

Statistical thermodynamics for MD and MC simulations

knowing 2 atoms and wishing to know 10^{23} of them

Marcus Elstner and Tomáš Kubař

2018, May 18

Thermodynamic properties of molecular systems

statistical mechanics: the way from the properties of particles
to the thermodynamic properties of ensembles
via the **partition function**

Thermodynamic properties of molecular systems

statistical mechanics: the way from the properties of particles
to the thermodynamic properties of ensembles
via the **partition function**

- how is the thermodynamic equilibrium characterized?
- which quantities are of interest?
- how may these be derived from the partition function?

Thermodynamic properties of molecular systems

statistical mechanics: the way from the properties of particles
to the thermodynamic properties of ensembles
via the **partition function**

- how is the thermodynamic equilibrium characterized?
- which quantities are of interest?
- how may these be derived from the partition function?
- how to derive the ensemble partition function
from the partition function of a single molecule?
- how is partition function connected to **phase-space density**?

Thermodynamic properties of molecular systems

statistical mechanics: the way from the properties of particles
to the thermodynamic properties of ensembles
via the **partition function**

- how is the thermodynamic equilibrium characterized?
- which quantities are of interest?
- how may these be derived from the partition function?
- how to derive the ensemble partition function
from the partition function of a single molecule?
- how is partition function connected to **phase-space density**?
- MD simulation provides an alternative way
to thermodynamic quantities
- it is difficult to obtain free energies
from normal simulations

Equilibrium and spontaneous processes

classical thermodynamics →

thermodynamic equilibrium and **spontaneous process**

says which quantities are maximized/minimized in equilibrium
and show a definite change during spontaneous processes

Equilibrium and spontaneous processes

classical thermodynamics →

thermodynamic equilibrium and **spontaneous process**

says which quantities are maximized/minimized in equilibrium
and show a definite change during spontaneous processes

microcanonical ensemble:

equilibrium reached if **entropy** S is maximized

process is spontaneous if entropy increases: $\Delta S > 0$

Equilibrium and spontaneous processes

classical thermodynamics →

thermodynamic equilibrium and **spontaneous process**

says which quantities are maximized/minimized in equilibrium
and show a definite change during spontaneous processes

microcanonical ensemble:

equilibrium reached if **entropy** S is maximized

process is spontaneous if entropy increases: $\Delta S > 0$

canonical ensemble – more complex because we need

to consider system of interest **together** with surroundings
(to identify equilibrium and spontaneity)

calculation on the supersystem impossible – an alternative needed

Free energy / enthalpy – fundamental property

How to keep our molecular system and drop the surroundings?
introduce a new thermodynamic function:

- Helmholtz free energy F in canonical NVT ensemble
- Gibbs free energy/enthalpy G in NPT ensemble

$$F = U - TS$$

$$G = H - TS = U + pV - TS$$

Free energy / enthalpy – fundamental property

How to keep our molecular system and drop the surroundings?
introduce a new thermodynamic function:

- Helmholtz free energy F in canonical NVT ensemble
- Gibbs free energy/enthalpy G in NPT ensemble

$$F = U - TS$$

$$G = H - TS = U + pV - TS$$

- $F = F(T, V)$ or $G = G(T, P)$ – depends
on experimentally controllable variables T and V/P

Free energy / enthalpy – fundamental property

How to keep our molecular system and drop the surroundings?
introduce a new thermodynamic function:

- Helmholtz free energy F in canonical NVT ensemble
- Gibbs free energy/enthalpy G in NPT ensemble

$$F = U - TS$$

$$G = H - TS = U + pV - TS$$

- $F = F(T, V)$ or $G = G(T, P)$ – depends
on experimentally controllable variables T and V/P
- an equilibrium state of the system
 - minimum of F or G rather than of U
 - equivalent to maximization of entropy of universe

Free energy / enthalpy – fundamental property

How to keep our molecular system and drop the surroundings?
introduce a new thermodynamic function:

- **Helmholtz free energy** F in canonical NVT ensemble
- **Gibbs free energy/enthalpy** G in NPT ensemble

$$F = U - TS$$

$$G = H - TS = U + pV - TS$$

- $F = F(T, V)$ or $G = G(T, P)$ – depends
on experimentally controllable variables T and V/P
- an equilibrium state of the system
 - minimum of F or G rather than of U
 - equivalent to maximization of entropy of universe
- F / G **decreases** in the course of spontaneous process
 - holy grail of MD simulation

Microscopic definition of entropy

W – number of ways ('microstates')
in which a certain 'macrostate' may be realized
(a certain occupation of energy levels of the system)

microscopic entropy:

$$S = k_B \cdot \ln W$$

(k_B – universal Boltzmann constant)

Microscopic definition of entropy

Ludwig Boltzmann – Austrian physicist



Microscopic definition of entropy

- S tells us something about the travel of the system through the configuration (phase) space
- equivalently, S can be related to the **order** in the system

Microscopic definition of entropy

- S tells us something about the travel of the system through the configuration (phase) space
- equivalently, S can be related to the **order** in the system
- low entropy – few states are occupied
 - only a small part of configuration space accessible
 - ordered system
- high entropy – many states are visited
 - extended regions of the configuration space covered
 - less ordered system

Microscopic definition of entropy

- S tells us something about the travel of the system through the configuration (phase) space
- equivalently, S can be related to the **order** in the system
- low entropy – few states are occupied
 - only a small part of configuration space accessible
 - ordered system
- high entropy – many states are visited
 - extended regions of the configuration space covered
 - less ordered system

Example – pile of books on a desk

Jan Černý (Charles University in Prague, Dept Cellular Biology):

anthropy – “entropy of human origin”

Introduction

“...series of molecular structures generated by an MD simulation with the Berendsen thermostat does not represent the correct canonical ensemble” – what does this mean?

The phase space may be sampled (walked through) in various ways
– just what is the correct way?

Closed system – canonical ensemble

system in **thermal contact** with the surroundings

- **temperature** rather than energy remains constant
- **Boltzmann distribution** of p_i applies:

$$p_i = \frac{\exp[-\beta \cdot \varepsilon_i]}{Q}$$

Closed system – canonical ensemble

system in **thermal contact** with the surroundings

- **temperature** rather than energy remains constant
- **Boltzmann distribution** of p_i applies:

$$p_i = \frac{\exp[-\beta \cdot \varepsilon_i]}{Q}$$

$$Q = \sum_j \exp[-\beta \cdot \varepsilon_j]$$

Q – **canonical partition function** (Zustandssumme)

Closed system – canonical ensemble

system in **thermal contact** with the surroundings

- **temperature** rather than energy remains constant
- **Boltzmann distribution** of p_i applies:

$$p_i = \frac{\exp[-\beta \cdot \varepsilon_i]}{Q}$$

$$Q = \sum_j \exp[-\beta \cdot \varepsilon_j]$$

Q – **canonical partition function** (Zustandssumme)

derive the meaning of β – fall back to basic thermodynamics. . . :

$$\beta = \frac{1}{k_B T}$$

Canonical partition function

Q – seems to be purely abstract. . .

but – to characterize the thermodynamics of a system

Q is completely sufficient,

all thermodynamic observables follow as functions of Q

how to obtain Q ?

– there will be 2 ways, depending on the system studied

Canonical partition function

Q – seems to be purely abstract. . .

but – to characterize the thermodynamics of a system

Q is completely sufficient,

all thermodynamic observables follow as functions of Q

how to obtain Q ?

– there will be 2 ways, depending on the system studied

partition function connects the **microscopic** and **macroscopic** world

Using the partition function

we can get all thermodynamic functions from Q :

$$U = \langle E \rangle = k_B T^2 \frac{\partial \ln Q}{\partial T}$$

$$S = k_B T \cdot \frac{\partial \ln Q}{\partial T} + k_B \cdot \ln Q$$

$$F = -k_B T \cdot \ln Q$$

$$P = k_B T \cdot \left(\frac{\partial \ln Q}{\partial V} \right)_T \quad (\text{equation of state})$$

$$H = U + pV$$

$$G = F + pV = H - TS$$

Using the partition function

we can get all thermodynamic functions from Q :

$$U = \langle E \rangle = k_B T^2 \frac{\partial \ln Q}{\partial T}$$

$$S = k_B T \cdot \frac{\partial \ln Q}{\partial T} + k_B \cdot \ln Q$$

$$F = -k_B T \cdot \ln Q$$

$$P = k_B T \cdot \left(\frac{\partial \ln Q}{\partial V} \right)_T \quad (\text{equation of state})$$

$$H = U + pV$$

$$G = F + pV = H - TS$$

Michal Otyepka (Palacký University Olomouc, Dept Physical Chemistry):

“Just grab the partition function at the tail, and then you have everything!”

