Atomistic Computer Simulation of Hard and Soft Matter

Eckhard Spohr, Stéphane Kenmoe, Faculty for Chemistry
University Duisburg-Essen, Germany
eckhard.spohr@uni-due.de

8th African School of Physics
Marrakech, July 16, 2024

CRC/TRR 247
HETEROGENEOUS OXIDATION CATALYSIS IN THE LIQUID PHASE

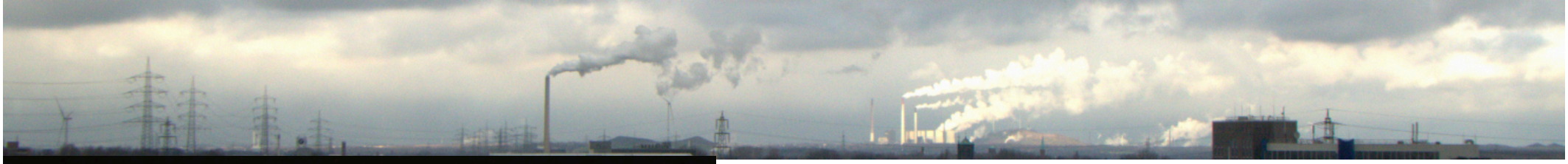
DFG Deutsche Forschungsgemeinschaft

UNIVERSITÄT DUISBURG ESSEN

CCSS of Uni Due

ZOLLVEREIN UNESCO WORLD HERITAGE SITE

THE HEART OF THE RUHR AREA



- Dr. Stéphane Kenmoe
- Cameroonian
- Science Communication
- CASESMA (since 2019) (Central African School of Electronic Structure of Materials)

Physics → Chemistry

Electrodynamics (Quantum Mechanics)

The rest:

- too weak
- too strong

Atomistic Simulation?

- modeling of real physical (chemical / biological) systems
- a subdiscipline of statistical mechanics
- calculates thermodynamic properties
- follows the 'real' trajectory of a system
- study structure, evolution thereof
- Molecular Dynamics: studies the dynamics
  - individual atoms
  - molecules
  - clusters
  - structural elements
  - ….
  - parts of living cells
  - materials
- chemical reaction dynamics

Atomistic Simulation?

- **modeling** of real physical (chemical / biological) systems
- a subdiscipline of statistical mechanics
- calculates thermodynamic properties
- follows the 'real' trajectory of a system
- studies structure, evolution thereof
- Molecular Dynamics: studies the dynamics
  - individual atoms
  - molecules
  - clusters
  - structural elements
  - ….
  - parts of living cells
  - materials
- chemical reaction dynamics

Modeling: Hamiltonians

- assume that we can write down a Hamiltonian of a system of atoms (usually assumed to be classical)
  - with or
  - without electronic degrees of freedom
- i.e., we often assume a first level of 'coarse graining'
  - we treat the atom cores as the moving entities
  - electrons follow (adiabatically?)
  - Born-Oppenheimer approximation:
- electronic degrees of freedom:
  - quantum mechanically: various approximations to the Schrödinger equation
    - most frequently: Density functional Theory (DFT)
    - …
    - empirical force field

Modeling

several ‘directions’ along which one can improve/ simplify Hamiltonian models:

- system size: as large as possible, as small as necessary
  - 10 atoms —— 10^8 atoms
- simulation time: long enough to observe the relevant processes
  - 1 ps —— 10 ps —— 1 ns —— … ——- ms
- realism / simplification of the Hamiltonian:
  - is electronic structure important?
  - ground state properties only?
  - does every atom count? ‘coarse graining’

Atomistic Simulation?

- modeling of real physical (chemical / biological) systems
- a subdiscipline of statistical mechanics
- calculates thermodynamic properties
- follows the 'real' trajectory of a system
- **Monte Carlo:** studies structure, evolution thereof
- **Molecular Dynamics**: also studies the dynamics
  - individual atoms
  - molecules
  - clusters
  - structural elements
  - ….
  - parts of living cells
  - materials
- chemical reaction dynamics

Monte Carlo

- principle: calculate multi-dimensional integrals

$$\langle \mathcal{A} \rangle_{NVT} = \int \mathrm{d}\Gamma \rho_{NVT}(\Gamma) \mathcal{A}(\Gamma) \qquad \langle \mathcal{A} \rangle_{NVT} = \langle \frac{\mathcal{A}\rho_{NVT}}{\rho} \rangle \quad \rho = \rho_{NVT}$$

- choose:

$$\rho_{NVT}(\Gamma) = \frac{1}{Z_{NVT}} \exp[-\beta \mathcal{V}(\Gamma)]$$

$$\frac{\rho_n}{\rho_m} = \frac{\exp(-\beta \mathcal{V}_n)}{\exp(-\beta \mathcal{V}_m)} = \exp[-\beta(\mathcal{V}_n - \mathcal{V}_m)] \qquad \boxed{\langle \mathcal{A} \rangle_{NVT} = \langle \mathcal{A} \rangle_{Versuche}}$$

- algorithm :

$$\Pi_{mn} = \alpha_{mn} \quad \text{wenn } \rho_n \geq \rho_m \quad \text{und } m \neq n$$

$$\Pi_{mn} = \alpha_{mn}(\rho_n/\rho_m) \quad \text{wenn } \rho_n < \rho_m \quad \text{und } m \neq n \quad (\text{also } \Pi_{mn} < \alpha_{mn})$$

$$\Pi_{mm} = 1 - \sum_{n \neq m} \Pi_{mn}$$

$$\rho_m \Pi_{mn} = \rho_n \Pi_{nm} \iff \frac{\rho_n}{\rho_m} = \frac{\Pi_{mn}}{\Pi_{nm}}$$

Molecular Dynamics

- solve Newton's equations of motions for 3N dofs

$$q_i(t+\delta t) = q_i(t) + \dot{q}_i(t)\cdot\delta t + \frac{1}{2}\ddot{q}_i(t)\cdot(\delta t)^2 \pm \ldots$$

$$q_i(t-\delta t) = q_i(t) + \dot{q}_i(t)\cdot(-\delta t) + \frac{1}{2}\ddot{q}_i(t)\cdot(-\delta t)^2 \pm \ldots$$

- add up: $\boxed{q_i(t+\delta t) = 2q_i(t) - q_i(t-\delta t) + \ddot{q}_i(t)\cdot(\delta t)^2}$

- generates configurations with equal a priori probability,
- i.e. a **microcanonical NVE ensemble**
- averages are simple sums divided by # of configurations

used already in first simulations in 1960s

A Simple MD Code

```
nstepmax = ????
step = 0
model = your_choice_of_hamiltonian
read_starting_configuration()

while not happy or step < nstepmax:
    propagate_coordinates_a_bit(delta_t)
    calculate_forces(model)
    collect_some_data_sometimes(your_question)
    write_conf_to_disk_every_now_and_then()
    happy = your_decision_based_on…()
    step += 1

write_final_configuration()
```

Open Source MD Codes

- Gromacs (> 50,000 lines)
- Amber
- Namd
- OpenMM
- Lammps (> 640,000 lines)
- CP2K (> 1,150,000 lines)
- ….

Open Source MD Codes

what makes these codes so large?

- support for different interaction models
- support for electronic structure calculations
- support for handling hundreds of atoms
- support for handling millions of atoms
- support for building geometric models of chemical systems
- I/O
- optimization
- on-the-fly analysis
- support for various ensembles / boundary conditions
- … 'infrastructure'

Open Source MD Codes

as of 2024:

- infrastructure well developped
- no (?) need to write own codes
- learn how to
  - ask a good scientific question
  - select a suitable code
  - use it
  - understand its limitations

What do you want to know?

- what is the physical/chemical process of interest?
- is the system ordered or disordered?
- spatial structure: assume it or produce it?
- relevant time scale of process?
- relevant time scale of relaxation to equilibrium?
- do you want to study a system
    - in equilibrium?
    - out of equilibrium?
- dominant interactions?

… choose simulation code / system size / interaction model / simulation length / ….

Atomistic Simulation?

