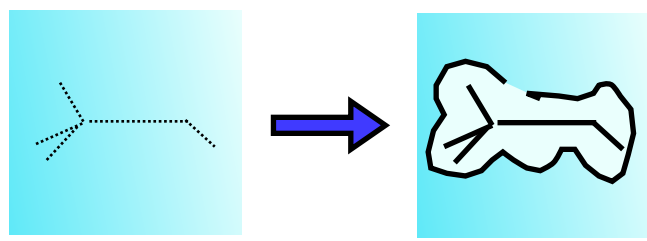
Order Parameters

Variables chosen to describe changes in the system

- **Reaction coordinate** : the order parameter corresponds to the pathway along which the transformation occur in nature
- **Collective variable** : fully represented as function of coordinates
- Indicating intermediate stages of the transformation: **mutation point**



torsion angle



annihilation non-bonded

- Different possible definitions of OP
- Effects on accuracy and efficiency of ΔA calculations
- Set up of system with desired values of OP
- Smoothness of the simulated path

Extended Ensemble

Select parameters, continuous functions of coordinates $\hat{\xi}_i(\mathbf{r}^N)$

Density of States

$$\Omega_\xi(N,V,E,\boldsymbol{\xi}) = \int \delta[\mathcal{U}(\mathbf{r}^N) - E] \left(\Pi_i \delta[\hat{\xi}_i(\mathbf{r}^N) - \xi_i]\right) d\mathbf{r}^N$$

$$\boldsymbol{\xi} = \{\xi_i\}$$

Canonical Partition Function

$$Q_\xi(N,V,T,\boldsymbol{\xi}) = \int e^{-\beta\mathcal{U}(\mathbf{r}^N)} \left(\Pi_i \delta[\hat{\xi}_i(\mathbf{r}^N) - \xi_i]\right) d\mathbf{r}^N$$

Free Energy

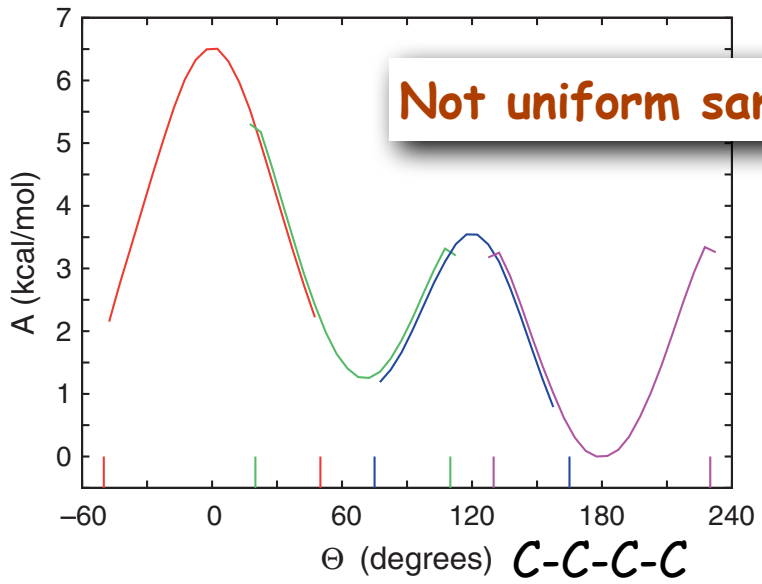
$$A_\xi = -\frac{1}{\beta} \ln Q_\xi$$

$\hat{\xi}_i(\mathbf{r}^N)$ must distinguish among metastable states

select specific configurations in the partition function

Stratification Scheme

Free energy butane isomerization



Probability distribution of the order parameter

$$\Delta A(\xi) = A(\xi_1) - A(\xi_0) = -\beta^{-1} \ln \frac{\mathcal{P}(\xi_1)}{\mathcal{P}(\xi_0)}$$

Histogram of M bins $\delta\xi = (\xi_1 - \xi_0)/M$

$$\mathcal{P}(\xi_0 + (i - 0.5)\delta\xi) = \frac{f_i}{\sum_j f_j}$$

Restrain the system within a window by harmonic potential

Overlapping windows

Efficient sampling $\tau = L\tau_w \propto \frac{(\xi_1 - \xi_0)^2}{LD_\xi}$

Reconstruct the full probability by matching

Importance Sampling

Non-Boltzmann sampling to enhance the probability of important regions

$$\mathcal{P}^{(w)}(\xi,T)=\frac{\int w(\hat{\xi}(\mathbf{r}^N))\exp[-\beta\mathcal{U}(\mathbf{r}^N)]\delta[\xi-\hat{\xi}(\mathbf{r}^N)]d\mathbf{r}^N}{\int w(\hat{\xi}(\mathbf{r}^N))\exp[-\beta\mathcal{U}(\mathbf{r}^N)]d\mathbf{r}^N}$$

positive bias w(ξ)

Free energy differences

$$\Delta A^{(w)}(\xi)=-\beta^{-1}\ln\frac{w(\xi_0)\mathcal{P}^{(w)}(\xi,T)}{w(\xi)\mathcal{P}^{(w)}(\xi_0,T)}=-\beta^{-1}\left[\ln\frac{\mathcal{P}^{(w)}(\xi,T)}{\mathcal{P}^{(w_0)}(\xi,T)}+\ln w(\xi_0)-\ln w(\xi)\right]$$

Biasing potential

$$w(\xi)=\exp[-\beta V(\xi)] \qquad \mathcal{U}^{(b)}(\mathbf{r}^N)=\mathcal{U}(\mathbf{r}^N)+V(\xi(\mathbf{r}^N))$$

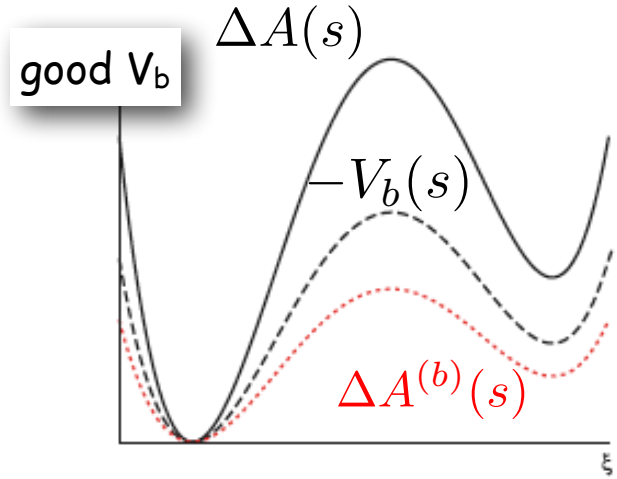
Umbrella Potential
connecting different regions of the phase space

Umbrella Sampling

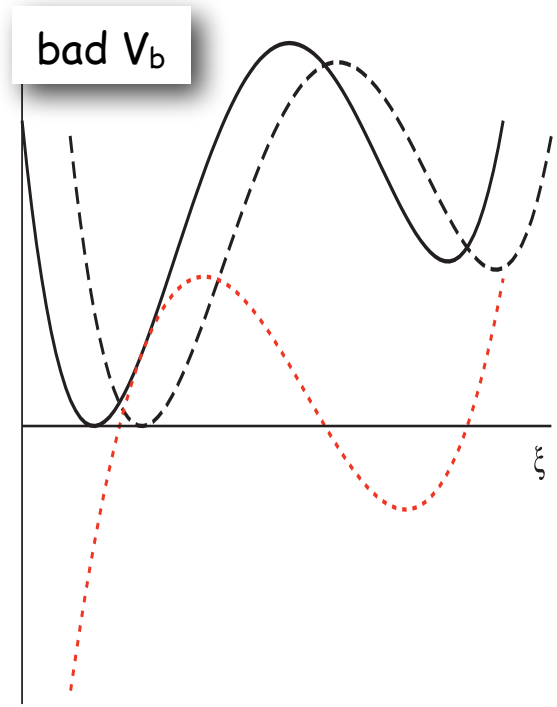
Modify the underlying potential to obtain a uniform sampling : $V_b(s) = -\Delta A(s)$

$$U^{(p)}(x) = U(x) + V_b(s(x))$$

First guess & iterative improvement



as flat as possible



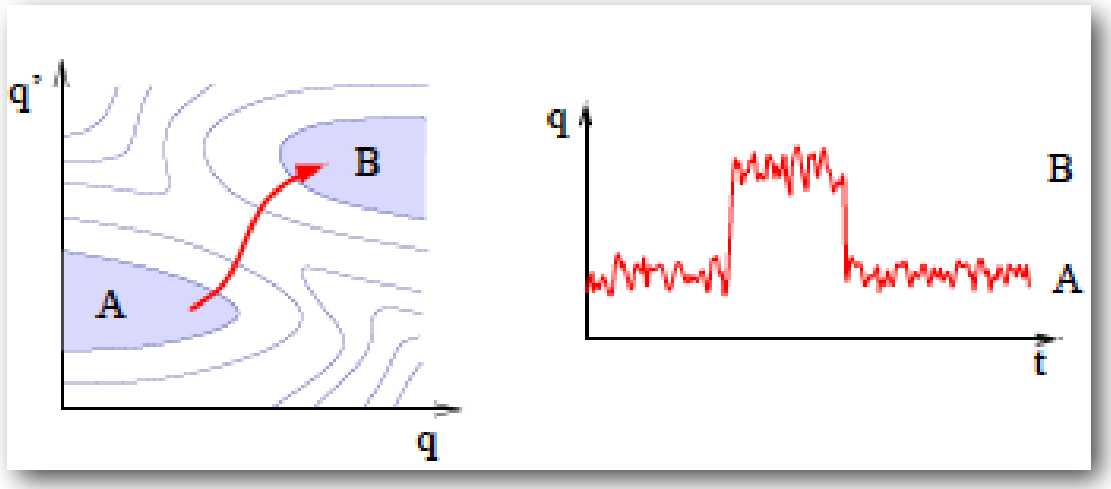
$$\begin{aligned}
\mathcal{P}^{(b)}(s) &= \frac{1}{Z^{(b)}} \int dx \exp\left(-\beta(U(x) + V_b(s(x)))\right) \delta(s - s(x)) \\
&= \frac{Z}{Z^{(b)}} \exp\left(-\beta V_b(s)\right) \frac{1}{Z} \int dx \exp\left(-\beta U(x)\right) \delta(s - s(x)) \\
&= \frac{Z}{Z^{(b)}} \exp\left(-\beta(V_b(s) + A(s))\right)
\end{aligned}$$

Probability in the presence of the bias & un-biasing

$$A(s) = -\frac{1}{\beta} \ln \mathcal{P}^{(b)}(s) - V_b(s) - \frac{1}{\beta} \ln\left(\frac{Z}{Z^{(b)}}\right)$$

Good Coordinates for Pathways

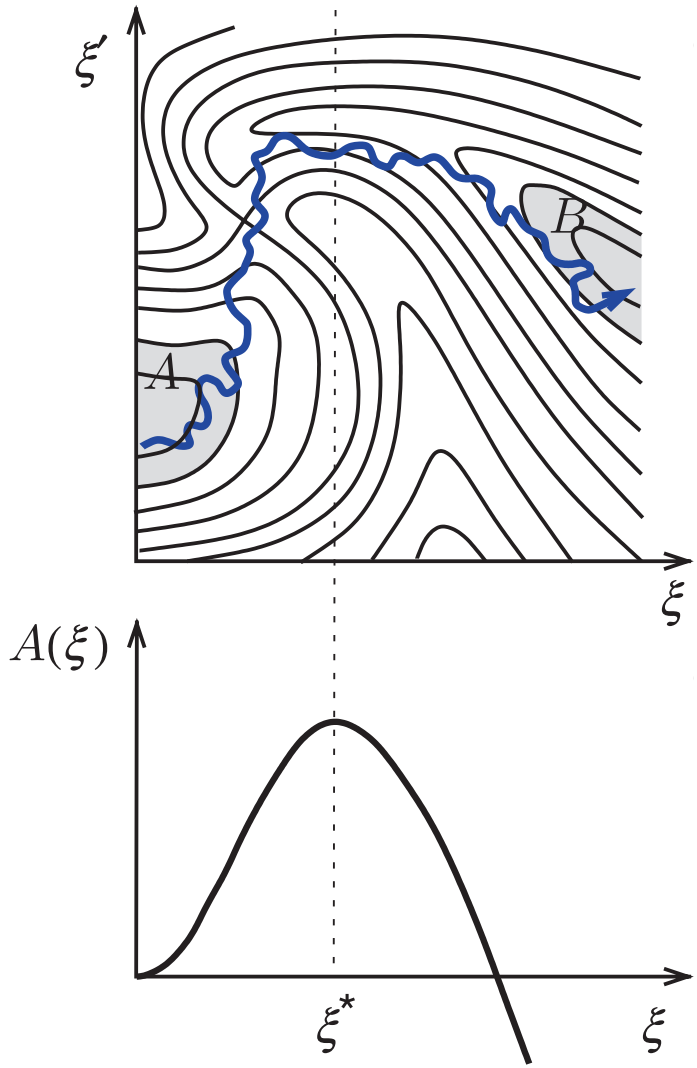
Capture the essential physics include all relevant DoF and properly describes the dynamics



q distinguishes between A and B but might fail in describing essential aspects of the transition

- Discriminate configurations of reactants and products and intermediates
- Characterisation of the mechanisms of transition
- Reversibility
- Fast equilibration of orthogonal DoF (no relevant bottlenecks)

Hypothetical 2D Free Energy Landscape



Not including important DoF leads to wrong interpretation

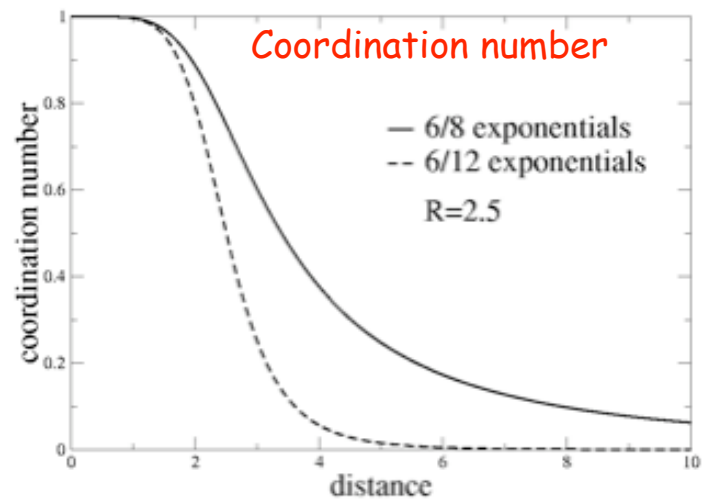
Some simple collective variables

Derivable function of the degrees of freedom to which a given value can be assigned

- Distance $|\mathbf{R}_I - \mathbf{R}_J|$
- Angle $\theta(\mathbf{R}_I, \mathbf{R}_J, \mathbf{R}_k)$
- Dihedral $\Theta(\mathbf{R}_I, \mathbf{R}_J, \mathbf{R}_k, \mathbf{R}_L)$
- Difference of distances $|\mathbf{R}_I - \mathbf{R}_J| - |\mathbf{R}_J - \mathbf{R}_K|$

