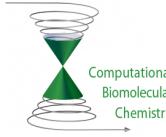




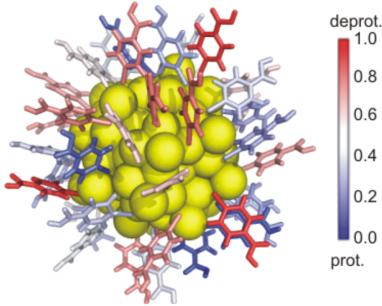
Finding the protons with molecular dynamics computer simulations

Gerrit Groenhof & Serena Donnini

Nanoscience center
University of Jyväskylä
Finland



Computational Biomolecular Chemistry



deprot.
1.0
0.8
0.6
0.4
0.2
0.0
prot.



Serena
Donnini

Contents & aims of the lecture

thermodynamics refresher

enthalpy, entropy & free energy

statistical mechanics refresher

from micro (atoms) to macro (bulk)

partition function

classical molecular dynamics simulations

ensembles

ergodicity theorem

force fields

potential energy function

free energy

thermodynamic integration

introducing pH in molecular dynamics simulations

see where the protons end up in a macro molecule

Thermodynamics what is/was that again?

boring
difficult
abstract

Thermodynamics what is/was that again?

important

in particular if you're a 19th century engineer (pre-molecules/atoms era)



how can we maximise work?

$$W^{\max} = \Delta G = \Delta U - T\Delta S$$

Thermodynamics

free energy

internal energy

entropy

$$G = U - TS$$

extensive properties

chemical/physical change at constant temperature

$$\Delta G = \Delta U - T\Delta S \quad \Delta G \equiv -T\Delta S^{\text{universe}}$$

spontaneous

$$\Delta G < 0$$

equilibrium

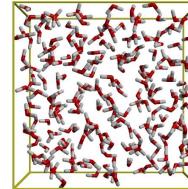
$$\Delta G = 0$$

Thermodynamics

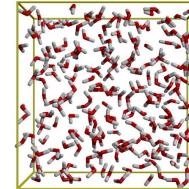
entropy of isolated system

micro-states (realizations)

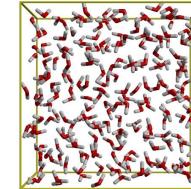
box with 216 waters:



microstate 1



microstate 2



microstate 3

total number of micro-states (realizations)

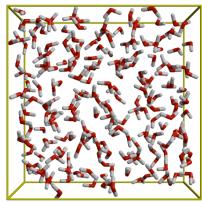
$$\Omega$$

Thermodynamics

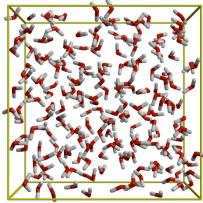
entropy of isolated system

micro-states (realizations)

two boxes of water:



box 1



box 2

total number of micro states (realizations)

$$\Omega^{\text{tot}} = \Omega_1 \cdot \Omega_2$$

entropy is extensive, so should be sum:

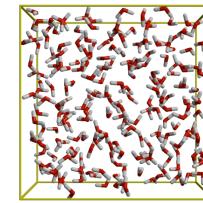
$$S = k \ln \Omega$$

Thermodynamics

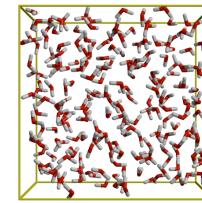
entropy of isolated system

micro-states (realizations)

two boxes of water:



box 1



box 2

total number of micro states (realizations)

$$\Omega^{\text{tot}} = \Omega_1 \cdot \Omega_2$$

entropy is extensive, so should be sum:

$$S^{\text{tot}} = k \ln[\Omega_1 \cdot \Omega_2] = k \ln \Omega_1 + k \ln \Omega_2 = S_1 + S_2$$

Thermodynamics

two systems in thermal equilibrium, isolated from world

diathermic walls (only energy can transfer)

$$\begin{array}{ccc} & N_1 + N_2 = N & \\ \text{box 1} & | & \text{box 2} \\ & E_1 + E_2 = E & \\ & \frac{dE_2}{dE_1} = -1 & \end{array}$$

equilibrium: no net changes in total entropy

$$\frac{\partial S^{\text{tot}}}{\partial E_1} = \frac{\partial S_1}{\partial E_1} + \frac{\partial S_2}{\partial E_2} \frac{dE_2}{dE_1} = 0$$

$$\frac{\partial S_1}{\partial E_1} = \frac{\partial S_2}{\partial E_2}$$

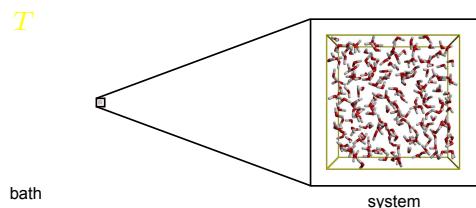
definition of temperature:

$$\frac{\partial S_1}{\partial E_1} = \frac{1}{T_1}$$

Statistical mechanics

canonical ensemble

system in thermal equilibrium with bath



micro-states of system

each with different energy

$$E_1 < E_2 < E_3 < \dots < E_i < \dots$$

probability of micro state i proportional to number of micro states of bath

$$p_i = \text{const} \cdot \Omega_{\text{bath}}(E^{\text{tot}} - E_i)$$

normalization (const):

$$p_i = \frac{\Omega_{\text{bath}}(E^{\text{tot}} - E_i)}{\sum_i \Omega_{\text{bath}}(E^{\text{tot}} - E_i)}$$

Thermodynamics

two systems in thermal equilibrium, isolated from world

diathermic walls (only energy can transfer)

$$\begin{array}{ccc} & N_1 + N_2 = N & \\ \text{box 1} & | & \text{box 2} \\ & E_1 + E_2 = E & \end{array}$$

no equilibrium: total entropy must increase

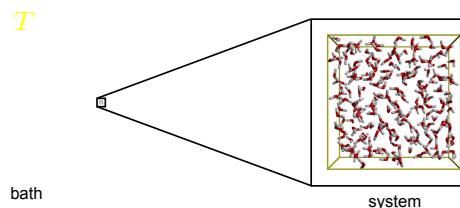
$$\frac{dS^{\text{tot}}}{dt} = \left(\frac{1}{T_1} - \frac{1}{T_2} \right) \frac{dE_1}{dt} > 0$$

energy flows from higher to lower temperature

Statistical mechanics

canonical ensemble

system in thermal equilibrium with bath



micro-states of system

probability of micro state i proportional to number of micro states of bath

$$p_i = \text{const} \cdot \Omega_{\text{bath}}(E^{\text{tot}} - E_i)$$

normalization (const):

$$p_i = \frac{\Omega_{\text{bath}}(E^{\text{tot}} - E_i)}{\sum_i \Omega_{\text{bath}}(E^{\text{tot}} - E_i)}$$

Statistical mechanics

canonical ensemble

system in thermal equilibrium with bath
micro-states of system

probability of micro state i proportional to number of micro states of bath

$$p_i = \text{const} \cdot \Omega_{\text{bath}}(E^{\text{tot}} - E_i)$$

with definition of entropy: $S = k \ln \Omega$

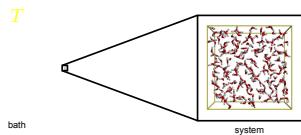
$$p_i = \text{const} \cdot \exp[S_{\text{bath}}(E^{\text{tot}} - E_i)/k]$$

bath much larger than system:

$$E^{\text{tot}} \gg E_i$$

Taylor expansion of S_{bath} around E^{tot}

$$\frac{1}{k} S_{\text{bath}}(E^{\text{tot}} - E_i) = \frac{1}{k} S_{\text{bath}}(E^{\text{tot}}) - \frac{E_i}{k} \frac{\partial S_{\text{bath}}(E)}{\partial E} \Big|_{E=E^{\text{tot}}} + \dots$$



Statistical mechanics

canonical ensemble

system in thermal equilibrium with bath
micro-states of system

probability of micro state i proportional to number of micro states of bath

$$p_i = \text{const} \cdot \exp[S_{\text{bath}}(E^{\text{tot}} - E_i)/k]$$

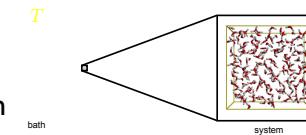
Taylor expansion of S_{bath} around E^{tot}

$$\frac{1}{k} S_{\text{bath}}(E^{\text{tot}} - E_i) = \frac{1}{k} S_{\text{bath}}(E^{\text{tot}}) - \frac{E_i}{k} \frac{\partial S_{\text{bath}}(E)}{\partial E} \Big|_{E=E^{\text{tot}}} + \dots$$

with definition of temperature

$$\frac{\partial S}{\partial E} = \frac{1}{T}$$

$$\frac{1}{k} S_{\text{bath}}(E^{\text{tot}} - E_i) = \frac{1}{k} S_{\text{bath}}(E^{\text{tot}}) - \frac{1}{kT} E_i$$



Statistical mechanics

canonical ensemble

system in thermal equilibrium with bath
micro-states of system

probability of micro-state i proportional to number of micro-states of bath

$$p_i = \text{const} \cdot \exp[S_{\text{bath}}(E^{\text{tot}} - E_i)/k]$$

Taylor expansion of S_{bath} around E^{tot}

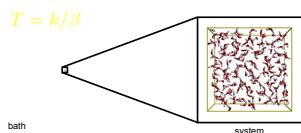
$$\frac{1}{k} S_{\text{bath}}(E^{\text{tot}} - E_i) = \frac{1}{k} S_{\text{bath}}(E^{\text{tot}}) - \frac{1}{kT} E_i$$

define

$$\beta \equiv \frac{1}{kT}$$

so that

$$\frac{1}{k} S_{\text{bath}}(E^{\text{tot}} - E_i) = \frac{1}{k} S_{\text{bath}}(E^{\text{tot}}) - \beta E_i$$