2 ways to thermodynamic properties

- simple molecules with 1 or few minima of energy
 - calculate the partition function (trans+vib+rot)
and employ some approximations (IG+HO+RR)
 - derive properties from Q

Discrete systems

system with **discrete energy levels** E_i – partition function:

$$Q = \sum_i \exp[-\beta E_i]$$

Boltzmann distribution function: (prob. of system in state E_i)

$$p_i = \frac{1}{Q} \exp[-\beta E_i]$$

Discrete systems

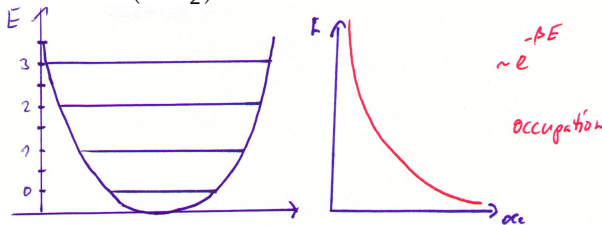
system with **discrete energy levels** E_i – partition function:

$$Q = \sum_i \exp[-\beta E_i]$$

Boltzmann distribution function: (prob. of system in state E_i)

$$p_i = \frac{1}{Q} \exp[-\beta E_i]$$

example – HO: $E_i = (i + \frac{1}{2}) \cdot \hbar\omega$



Using the partition function of 1 molecule

partition function of a large system – simplifications possible:

- system of n **identical and indistinguishable** particles (gas):

$$Q = \frac{q^n}{n!}$$

Using the partition function of 1 molecule

partition function of a large system – simplifications possible:

- system of n **identical and indistinguishable** particles (gas):

$$Q = \frac{q^n}{n!}$$

- necessary effort reduced greatly!
 - get the **molecular partition function** q
(calculate 1 molecule, or 2)
 - obtain the ensemble partition function Q
and all thermodynamic quantities

Simple molecules

... with 1 or few well characterized minima
for a certain minimum – consider contributions to energy:

$$E = E^{\text{el}} + E^{\text{trans}} + E^{\text{rot}} + E^{\text{vib}}$$

Simple molecules

... with 1 or few well characterized minima
for a certain minimum – consider contributions to energy:

$$E = E^{\text{el}} + E^{\text{trans}} + E^{\text{rot}} + E^{\text{vib}}$$

partition function follows as

$$\begin{aligned} Q &= \exp \left[-\beta \left(E^{\text{el}} + E^{\text{trans}} + E^{\text{rot}} + E^{\text{vib}} \right) \right] = \\ &= \exp[-\beta E^{\text{el}}] \cdot \exp[-\beta E^{\text{trans}}] \cdot \exp[-\beta E^{\text{rot}}] \cdot \exp[-\beta E^{\text{vib}}] = \\ &= Q^{\text{el}} \cdot Q^{\text{trans}} \cdot Q^{\text{rot}} \cdot Q^{\text{vib}} \end{aligned}$$

or

$$\ln Q = \ln Q^{\text{el}} + \ln Q^{\text{trans}} + \ln Q^{\text{rot}} + \ln Q^{\text{vib}}$$

Electronic partition function

usually: quite high excitation energy

→ electronic ground state only populated:

$$E^{\text{el}}(0) = 0 \quad \text{arbitrarily}$$

electronic partition function:

$$Q^{\text{el}} = \exp[-\beta E^{\text{el}}(0)] + \exp[-\beta E^{\text{el}}(1)] + \dots \approx 1 + 0 + \dots = 1$$