- Generalised coordination number

$$C_{L_1 L_2} = \frac{1}{N_{L_1}} \sum_{j=1}^{N_{L_1}} \left\{ \sum_{i=1}^{N_{L_2}} \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^n}{1 - \left(\frac{r_{ij}}{r_0}\right)^m} \right\}$$

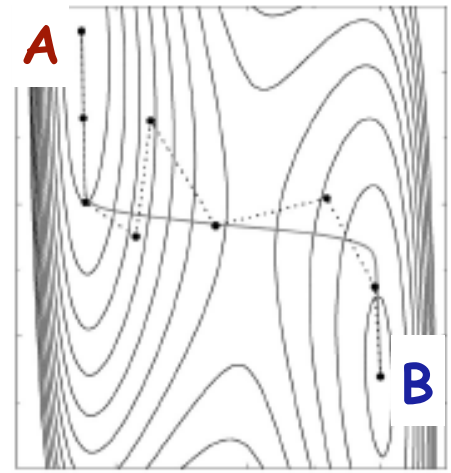


- Generalised displacement

$$D^{[klm]}_{L_1 L_2} = \frac{1}{N_{L_1}} \sum_{i \in L_1} \mathbf{d}_i \cdot \hat{\mathbf{v}}_{[klm]} - \frac{1}{N_{L_2}} \sum_{j \in L_2} \mathbf{d}_j \cdot \hat{\mathbf{v}}_{[klm]}$$

Path Collective Variables

Knowing end-points or a full approximate path (NEB)



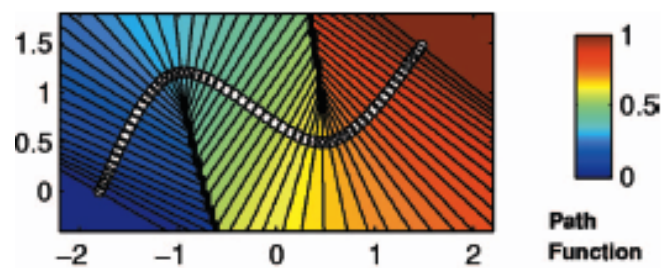
RMSD

$$\mathcal{R}_{\rm k}(\mathbf{R}) = \sqrt{\frac{\sum_i (\mathbf{R}_i - \mathbf{R}_i^{(\rm k)})^2}{N}} \qquad \mathcal{R}_{\rm k}(\mathbf{R}) = \sqrt{\frac{\sum_j^{N_{\rm dist}} (d_j(\mathbf{R}) - d_j^{(\rm k)}(\mathbf{R}))^2}{N_{\rm dist}}}$$

$$s(\mathbf{R}) = \sum_k [\mathcal{R}_{\rm k}(\mathbf{R})]^{n_k} c_k$$

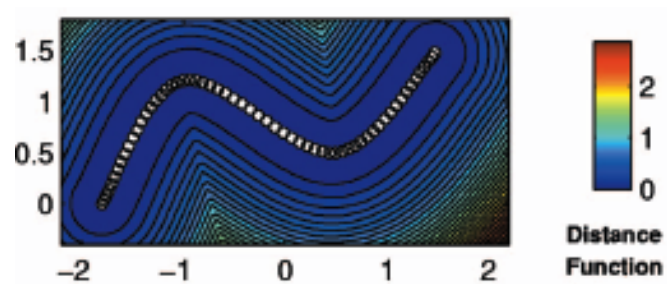
Position along the path

$$s(\mathbf{R}) = \frac{1}{P-1} \frac{\sum_{\rm k=1}^{P} {\rm k}\, e^{-\lambda ||\mathbf{S}(\mathbf{R}) - \mathbf{S}({\rm k})||^2}}{\sum_{\rm k=1}^{P} e^{-\lambda ||\mathbf{S}(\mathbf{R}) - \mathbf{S}({\rm k})||^2}}$$

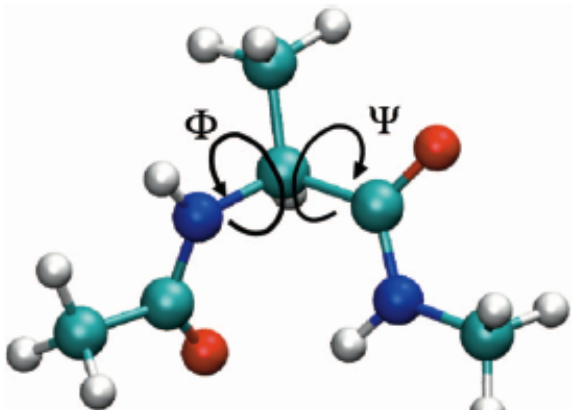


Distance from the path

$$z(\mathbf{R}) = \frac{-1}{\lambda} \ln \left(\sum_{\rm k=1}^{P} e^{-\lambda ||\mathbf{S}(\mathbf{R}) - \mathbf{S}({\rm k})||^2} \right)$$



Dialanine in vacuum

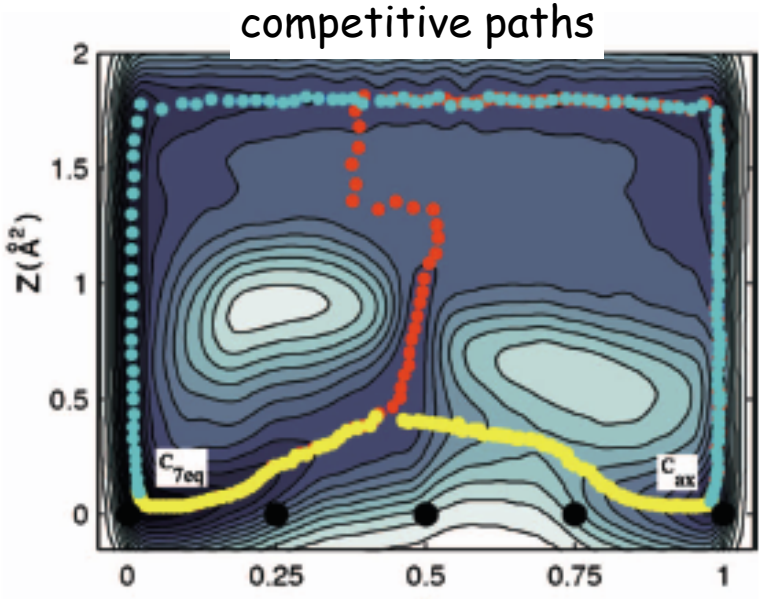


linear interpolation for the initial path

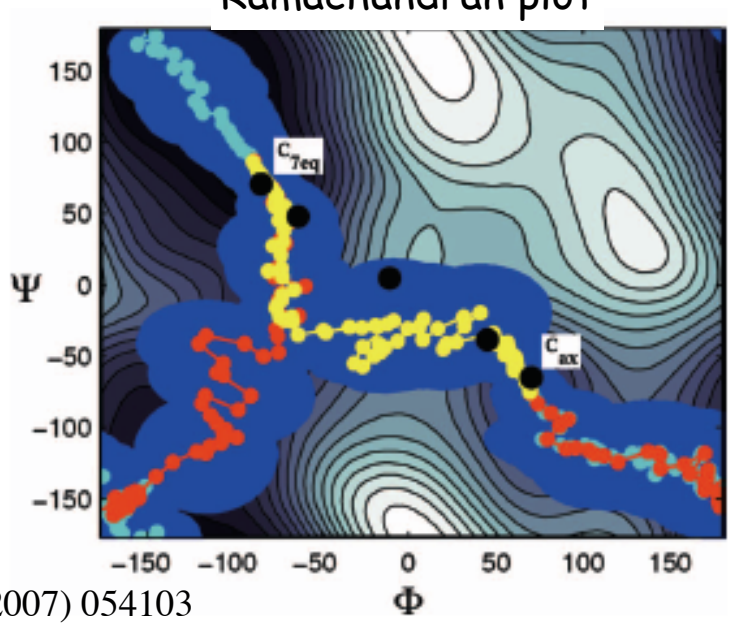
s and z in terms of coordinates

multidimensional path

competitive paths



Ramachandran plot



D. Branduardi, F.L. Gervasio, M. Parrinello, J. Chem. Phys. 126 (2007) 054103

CP2K input for CV

In SUBSYS add one section per CV

```
&COLVAR
 &DISTANCE
  AXIS  X
  ATOMS 1 4
 &END DISTANCE
&END COLVAR
```

```
&COLVAR
 &DISTANCE_FUNCTION
  ATOMS 4 6 6 1
  COEFFICIENT -1.00000
  # distance 1 = ( 4 - 6 )
  # distance 2 = ( 6 - 1 )
 &END DISTANCE_FUNCTION
&END COLVAR
```

```
&COLVAR
 &COORDINATION
  KINDS_FROM  N
  KINDS_TO   O
  R_0 [angstrom]  1.8
  NN  8
  ND  14
 &END COORDINATION
&END COLVAR
```

```
&COLVAR
 &RMSD
  &FRAME
   COORD_FILE_NAME planar.xyz
  &END
  &FRAME
   COORD_FILE_NAME  cage.xyz
  &END
  SUBSET_TYPE LIST
  ATOMS 1 3 5 6 8 9
  ALIGN_FRAMES  T
 &END
&END
```

Constraints and Restraints

In MOTION add one section per constraint

```
&CONSTRAINT
  &COLLECTIVE
    COLVAR 1
    INTERMOLECULAR
    TARGET       5.
    TARGET_GROWTH   1.1
    TARGET_LIMIT    10.
  &END COLLECTIVE
&END CONSTRAINT
```

```
&COLLECTIVE
  TARGET [deg] 0.0
  MOLECULE 1
  COLVAR 1
  &RESTRAINT
     K   [kcalmol] 4.90
  &END
&END COLLECTIVE
```

Geometrical Constraints

Implicit functions of the degrees of freedom of the system

$$\sigma(\{\mathbf{R}_I\}, \mathbf{h}, \Psi) = 0 \qquad \dot{\sigma}(\{\mathbf{R}_I\}, \mathbf{h}, \Psi) = 0$$

- To freeze fast degrees of freedom and increase the time step: e.g., intra-molecular bonds by distance constraints
- To explore only a sub-region of the conformational space
- As reaction coordinates : constrained dynamics and thermodynamic integration
- To prevent specific events or reactions

Lagrange formulation for simple constraints, functions of $\mathbf{R}_I$

$$\mathcal{L}'(\{\mathbf{R}_I\}, \{\mathbf{P}_I\}) = \mathcal{L}(\{\mathbf{R}_I\}, \{\mathbf{P}_I\}) - \sum_\alpha \lambda_\alpha \sigma(\{\mathbf{R}_I\})$$

The Lagrange multipliers ensure that positions and velocities satisfy the constraints

Shake-Rattle algorithm

Modified velocity Verlet scheme by additional constraint forces

- First update of velocities (first half step) and positions

$$V_I' = V_I(t) + \frac{\delta t}{2M_I} F_I(t) \qquad R_I' = R_I(t) + \delta t V_I'$$

- Positions' correction by **constraint forces**

$$R_I(t+\delta t) = R_I' + \frac{\delta t^2}{2M_I} g_I^{(p)}(t)$$

- Calculation of the new forces $F_I(t+\delta t)$

- Update of velocity (second half step)

$$V_I(t+\delta t) = V_I' + \frac{\delta t}{2M_I}[F_I(t+\delta t) + g_I^{(v)}(t+\delta t)]$$

- Correction by the constraint forces

Constraint Forces

$$g_I^{(p)}(t) = -\sum_\alpha \lambda_\alpha^{(p)} \frac{\partial \sigma_\alpha(\{R_I\})}{\partial R_I}$$

$$e_\alpha(\{\lambda_\gamma\}) = -\sum_\beta \mathbf{J}^{-1}_{\alpha\beta}\, \sigma_\beta(\{\lambda_\gamma\}) \qquad \mathbf{J}_{\alpha\beta} = \frac{\partial \sigma_\alpha(\{\lambda_\gamma\})}{\partial \lambda_\beta}$$

Set of non-linear equations **solved iteratively**

$$g_I^{(v)}(t) = -\sum_\alpha \lambda_\alpha^{(v)} \frac{\partial \sigma_\alpha(\{R_I\})}{\partial R_I}$$

$$\sum_I \frac{\partial \sigma_\alpha}{\partial \mathbf{R}_I} \mathbf{V}_I = \sum_I \frac{\partial \sigma_\alpha}{\partial \mathbf{R}_I} \cdot \mathbf{V}_I' + \sum_\beta \left(\sum_I \frac{\delta t^2}{2M_I} \frac{\partial \sigma_\alpha}{\partial \mathbf{R}_I} \frac{\partial \sigma_\beta}{\partial \mathbf{R}_I} \right) \lambda_\beta^v = 0$$

H.C. Andersen, J. Comp. Phys., **52**, 24 (1983)

Thermodynamic Integration

$$A(\xi_1) - A(\xi_0) = \int_{\xi_0}^{\xi_1} \frac{dA}{d\xi} d\xi$$

along a one dimensional generalized coordinate ξ(x)
Path-independent

Potential of Mean Force exerted on ξ

$(\mathbf{x}, \mathbf{p}) \Rightarrow (\xi, q_1 .. q_{N-1}, p^{\xi} .. p^{q}_{N-1})$ generalized coordinate to simplify derivative

$$\frac{dA}{d\xi} = \frac{\int \frac{\partial \mathcal{H}}{\partial \xi} e^{-\beta \mathcal{H}} dq_1 .. dq_{N-1} dp^{\xi} .. dp^{q}_{N-1}}{\int e^{-\beta \mathcal{H}} dq_1 .. dq_{N-1} dp^{\xi} .. dp^{q}_{N-1}} = \left\langle \frac{\partial \mathcal{H}}{\partial \xi} \right\rangle_{\xi}$$

instantaneous force on ξ

force at ξ, averaged over fluctuations of other DoF

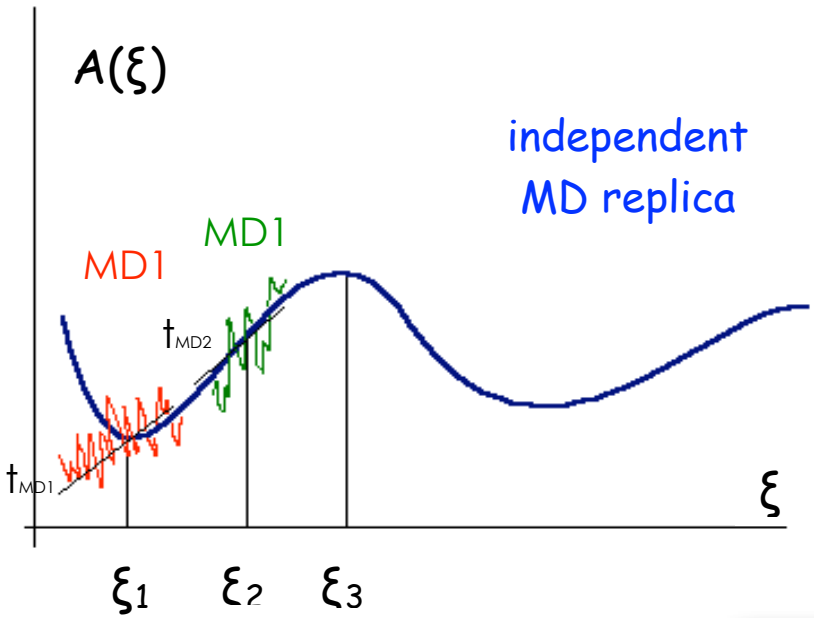
$$\left\langle \frac{\partial \mathcal{H}}{\partial \xi} \right\rangle_{\xi} = \left\langle \frac{\partial U}{\partial \xi} - \frac{1}{\beta} \frac{\partial \ln |\mathbf{J}|}{\partial \xi} \right\rangle$$

$$[\mathbf{J}]_{ij} = \frac{\partial x_i}{\partial q_j}$$

mechanical + entropic

Blue Moon Ensemble

Series of constrained MD simulations $\mathcal{H}^\lambda = \mathcal{H} + \lambda(\xi - \xi(\mathbf{r}))$