Statistical mechanics

canonical ensemble

system in thermal equilibrium with bath
micro-states of system

probability of micro state i proportional to number of micro-states of bath

$$p_i = \text{const} \cdot \exp[S_{\text{bath}}(E^{\text{tot}} - E_i)/k]$$

Taylor expansion of S_{bath} around E^{tot}

$$\frac{1}{k} S_{\text{bath}}(E^{\text{tot}} - E_i) = \frac{1}{k} S_{\text{bath}}(E^{\text{tot}}) - \beta E_i$$

probability of micro state i

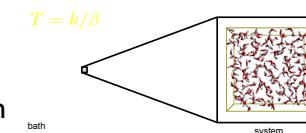
Boltzmann distribution

$$p_i = \frac{1}{Z} e^{-\beta E_i}$$

partition function

$$Z = \sum_i e^{-\beta E_i}$$

from microscopic to macroscopic with partition function



Statistical mechanics

canonical ensemble

system in thermal equilibrium with bath
entropy of system

probability of micro state i

$$p_i = \frac{1}{Z} e^{-\beta E_i} \quad Z = \sum_i e^{-\beta E_i}$$

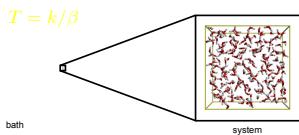
average energy of system

$$\langle E \rangle = \sum_i p_i E_i = \frac{\sum_i E_i e^{-\beta E_i}}{\sum_i e^{-\beta E_i}} = -\frac{\partial \ln Z}{\partial \beta}$$

entropy of system

$$S = -k \sum_i p_i \ln p_i$$

why?



Statistical mechanics

canonical ensemble

system in thermal equilibrium with bath
entropy of system

ensemble of N replicas

$$n_i = N p_i \quad \sum_i p_i = 1$$

total number of micro-states

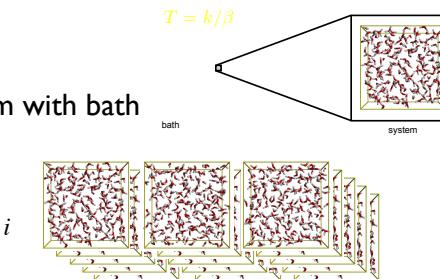
$$\Omega_N = \frac{N!}{n_1! n_2! \dots n_i! \dots}$$

entropy of ensemble

$$S_N = k \ln \Omega_N = k \ln \left[\frac{N!}{n_1! n_2! \dots n_i! \dots} \right]$$

Stirling approximation:

$$\lim_{x \rightarrow \infty} \ln x! = x \ln x - x$$



Statistical mechanics

canonical ensemble

system in thermal equilibrium with bath
entropy of system

ensemble of N replicas

number of replicas in micro-state i

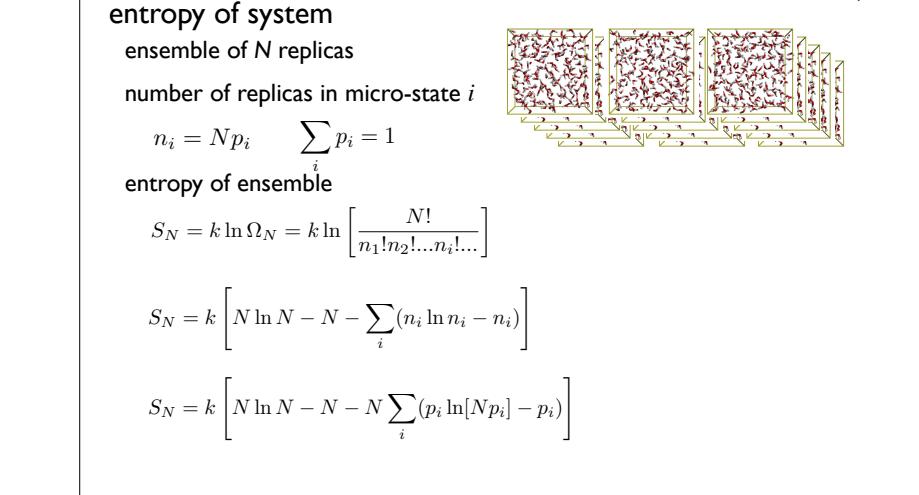
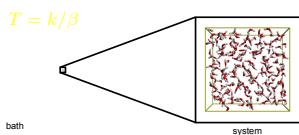
$$n_i = N p_i \quad \sum_i p_i = 1$$

entropy of ensemble

$$S_N = k \ln \Omega_N = k \ln \left[\frac{N!}{n_1! n_2! \dots n_i! \dots} \right]$$

$$S_N = k \left[N \ln N - N - \sum_i (n_i \ln n_i - n_i) \right]$$

$$S_N = k \left[N \ln N - N - N \sum_i (p_i \ln [N p_i] - p_i) \right]$$



Statistical mechanics

canonical ensemble

system in thermal equilibrium with bath
entropy of system

ensemble of N replicas

number of replicas in micro-state i

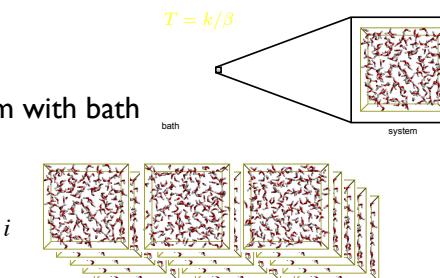
$$n_i = N p_i \quad \sum_i p_i = 1$$

entropy of ensemble

$$S_N = k \left[N \ln N - N - N \sum_i (p_i \ln [N p_i] - p_i) \right]$$

$$S_N = k \left[N \ln N - N - N \ln N - N \sum_i p_i \ln p_i - N \right]$$

$$S_N = -Nk \sum_i p_i \ln p_i$$



Statistical mechanics

canonical ensemble

system in thermal equilibrium with bath

entropy of system

ensemble of N replicas

number of replicas in micro-state i

$$n_i = N p_i \quad \sum_i p_i = 1$$

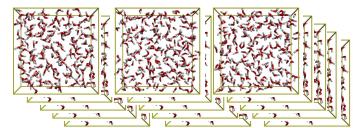
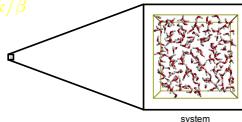
entropy of ensemble

$$S_N = -Nk \sum_i p_i \ln p_i$$

entropy of replica

$$S = -k \sum_i p_i \ln p_i$$

$$T = k/\beta$$



Statistical mechanics

canonical ensemble

system in thermal equilibrium with bath

free energy of system

microscopic entropy

$$S = \frac{1}{T} \langle E \rangle + k \ln Z$$

microscopic free energy

$$-kT \ln Z = \langle E \rangle - TS$$

$$G = -kT \ln Z$$

macroscopic free energy

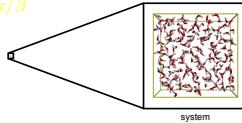
$$G = U - TS$$

from molecules to bulk: generate partition function

Monte Carlo

molecular dynamics simulations

$$T = k/\beta$$



Statistical mechanics

canonical ensemble

system in thermal equilibrium with bath

entropy of system

$$S = -k \sum_i p_i \ln p_i$$

Boltzmann distribution

$$p_i = \frac{1}{Z} e^{-\beta E_i} \quad Z = \sum_i e^{-\beta E_i} \quad \beta \equiv \frac{1}{kT}$$

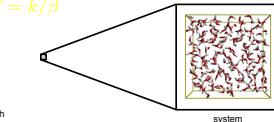
substituting and rearranging

$$S = \frac{k}{Z} \sum_i e^{-\beta E_i} \beta E_i + \frac{k}{Z} \sum_i e^{-\beta E_i} \ln Z$$

an almost familiar expression

$$S = \frac{1}{T} \langle E \rangle + k \ln Z$$

$$T = k/\beta$$



Isaac Newton

Molecular dynamics simulations

classical nuclei

$$\mathbf{F}_i = -\nabla_{\mathbf{R}_i} V(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) = m_i \frac{d^2}{dt^2} \mathbf{R}_i$$

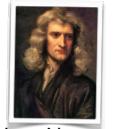
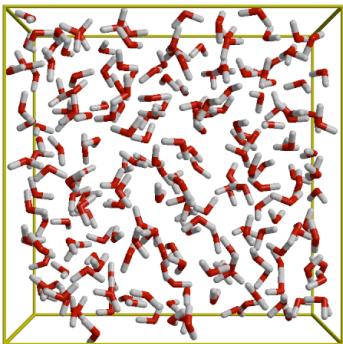
Molecular dynamics simulations

classical nuclei

$$\mathbf{F}_i = -\nabla_{\mathbf{R}_i} V(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) = m_i \frac{d^2}{dt^2} \mathbf{R}_i$$

trajectory

jiggling & wiggling



Isaac Newton

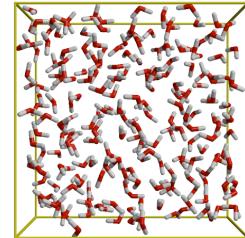
Molecular dynamics simulations

classical nuclei

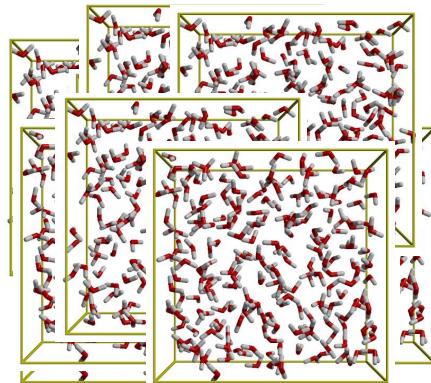
$$\mathbf{F}_i = -\nabla_{\mathbf{R}_i} V(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) = m_i \frac{d^2}{dt^2} \mathbf{R}_i$$

trajectory

ergodicity



≡



ensemble



Isaac Newton

Molecular dynamics simulations

classical nuclei

$$\mathbf{F}_i = -\nabla_{\mathbf{R}_i} V(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) = m_i \frac{d^2}{dt^2} \mathbf{R}_i$$

potential energy functions



Isaac Newton

Molecular dynamics simulations

classical nuclei

$$\mathbf{F}_i = -\nabla_{\mathbf{R}_i} V(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) = m_i \frac{d^2}{dt^2} \mathbf{R}_i$$

potential energy functions

quantum chemistry

$$V^{\text{QM}}(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) =$$

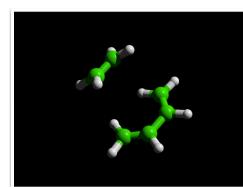
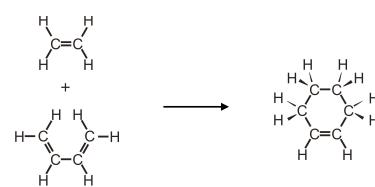