Electronic partition function

usually: quite high excitation energy

→ electronic ground state only populated:

$$E^{\text{el}}(0) = 0 \quad \text{arbitrarily}$$

electronic partition function:

$$Q^{\text{el}} = \exp[-\beta E^{\text{el}}(0)] + \exp[-\beta E^{\text{el}}(1)] + \dots \approx 1 + 0 + \dots = 1$$

so this may be neglected ☺

Translational partition function

calculated for quantum-mechanical particle (mass m) in a 3D box:
energy levels:

$$E_{n_x, n_y, n_z} = \frac{h^2}{8m} \left(\frac{n_x^2}{L_x^2} + \frac{n_y^2}{L_y^2} + \frac{n_z^2}{L_z^2} \right)$$

quantum numbers $n_i = 1, 2, \dots$

Translational partition function

calculated for quantum-mechanical particle (mass m) in a 3D box:
energy levels:

$$E_{n_x, n_y, n_z} = \frac{h^2}{8m} \left(\frac{n_x^2}{L_x^2} + \frac{n_y^2}{L_y^2} + \frac{n_z^2}{L_z^2} \right)$$

quantum numbers $n_i = 1, 2, \dots$

partition function:

$$Q^{\text{trans}} = \left(\frac{2\pi m k_B T}{h^2} \right)^{\frac{3}{2}} \cdot V$$

Rotational partition function

calculated for a rigid rotor (moments of inertia I_x):
energy levels:

$$E_J = B \cdot J(J + 1)$$

quantum number $J = 0, 1, 2, \dots$, degeneracy of levels $2J + 1$

rotational constant $B = \frac{h^2}{8\pi^2 I}$ (I – moment of inertia)

Rotational partition function

calculated for a rigid rotor (moments of inertia I_x):
energy levels:

$$E_J = B \cdot J(J + 1)$$

quantum number $J = 0, 1, 2, \dots$, degeneracy of levels $2J + 1$

rotational constant $B = \frac{h^2}{8\pi^2 I}$ (I – moment of inertia)

$$Q^{\text{rot}} = \sum_{J=0}^{\infty} (2J + 1) \exp \left[-J(J + 1) \cdot \frac{B}{k_B T} \right]$$

Rotational partition function

calculated for a rigid rotor (moments of inertia I_x):
energy levels:

$$E_J = B \cdot J(J + 1)$$

quantum number $J = 0, 1, 2, \dots$, degeneracy of levels $2J + 1$

rotational constant $B = \frac{h^2}{8\pi^2 I}$ (I – moment of inertia)

$$Q^{\text{rot}} = \sum_{J=0}^{\infty} (2J + 1) \exp \left[-J(J + 1) \cdot \frac{B}{k_B T} \right]$$

for asymmetric top with rotational constants B_x , B_y , B_z :

$$Q^{\text{rot}} = \sqrt{\frac{\pi (k_B T)^3}{B_x B_y B_z}}$$

Vibrational partition function

calculated with harmonic vibrational frequencies ω_k of the molecule
(computation of Hessian in the minimum of potential energy)
– each vibrational mode k is one harmonic oscillator

Vibrational partition function

calculated with harmonic vibrational frequencies ω_k of the molecule
(computation of Hessian in the minimum of potential energy)
– each vibrational mode k is one harmonic oscillator

energy levels: $E_k^m = \left(m + \frac{1}{2}\right) \cdot \hbar\omega_k$

where $E_k^0 = \frac{1}{2}\hbar\omega_k$ is **zero point vibrational energy**

Vibrational partition function

calculated with harmonic vibrational frequencies ω_k of the molecule
 (computation of Hessian in the minimum of potential energy)
 – each vibrational mode k is one harmonic oscillator

energy levels: $E_k^m = \left(m + \frac{1}{2}\right) \cdot \hbar\omega_k$

where $E_k^0 = \frac{1}{2}\hbar\omega_k$ is **zero point vibrational energy**

partition function (using $\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$):

$$Q_k^{\text{vib}} = \sum_{m=0}^{\infty} \exp \left[-\beta \left(m + \frac{1}{2} \right) \hbar\omega_k \right] = \frac{\exp \left[-\frac{1}{2}\beta\hbar\omega_k \right]}{1 - \exp \left[-\beta\hbar\omega_k \right]}$$