$-\lambda\nabla(\xi - \xi(\mathbf{r}))$ mean force acting on the system to hold ξ constant

$\dot{\xi} = 0 \quad : \quad p^\xi(\mathbf{q}, \mathbf{p}^q)$

un-constrained <..> = constrained corrected $\langle .. \rangle^F$

Fixman Potential

$$\mathcal{H}_F^\lambda = \mathcal{H}^\lambda + \frac{1}{2\beta} \ln Z_\xi \qquad Z_\xi = \sum_i \frac{1}{m_i} \left(\frac{\partial \xi}{\partial x_i} \right)^2$$

$$\frac{dA}{d\xi} = \langle \lambda_F \rangle_{\xi\dot{\xi}}^F$$

mean force acting on ξ related to external force to hold ξ constant

...some difficulties

MD performed at fixed ξ, collecting statistics of the force acting on ξ $\Rightarrow \lambda\nabla\xi$, **constraint force**

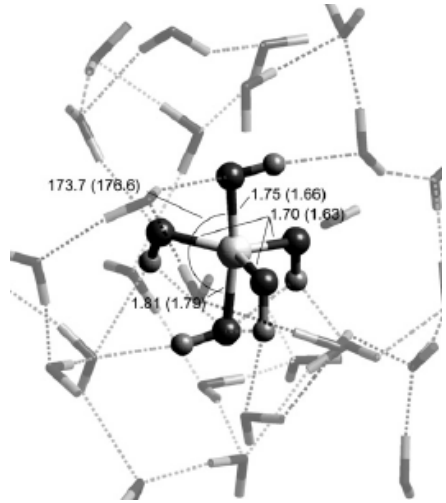
- Many estimates at ξ to reduce statistical errors
- Many windows to get accurate integrals
- May not be easy to prepare by hand the system at given ξ
- Different possible pathways: the starting configuration selects one path, but crossing is rare, ξ(r)=ξ partially sampled or rate limiting
- Multidimensionality (more coordinates) too expensive

De-protonation of $P(OH)_5$

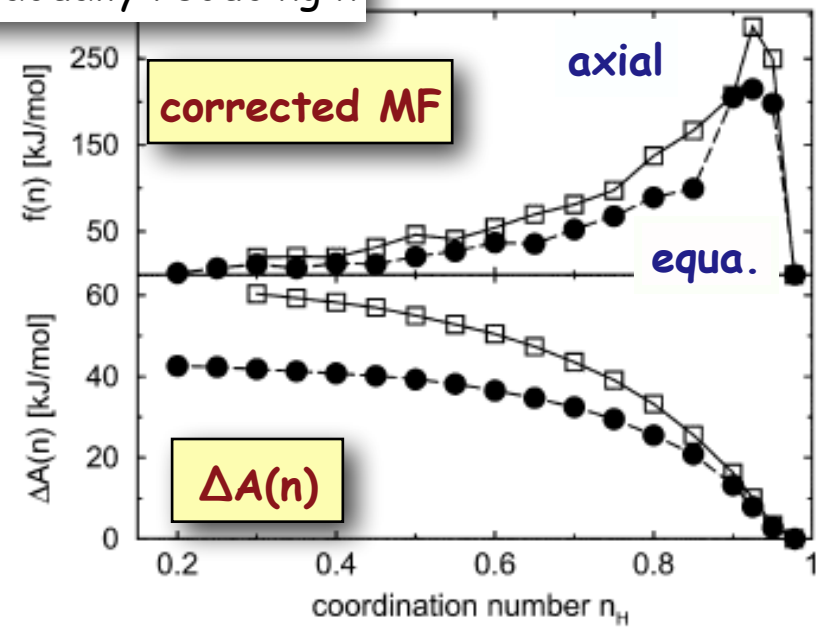
Order Parameter : coordination of a specific donor phosphorane O (r_0=1.3 A)

equivalent PO in solvated $P(OH)_5$

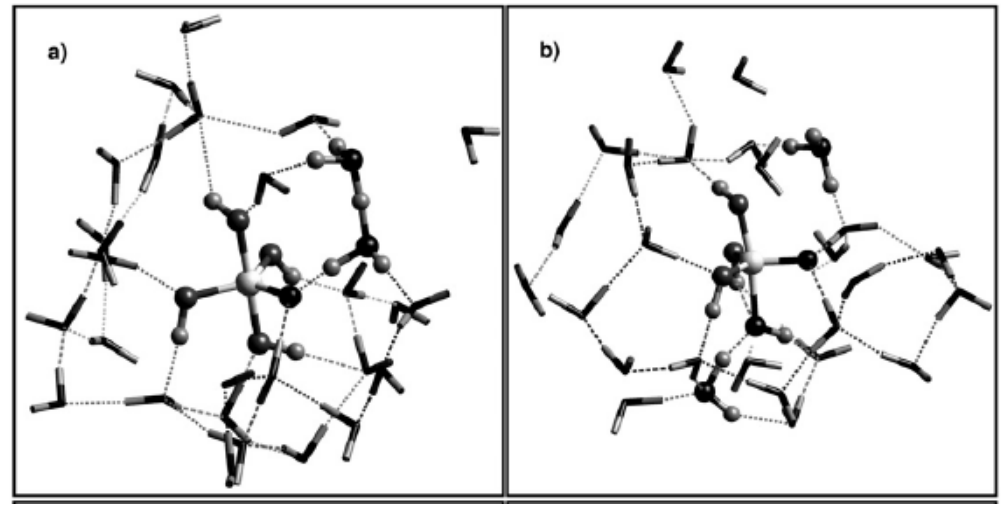
Pronounced equatorphilicity of O^- in $PO_5H_4^-$



gradually reducing n



- de-protonation and formation of Zundel ion
- H_3O^+ breaks loose
- chain of proton transfers
- protonation of axial OH
- formation of H_3PO_4 and subsequent de-protonation



N. Doltsinis et al, PCCP, **5**, 2612 (2003)

Parallel Tempering

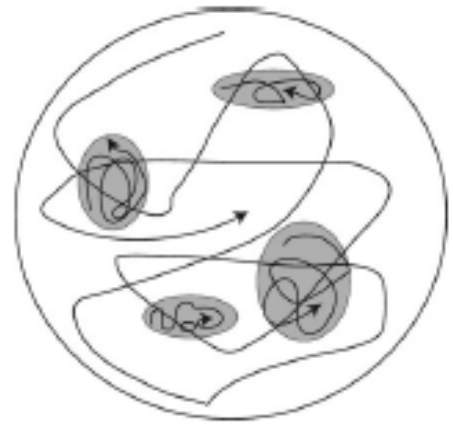
Running **M replica** at different T

high T to explore large part of the phase space

low T to sample precisely local minima

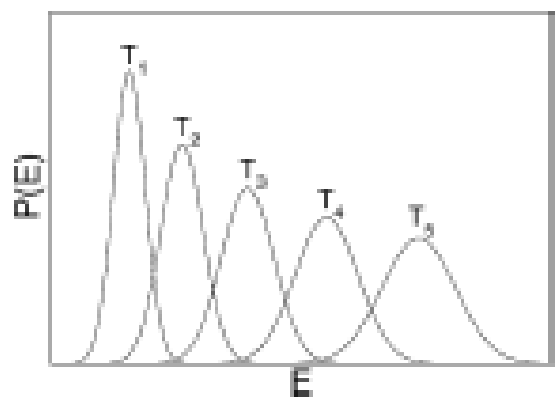
$$T_1 < T_2 < .. < T_M$$

Equilibration + swap attempt + rescaling of velocities



exchange ensures access to all local minima at low T

Energy histogram



$$\sigma \propto (N)^{1/2}$$

Swapping acceptance

$$\alpha_{IJ} = min\left\{1, e^{-(\beta_I - \beta_J)(U(\mathbf{r}_I) - U(\mathbf{r}_J))}\right\}$$

likelihood that two replicas have overlap in phase space

M and ΔT strongly affect sampling efficiency and computational costs

Distribution dihedral of Gly-2

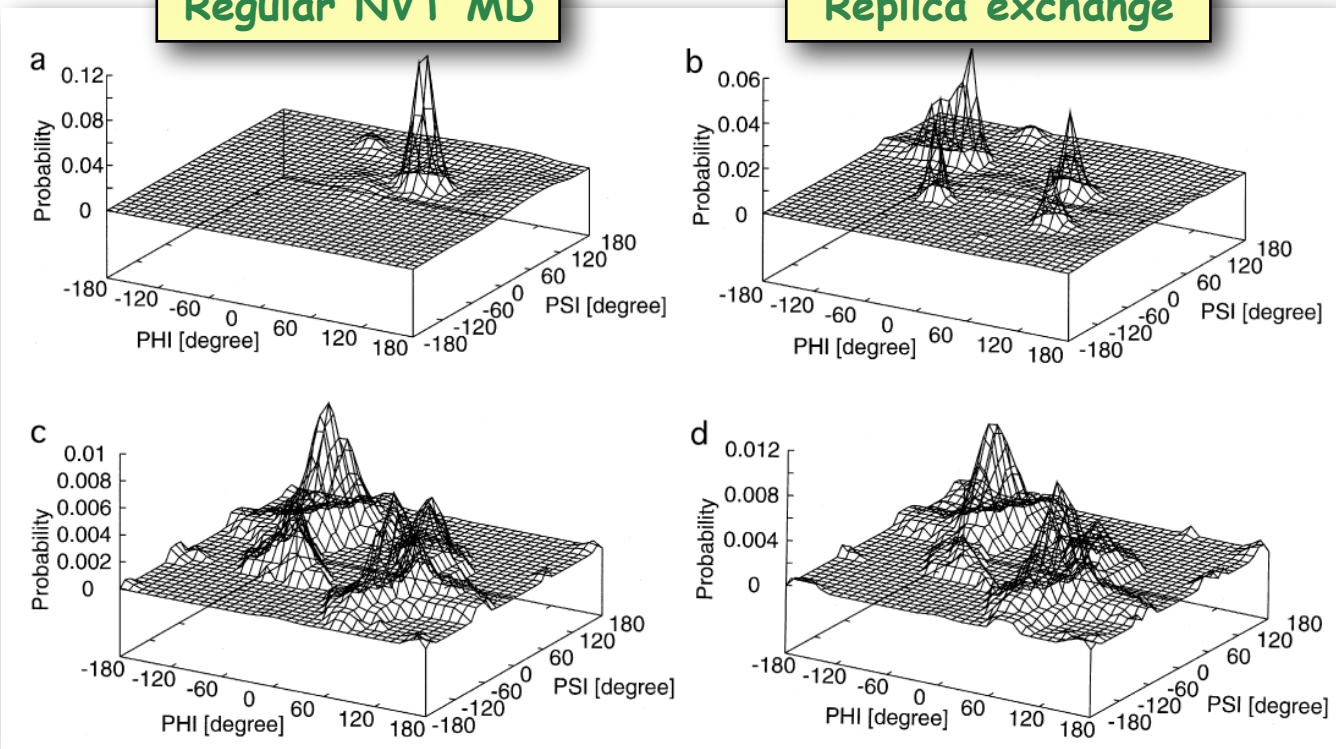
Penta-peptide system in gas phase(Try-Gly-Gly-Phe-Met), all-atom AMBER

Eight temperatures replica exchange simulation, over 1 ns

Pair of temperatures	Acceptance ratio
200 ↔ 239 K	0.160
239 ↔ 286 K	0.149
286 ↔ 342 K	0.143
342 ↔ 409 K	0.139
409 ↔ 489 K	0.142
489 ↔ 585 K	0.146
585 ↔ 700 K	0.146

Regular NVT MD

Replica exchange



200 K

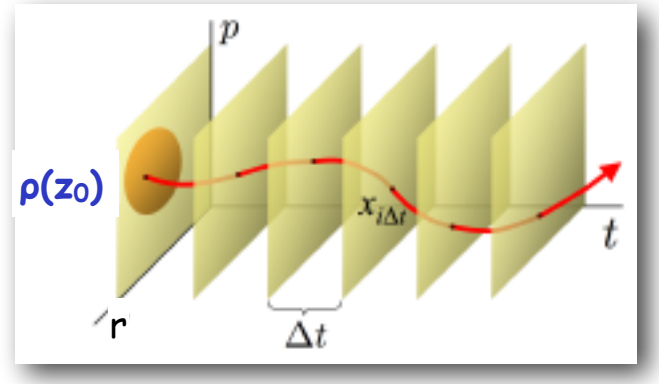
700 K

Y. Sugita et al, CPL, **314**, 141 (1999)

Transition Path Sampling

- Statistical, reaction-coordinate free description of pathways connecting long-lived stable states

  $z(T) \equiv \{z_0, z_{\Delta t}, \ldots z_T\}$

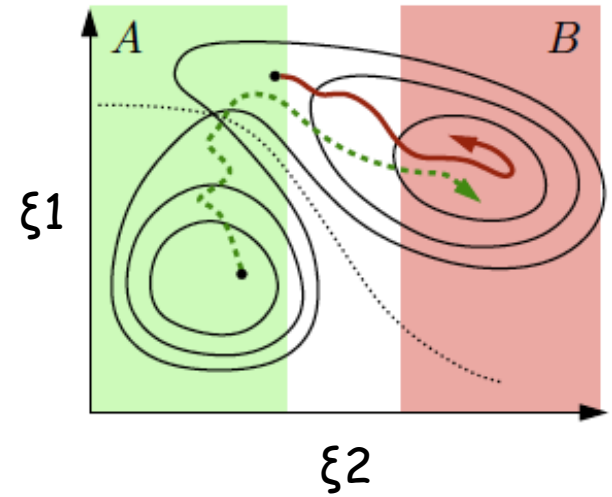


- The probability distribution in the pathway ensemble depends on $\rho(z_0)$ and the dynamics

$$\mathcal{P}[z(\tau)] = \rho(z_0) \prod_{i=0}^{\tau/\Delta t - 1} p(z_{i\Delta t} \rightarrow z_{(i+1)\Delta t})$$