Isaac Newton



Erwin Schrödinger

$$\langle \Psi_e(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) | \hat{H}_e(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) | \Psi_e(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) \rangle$$



Molecular dynamics simulations

classical nuclei

$$\mathbf{F}_i = -\nabla_{\mathbf{R}_i} V(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) = m_i \frac{d^2}{dt^2} \mathbf{R}_i$$



potential energy functions

quantum chemistry (QM)



$$V^{\text{QM}}(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) =$$

$$\langle \Psi_e(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) | \hat{H}_e(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) | \Psi_e(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n) \rangle$$

molecular mechanics (MM)

empirical functions with parameters

$$V^{\text{MM}}(R_1, R_2, \dots, R_N) = \sum_k v_k(R_i, R_j, R_k, R_l; \{p_k\})$$

Molecular Mechanics Force Field

throw away the electrons!!!

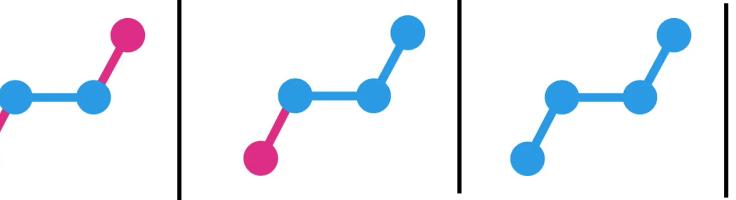
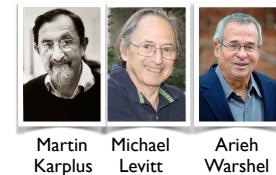
low-dimensional functions

empirical parameters

experimental data & *ab initio* calculations

$$V^{\text{MM}}(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) = \sum_k v_k(\mathbf{R}_A, \mathbf{R}_B, \mathbf{R}_C, \mathbf{R}_D; \mathbf{p}_k)$$

bonded interactions



$$v_b(\mathbf{R}_A, \mathbf{R}_B) = \frac{1}{2} k_b (R_{AB} - r_0)^2 \quad v_\theta(\mathbf{R}_A, \mathbf{R}_B, \mathbf{R}_C) = \frac{1}{2} k_a (\theta_{ABC} - \theta_0)^2 \quad v_d(\mathbf{R}_A, \mathbf{R}_B, \mathbf{R}_C, \mathbf{R}_D) = \sum_n^5 k_n (\cos(\varphi_{ABCD}))^n$$

Molecular Mechanics Force Field

throw away the electrons!!!

low-dimensional functions

empirical parameters

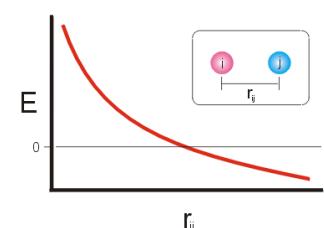
experimental data & *ab initio* calculations

$$V^{\text{MM}}(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N) = \sum_k v_k(\mathbf{R}_A, \mathbf{R}_B, \mathbf{R}_C, \mathbf{R}_D; \mathbf{p}_k)$$

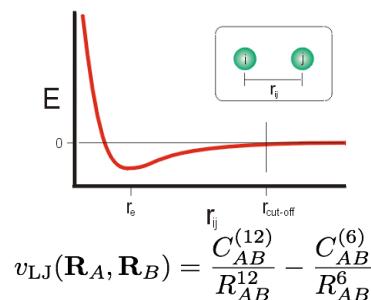
bonded interactions



Martin Karplus
Michael Levitt
Arieh Warshel



$$v_{\text{Coul}}(\mathbf{R}_A, \mathbf{R}_B) = \frac{e^2 q_A q_B}{4\pi\epsilon_0 R_{AB}}$$

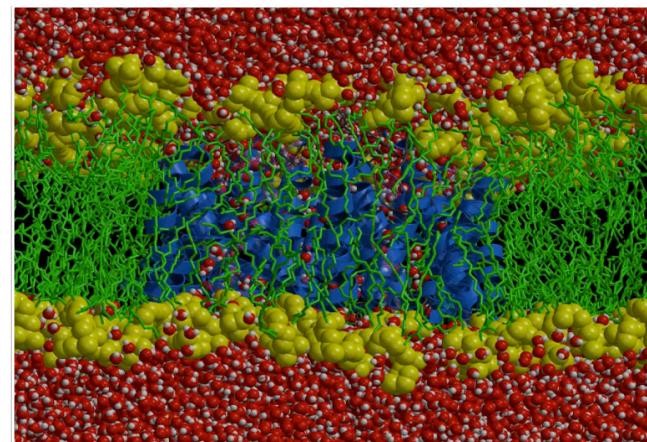


$$v_{\text{LJ}}(\mathbf{R}_A, \mathbf{R}_B) = \frac{C_{AB}^{(12)}}{R_{AB}^{12}} - \frac{C_{AB}^{(6)}}{R_{AB}^6}$$

Chemistry without test-tubes

observe jiggling and wiggling

MD simulation of aquaporin water channel

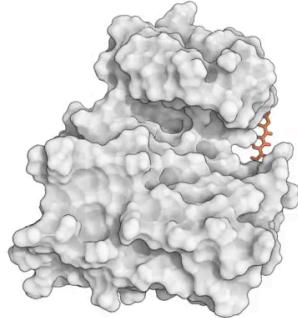


B. L. de Groot & H. Grubmüller. Science 294 (2001): 2353

Chemistry without test-tubes

how drugs bind their targets

all-atom simulations of 2.3 μ s (in total 35 μ s)



Y Shan, E. T. Kim, M. P. Eastwood, R. O. Dror, M. A. Seeliger & D. E. Shaw: *J. Am. Chem. Soc.* **133** (2011): 9181

Chemistry without test-tubes

starting structures

x-ray crystallography (www.rcsb.org)

ready to go... ... wait, where are the protons???

x-ray scatter/diffract off electrons

hydrogens have only one electron

hydrogens difficult to 'see' in density maps

infer the proton positions from pK_a estimates

$$A_i H \rightleftharpoons A_i^- + H^+ (\text{aq}) \quad K_a = \frac{[A_i^-][H^+]}{[A_i H]}$$

$$\Delta G = -RT \ln(K_a) = RT \ln(10)(pK_a - pH)$$

calculate free energy

Chemistry without test-tubes

starting structures

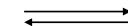
x-ray crystallography (www.rcsb.org)

ready to go...

Chemistry without test-tubes

calculating free energies in MD simulations

e.g. binding



$$G_A = -kT \ln Z_A$$

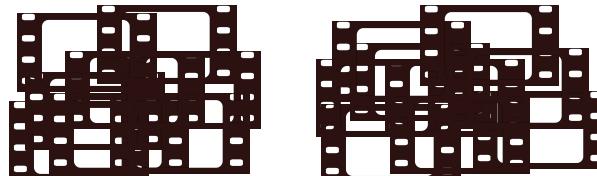
$$G_B = -kT \ln Z_B$$

$$\Delta G = G_B - G_A = -kT \ln Z_B/Z_A$$

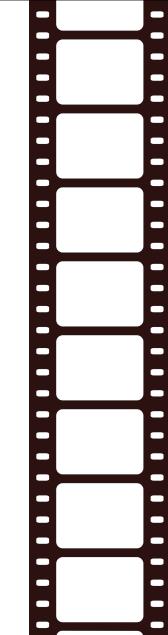
most naive approach: counting

ergodicity: time average same as ensemble average

sort frames of a trajectory and compute probabilities



$$\frac{Z_B}{Z_A} = \frac{\sum_{i \in B} e^{-\beta E_i}}{\sum_{i \in A} e^{-\beta E_i}} = \frac{\sum_{i \in B} p_i}{\sum_{i \in A} p_i} = \frac{N_B^{\text{snapshot}}}{N_A^{\text{snapshot}}}$$



Chemistry without test-tubes

calculating free energy differences

e.g. competitive binding



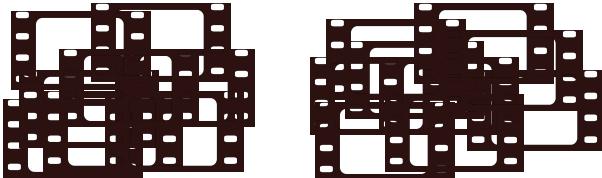
$$G_A = -kT \ln Z_A \quad G_0 = -kT \ln Z_0 \quad G_B = -kT \ln Z_B$$

$$\Delta\Delta G_{AB} = (G_B - G_0) - (G_A - G_0) = -kT \ln \left[\frac{Z_B}{Z_0} \frac{Z_0}{Z_A} \right] = -kT \ln Z_B/Z_A$$