Vibrational partition function

calculated with harmonic vibrational frequencies ω_k of the molecule
 (computation of Hessian in the minimum of potential energy)
 – each vibrational mode k is one harmonic oscillator

energy levels: $E_k^m = \left(m + \frac{1}{2}\right) \cdot \hbar\omega_k$

where $E_k^0 = \frac{1}{2}\hbar\omega_k$ is **zero point vibrational energy**

partition function (using $\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$):

$$Q_k^{\text{vib}} = \sum_{m=0}^{\infty} \exp \left[-\beta \left(m + \frac{1}{2} \right) \hbar\omega_k \right] = \frac{\exp \left[-\frac{1}{2}\beta\hbar\omega_k \right]}{1 - \exp \left[-\beta\hbar\omega_k \right]}$$

each molecule: $N - 6$ vibrational modes $\rightarrow N - 6$ HOs

example – H₂O: 3 modes (2 stretches, 1 bend)

Thermodynamic properties

for enthalpy – pV is needed – simple for IG:

$$pV = Nk_{\text{B}}T$$

then, enthalpy and Gibbs free energy follow:

$$H = U + pV = U + Nk_{\text{B}}T$$

$$G = F + pV = F + Nk_{\text{B}}T$$

Thermodynamic properties

for enthalpy – pV is needed – simple for IG:

$$pV = Nk_B T$$

then, enthalpy and Gibbs free energy follow:

$$H = U + pV = U + Nk_B T$$

$$G = F + pV = F + Nk_B T$$

thermal contributions – calculated by default

with many QCh and MD programs

whenever vibrational analysis is requested

reason – vibrational frequencies are computationally costly

while the thermodynamics is done ‘for free’

Example – vibrational contributions

$$\ln Q_k = -\frac{1}{2}\beta\hbar\omega_k - \ln[1 - \exp[-\beta\hbar\omega_k]]$$
$$U_k = -\frac{\partial \ln Q_k}{\partial \beta} = \hbar\omega_k \left(\frac{1}{2} + \frac{1}{\exp[\beta\hbar\omega_k] - 1} \right)$$

Example – vibrational contributions

$$\ln Q_k = -\frac{1}{2}\beta\hbar\omega_k - \ln[1 - \exp[-\beta\hbar\omega_k]]$$

$$U_k = -\frac{\partial \ln Q_k}{\partial \beta} = \hbar\omega_k \left(\frac{1}{2} + \frac{1}{\exp[\beta\hbar\omega_k] - 1} \right)$$

consider this for all of $N - 6$ vibrational DOFs, dropping **ZPVE**:

$$U^{\text{vib}} = \sum_{k=1}^{N-6} \left(\frac{\hbar\omega_k}{\exp[\beta\hbar\omega_k] - 1} \right)$$

Example – vibrational contributions

$$\ln Q_k = -\frac{1}{2}\beta\hbar\omega_k - \ln[1 - \exp[-\beta\hbar\omega_k]]$$

$$U_k = -\frac{\partial \ln Q_k}{\partial \beta} = \hbar\omega_k \left(\frac{1}{2} + \frac{1}{\exp[\beta\hbar\omega_k] - 1} \right)$$

consider this for all of $N - 6$ vibrational DOFs, dropping **ZPVE**:

$$U^{\text{vib}} = \sum_{k=1}^{N-6} \left(\frac{\hbar\omega_k}{\exp[\beta\hbar\omega_k] - 1} \right)$$

$$F^{\text{vib}} = -k_B T \ln Q^{\text{vib}} = \sum_{k=1}^{N-6} k_B T \ln[1 - \exp[-\beta\hbar\omega_k]]$$

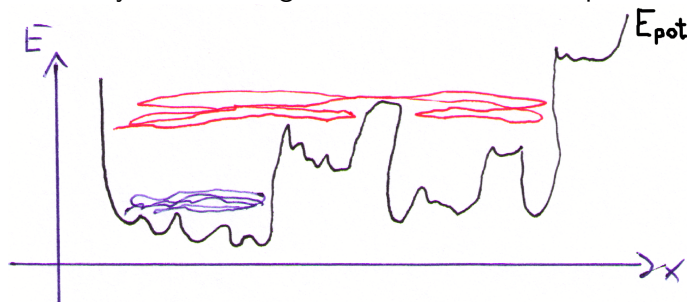
$$\frac{S^{\text{vib}}}{k_B} = \frac{U^{\text{vib}} - F^{\text{vib}}}{k_B T} = \sum_{k=1}^{N-6} \left(\frac{\beta\hbar\omega_k}{\exp[\beta\hbar\omega_k] - 1} - \ln[1 - \exp[-\beta\hbar\omega_k]] \right)$$