Markovian processes

- Good description of initial and final states required by definition of multidimensional $\xi(\mathbf{r})$ to identify A and B
  - large enough to contain equilibrium fluctuations
  - no overlap with opposite basin of attraction



Path generation

Transition-path-ensemble

- Focused on the sub-ensemble of pathways containing barrier-crossing events

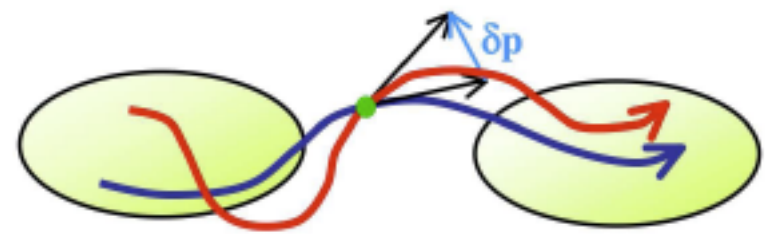
$$\mathcal{P}_{AB}[z(\tau)] = h_A(z_0)\mathcal{P}[z(\tau)]h_B(z_\tau)/Z_{AB}(\tau)$$

Efficiency:

- z^n and z^o as different as possible
- $p_{acc}[z^n \rightarrow z^o]$ as large as possible

- Random walk in TP-ensemble by Metropolis MC : **detailed balance criterion**

$$p_{\rm acc}[z^{\rm n}(\tau) \rightarrow z^{\rm o}(\tau)] = \min\left\{1, \frac{\mathcal{P}_{AB}[z^{\rm n}(\tau)]p_{\rm g}[z^{\rm n} \rightarrow z^{\rm o}]}{\mathcal{P}_{AB}[z^{\rm o}(\tau)]p_{\rm g}[z^{\rm n} \rightarrow z^{\rm o}]}\right\}$$



Shooting

$$p_{\rm acc}[z^{\rm n}(\tau) \rightarrow z^{\rm o}(\tau)] = h_A(z_0^{\rm n})h_B(z_\tau^{\rm n})\min\left\{1, \frac{\rho(z_0^{\rm n})}{\rho(z_0^{\rm o})}\right\}$$

- The most probable reactive trajectories are identified and transition mechanisms are characterised

Shifting

Metadynamics

- Canonical equilibrium distribution under potential V(**r**)
- Choose a set of relevant Collective Variables **S**(**r**):{S_α(**r**)}, such that the process is well defined in the reduced space Σ(**S**)

$$P(\mathbf{S}) = \frac{e^{-\beta A(\mathbf{S})}}{\int \mathrm{d}\mathbf{S}\; e^{-\beta A(\mathbf{S})}} \qquad A(\mathbf{S}) = -\frac{1}{\beta}\ln\left(\int \mathrm{d}\mathbf{r}\; e^{-\beta V(\mathbf{r})}\delta(\mathbf{S}-\mathbf{S}(\mathbf{r}))\right)$$

- Perform MD and re-map each micro-state by projecting the trajectory into the configuration space Σ(**S**): meta-trajectory **S**(**r**(t))
- Enhance the exploration by adding a penalty potential that discourages the system to visit already explored macro-states

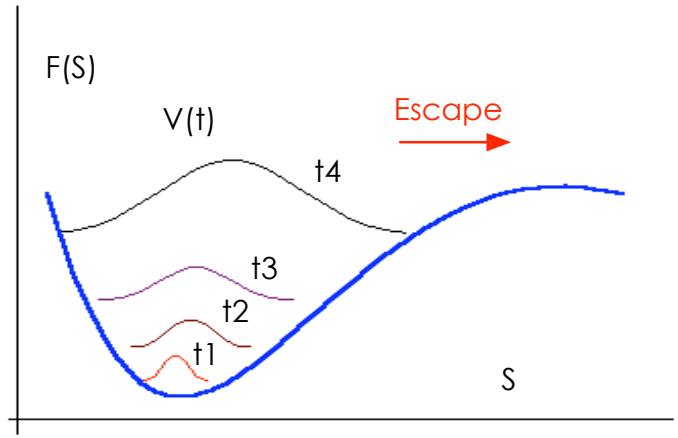
$$V_{\mathrm{MTD}}(\mathbf{S}(\mathbf{r}),t) = \sum_{t'=\tau_G,2\tau_G,\ldots} W_{t'} e^{-\frac{[\mathbf{S}(\mathbf{r})-\mathbf{S}(\mathbf{r}_G(t'))]^2}{2\Delta\mathbf{S}^2}}$$

- New probability distribution generated under the action of $V+V_{MTD}$

A Laio et al. Proc. Natl. Acad. Sci. U.S.A., 99, 12562 (2002)
M Iannuzzi et al, PRL, 90, 238302 (2003)

History Dependent Potential

Non Markovian Coarse-grained MD



Fill-up by successive contributions
Keeping track of the explored space
Direct estimate of FES topology

Eliminate metastability and reconstruct A(S) within Σ(S)

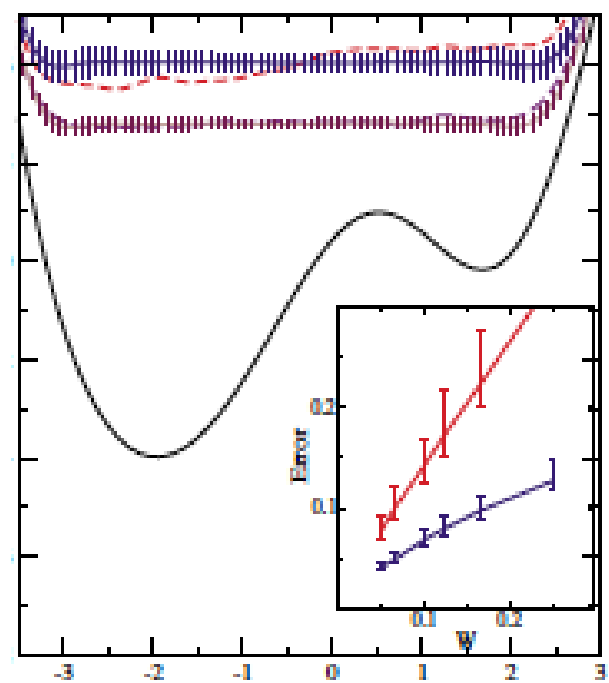
$$A_G(\mathbf{S},t) = -V_{\mathrm{MTD}}(\mathbf{S}(\mathbf{r}),t)$$

Flattening of free energy surface

$$W/\tau_G \to 0$$

$$P(\mathbf{S}) \propto e^{-\beta[A(\mathbf{S})-A_G(\mathbf{S},t)]}$$

$$\delta t_{\mathrm{MD}} \ll \tau_{\mathrm{G}} \ll \tau_{\mathbf{s}}$$



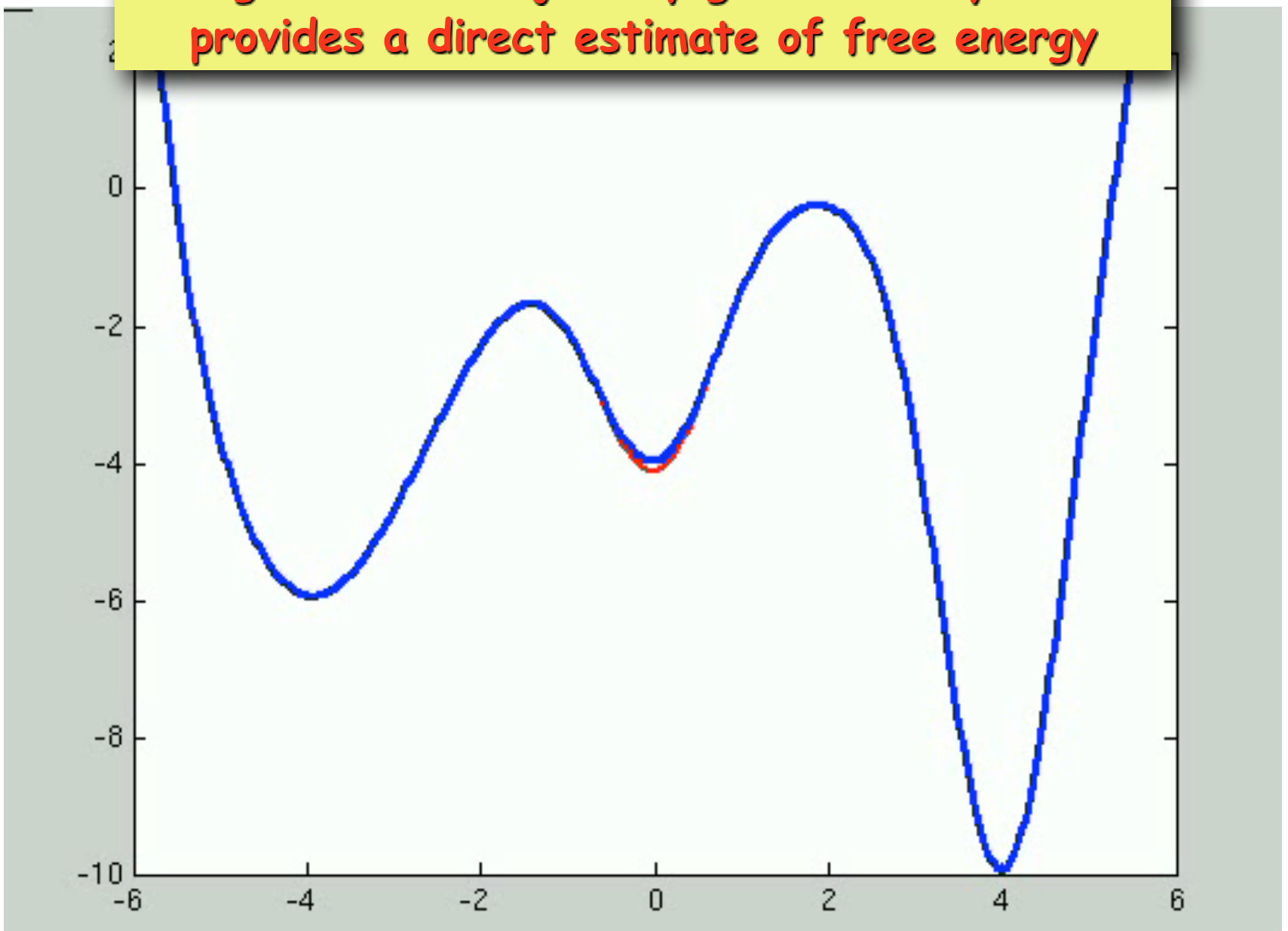
$$\delta A(\mathbf{S}) = A(\mathbf{S}) - A_G(\mathbf{S},t)$$

$$\langle \delta A(\mathbf{S}) \rangle$$

C Micheletti et al, PRL, 92, 170601 (2004)

2D FES

Histogram of a trajectory generated by $V+V_{MTD}$ provides a direct estimate of free energy



Extended Lagrangian MTD

Enforcing adiabatic separation, $\tau_{\mathbf{s}} \gg$ other time scales and memory effects

Introduction of auxiliary variables **s** : $\{s_\alpha\}$ one for each $S_\alpha(\mathbf{r})$ with large enough **M**

$$\mathcal{L} = K - V(\mathbf{r}) + \sum_\alpha \frac{1}{2} M_\alpha \dot{s}_\alpha - \sum_\alpha \frac{1}{2} k_\alpha (s_\alpha - S_\alpha(\mathbf{r}))^2 - V_{\mathrm{G}}(\mathbf{s}, t)$$

thermalization

coupling to system DoF

sampling enhancement

$$V_{\mathrm{G}}(\mathbf{s}, t) = \sum_{t'<t} W_{t'} e^{-\frac{(\mathbf{s}-\mathbf{s}_{\mathrm{G}}(t'))^2}{2\Delta \mathbf{s}^2}}$$

For large t and slow deposition rate, V_G approximates the free energy and the meta-trajectory **s**(t) follows the MEP

$$A_{\mathbf{k}}(\mathbf{s}) = -\frac{1}{\beta} \ln \left(\int \mathrm{d}\mathbf{r}\, e^{-\beta \left[V(\mathbf{r}) + \frac{1}{2} \sum_\alpha k_\alpha (s_\alpha - S_\alpha(\mathbf{r}))^2 \right]} \right)$$

$$\lim_{\mathbf{k}\to\infty} A_{\mathbf{k}}(\mathbf{s}) = A(\mathbf{s})$$

M Iannuzzi et al, PRL, 90, 238302 (2003)

Equations of Motion

Smooth and continuous meta-trajectories that follow the minimum energy pathway are obtained by a modified Velocity Verlet algorithm

By the assumption that S_α is function only of $\mathbf{R}_I$

Force on the meta-variable

$$f_\alpha(t) = k_\alpha(S_\alpha(\{\mathbf{R}_I(t)\}) - s_\alpha(t)) - \frac{\partial}{\partial s_\alpha} V_{\mathrm{MTD}}(\mathbf{s}, t)$$

V_{harm} and V_{MTD} generate opposite contributions

Modified force on ions due to the coupling with the dynamics of the meta-variables

$$f_I(t) = f_I^{(0)}(t) - k_\alpha(S_\alpha(\{\mathbf{R}_I(t)\}) - s_\alpha(t)) \frac{\partial S_\alpha}{\partial R_I}$$

The Metadynamics Lagrangian generates fictitious dynamics describing the most probable pathways

Gaussian anisotropic shape

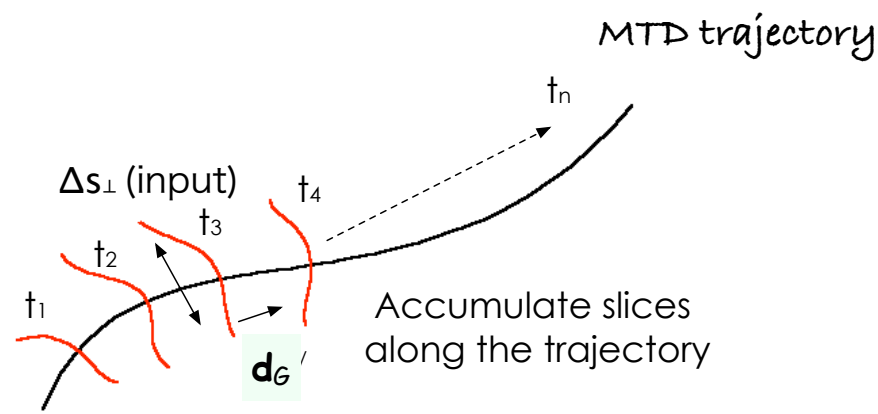
Different CVs, anisotropic A(s), different diffusion coefficients

$$\tau_{s_\alpha} = L_\alpha / D_\alpha \qquad\qquad \sigma_\alpha = s_\alpha / L_\alpha$$

$$V_{\mathrm{G}}(\mathbf{s},t) = \sum_{t'<t} W_{t'} e^{-\sum_\alpha \frac{(s_\alpha - s_\alpha(t'))^2}{2(\Delta s_\alpha)^2}}$$

$$V_{\mathrm{G}}(\mathbf{s},t) = \sum_{t'<t} W_{t'} e^{-\frac{(\mathbf{s}-\mathbf{s}_{\mathrm{G}}(t'))^2}{2\Delta \mathbf{s}^2}} \times e^{-\frac{[(\mathbf{s}-\mathbf{s}_{\mathrm{G}}(t'))\cdot \mathbf{d}_{\mathrm{G}}(t')]^2}{2||\mathbf{d}_{\mathrm{G}}(t')||^4}}$$

re-shaped Gaussian tube



Accuracy of FES profile

Dynamics generating the equilibrium distribution associated with $A(\mathbf{s})$-$A_G(\mathbf{s},t)$