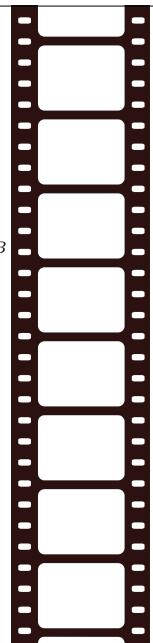
most naive approach: counting

ergodicity: time average same as ensemble average

sort frames of a trajectory and compute probabilities



$$\frac{Z_B}{Z_A} = \frac{\sum_{i \in B} e^{-\beta E_i}}{\sum_{i \in A} e^{-\beta E_i}} = \frac{\sum_{i \in B} p_i}{\sum_{i \in A} p_i} = \frac{N_B^{\text{snapshot}}}{N_A^{\text{snapshot}}}$$



Chemistry without test-tubes

calculating free energy differences

e.g. competitive binding



$$G_A = -kT \ln Z_A \quad G_0 = -kT \ln Z_0 \quad G_B = -kT \ln Z_B$$

$$\Delta\Delta G_{AB} = (G_B - G_0) - (G_A - G_0) = -kT \ln \left[\frac{Z_B}{Z_0} \frac{Z_0}{Z_A} \right] = -kT \ln Z_B/Z_A$$

most naive approach: counting

ergodicity: time average same as ensemble average

sort frames of a trajectory and compute probabilities

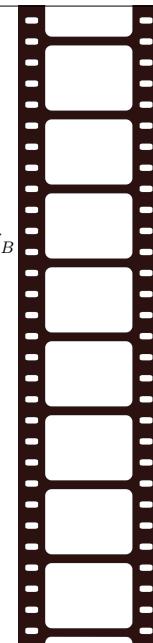
converged MD ensemble?

sufficiently long trajectory

very difficult in practice

alternative approaches

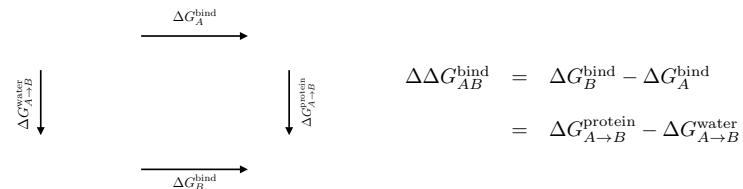
thermodynamic integration



Thermodynamic integration

calculating free energy differences

thermodynamic cycle



free energy is a state function: independent of path

$$\Delta G_A^{\text{bind}} + \Delta G_{A \rightarrow B}^{\text{protein}} = \Delta G_{A \rightarrow B}^{\text{water}} + \Delta G_B^{\text{bind}}$$

difficult to evaluate with MD simulations

$$\Delta G_A^{\text{bind}} \quad \Delta G_B^{\text{bind}}$$

'easy' (sometimes) to evaluate with MD simulations

$$\Delta G_{A \rightarrow B}^{\text{water}} \quad \Delta G_{A \rightarrow B}^{\text{protein}}$$

Thermodynamic integration

why 'easy'?

interpolate the energy function

in silico everything is possible



© 2023 The University of Texas at Austin. All rights reserved.

Thermodynamic integration

why ‘easy’?

interpolate the energy function

in silico everything is possible

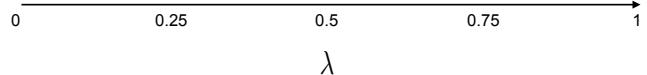
$$E^A(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{r}_1, \mathbf{r}_2)$$

$$E^B(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{r}_1, \mathbf{r}_2)$$

$$E(\lambda, \mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{r}_1, \mathbf{r}_2) = (1 - \lambda)E^A(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{r}_1, \mathbf{r}_2) + \lambda E^B(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{r}_1, \mathbf{r}_2)$$

A

B



$$\Delta G_{A \rightarrow B}^{\text{protein}}$$

Thermodynamic integration

why ‘easy’?

interpolate the energy function

in silico everything is possible

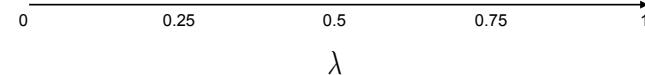
$$E^A(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{r}_1, \mathbf{r}_2)$$

$$E^B(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{r}_1, \mathbf{r}_2)$$

$$E(\lambda, \mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{r}_1, \mathbf{r}_2) = (1 - \lambda)E^A(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{r}_1, \mathbf{r}_2) + \lambda E^B(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{r}_1, \mathbf{r}_2)$$

A

B



$$\Delta G_{A \rightarrow B}^{\text{protein}}$$

Thermodynamic integration

why ‘easy’?

interpolate the energy function

in silico everything is possible

$$E^A(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{r}_1, \mathbf{r}_2)$$

$$E^B(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{r}_1, \mathbf{r}_2)$$

$$E(\lambda, \mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{r}_1, \mathbf{r}_2) = (1 - \lambda)E^A(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{r}_1, \mathbf{r}_2) + \lambda E^B(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{r}_1, \mathbf{r}_2)$$

partition function

$$Z(\lambda) = \sum_i e^{-\beta E(\lambda, \mathbf{P}_i, \mathbf{R}_i)}$$

$$= \sum_i e^{-\beta[(1-\lambda)E^A(\mathbf{P}_i, \mathbf{R}_i) + \lambda E^B(\mathbf{P}_i, \mathbf{R}_i)]}$$

free energy

$$G(\lambda) = -kT \ln Z(\lambda)$$

how does this help?

$$\mathbf{P}_i \equiv \begin{pmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \dots \end{pmatrix}_i \quad \mathbf{R}_i \equiv \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \dots \end{pmatrix}_i$$

Thermodynamic integration

why ‘easy’?

interpolate the energy function

partition function

replace sum by integration (classical)

$$Z(\lambda) = \sum_i e^{-\beta E(\lambda, \mathbf{P}, \mathbf{R})}$$

$$= \int \int e^{-\beta E(\lambda, \mathbf{P}, \mathbf{R})} d\mathbf{P} d\mathbf{R}$$

free energy

$$G(\lambda) = -kT \ln Z(\lambda)$$

$$\mathbf{P}_i \equiv \begin{pmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \dots \end{pmatrix}_i \quad \mathbf{R}_i \equiv \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \dots \end{pmatrix}_i$$

derivative of free energy

$$\frac{\partial G}{\partial \lambda} = -\frac{kT}{Z(\lambda)} \frac{\partial Z}{\partial \lambda}$$

$$= \int \int \frac{\partial E(\lambda)}{\partial \lambda} \frac{e^{-\beta E(\lambda, \mathbf{P}, \mathbf{R})}}{Z} d\mathbf{P} d\mathbf{R}$$

$$= \left\langle \frac{\partial E(\lambda)}{\partial \lambda} \right\rangle_\lambda$$

$$\Delta G_{A \rightarrow B}^{\text{protein}}$$

$$\Delta G_{A \rightarrow B}^{\text{protein}}$$

Thermodynamic integration

why ‘easy’?

interpolate the energy function

derivative of free energy

$$\frac{\partial G}{\partial \lambda} = -\frac{kT}{Z(\lambda)} \frac{\partial Z}{\partial \lambda}$$

$$= \int \int \frac{\partial E(\lambda) e^{-\beta E(\lambda, \mathbf{P}, \mathbf{R})}}{Z} d\mathbf{P} d\mathbf{R}$$

$$= \left\langle \frac{\partial E(\lambda)}{\partial \lambda} \right\rangle_{\lambda}$$

free energy

$$\Delta G_{A \rightarrow B} = \int_0^1 \left\langle \frac{\partial E}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$



Thermodynamic integration

why ‘easy’?

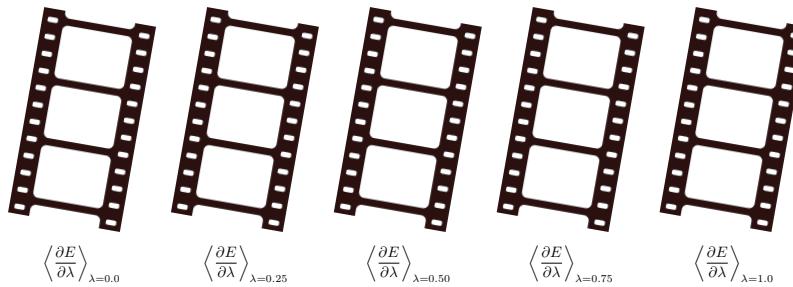
interpolate the energy function

free energy

$$\Delta G_{A \rightarrow B} = \int_0^1 \left\langle \frac{\partial E}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

multiple MD trajectories

in protein



Thermodynamic integration

why ‘easy’?