2 ways to thermodynamic properties

- simple molecules with 1 or few minima of energy
 - calculate the partition function (trans+vib+rot)
(probably employ approximations IG+HO+RR)
 - derive properties from Q
- flexible molecules, complex molecular systems
 - quantum mechanical energy levels cannot be calculated
 - a single minimum of energy not meaningful
 - do MD (MC) simulation instead, to sample phase space
 - evaluate time averages of thermodynamic quantities

Systems with continuous distribution of energy

dynamics of molecules – at different tot. energies or **temperatures**, differently extended regions of conformational space are sampled



complex energy landscape $E_{\text{pot}}(x)$

blue and red – trajectories at different total energies

– **different phase-space densities**

Continuous systems – canonical ensemble

- every point in phase space – a certain value of energy
composed of $E_{\text{pot}} = E_{\text{pot}}(\vec{r})$ (force field), $E_{\text{kin}} = E_{\text{kin}}(\vec{p})$
- continuous energy levels – infinitesimally narrow spacing

Continuous systems – canonical ensemble

- every point in phase space – a certain value of energy composed of $E_{\text{pot}} = E_{\text{pot}}(\vec{r})$ (force field), $E_{\text{kin}} = E_{\text{kin}}(\vec{p})$
- continuous energy levels – infinitesimally narrow spacing
- canonical **probability distribution function**
 - probability to find the system in state with E :

$$P(\vec{r}, \vec{p}) = \rho(\vec{r}, \vec{p}) = \frac{1}{Q} \cdot \exp \left[-\frac{E(\vec{r}, \vec{p})}{k_B T} \right]$$

Continuous systems – canonical ensemble

- every point in phase space – a certain value of energy composed of $E_{\text{pot}} = E_{\text{pot}}(\vec{r})$ (force field), $E_{\text{kin}} = E_{\text{kin}}(\vec{p})$
- continuous energy levels – infinitesimally narrow spacing
- canonical **probability distribution function**
 - probability to find the system in state with E :

$$P(\vec{r}, \vec{p}) = \rho(\vec{r}, \vec{p}) = \frac{1}{Q} \cdot \exp \left[-\frac{E(\vec{r}, \vec{p})}{k_B T} \right]$$

- perform MD simulation with a correct thermostat, or MC sim.
 - the goal is to obtain the right probability density ρ

Continuous systems – canonical ensemble

- every point in phase space – a certain value of energy composed of $E_{\text{pot}} = E_{\text{pot}}(\vec{r})$ (force field), $E_{\text{kin}} = E_{\text{kin}}(\vec{p})$
- continuous energy levels – infinitesimally narrow spacing
- canonical **probability distribution function**
 - probability to find the system in state with E :

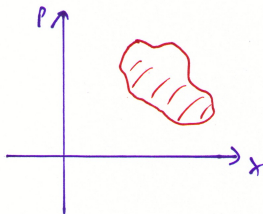
$$P(\vec{r}, \vec{p}) = \rho(\vec{r}, \vec{p}) = \frac{1}{Q} \cdot \exp \left[-\frac{E(\vec{r}, \vec{p})}{k_B T} \right]$$

- perform MD simulation with a correct thermostat, or MC sim.
 - the goal is to obtain the right probability density ρ
- **partition fction** Q – integral over phase space rather than sum

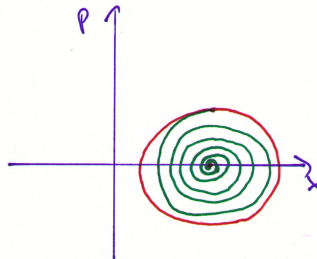
$$Q = \int \exp \left[-\frac{E(\vec{r}, \vec{p})}{k_B T} \right] d\vec{r} d\vec{p}$$