Averaging over many independent trajectories

$$\epsilon(\mathbf{s},t) = \sqrt{\langle (A_{\mathrm{G}}(\mathbf{s},t) - A(\mathbf{s}) - \langle A_{\mathrm{G}}(\mathbf{s},t) - A(\mathbf{s})\rangle)^2\rangle}$$

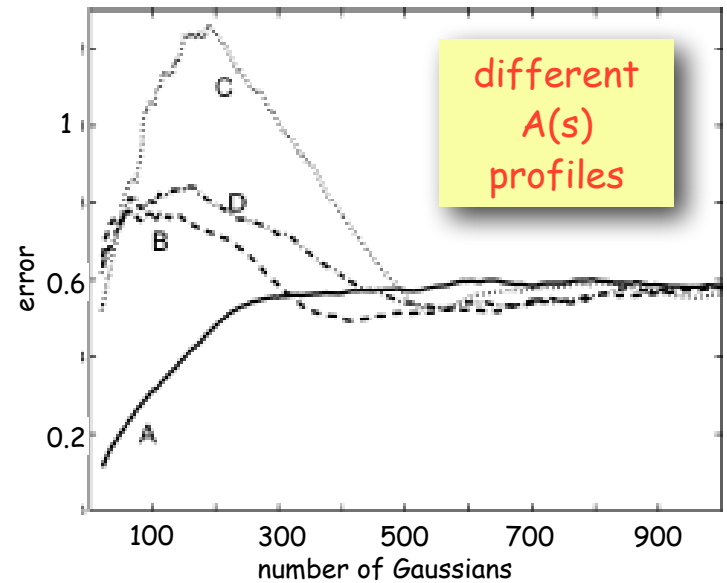
$$\bar{\epsilon}(t) = \frac{\int_\Omega \mathrm{d}\mathbf{s}\ \epsilon(\mathbf{s},t)}{\int_\Omega \mathrm{d}\mathbf{s}}$$

Empirical error estimate

$$\bar{\epsilon} = C(d)\sqrt{\frac{\mathcal{V}_\Omega \|\Delta\mathbf{s}\| W}{\|\mathbf{D}\|\tau_{\mathrm{G}}\beta}}$$

$$t_{\mathrm{tot}} = \tau_{\mathrm{G}} \frac{\int_{\Omega: A(\mathbf{s})<A_{\max}} \mathrm{d}\mathbf{s}(A_{\max} - A(\mathbf{s}))}{(2\pi)^{d/2} W \prod_\alpha \Delta s_\alpha}$$

$$\bar{\epsilon} \propto \sqrt{\frac{\|\tau_{\mathbf{s}}\| \bar{A}}{t_{\mathrm{tot}}\beta} \frac{\mathcal{V}_\Omega}{\prod_\alpha \Delta s_\alpha}}$$



too large Δs would smear out A(s) details : Δs/L<0.1

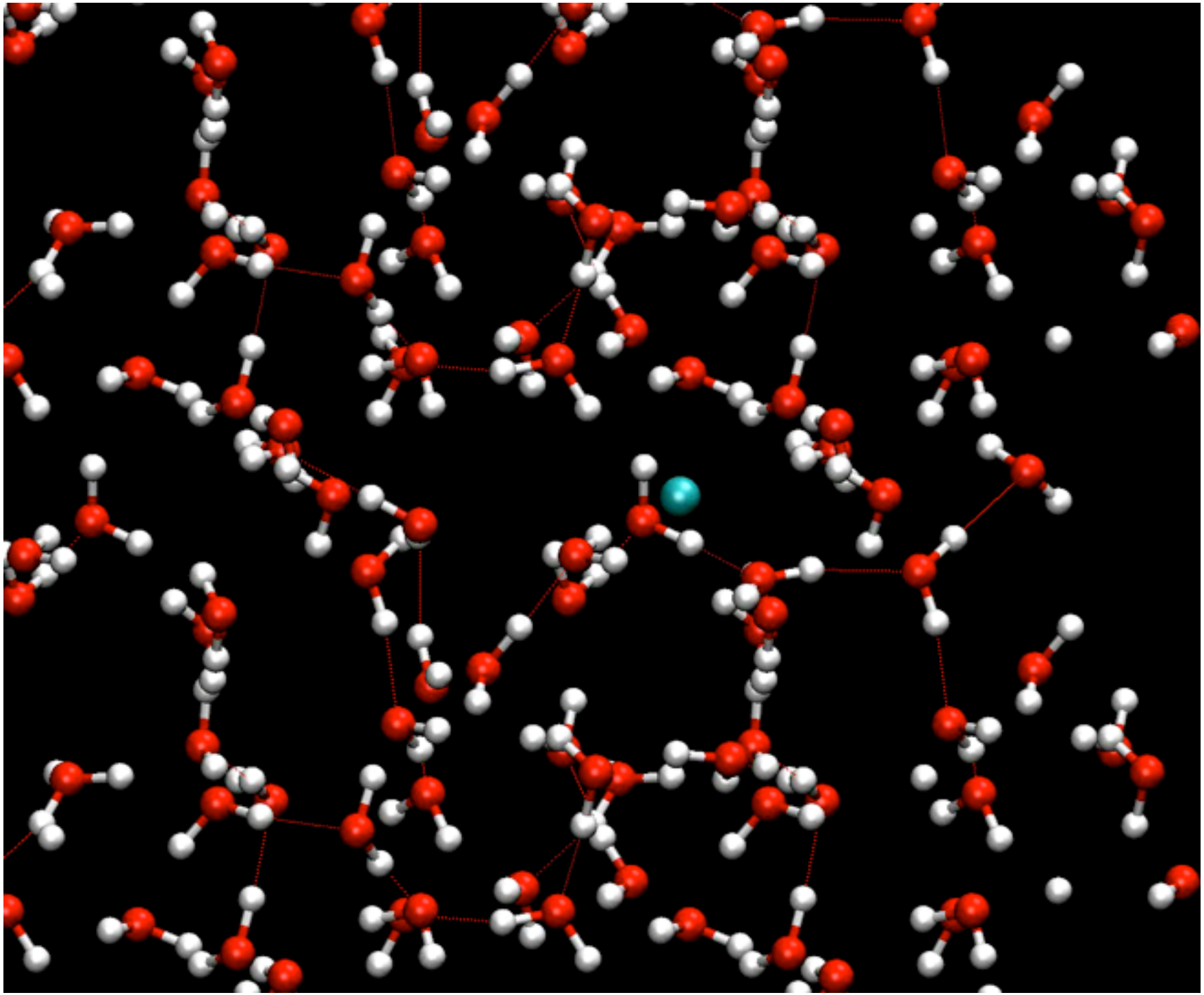
Only relevant time scale is τs

the error depends on τ_G/W
small Gaussians more frequently is better

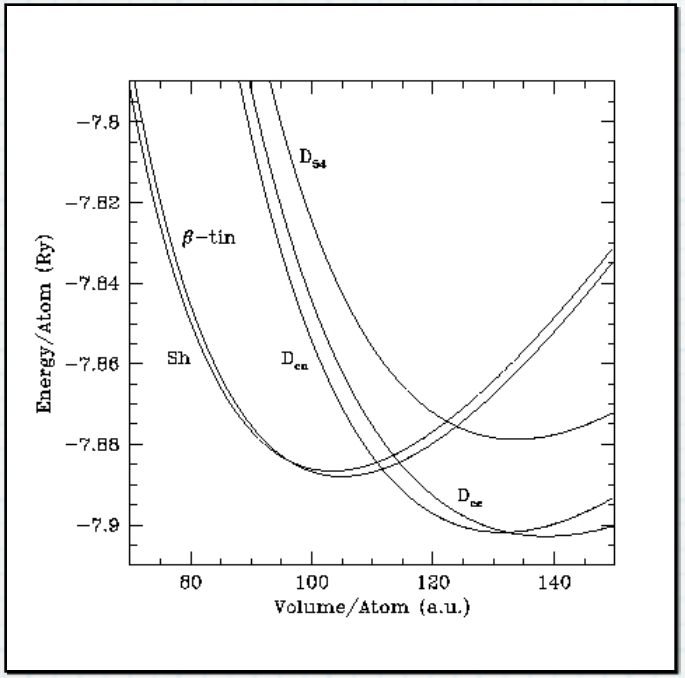
C Micheletti et al, PRL, 92, 170601 (2004)
A Laio et al, JPC-B 109, 6714 (2005)

To be considered ....

- The selected CV must **discriminate among the relevant states** (reactants, products, TS)
- The **number of hills** required to fill the well is proportional to $1/(\Delta)^{NCV}$
- The sampling of large variations of the CV over almost **flat energy regions** is expensive: **diffusive behavior**
- **MTD is not the true dynamics**. Reaction rates are derived a posteriori from the estimated FES
- The analysis of the trajectory is needed to isolate the TS
- With proper choices of CV and parameters, the MTD trajectory describes the **most probable pathway** taking into account also possible kinetic effects (lager and shallower channels are preferred)
- The accuracy in the evaluation of the FES depends on hills' shape and size, and on the deposition rate. The **ideal coverage $V_G(\{S\}) = -A(\{S\})$ (flat surface)**



Phase Transitions



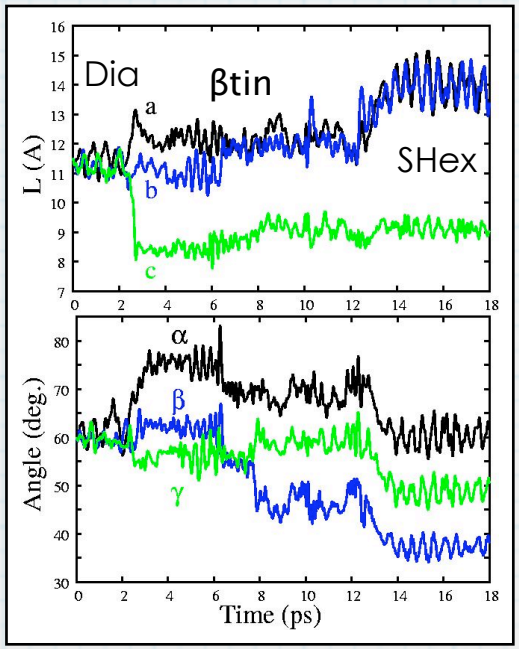
Crystalline Silicon can assume different lattice structures. Phase transitions induced by external pressure are known from experiment

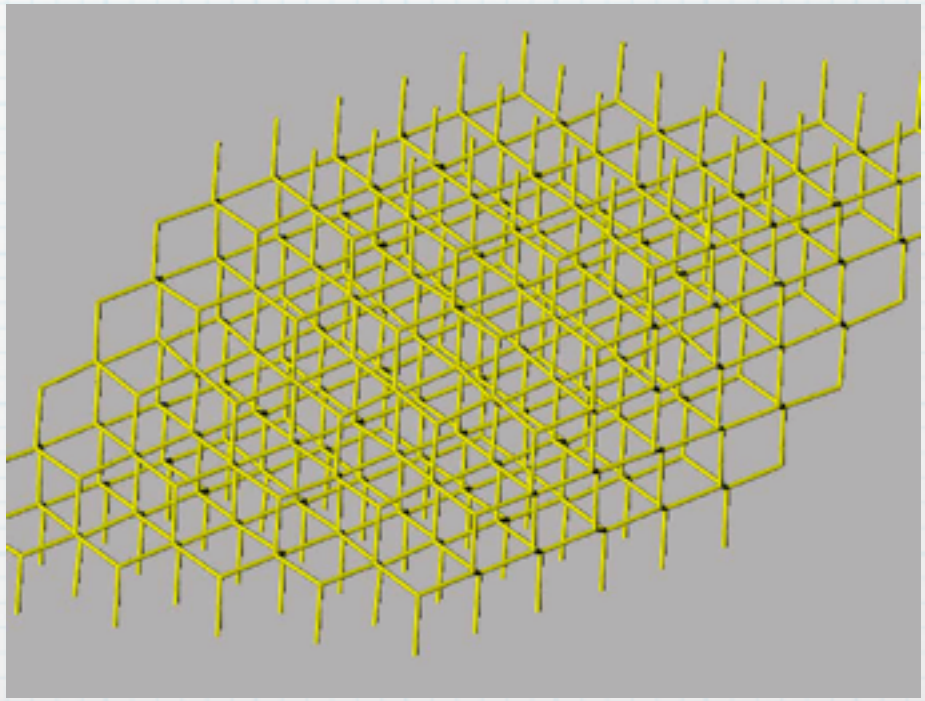
Metadynamics is able to explore **all the accessible metastable states**, without requiring any over-pressurization

Collective variables: the 3 vectors that define the simulation box

$$\mathbf{h} = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3]$$

Evolution of the cell parameters during the metadynamics run





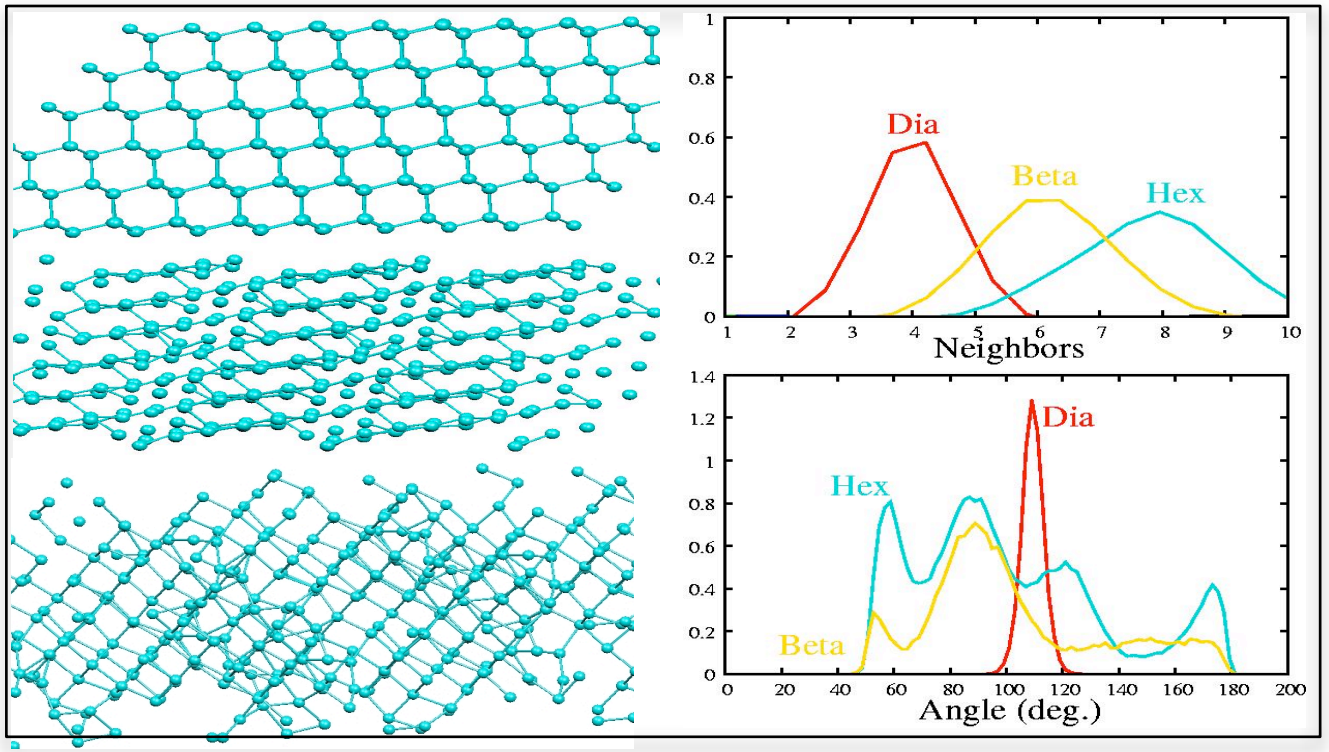
Silicon: Diamond, B-tin, Simple Hexagonal