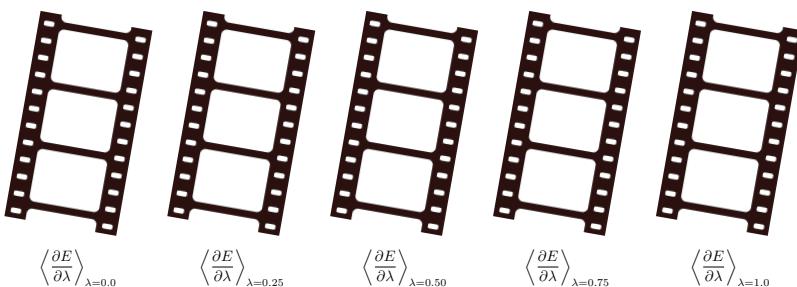
interpolate the energy function

free energy

$$\Delta G_{A \rightarrow B} = \int_0^1 \left\langle \frac{\partial E}{\partial \lambda} \right\rangle_{\lambda} d\lambda$$

multiple MD trajectories

in water



Thermodynamic integration

in practice

energy function (Hamiltonian)

$$E = H = E_{\text{kin}} + E_{\text{pot}}$$

leave kinetic energy untouched

$$E_{\text{kin}} = \frac{1}{2} \sum_i^N \frac{p_i^2}{m_i}$$

interpolate only potential energy

$$E_{\text{pot}} = V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N, \lambda)$$

partition function

$$Z(\lambda) = \frac{\int e^{-\beta E_{\text{kin}}(\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N)} d\mathbf{p}_1 d\mathbf{p}_2 \dots d\mathbf{p}_N}{\int e^{-\beta E_{\text{pot}}(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N, \lambda)} d\mathbf{r}_1 d\mathbf{r}_2 \dots d\mathbf{r}_N}$$

always same!

derivative of free energy

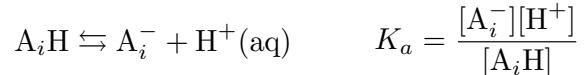
$$\frac{\partial G}{\partial \lambda} = \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_{\lambda}$$



Where are the protons in a protein?

pH is common experimental parameter

controls the protonation equilibria of amino acids side chains



$$\Delta G = -RT \ln(K_a) = RT \ln(10)(pK_a - pH)$$

how to model pH?

estimate *a priori* and choose protonation pattern

protonation stays same throughout simulation

recompute pKa's during simulations

adapt protonation

costly: 2^N free energy computations

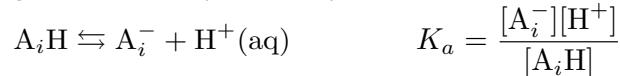
treat protonation as dynamic variable

extended Hamiltonian

Where are the protons in a protein?

measuring pKa

single amino acid (side chain) in water



$$\Delta G = -RT \ln(K_a) = RT \ln(10)(pK_a - pH)$$

at pH = pKa

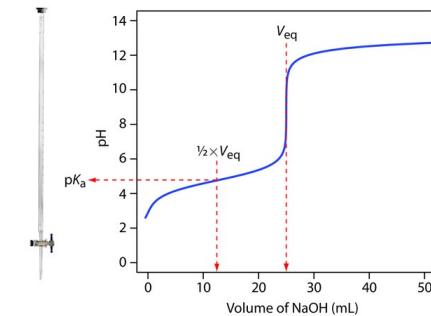
$$[A_i H] = [A_i^-]$$

titration

Henderson Hasselbalch

$$pH = pK_a - \log_{10} \frac{[A_i H]}{[A_i^-]}$$

literature/internet/...

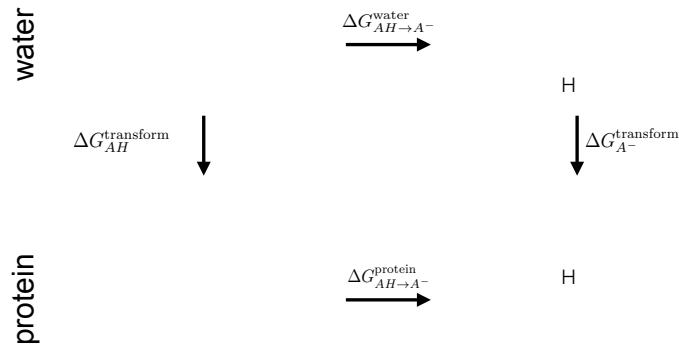


Where are the protons in a protein?

calculating pKa

single amino acid (side chain) in a protein

thermodynamic cycle



Where are the protons in a protein?

calculating pKa

single amino acid in a protein

thermodynamic cycle

$$\begin{aligned} \Delta \Delta G &= \Delta G_{AH \rightarrow A^-}^{\text{protein}} - \Delta G_{AH \rightarrow A^-}^{\text{water}} \\ &= \Delta G_{A^-}^{\text{transform}} - \Delta G_{AH}^{\text{transform}} \end{aligned}$$

amino acid in water is reference

assuming force field and sampling errors are identical!

only interactions with protein environment change affect pKa

shift in pKa with respect to water

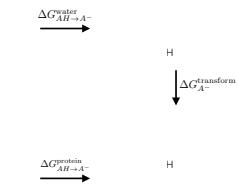
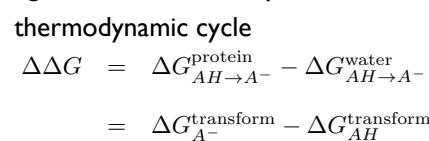
pKa in water known!

$$\Delta pK_a = \frac{\Delta \Delta G}{RT \ln(10)}$$

Where are the protons in a protein?

calculating pKa

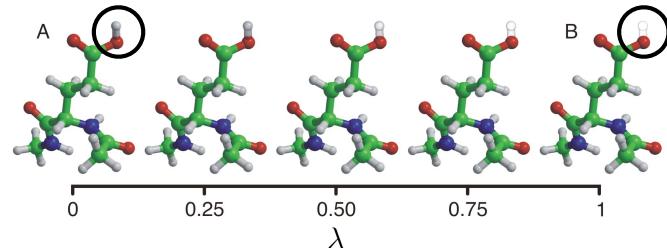
single amino acid in a protein



two thermodynamic integrations

in water

in protein



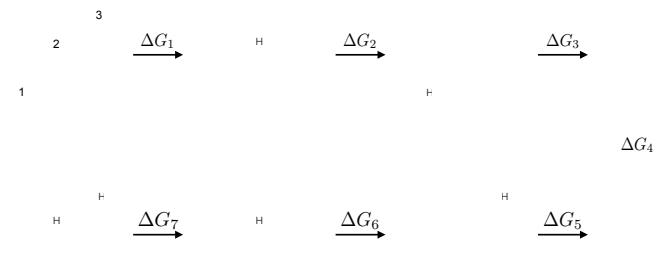
Where are the protons in a protein?

calculating pKa

multiple amino acid in a protein

lower branch of thermodynamic cycle(s)

all possible protonation states: combinatorial problem



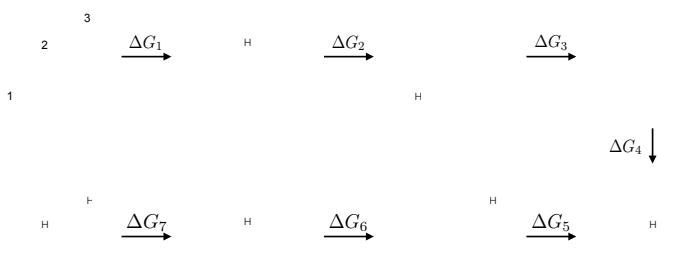
Where are the protons in a protein?

simulation at constant pH

visit only relevant protonation states

include protonation as dynamic variable

extended energy function (Hamiltonian)



$$E(p_1, p_2, \dots, r_1, r_2, \dots, H_1^+, H_2^+, \dots) = H(p_1, p_2, \dots, r_1, r_2, \dots, H_1^+, H_2^+, \dots)$$

$$= \sum_i^N \frac{p_i^2}{2m_i} + V(r_1, r_2, \dots, H_1^+, H_2^+, \dots) + E_{\text{kin}}(H_1^+, H_2^+, \dots)$$

Where are the protons in a protein?

constant pH molecular dynamics simulations

pH parameter in simulations

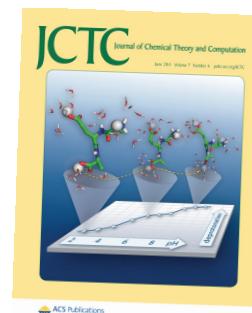
simultaneously sampling of configurations & protonations

fully atomistic description

no force field dependencies

atomistic & coarse grained

pH gradients



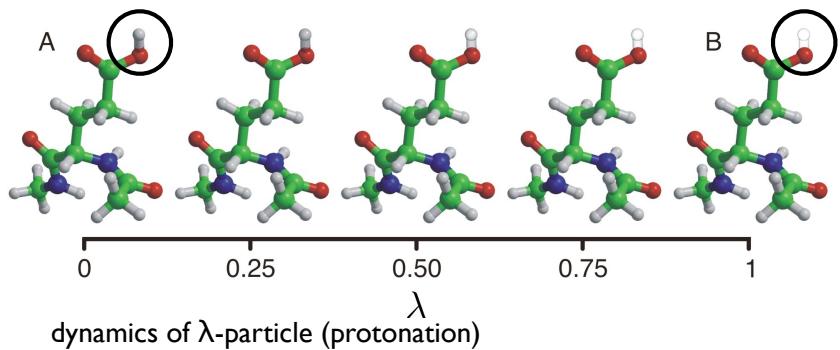
Donnini, Tegeler, Groenhof, Grubmüller, *J. Chem. Theory Comput.* **7** (2011) 1962

Donnini, Ullmann, Groenhof, Grubmüller, *J. Chem. Theory Comput.* **12** (2016) 1040

Dobrev, Donnini, Groenhof, Grubmüller, *J. Chem. Theory Comput.* **13** (2017) 147

Molecular dynamics at constant pH: how?