Thermodynamic quantities – sampling

$\rho(\vec{r}, \vec{p})$ – gives the probability of finding the system at (\vec{r}, \vec{p})
typically: system is sampling only a part of phase space ($P \neq 0$):



sampling in MD or MC



undamped and damped classical HO

Thermodynamic quantities – sampling

thermodynamic quantities – weighted averages:

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

Thermodynamic quantities – sampling

thermodynamic quantities – weighted averages:

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

why do MD? ... obtain the correct phase-space density ρ

– the density ρ is present in the the trajectory inherently

→ e.g. structure with high ρ occur more often

Thermodynamic quantities – sampling

thermodynamic quantities – weighted averages:

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

why do MD? ... obtain the correct phase-space density ρ

– the density ρ is present in the the trajectory inherently

→ e.g. structure with high ρ occur more often

thermodynamic quantities in MD simulation – time averages:

$$\langle A \rangle = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} A(t) dt$$

– valid only if the simulation has sampled the canonical ensemble

→ phase-space density is correct – **ergodic simulation**

Thermodynamic quantities – sampling

Thus, we have the following to do:

- perform MD simulation (with correct thermostat!)
 - trajectory in phase space
(simulation has ‘taken care’ of the phase-space density)
- get time averages of desired thermodynamic properties

$$\langle A \rangle = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} A(t) dt$$

Thermodynamic quantities – sampling

MD simulation serves to provide the phase-space density

$$\rho(\vec{r}, \vec{p}) = \frac{\exp[-\beta E(\vec{r}, \vec{p})]}{Q} \quad \vec{r} = \{r_1, \dots, r_{3N}\}, \vec{p} = \{p_1, \dots, p_{3N}\}$$

which is the probability of system occurring at point (\vec{r}, \vec{p})

Thermodynamic quantities – sampling

MD simulation serves to provide the phase-space density

$$\rho(\vec{r}, \vec{p}) = \frac{\exp[-\beta E(\vec{r}, \vec{p})]}{Q} \quad \vec{r} = \{r_1, \dots, r_{3N}\}, \vec{p} = \{p_1, \dots, p_{3N}\}$$

which is the probability of system occurring at point (\vec{r}, \vec{p})

How long an MD simulation can we perform?

1 ps \rightarrow 1,000 points in trajectory

100 ns \rightarrow 100M points – we cannot afford much more

Thermodynamic quantities – sampling

MD simulation serves to provide the phase-space density

$$\rho(\vec{r}, \vec{p}) = \frac{\exp[-\beta E(\vec{r}, \vec{p})]}{Q} \quad \vec{r} = \{r_1, \dots, r_{3N}\}, \vec{p} = \{p_1, \dots, p_{3N}\}$$

which is the probability of system occurring at point (\vec{r}, \vec{p})

How long an MD simulation can we perform?

1 ps \rightarrow 1,000 points in trajectory

100 ns \rightarrow 100M points – we cannot afford much more

example – we have 1,000 points

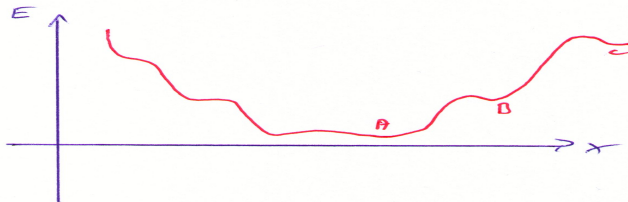
then, we have **not** sampled (\vec{r}, \vec{p}) for which $\rho(\vec{r}, \vec{p}) < \frac{1}{1000}$

\rightarrow points with high energy will be never reached!

(while low-energy region may be sampled well)

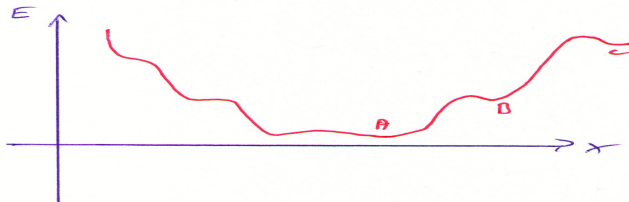
– actually, most of the points accumulate in low-energy regions,
so this is mmore serious!