- modeling of real physical (chemical / biological) systems
- a subdiscipline of statistical mechanics
- calculates **thermodynamic** properties
- follows the 'real' trajectory of a system
- studies structure, evolution thereof
- Molecular Dynamics: studies the dynamics of
  - individual atoms
  - molecules
  - clusters
  - structural elements
  - ….
  - parts of living cells
  - materials
- chemical reaction dynamics

Thermostats / Barostats

- Nosé-Hoover thermostat: NVT ensemble

$$H = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) + \frac{p_\xi^2}{2Q} + N_f kT\xi$$

- Parrinello-Rahman barostat: NPH ensemble

$$H = E_{\text{pot}} + E_{\text{kin}} + \sum_i P_{ii} V + \sum_{i,j} \frac{1}{2} W_{ij} \left(\frac{\mathrm{d}b_{ij}}{\mathrm{d}t} \right)^2$$

- combination: NPT ensemble

extended system Hamiltonians

Thermodynamics: Eq. of State

- pressure, volume, temperature are also simple averages
- establish a simulation procedure that allows
  - calculation of <V>(T, P)
  - or <P>(T,V)
- ===> equation of state
  - very important in technical thermodynamics
  - also for matter under extreme conditions

Atomistic Simulation?

- modeling of real physical (chemical / biological) systems
- a subdiscipline of statistical mechanics
- calculates thermodynamic properties
- follows the 'real' trajectory of a system
- studies structure, evolution thereof
- **Molecular Dynamics**: studies the dynamics of
  - individual atoms
  - **molecules**
  - clusters
  - structural elements
  - ….
  - parts of living cells
  - materials
- chemical reaction dynamics

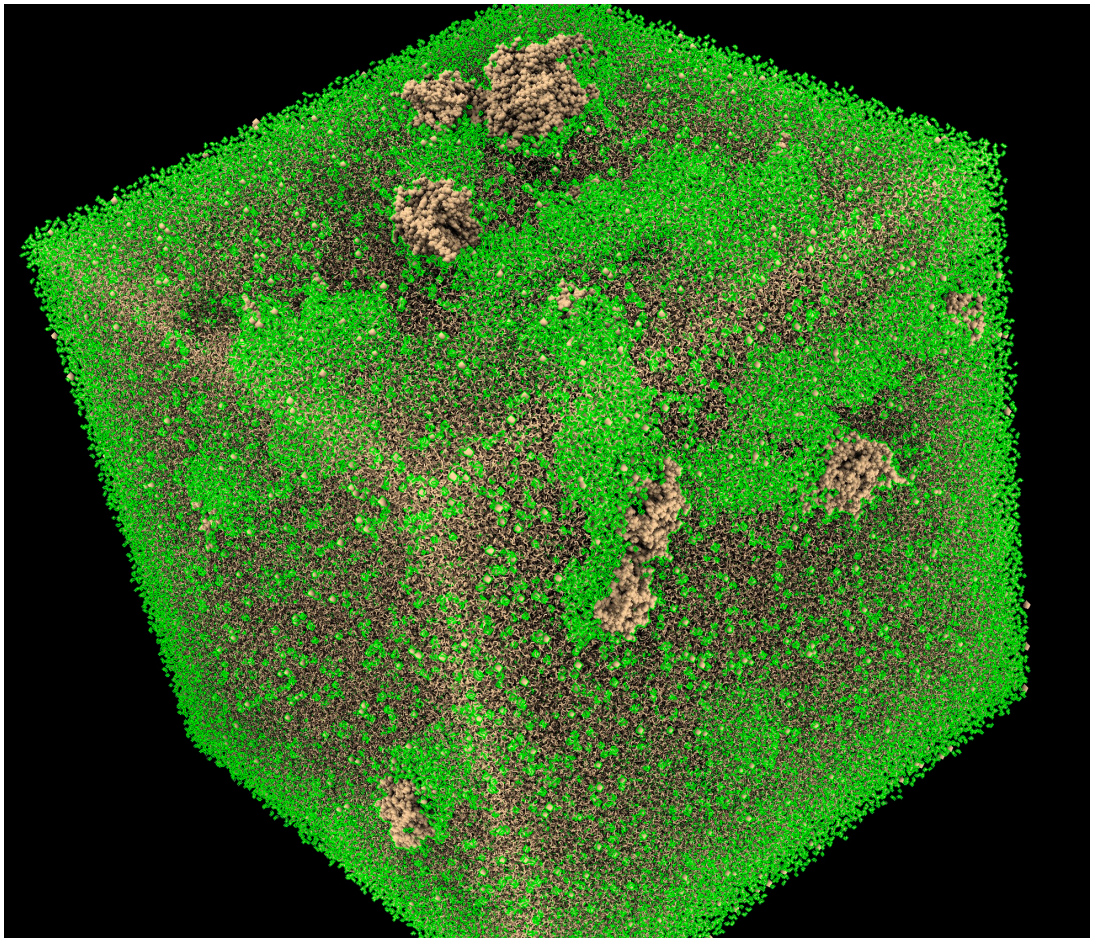
A Simple Protein Force Field

$$
\begin{aligned}
V(\{r\}) = & \\
& \sum_{\text{bonds}} \frac{k_i}{2}(r_i - r_{eq,i})^2 \\
+ & \sum_{\text{angles}} \frac{k_i}{2}(\theta_i - \theta_{eq,i})^2 \\
+ & \sum_{\text{torsions}} \frac{v_i}{2}\left[1 + \cos(n_i\omega_i - \gamma_i)\right] \\
+ & \sum_{\text{pairs i-j}} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \\
+ & \sum_{\text{pairs i-j}} \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}}
\end{aligned}
$$

Simple Averages

A Model for a Protein Solution

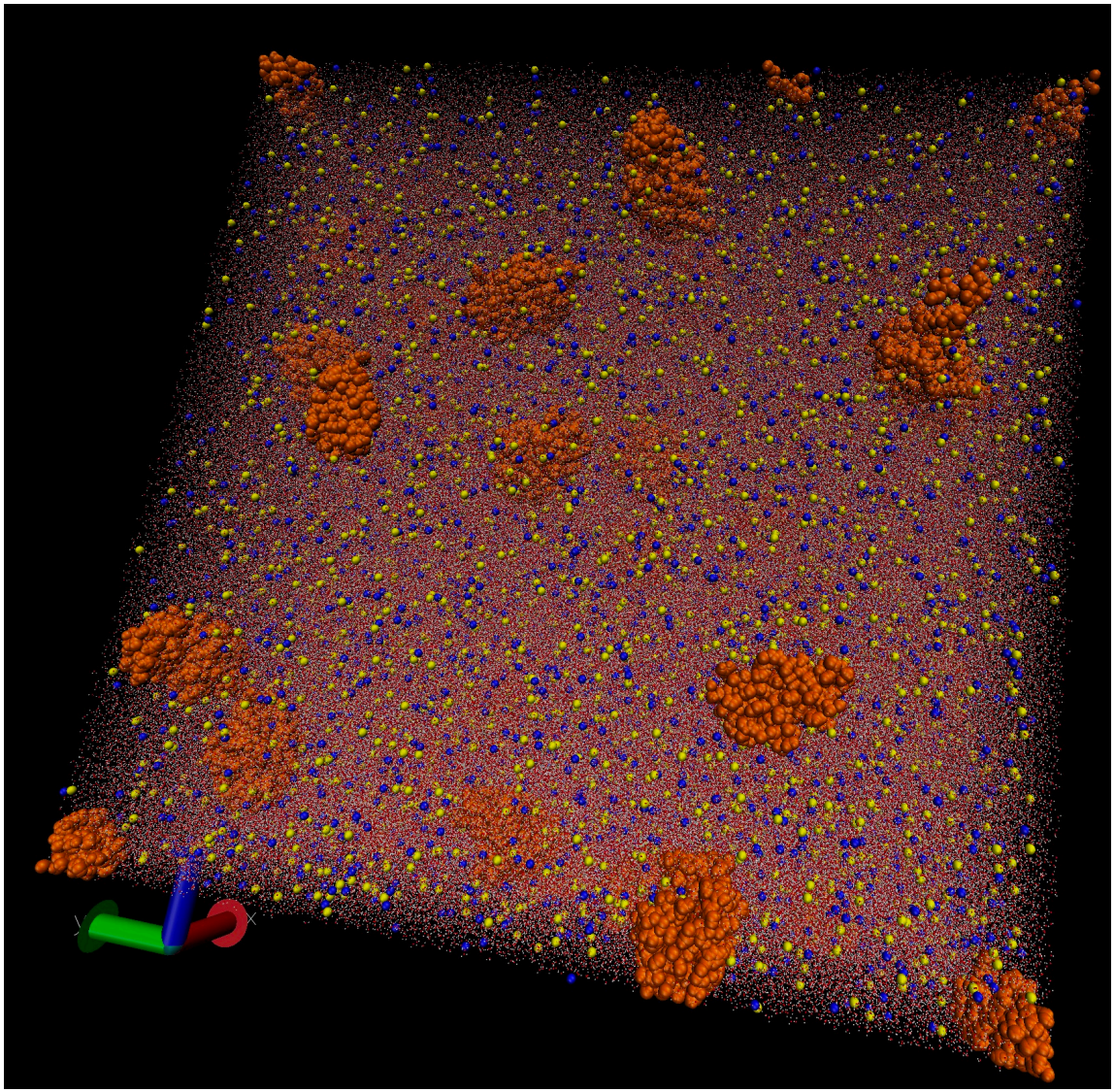
- solution:
  - pure water or
  - 0.25 mol/l NaCl in water
  - or any other solution of electrolyte
- 64 HEWL protein molecules
- altogether 3.4 million atoms
- running Gromacs with
  - Gromacs force field
  - Amber force field
- vary the nature of the electrolyte