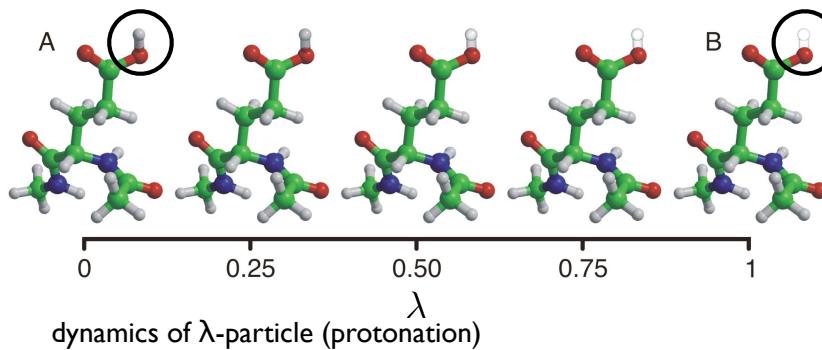
protons as extra degrees of freedom



X. Kong & C.L. Brooks *J. Chem. Phys.* 105 (1996) 2414
 M.S. Lee, J.F.R. Salsbury, C.L. Brooks *Proteins* 56 (2004) 738
 S. Donnini, F. Tegeler, G. Groenhof, H. Grubmüller *JCTC* 7 (2011) 1962

Molecular dynamics at constant pH: how?

protons as extra degrees of freedom

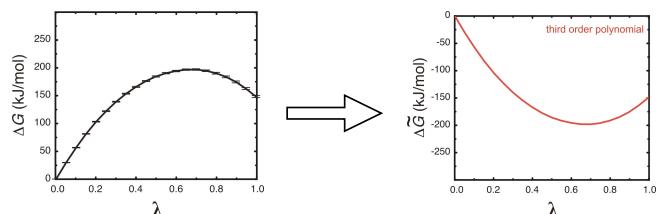


$$V(\mathbf{x}, \lambda) = V(\mathbf{x}, (1-\lambda)\{q\}^A + \lambda\{q\}^B) + U(\lambda) + \lambda RT \ln(10)[pK_{a,\text{ref}}^{\text{exp}} - pH] + \Delta\tilde{G}_{\text{MM}}^{\text{corr}}(\lambda)$$

$$V(\mathbf{x}, \lambda) = V(\mathbf{x}, (1-\lambda)\{q\}^A + \lambda\{q\}^B) +$$

$$U(\lambda) + \lambda RT \ln(10)[pK_{a,\text{ref}}^{\text{exp}} - pH] + \boxed{\Delta\tilde{G}_{\text{MM}}^{\text{corr}}(\lambda)}$$

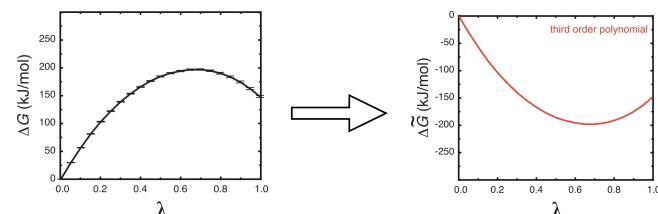
$\Delta\tilde{G}_{\text{MM}}^{\text{corr}}(\lambda)$ obtained by thermodynamic integration (pH = pK_a)



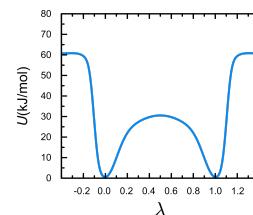
$$V(\mathbf{x}, \lambda) = V(\mathbf{x}, (1-\lambda)\{q\}^A + \lambda\{q\}^B) +$$

$$U(\lambda) + \lambda RT \ln(10)[pK_{a,\text{ref}}^{\text{exp}} - pH] + \boxed{\Delta\tilde{G}_{\text{MM}}^{\text{corr}}(\lambda)}$$

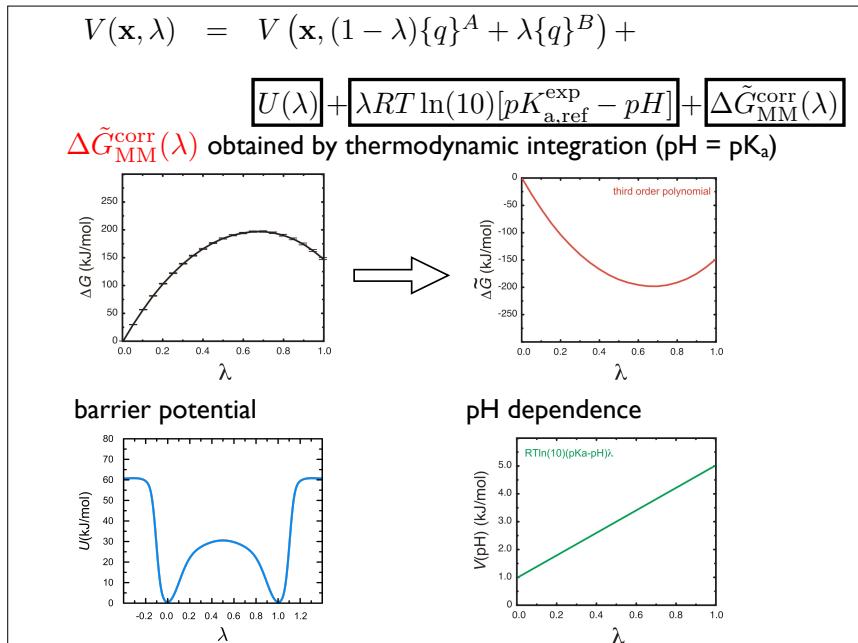
$\Delta\tilde{G}_{\text{MM}}^{\text{corr}}(\lambda)$ obtained by thermodynamic integration (pH = pK_a)



barrier potential

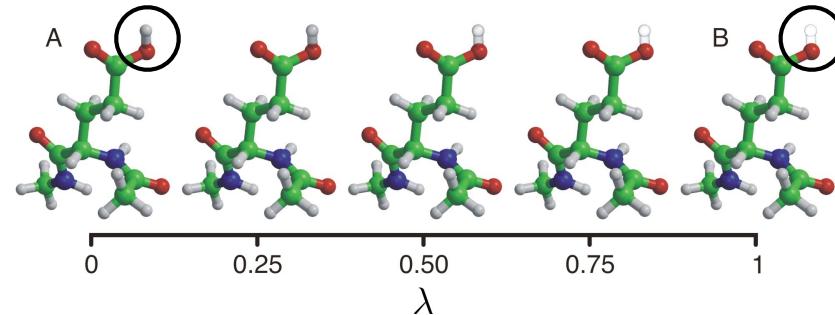


S. Donnini, R. T. Ullman, G. Groenhof, H. Grubmüller *JCTC* 12 (2016) 1040



Molecular dynamics at constant pH: how?

protons as extra degrees of freedom



straightforward extension to multiple titratable sites

$$m_{\lambda_i} d^2 \lambda_i / dt^2 = -\partial V(\mathbf{x}, \lambda) / \partial \lambda_i$$

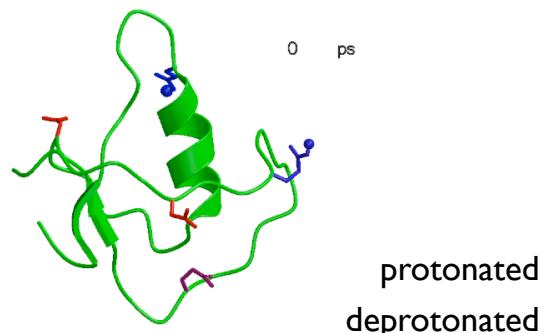
$$V(\mathbf{x}, \lambda) = V(\dots, (1 - \lambda_i)\{q\}_i^A + \lambda_i\{q\}_j^B, (1 - \lambda_j)\{q\}_j^A + \lambda_i\{q\}_i^B, \dots) +$$

$$\sum_i [U(\lambda_i) + \lambda_i RT \ln(10)[pK_{a,\text{ref}_i}^{\text{exp}} - pH] + \Delta\tilde{G}_{\text{MM},i}^{\text{corr}}(\lambda_i)]$$

Constant pH molecular dynamics

protonation states are variable

e.g. third domain of turkey ovomucoid inhibitor at pH = 4



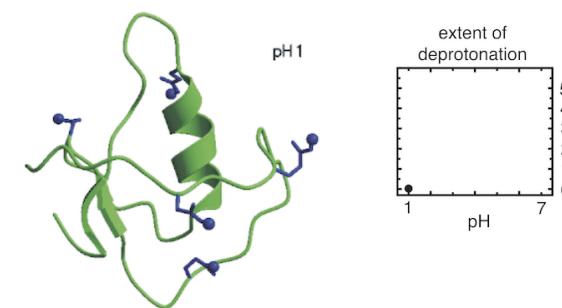
Constant pH molecular dynamics

in silico titration experiment

MD simulations at different solvent pH values

Henderson-Hasselbalch

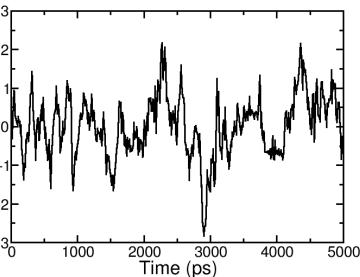
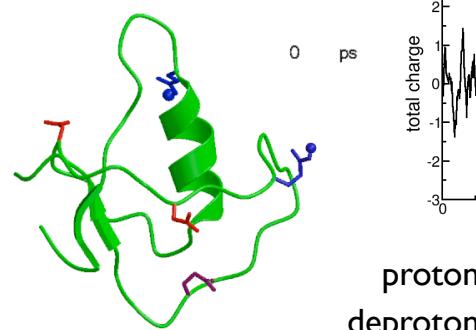
$$\frac{[\text{A}^-]}{[\text{A}^-] + [\text{AH}]} = \frac{1}{10^{n(\text{pK}_a - \text{pH})} + 1}$$



Constant pH molecular dynamics

protonation states are variable

e.g. third domain of turkey ovomucoid inhibitor at pH = 4



protonated
deprotonated

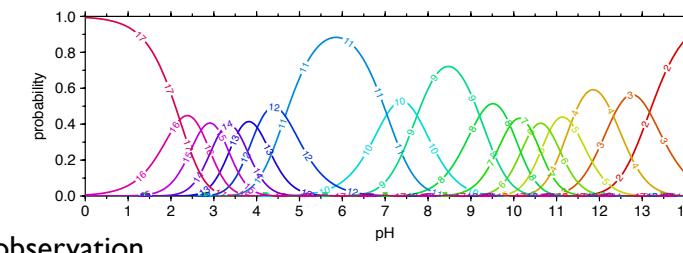
Constant pH simulations

a priori estimate of charge fluctuations

distribution of macroscopic protonation states

3rd domain of turkey ovomucoid inhibitor

GMCT (PB based)¹



observation

number of macroscopic states much smaller than number of titration sites

sampling problem is not as big as it seems

when interested in one pH point

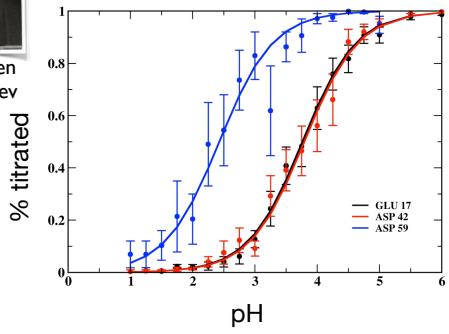
¹J. Comput. Chem. 33, 887-900 (2012)