Structural analysis by radial correlation function

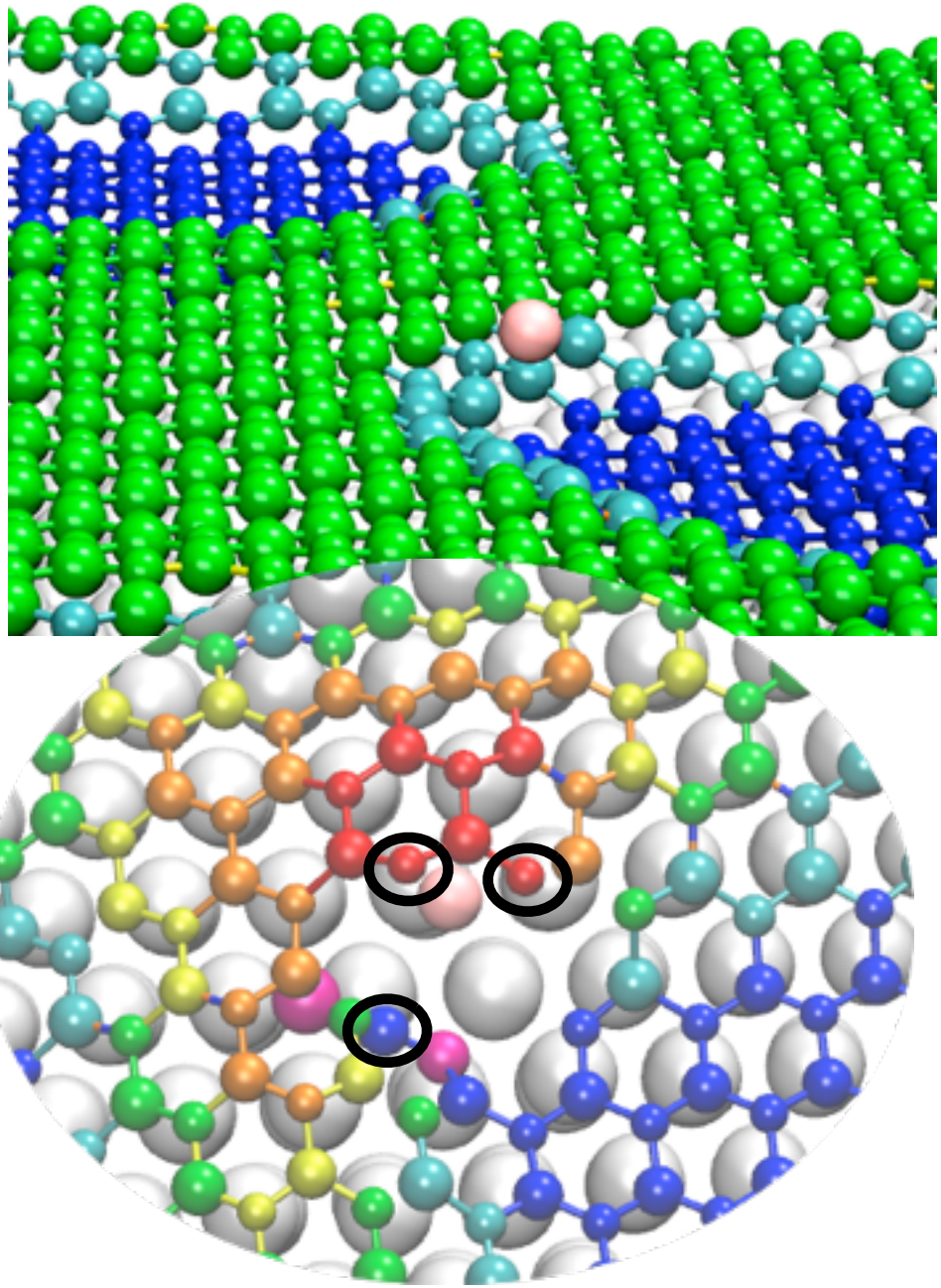
Dia

SHex

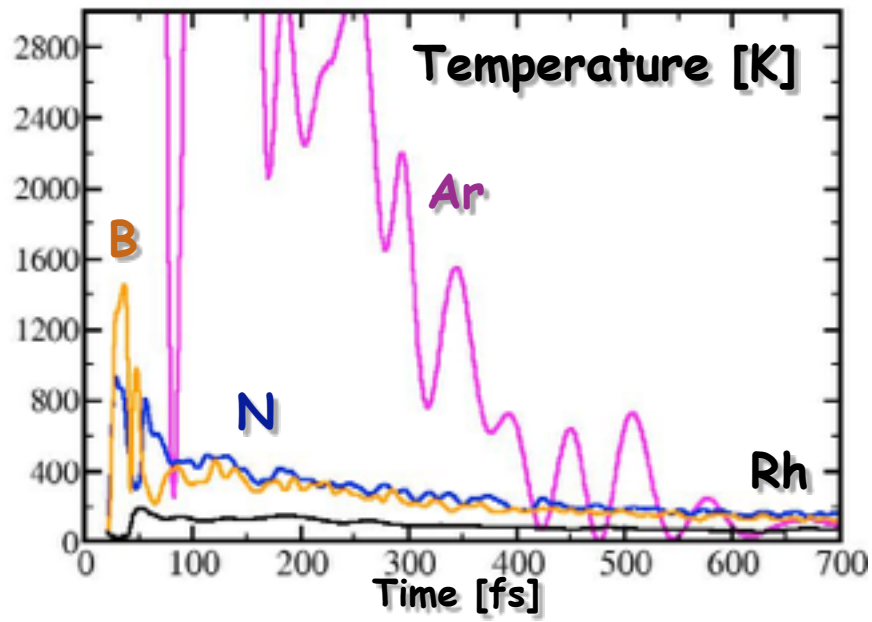
βtin

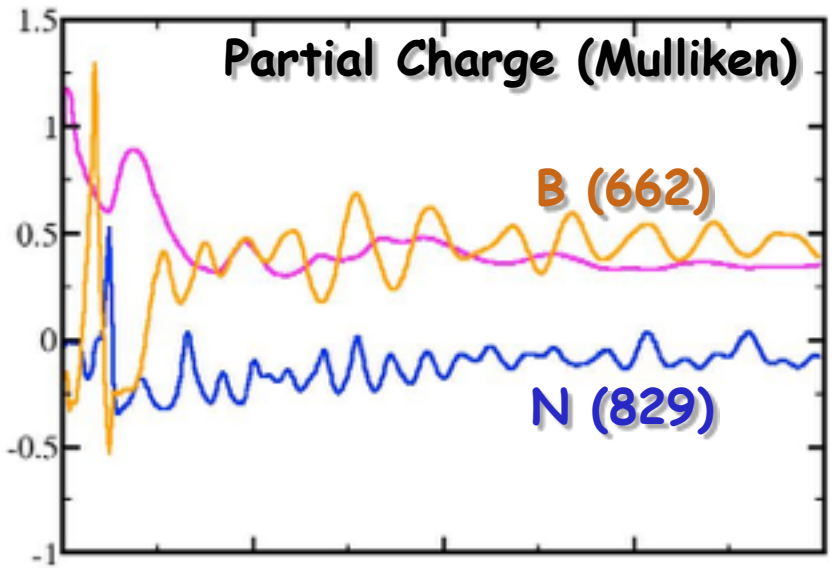


Implantation Process



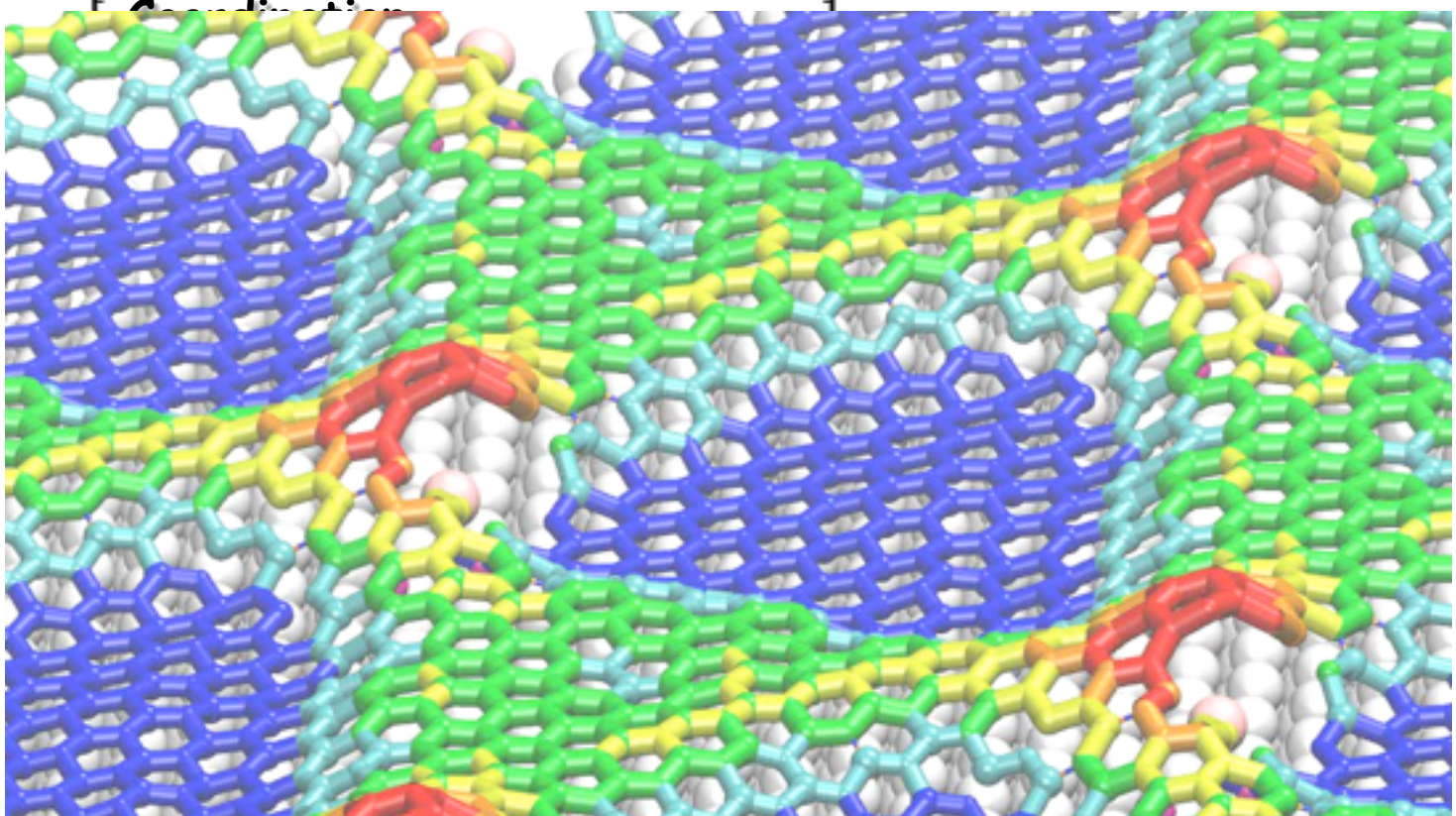
Ar+ at 50 eV striking the rim

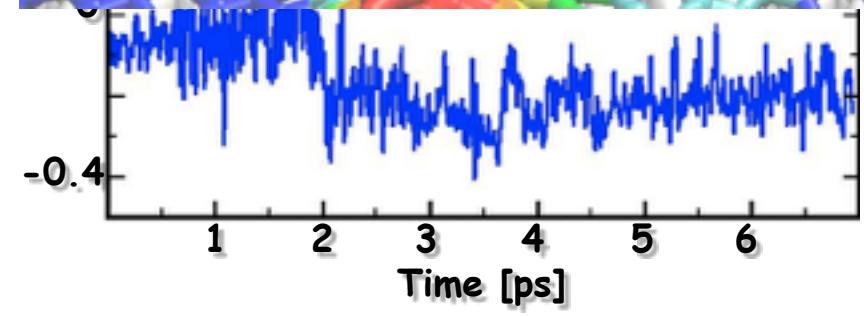


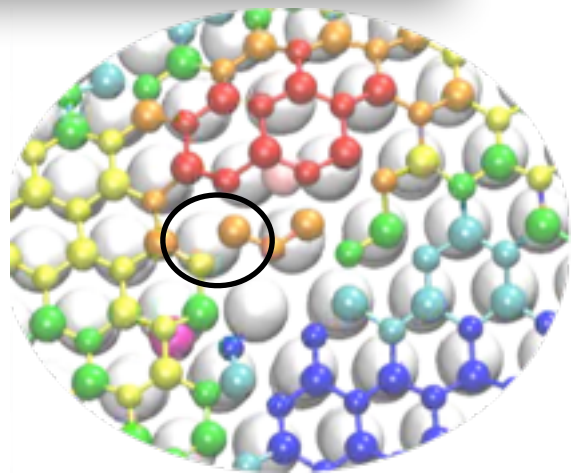


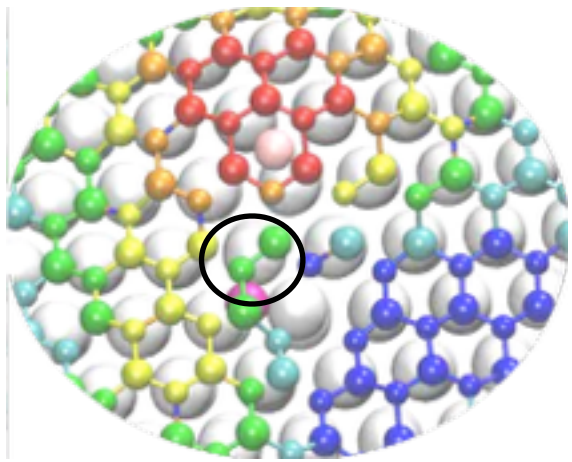
hBN Reconstruction

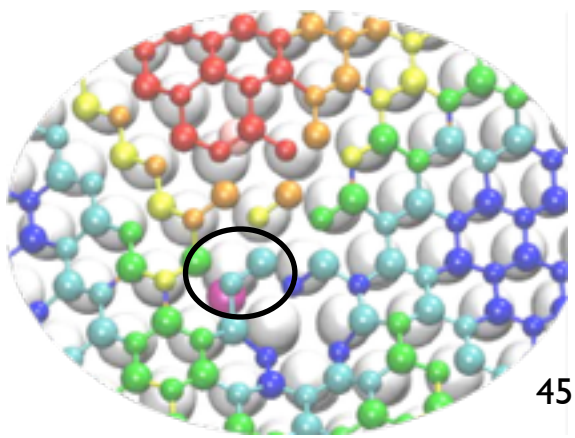
Metadynamics to accelerate the simulation of the healing process