Simple Averages

A Model for a Protein Solution

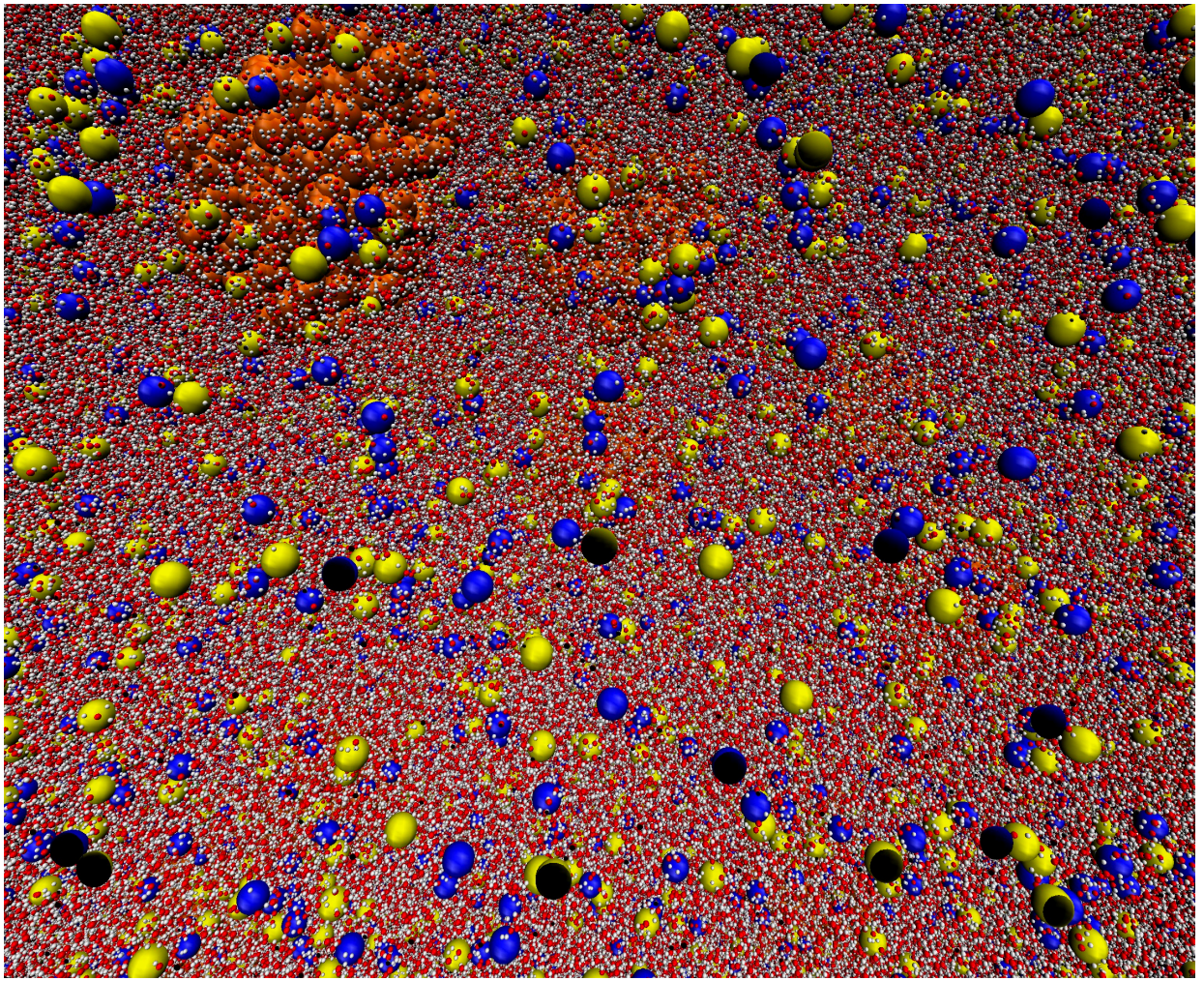
- solution:
  - pure water or
  - 0.25 mol/l NaCl in water
  - or any other solution of electrolyte
- 64 protein molecules
- altogether 3.4 million atoms
- running Gromacs with
  - Gromacs force field
  - Amber force field
- vary the nature of the electrolyte



Simple Averages

A Model for a Protein Solution

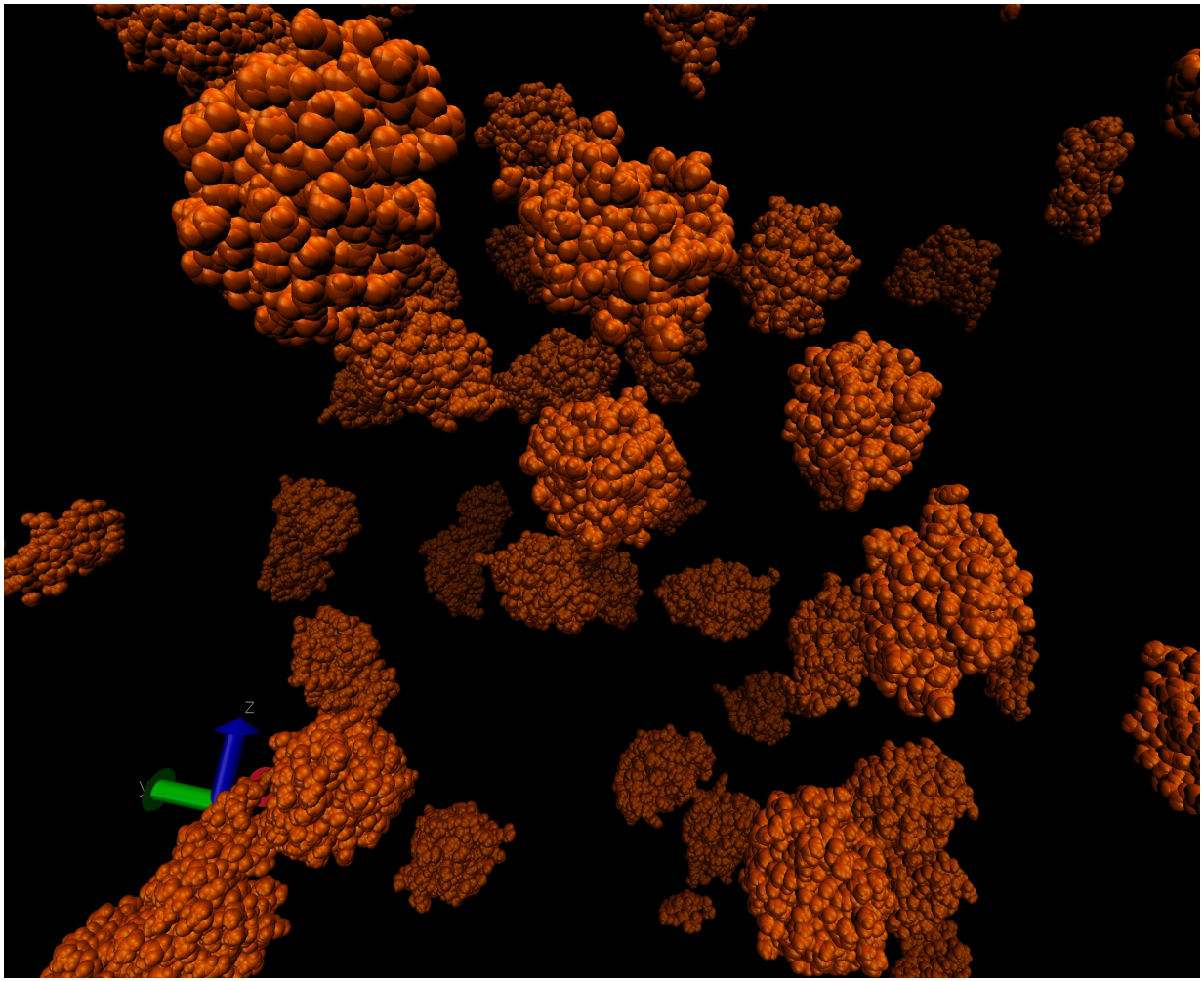
- solution:
  - pure water or
  - 0.25 mol/l NaCl in water
  - or any other solution of electrolyte
- 64 protein molecules
- altogether 3.4 million atoms
- running Gromacs with
  - Gromacs force field
  - Amber force field
- vary the nature of the electrolyte