Thomas
Ullmann

pK_a's of proteins

cardiotoxin V



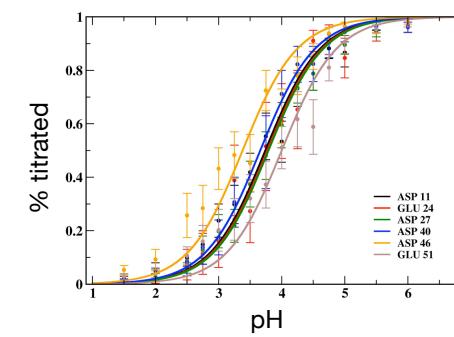
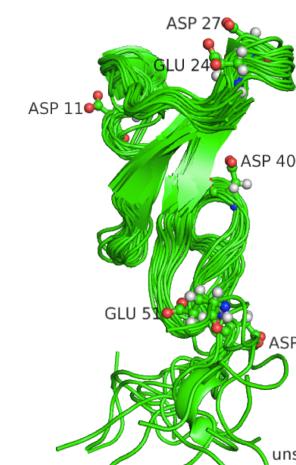
ASP 42
GLU 17
ASP 59

residue	CPHMD	PB	NMR
Glu17	3.77 (0.06)	3.73	4
Asp42	3.8 (0.07)	3.64	3.2
Asp59	2.44 (0.12)	2.84	< 2.3

NMR data: Biochemistry 35 (1996) 9167

pK_a's of proteins

epidermal growth factor



residue	CPHMD	PB	NMR
Asp11	3.74 (0.07)	4.2 (0.13)	3.9
Glu24	3.77 (0.12)	3.78 (0.15)	4.1
Asp27	3.79 (0.07)	3.8 (0.08)	4
Asp40	3.66 (0.09)	5.33 (0.28)	3.6
Asp46	3.38 (0.09)	4.1 (0.34)	3.8
Glu51	4.0 (0.09)	4.1 (0.24)	4

NMR data: Biochemistry 30 (1991) 4896

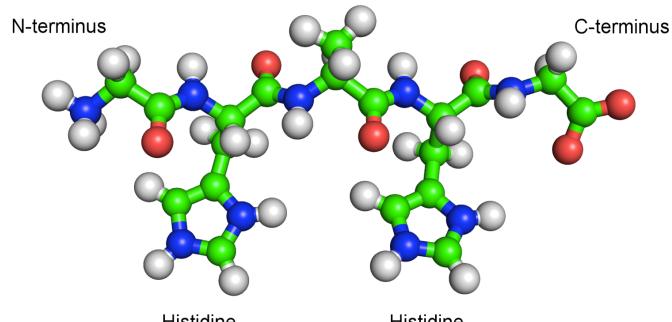
Titration of pentapeptides

constant MD & NMR (double blind)

challenges

high intrinsic conformational flexibility

coupling between side chains sites at 2nd and 4th position



NMR-based Structural Biology



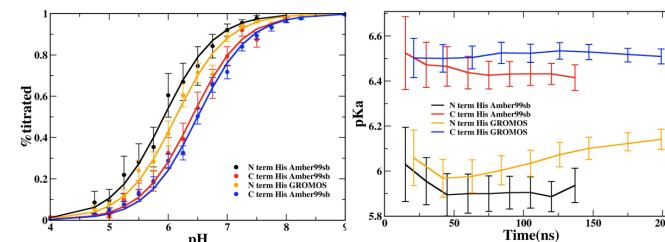
Christian Griesinger Nilamoni Nath Plamen Dobrev

Titration of pentapeptides

Results of MD and NMR

different forcefields

Amber99SB & Gromos53A2



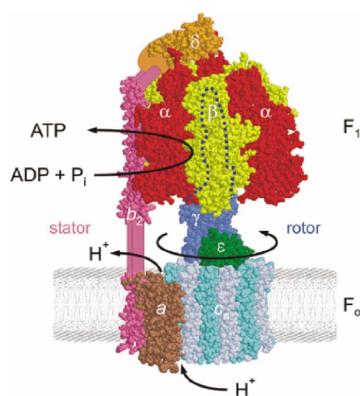
	AMBER	GROMOS	NMR
His2	6.09 (0.08)	6.30 (0.05)	6.24
His4	6.56 (0.06)	6.67 (0.03)	6.66

pH gradients in MD simulations

why gradients: ATP synthesis

ATP synthase: a Brownian motor

Mitchell's hypothesis



pH gradients in MD simulations

rectifying Brownian motion with proton gradient

(AlaAlaAlaAlaAsp)₉ peptide in a hexadecane slab

simulation parameters

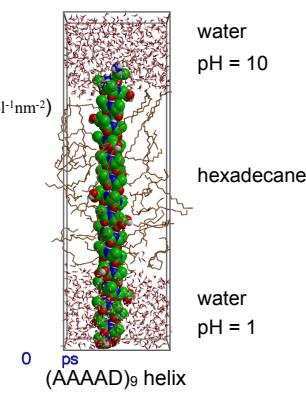
Gromos54a7/SPC

5 ns leap-frog @ 303 K

reaction-field ($r_c = 1.4$ nm)

harmonic restraints on C_α's in x-y plane ($f=1000 \text{ kJmol}^{-1}\text{nm}^{-2}$)

Gromacs-5.1 with constant pH

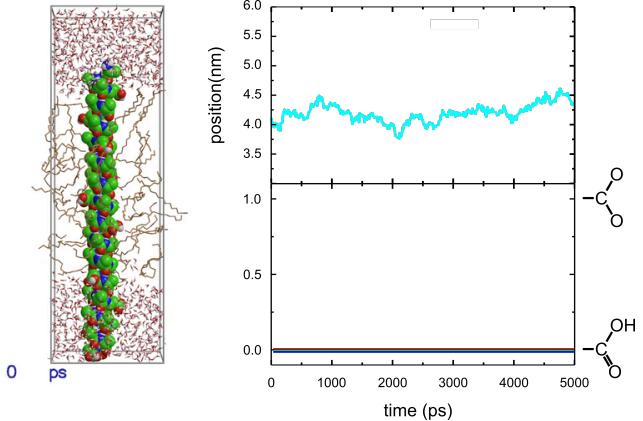


Effect of pH gradients over an interface

rectifying Brownian motion with proton gradient

(AlaAlaAlaAlaAsp)₉ peptide in a hexadecane slab

reference simulations: no constant pH

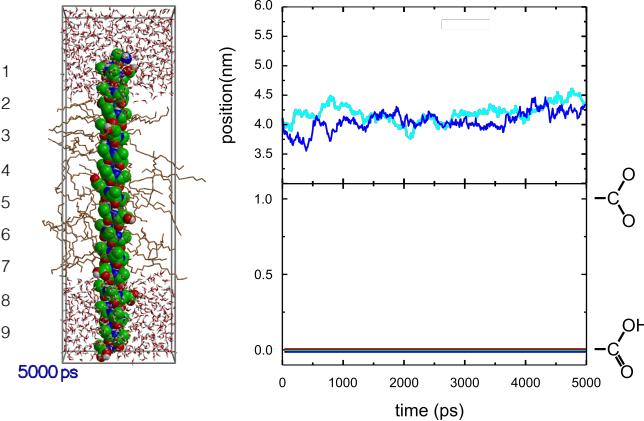


Effect of pH gradients over an interface

rectifying Brownian motion with proton gradient

(AlaAlaAlaAlaAsp)₉ peptide in a hexadecane slab

reference simulations: no constant pH

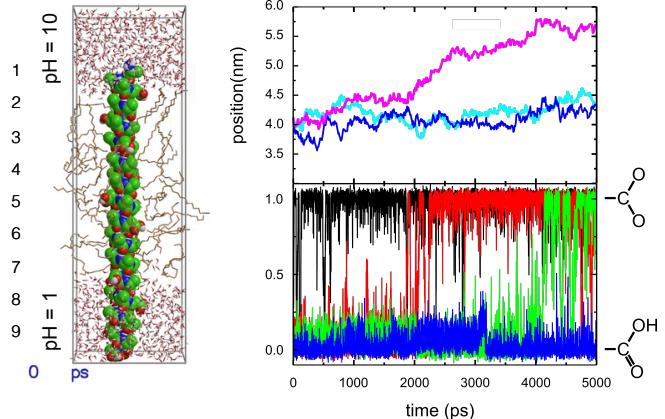


Effect of pH gradients over an interface

rectifying Brownian motion with proton gradient

(AlaAlaAlaAlaAsp)₉ peptide in a hexadecane slab

constant pH simulations

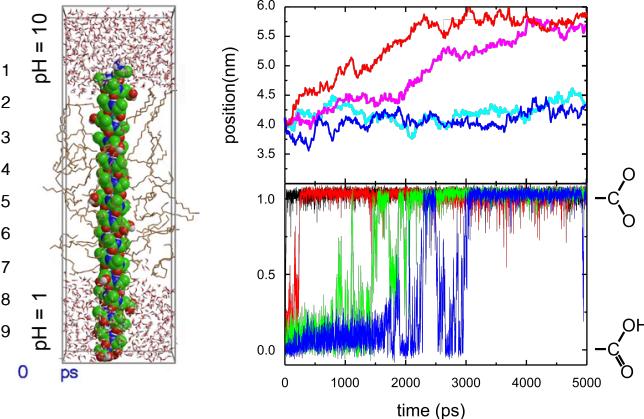


Effect of pH gradients over an interface

rectifying Brownian motion with proton gradient

(AlaAlaAlaAlaAsp)₉ peptide in a hexadecane slab

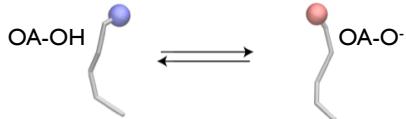
constant pH simulations



Coarse grained simulations at constant pH

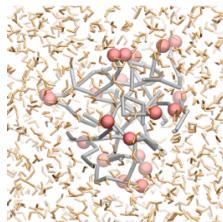
Oleic acid aggregates

MARTINI 2.1 coarse grained force field

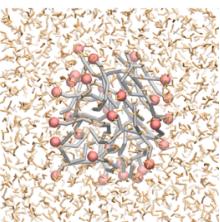


Drew
Bennett

micelle systems (self-aggregation)