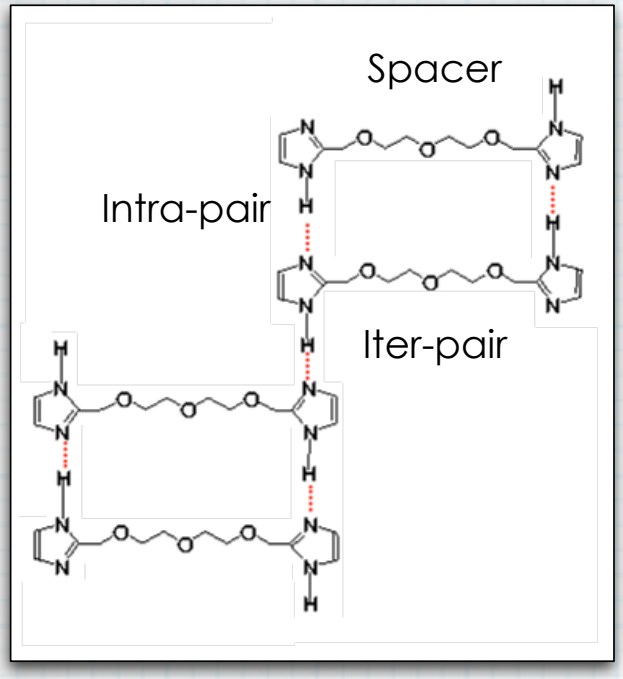


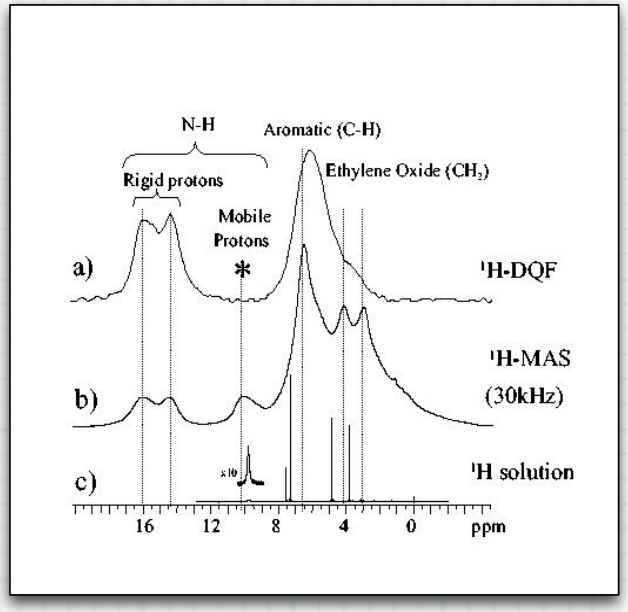






Imidazole-2EO: proton conductor

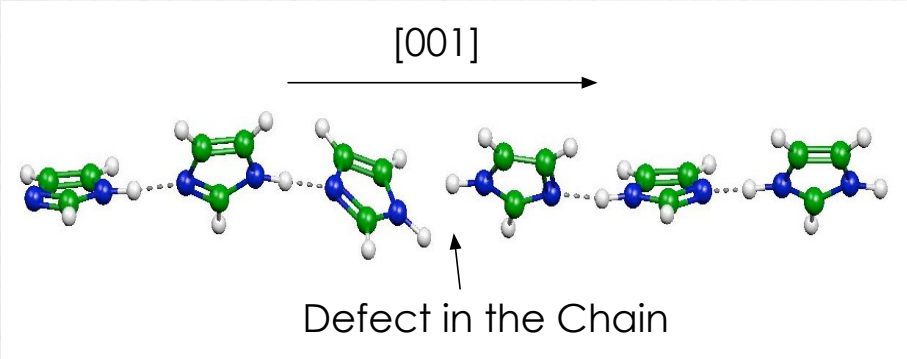




Good conductivity in doped samples

Dynamical disorder of local hydrogen-bond patterns

Fast microscopic rearrangements



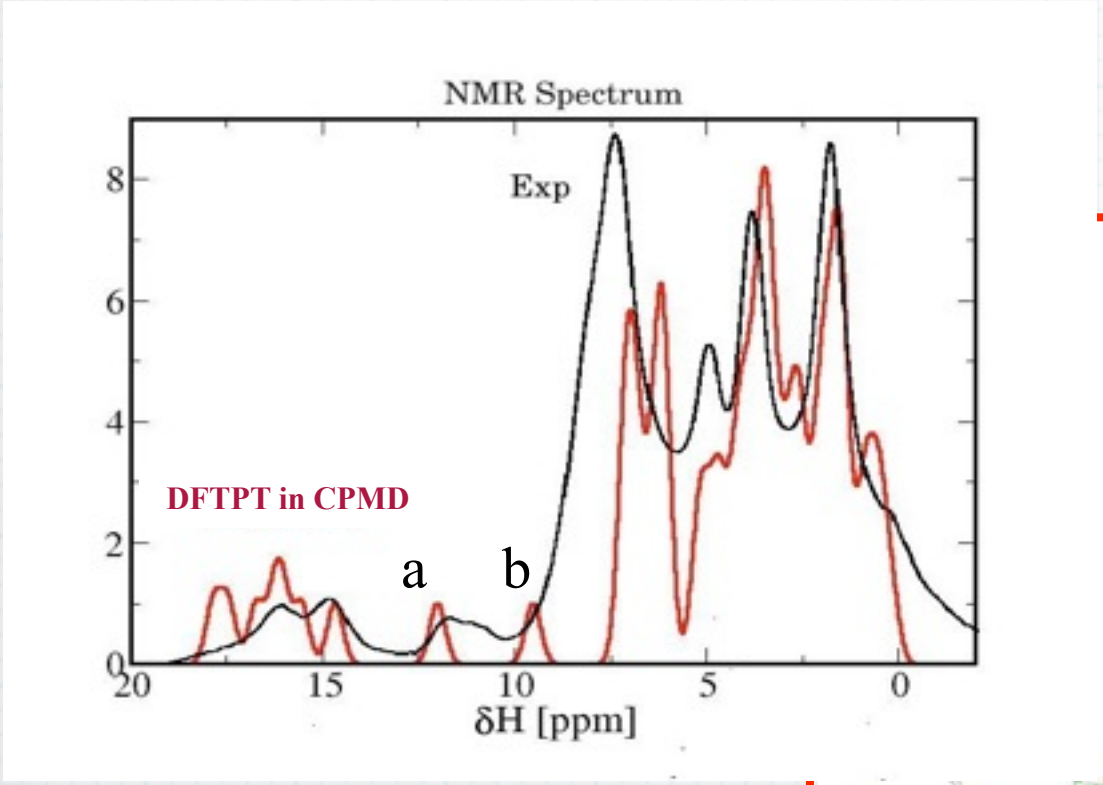
Generalized relative displacement

$$D^{[001]}_{H,N}(\{\mathbf{R}_I\}) = \frac{1}{N_H}\sum_{i\in H}\mathbf{d}_i\cdot\hat{\mathbf{v}}_{[001]} - \frac{1}{N_N}\sum_{j\in N}\mathbf{d}_j\cdot\hat{\mathbf{v}}_{[001]}$$

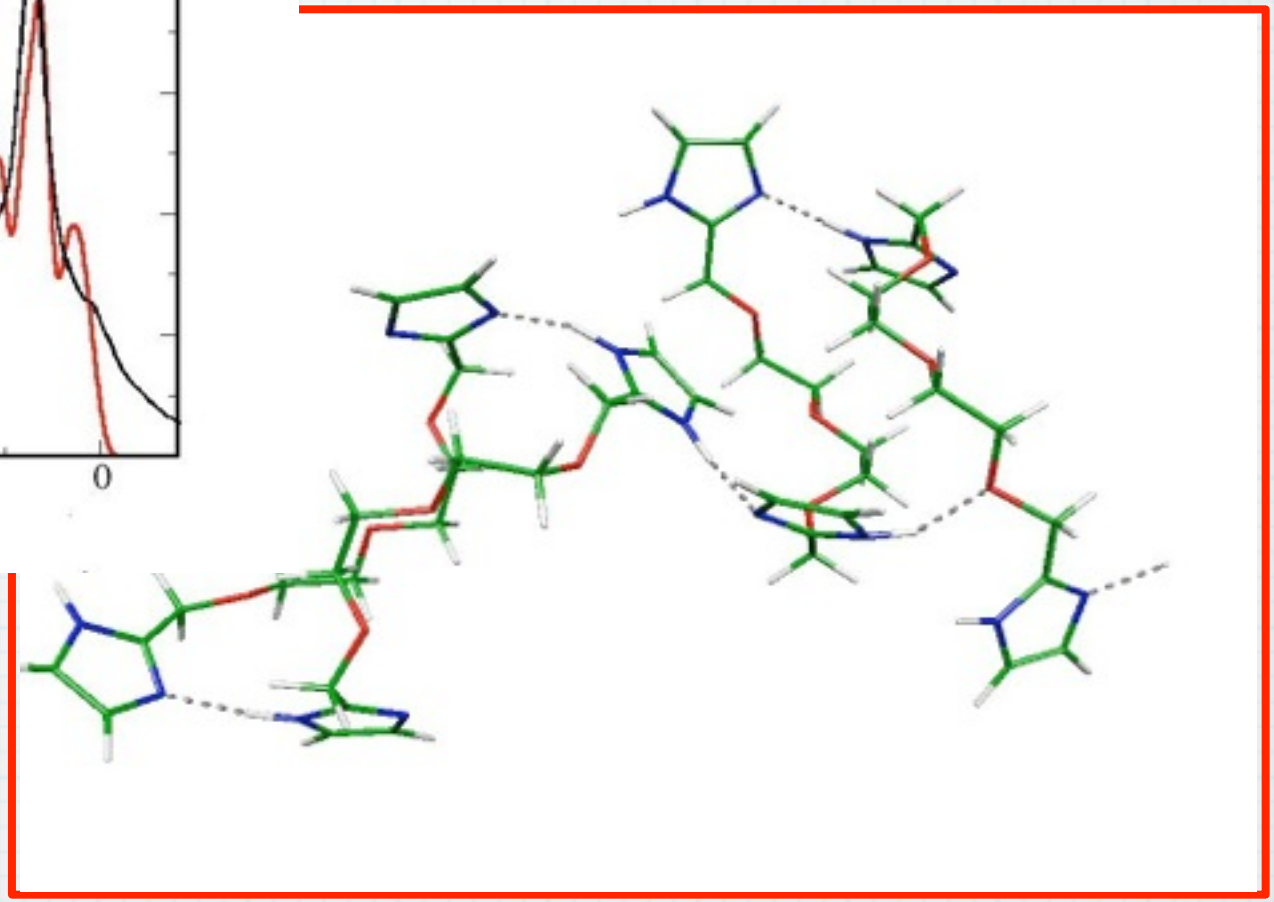
Coordination numbers and/or rotation angles

G.R. Gowers, J. Phys. Chem. B, **106**, 9322 (2002)

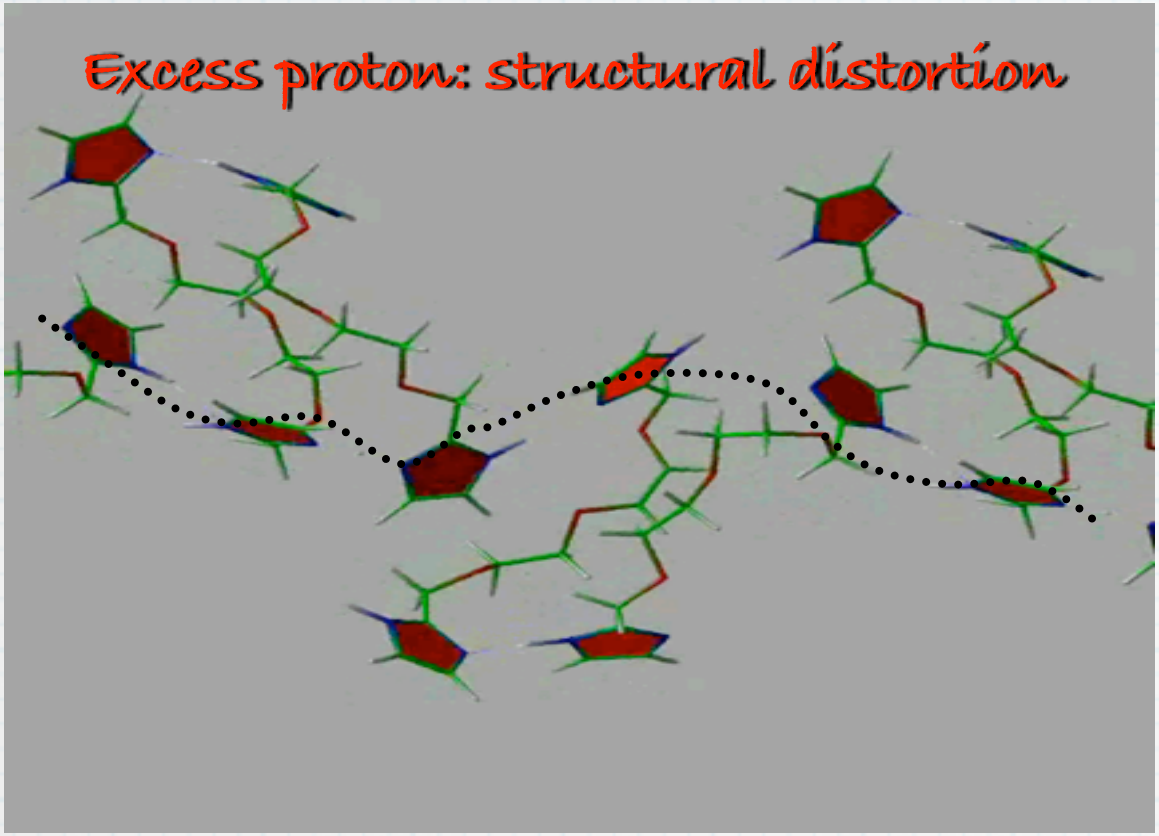
Excess-proton



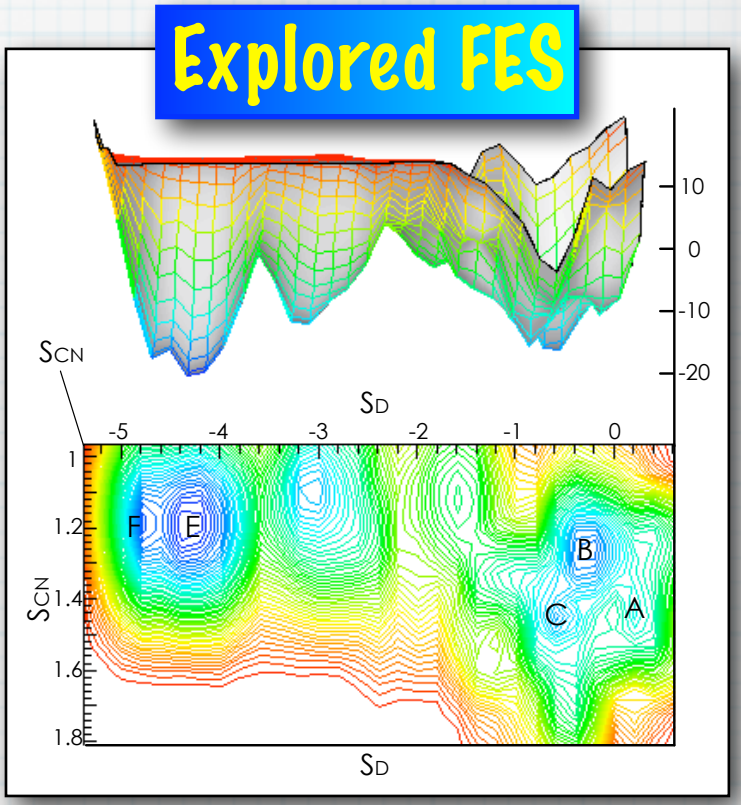
H-Doped Optimized Structure



Structure Diffusion



- Fluctuating hydrogen-bond network: structure diffusion
- Mobility induced by excess protons
- Monitoring defect diffusion by activating 5 Collective Variables