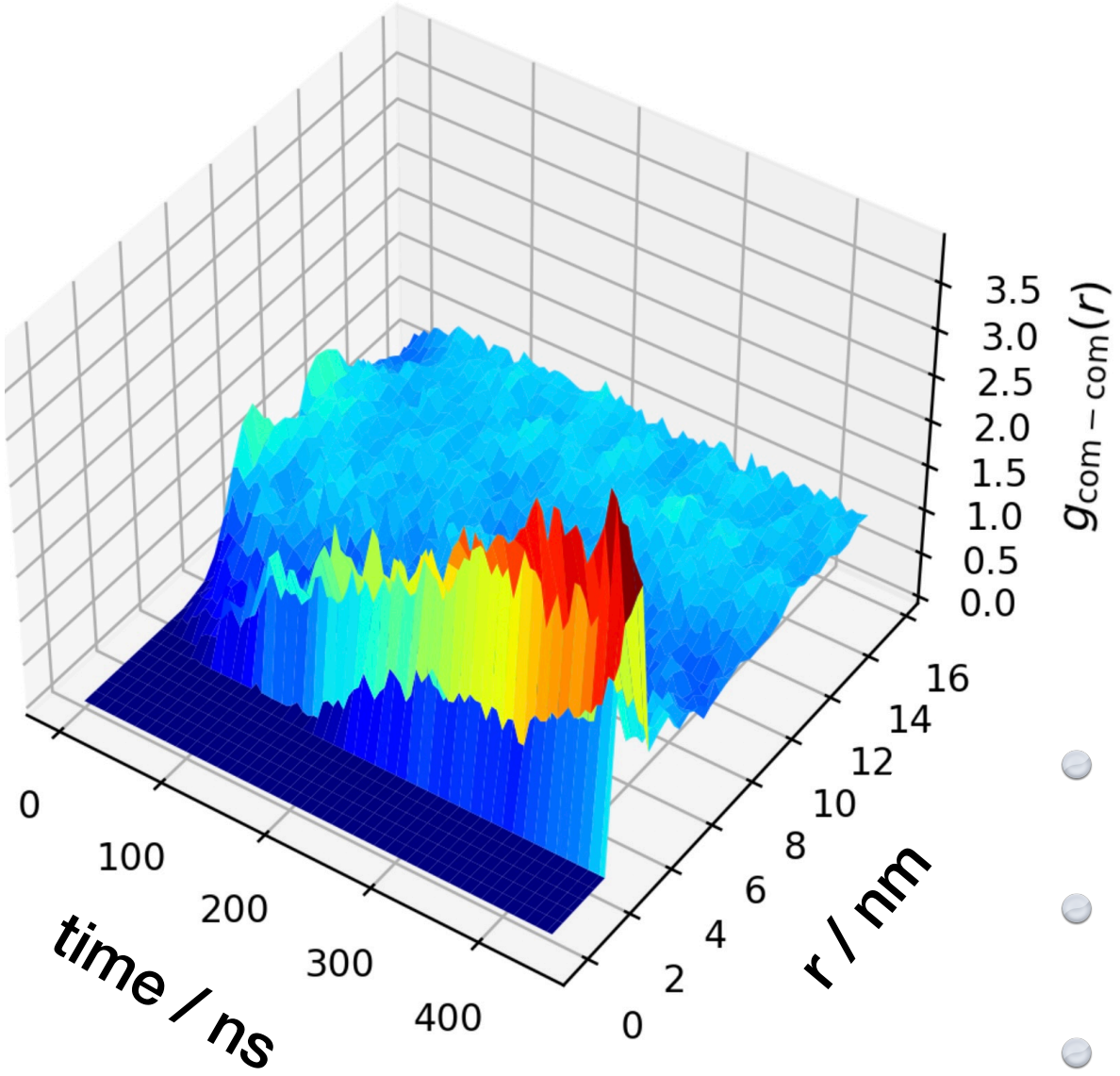
Simple Averages

A Model for a Protein Solution

- solution:
  - pure water or
  - 0.25 mol/l NaCl in water
  - or any other solution of electrolyte
- 64 protein molecules
- altogether 3.4 million atoms
- running Gromacs with
  - Gromacs force field
  - Amber force field
- vary the nature of the electrolyte



Simple Averages



radial distribution function

equilibration:

nuisance or objective?

- **Simulation was started from cubic 4x4x4 lattice of HEWL molecules**
- **Cubic structure vanishes over 50 ns**
- **Slow buildup of nearest neighbour structure over 400+ ns**

Atomistic Simulation?

- modeling of real physical (chemical / biological) systems
- a subdiscipline of **statistical mechanics**
- calculates thermodynamic properties
- follows the 'real' trajectory of a system
- studies structure, evolution thereof
- Molecular Dynamics: studies the dynamics of
  - individual atoms
  - molecules
  - clusters
  - structural elements
  - ….
  - parts of living cells
  - materials
- chemical reaction dynamics

Internal Energy

$$A = -kT \ln Q$$

$$\begin{aligned}
U &= A + TS \\
&= \frac{\int \mathrm{d}\Gamma E(\{q_k, p_k\}) \exp\left[-\frac{E(\{q_k,p_k\})}{kT}\right]}{\int \mathrm{d}\Gamma \exp\left[-\frac{E(\{q_k,p_k\})}{kT}\right]} \\
&= \frac{\int \mathrm{d}\Gamma \left(\mathcal{T}(\{p_k\}) + \mathcal{V}(\{q_k\})\right) \exp\left[-\frac{\mathcal{T}(\{p_k\})+\mathcal{V}(\{q_k\})}{kT}\right]}{\int \mathrm{d}\Gamma \exp\left[-\frac{\mathcal{T}(\{p_k\})+\mathcal{V}(\{q_k\})}{kT}\right]} \\
&= \langle \mathcal{T} \rangle + \frac{\int \mathrm{d}q_1 \mathrm{d}q_2 \ldots \mathrm{d}q_{3N} \mathcal{V}(\{q_k\}) \exp\left[-\frac{\mathcal{V}(\{q_k\})}{kT}\right]}{Z_N} \\
&= \langle \mathcal{T} \rangle + \langle \mathcal{V} \rangle \quad = \quad \frac{3N}{2} kT + \langle \mathcal{V} \rangle
\end{aligned}$$

internal energy is a simple average: a single simulation suffices

Free Energy

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

$$= -kT \ln \left\{ \frac{\int d\Gamma \exp\left[-\mathcal{H}_Y(\{q_k, p_k\})/kT\right]}{\int d\Gamma \exp\left[-\mathcal{H}_X(\{q_k, p_k\})/kT\right]} \right\}$$

$$= -kT \ln \left\{ \frac{\int d\Gamma \exp\left[-\mathcal{H}_Y/kT\right] \exp\left[+\mathcal{H}_X/kT\right] \exp\left[-\mathcal{H}_X/kT\right]}{\int d\Gamma \exp\left[-\mathcal{H}_X(\{q_k, p_k\})/kT\right]} \right\}$$

$$= -kT \ln \left\{ \frac{\int d\Gamma \exp\left[-\mathcal{H}_Y/kT\right] \exp\left[+\mathcal{H}_X/kT\right] \exp\left[-\mathcal{H}_X/kT\right]}{\int d\Gamma \exp\left[-\mathcal{H}_X(\{q_k, p_k\})/kT\right]} \right\}$$

$$= -kT \ln \left\{ \int d\Gamma \exp\left[-\mathcal{H}_Y/kT\right] \exp\left[+\mathcal{H}_X/kT\right] \rho_X(\{q_k, p_k\}) \right\}$$

$$= -kT \ln \left\langle \exp\left[-(\mathcal{H}_Y - \mathcal{H}_X)/kT\right] \right\rangle_X$$

free energy: exponential average (Boltzmann factor): difficult to converge

Free Energy: exponential averages

$$A = -kT \ln Q$$

$$= -kT \ln \int \int \mathrm{d}q_1 \mathrm{d}q_2 \ldots \mathrm{d}q_{3N} \int \mathrm{d}p_1 \mathrm{d}p_2 \ldots \mathrm{d}p_{3N} \exp\left[-\mathcal{H}(\{q_k, p_k\})/kT\right]$$

$$\Delta A = A_Y - A_X$$

$$= -kT \ln \langle \exp\left[-(\mathcal{H}_Y - \mathcal{H}_X)/kT\right] \rangle_X$$

$$-\Delta A = A_X - A_Y$$

$$= -kT \ln \langle \exp\left[-(\mathcal{H}_X - \mathcal{H}_Y)/kT\right] \rangle_Y$$

===> a new 'discipline': computational alchemy

Transition State Theory

- important concept in physical chemistry
- basic assumption
  - A chemical reaction proceeds to completion once the
  - activation barrier (maximum of the minimum energy path)
- has been reached
- probability proportional to Boltzmann factor $e^{-\Delta Ea/kT}$
- if activation energy ΔE_a is large, getting there is a rare event
- ??? (get there stepwise, somehow) ???
- dynamic corrections: Transmission factor $\kappa < 1$

Free Energy Differences

- free energy differences between 2 chemical compounds, e.g.