20-mer



30-mer



Peter
Tieleman

Bennet, Chen, Donnini, Groenhof, Tieleman, *Can. J. Chem.* **91** (2013) 839

Coarse grained constant pH simulations

Oleic acids aggregates

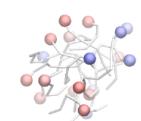
micelles at different pH

pH=2.0

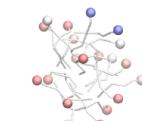


20mer

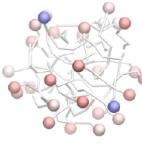
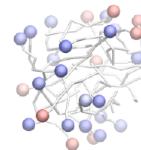
pH=5.5



pH=9.0



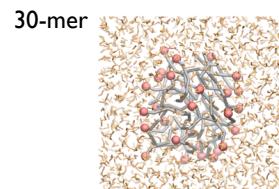
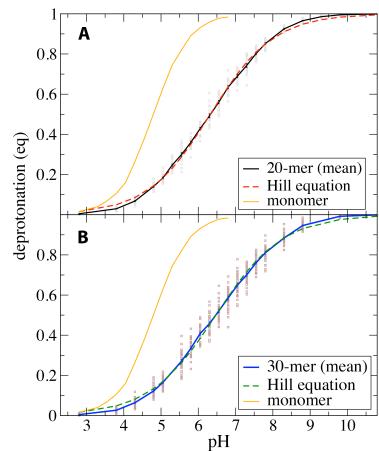
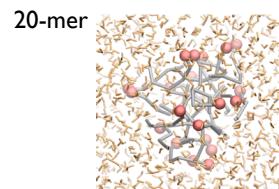
30mer



Coarse grained constant pH simulations

Oleic acids aggregates

upward pK_a shift of OA in micelles

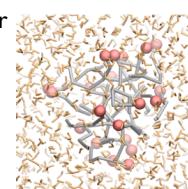


Coarse grained constant pH simulations

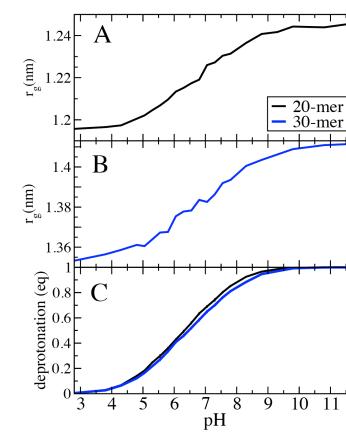
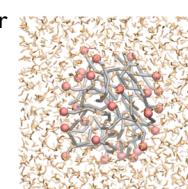
Oleic acids aggregates

pK_a shift correlates with micelle radius

20-mer



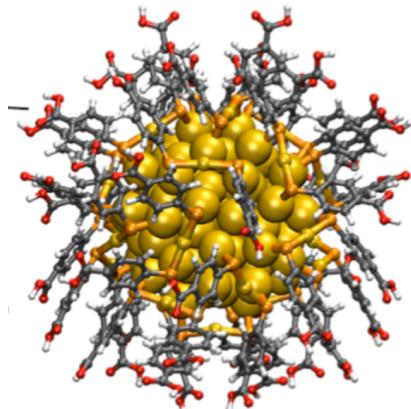
30-mer



Charge state of $\text{Au}_{102}(p\text{MBA})_{44}$ cluster

monolayer protected gold nanoparticle

x-ray structure



Jadzinsky, Calero, Ackerson, Bushnell, Kornberg, *Science* **318** (2007) 430

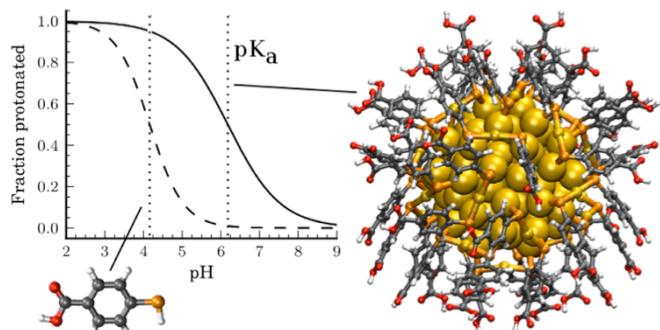
Charge state of $\text{Au}_{102}(p\text{MBA})_{44}$ cluster

determining pK_a value(s)

potentiometric

infrared spectra at various pH

carbonyl/carboxyl intensity ratio



Koivisto, Chen, Donnini, Lahtinen, Häkkinen, Groenhof, Pettersson *J. Phys. Chem. C* **120** (2016) 10041



Jaako
Koivisto



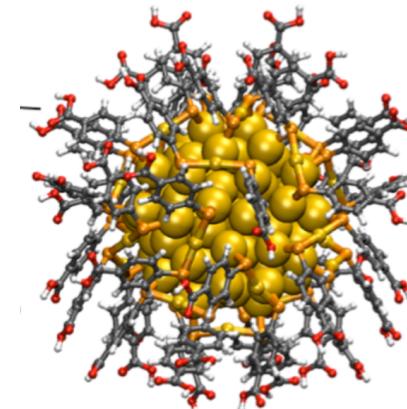
Mika
Pettersson

Charge state of $\text{Au}_{102}(p\text{MBA})_{44}$ cluster

monolayer protected gold nanoparticle

x-ray structure:

where are the protons ?



Jadzinsky, Calero, Ackerson, Bushnell, Kornberg, *Science* **318** (2007) 430

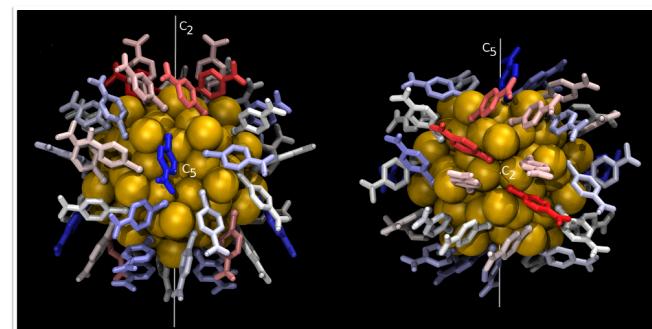
Charge state of $\text{Au}_{102}(p\text{MBA})_{44}$ cluster

surface charge distribution in MD at pH = pK_a

non-uniform



Serena
Donnini



Koivisto, Chen, Donnini, Lahtinen, Häkkinen, Groenhof, Pettersson *J. Phys. Chem. C* **120** (2016) 10041

Charge state of Au₁₀₂(pMBA)₄₄ cluster

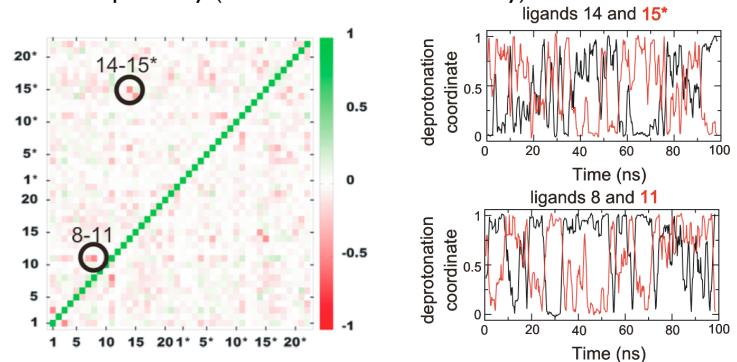
surface charge distribution in MD at pH = pK_a

non-uniform

pearson correlation

weak to moderate anti correlation

anti-cooperativity (Hill coefficient less than unity)



Serena
Donnini

Charge state of Au₁₀₂(pMBA)₄₄ cluster

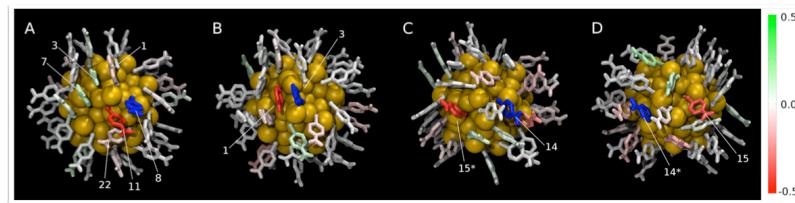
surface charge distribution in MD at pH = pK_a

non-uniform

pearson correlation

weak to moderate anti correlation

anti-cooperativity (Hill coefficient less than unity)



Serena
Donnini

Constant pH simulations

increased realism in MD simulations

charge (protonation) fluctuations

low computational overhead

linear scaling in number of titratable sites

challenges/problems

too many still....

... as you will see in the practical...