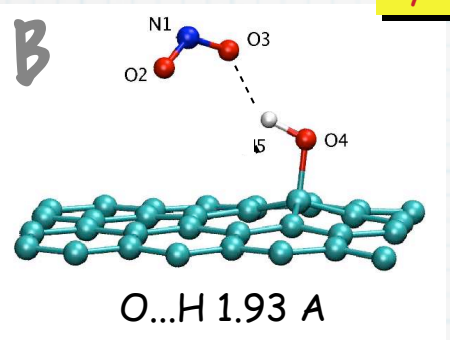
- Facile intermolecular proton hopping
- Proton transferred over 4 molecules
- Complex pathway with several intermediates
- Rearrangement of the h-bond network
- Stabilizing role of O...H interactions

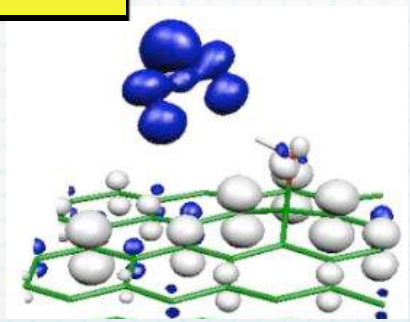
M Iannuzzi, Parrinello, PRL, **93**, 025901(2004);
M. Iannuzzi, JCP, **124**, 124710 (2006)

On Pristine Graphite

Three CN as CVs : N vs O, O vs C, N vs C

hydroxilation

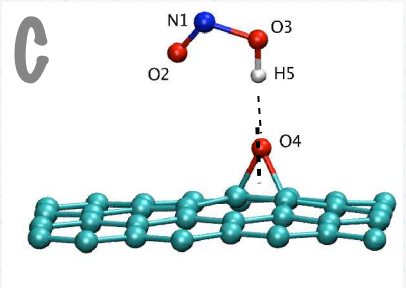




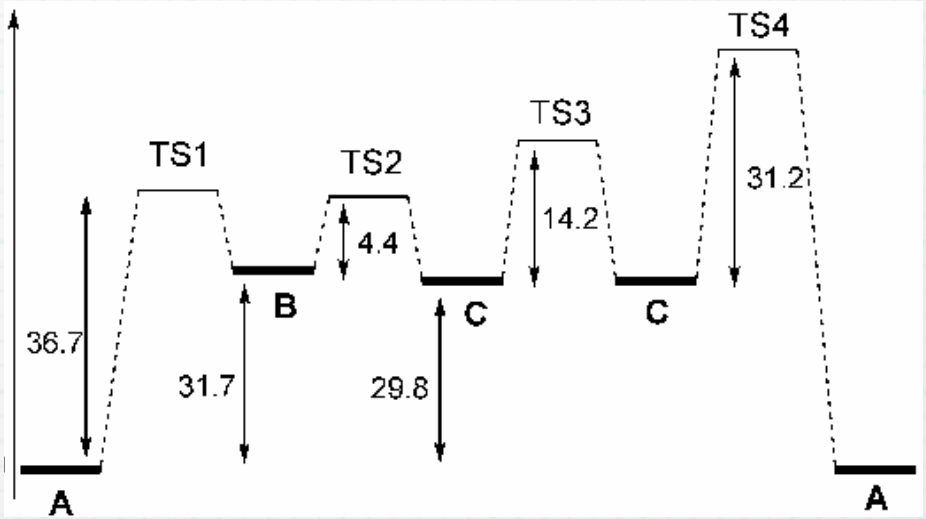
The first event is OH dissociation and hydroxilation of the surface (~35 kcal/mol), with formation of bi-radical system. System stabilized by the formation of H-bond and delocalization of the unpaired electron transferred to the surface. sp3 hybridization of C'

H transfer

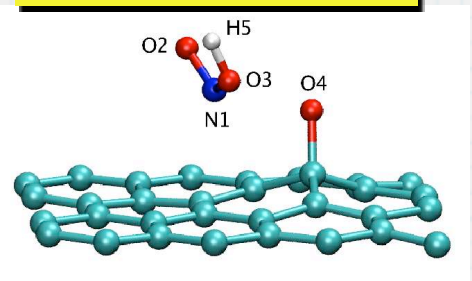
reactive NO_2 system is slightly stabilized; closed shell, only 1.9 kcal/mol more stable than open shell.
Epoxide structure migrates over the surface



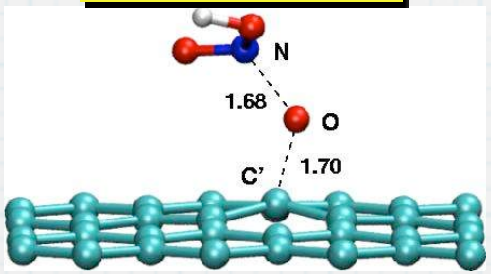
MEP energy profile



TS Epoxide migration



TS back transfer



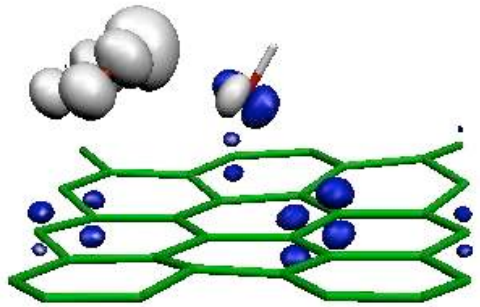
The direct epoxidation is a much less probable process

Graphite's role

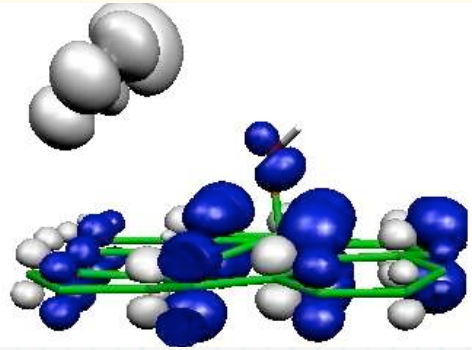
The OH-graphite interaction starts before the dissociation is complete

Early interaction between the incipient radical and the surface.
N-OH 2.1 A
OH-graphite 2.5 A
interaction with π el.

MD snapshot during the transition



MD snapshot at dissociation completed



As soon as the OH radical is adsorbed on the surface, its unpaired electron is delocalized among the atoms of the graphite layer.

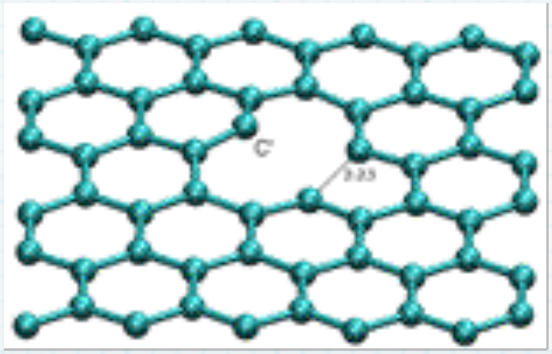
The dissociation barrier is 12 kcal/mol lower than in gas phase

Graphite catalyzes the dissociation of nitric acid
if the partial pressure is such as to guarantee a high probability of having HNO_3 in the vicinity of the surface.

On Defective Graphite : vacancy

Simplest type of defects, predominant in carbon layers

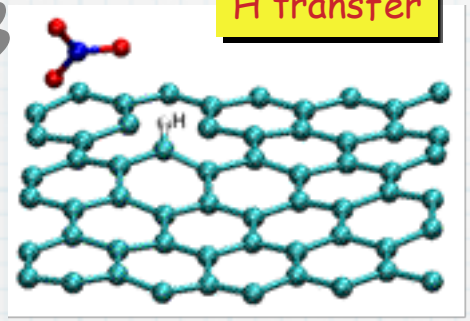
σ type sp2 dangling bonds at the unsaturated C'



4 CN as CVs:
N->O; O->C;
N->C;H->C

B

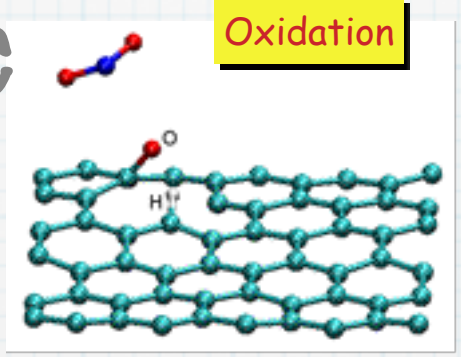
H transfer



unpaired H el. saturates σ type dangling,
π spin density delocalized

C

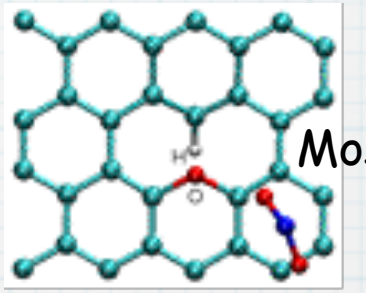
Oxidation



C=O, unpaired π el. engaged
no π spin density

D

Full saturation



Most stable

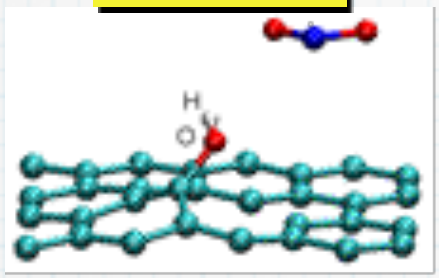
Ether group

Important effects of translational entropy on the free energy barriers

Alternative Paths

Hydroxilation

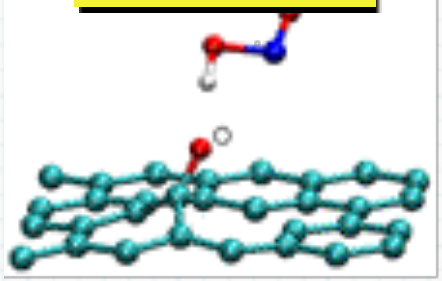
E



C-O σ type dangling,
π spin density delocalized

Direct oxidation

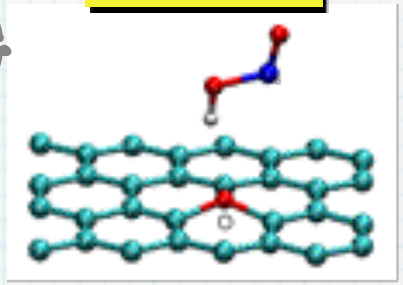
F



weak hydrogen bond
(1.88A); closed shell

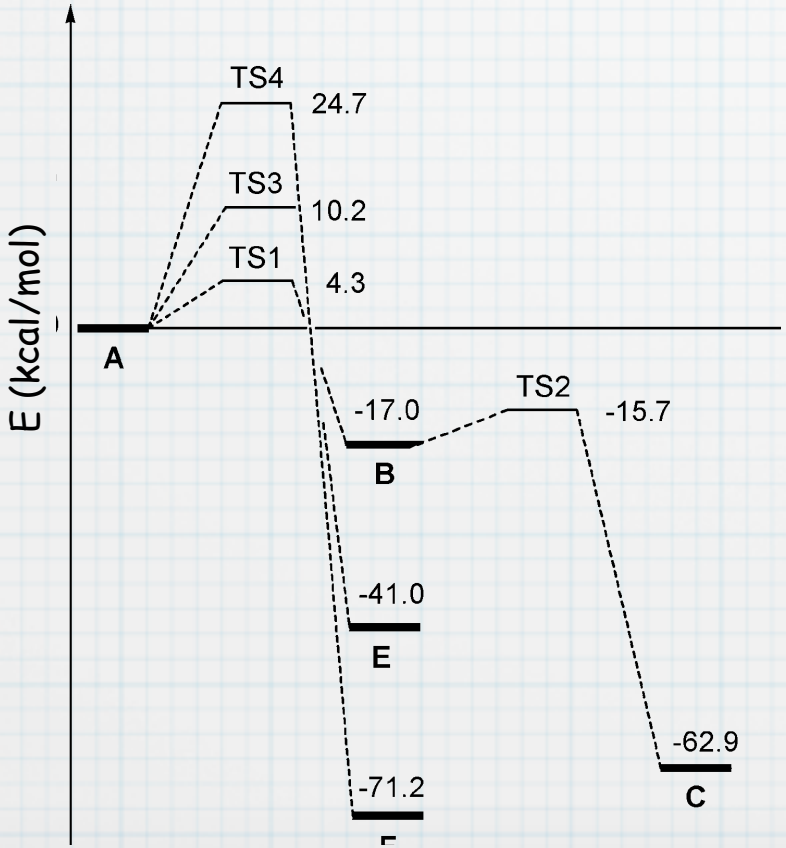
Replacement

G



sp3 character, not
symmetric, electropositive
O, no H-bond

the direct oxidation leads
to more stable
configurations but requires
higher energy barriers



The estimated activation energies are much smaller than what obtained in gas phase as well as on pristine graphite

MEP: ABC