X = ethanol (C_2H_5OH) and Y = thioethanol (C_2H_5SH

- Problem:

  distribution of states X and Y in phase space do NOT overlap

  (S is 'larger' than 'O')

  ===> bad (inefficient) sampling

This problem typically occurs when for a given phase space point

$|H_X - H_Y| >> kT$

problem solution: simulate intermediate steps

Free Energy Differences

$$X_1 = \text{ethanol} \qquad Y_1 = 0.9 * \text{Ethanol} + 0.1 * \text{Ethanthiol}$$

$$X_2 = 0.9 * \text{Ethanol} + 0.1 * \text{Ethanthiol} \qquad Y_2 = 0.8 * \text{Ethanol} + 0.2 * \text{Ethanthiol}$$

$$X_3 = 0.8 * \text{Ethanol} + 0.2 * \text{Ethanthiol} \qquad Y_3 = 0.7 * \text{Ethanol} + 0.3 * \text{Ethanthiol}$$

$$\ldots$$

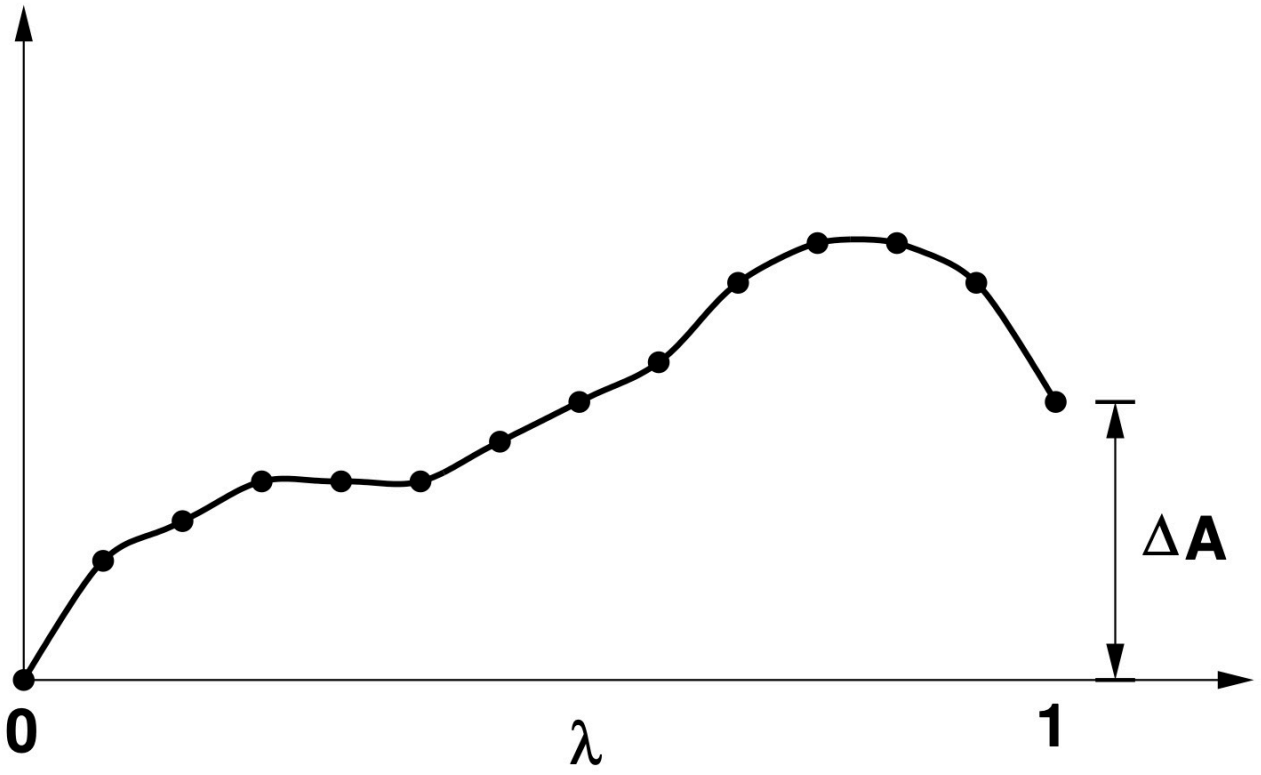
$$X_{11} = 0.1 * \text{Ethanol} + 0.9 * \text{Ethanthiol} \qquad Y_{11} = \text{Ethanthiol}$$

$$\begin{aligned} \Delta A &= A(Y_1) - A(X_1) + A(Y_2) - A(X_2) + \ldots \\ &= A(Y_{11}) - A_(X_1) \end{aligned}$$

- generate unphysical (intermediate) ensembles
- need to interpolate Hamiltonian somehow

Free Energy Differences

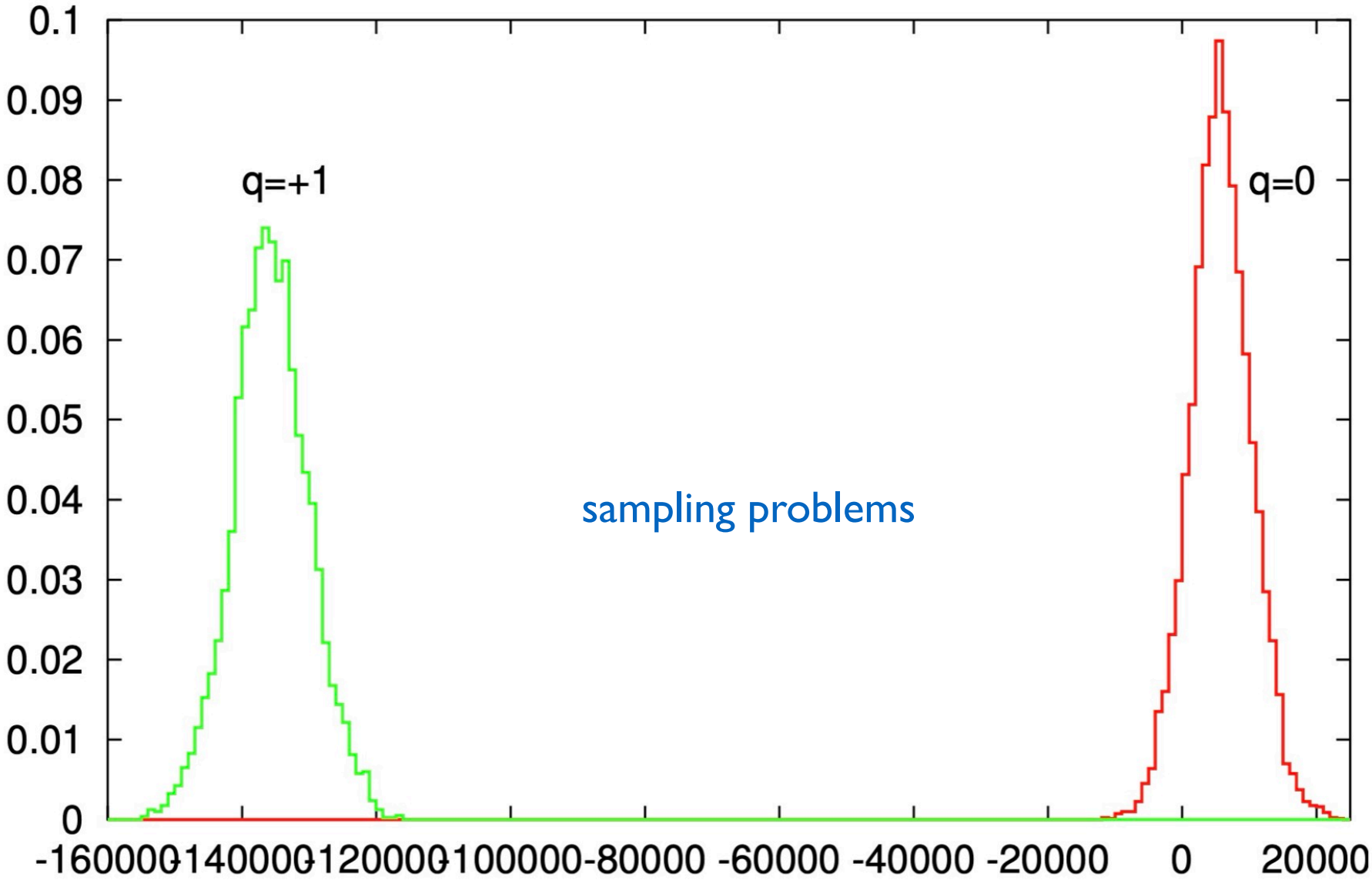
$\Delta A(\lambda_i \rightarrow \lambda_{i+1})$	"forward sampling"
$\Delta A(\lambda_i \rightarrow \lambda_{i-1})$	"backward sampling"
beide	"double-wide sampling"