Missing high-energy points in sampling



High-energy points B and C may be sampled badly

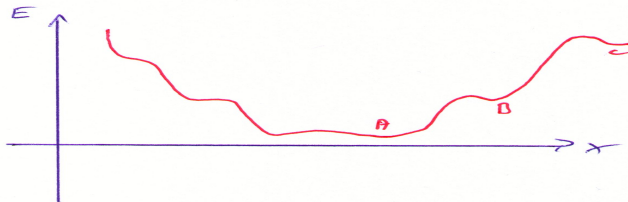
Missing high-energy points in sampling



High-energy points B and C may be sampled badly

- a typical problem in MD simulations of limited length
- the corresponding large energies are missing in averaging
- when does this matter?

Missing high-energy points in sampling



High-energy points B and C may be sampled badly

- a typical problem in MD simulations of limited length
- the corresponding large energies are missing in averaging
- when does this matter?
- no serious error for the internal energy – exponential dependence of phase-space density kills the contribution

$$\rho(\vec{r}, \vec{p}) = \frac{\exp[-\beta E(\vec{r}, \vec{p})]}{Q}$$

Missing high-energy points in sampling

Free energies

- determine the spontaneity of process
- NVT canonical – Helmholtz function F
- NPT ‘canonical’ – Gibbs function G
- the relevant quantity always obtained depending on whether NVT or NPT simulation is performed
- much more pronounced sampling issues than e.g. for internal energy!

Missing high-energy points in sampling

$$\begin{aligned}
 F &= -k_B T \ln Q = k_B T \ln \frac{1}{Q} = \\
 &= k_B T \ln \frac{c^{-1} \cdot \iint \exp[\beta E(\vec{r}, \vec{p})] \cdot \exp[-\beta E(\vec{r}, \vec{p})] d\vec{r} d\vec{p}}{Q} = \\
 &= k_B T \ln \iint \exp[\beta E(\vec{r}, \vec{p})] \cdot \rho(\vec{r}, \vec{p}) d\vec{r} d\vec{p} - \ln c \\
 &= k_B T \cdot \ln \left\langle \exp \left[\frac{E}{k_B T} \right] \right\rangle - \ln c
 \end{aligned}$$

Missing high-energy points in sampling

$$\begin{aligned}
 F &= -k_B T \ln Q = k_B T \ln \frac{1}{Q} = \\
 &= k_B T \ln \frac{c^{-1} \cdot \iint \exp[\beta E(\vec{r}, \vec{p})] \cdot \exp[-\beta E(\vec{r}, \vec{p})] d\vec{r} d\vec{p}}{Q} = \\
 &= k_B T \ln \iint \exp[\beta E(\vec{r}, \vec{p})] \cdot \rho(\vec{r}, \vec{p}) d\vec{r} d\vec{p} - \ln c \\
 &= k_B T \cdot \ln \left\langle \exp \left[\frac{E}{k_B T} \right] \right\rangle - \ln c
 \end{aligned}$$

serious issue – the large energy values enter an exponential,
 and so the high-energy regions may contribute significantly!
 → if these are undersampled, then free energies are wrong

Missing high-energy points in sampling

$$\begin{aligned}
 F &= -k_B T \ln Q = k_B T \ln \frac{1}{Q} = \\
 &= k_B T \ln \frac{c^{-1} \cdot \iint \exp[\beta E(\vec{r}, \vec{p})] \cdot \exp[-\beta E(\vec{r}, \vec{p})] d\vec{r} d\vec{p}}{Q} = \\
 &= k_B T \ln \iint \exp[\beta E(\vec{r}, \vec{p})] \cdot \rho(\vec{r}, \vec{p}) d\vec{r} d\vec{p} - \ln c \\
 &= k_B T \cdot \ln \left\langle \exp \left[\frac{E}{k_B T} \right] \right\rangle - \ln c
 \end{aligned}$$

serious issue – the large energy values enter an exponential,
and so the high-energy regions may contribute significantly!

→ if these are undersampled, then free energies are wrong

– calculation of free energies **impossible! special methods needed!**