Example: Discharge of Na^+ (Born)

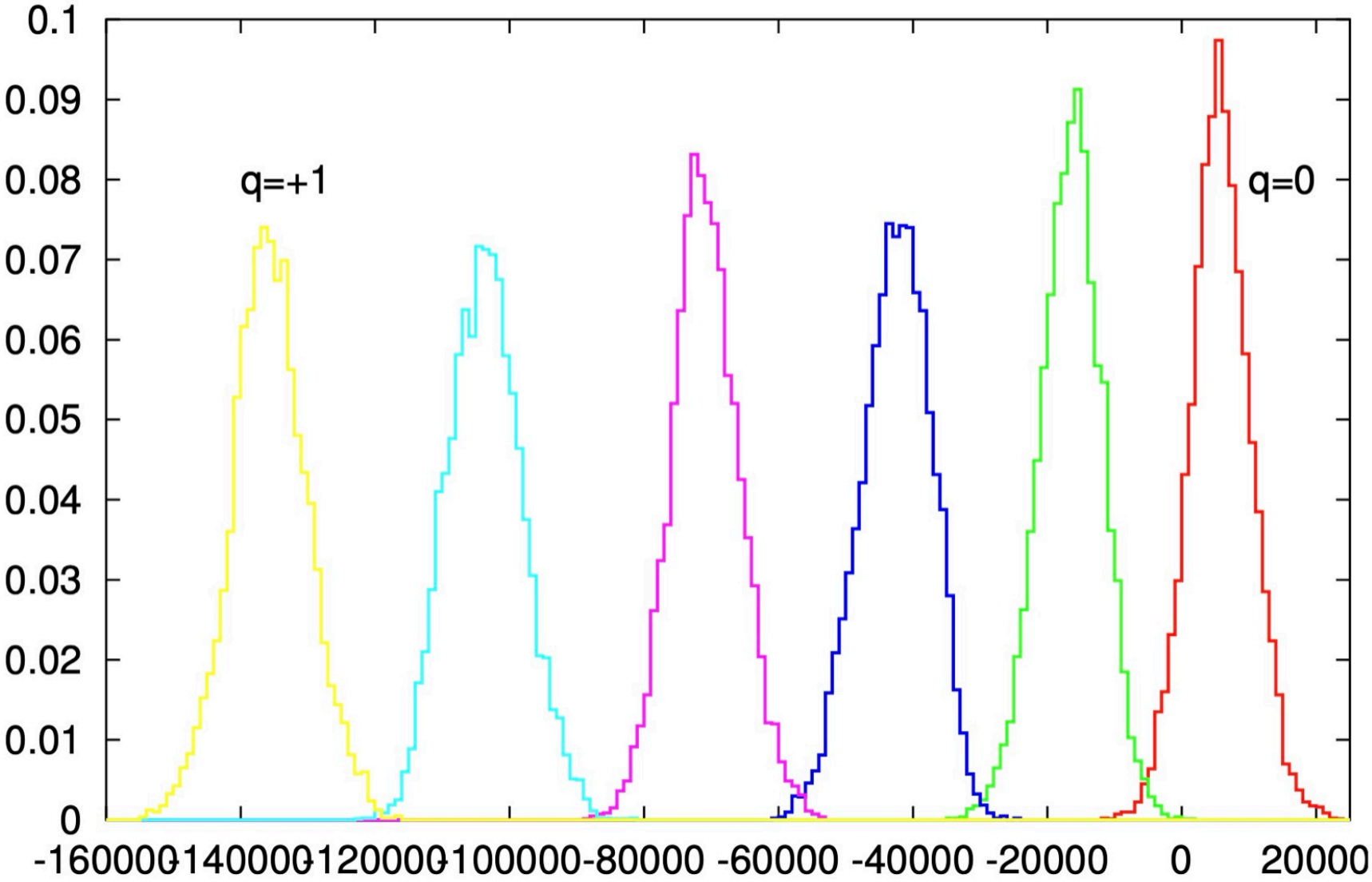
- generate 11 simulations where the sodium ion has charge 1e, 0.9e, 0.8e, … 0.1e, 0e
- in each simulation, water solvates the partially charged ion differently, creating a different electrostatic potential (and distribution)
- the potential distribution drives the charging/discharging of the ion (adiabatic principle)
- these 11 simulations can be analysed in 3 series:
- q = 0, 1 $\Delta H = \phi$
- q = 0, 0.2, 0.4, 0.6, 0.8, 1.0 $\Delta H = 0.2\phi$
- q = 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 $\Delta H = 0.1\phi$
- ΔH —> kT

Electrostatic Potential Histograms



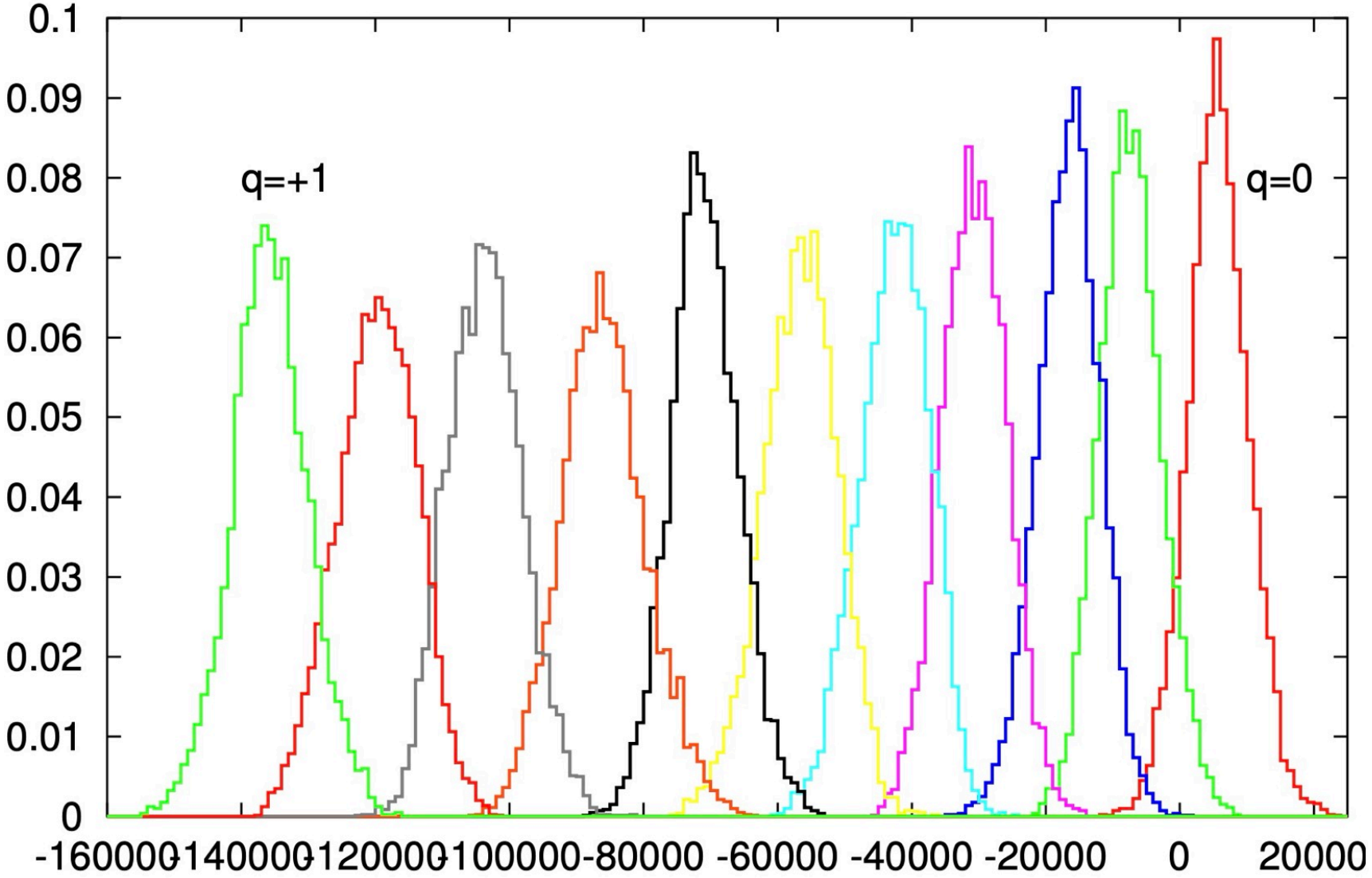
no overlap between initial and final configurations: impossible to calculate exp. average

Potential Histograms



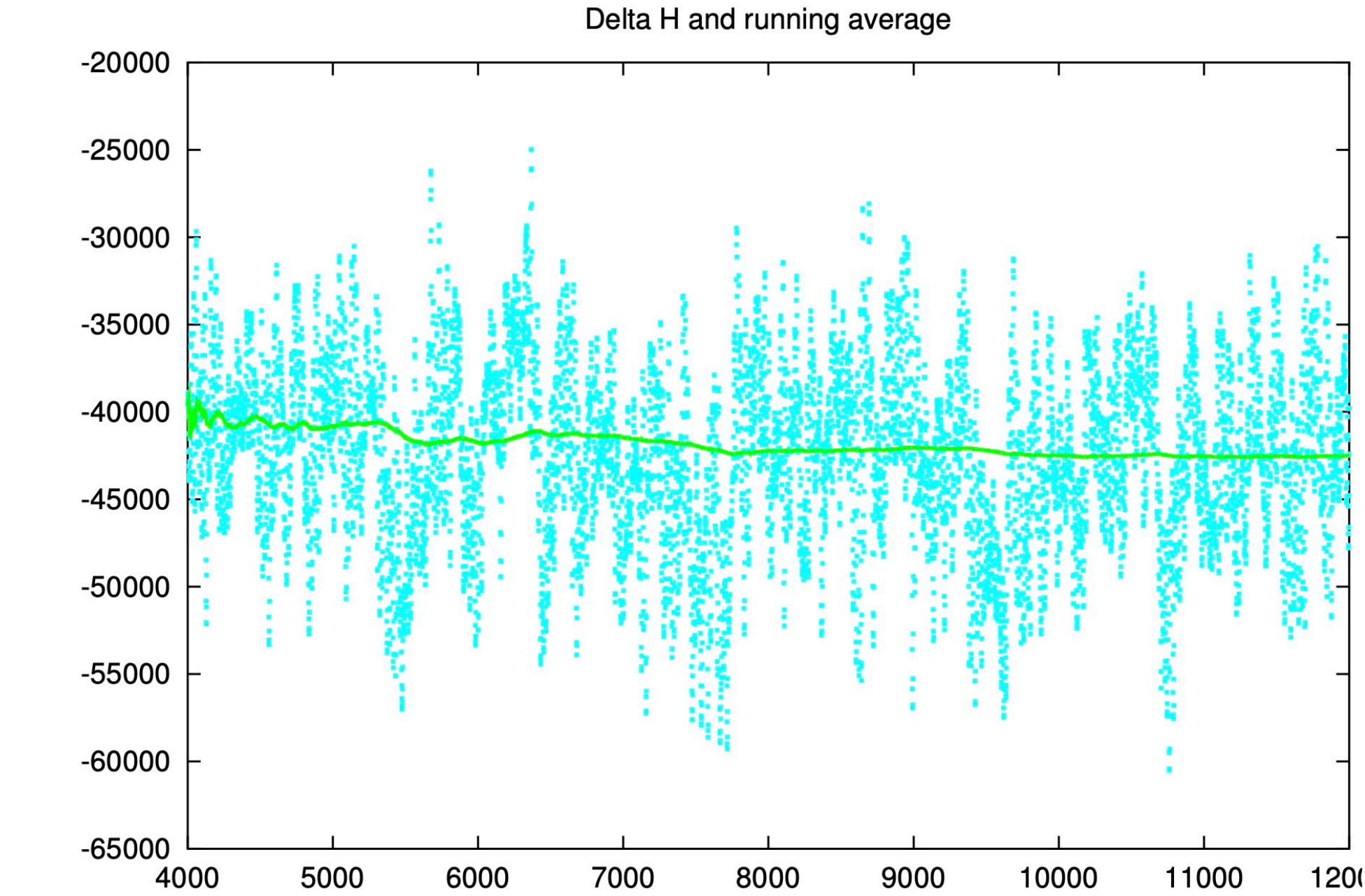
almost no overlap between initial and final configurations: average will be poor!

Potential Histograms

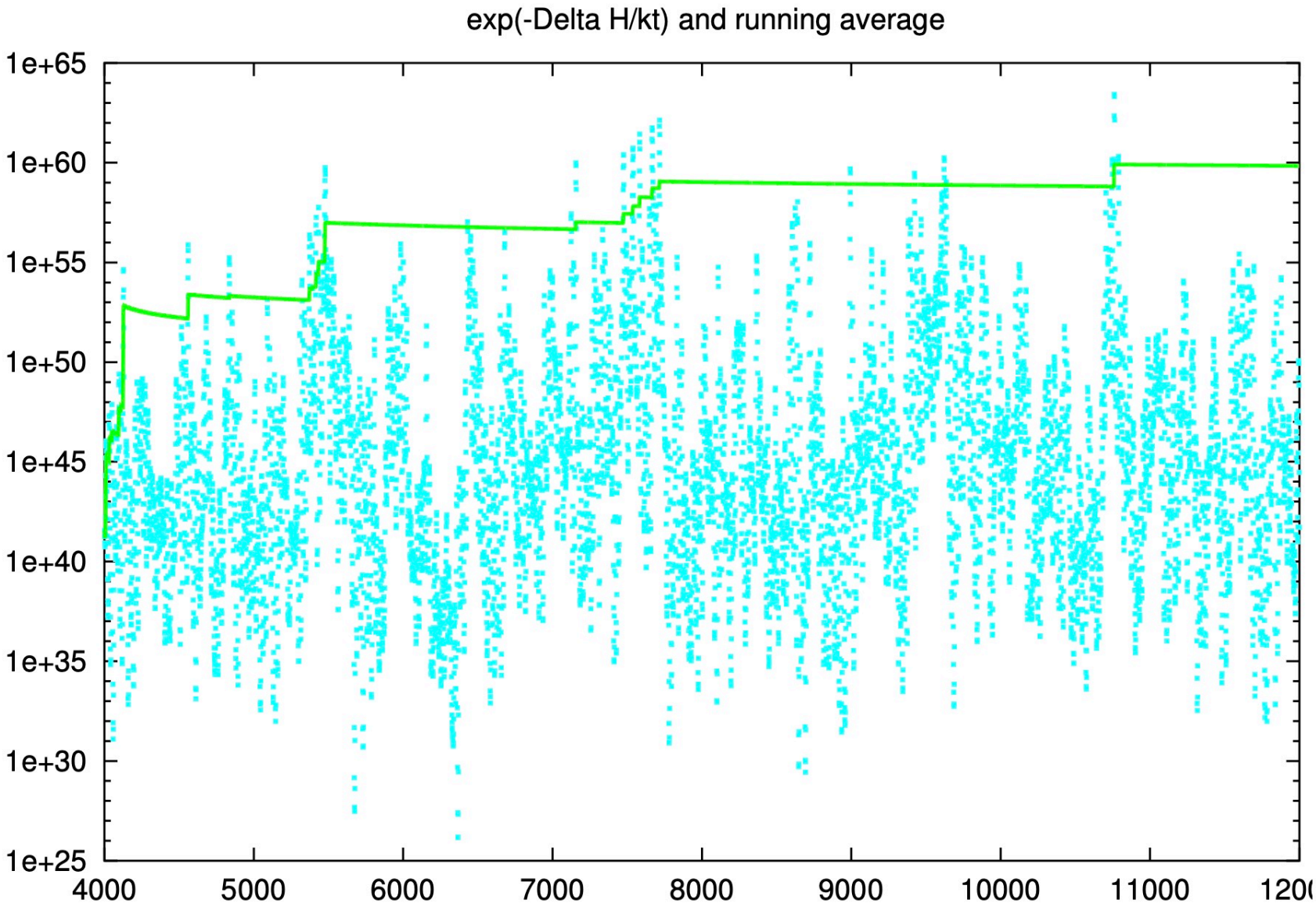


sufficient overlap to calculate decent average

Running Averages

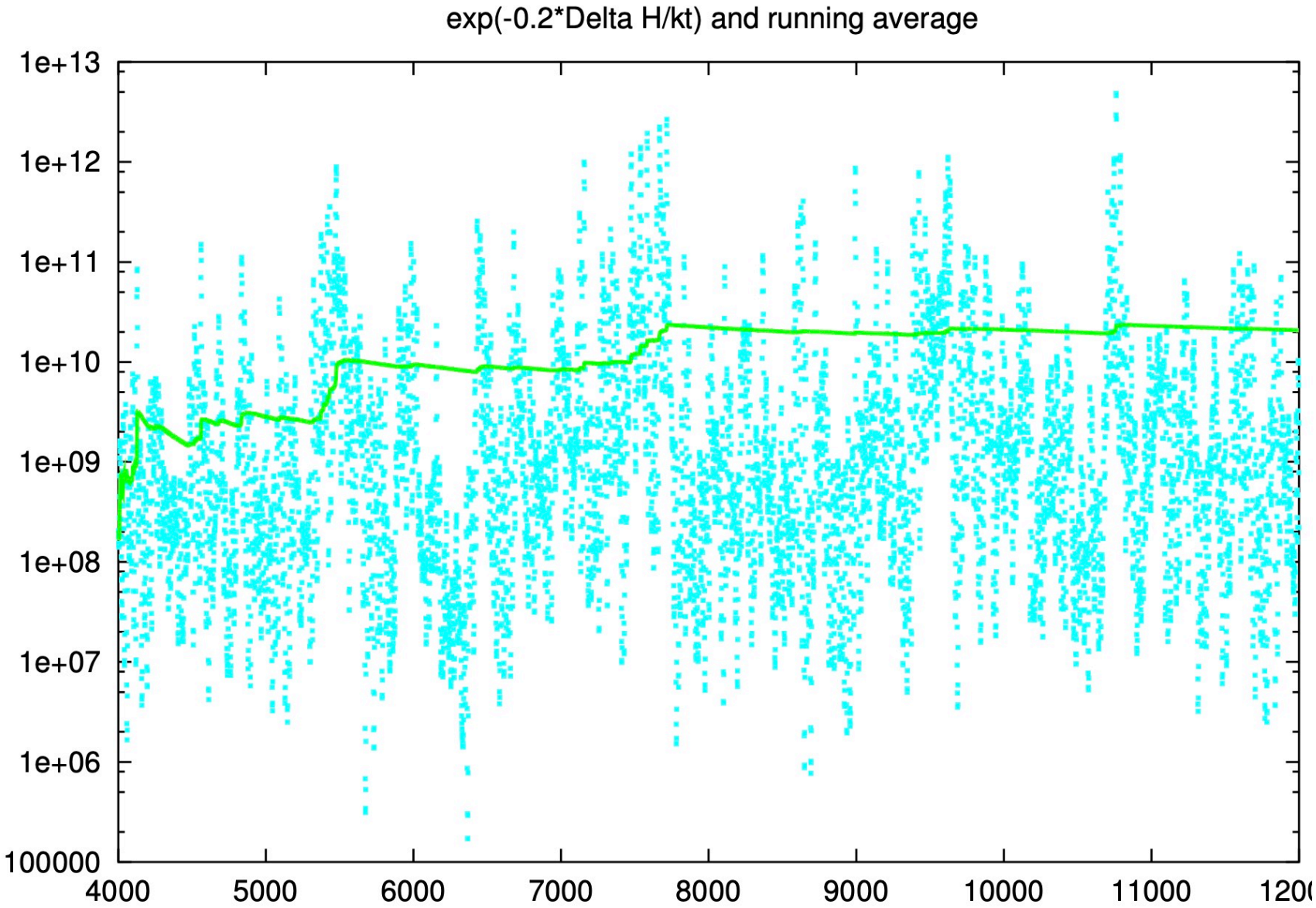


Running Averages



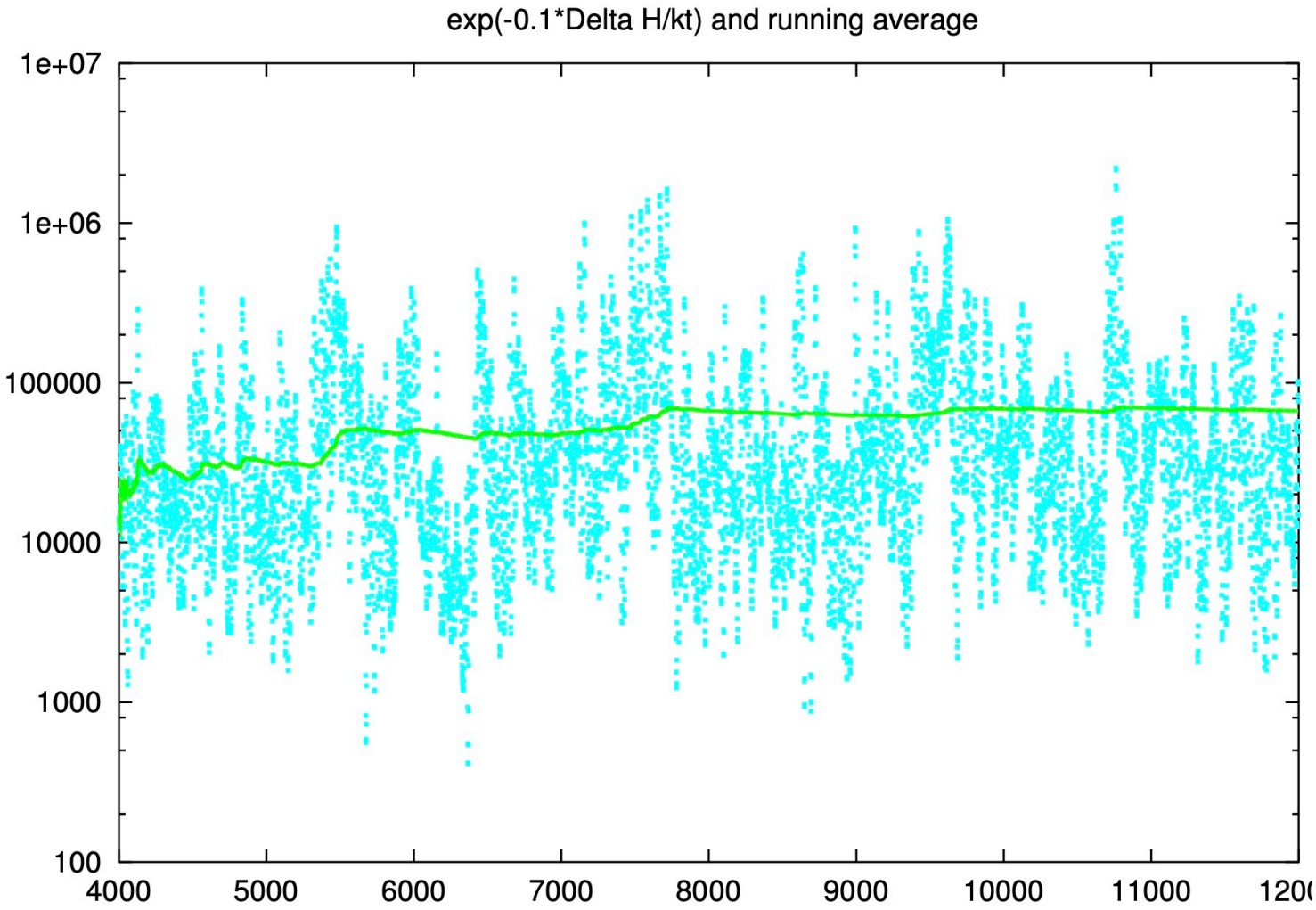
few configurations dominate the average!

Running Averages



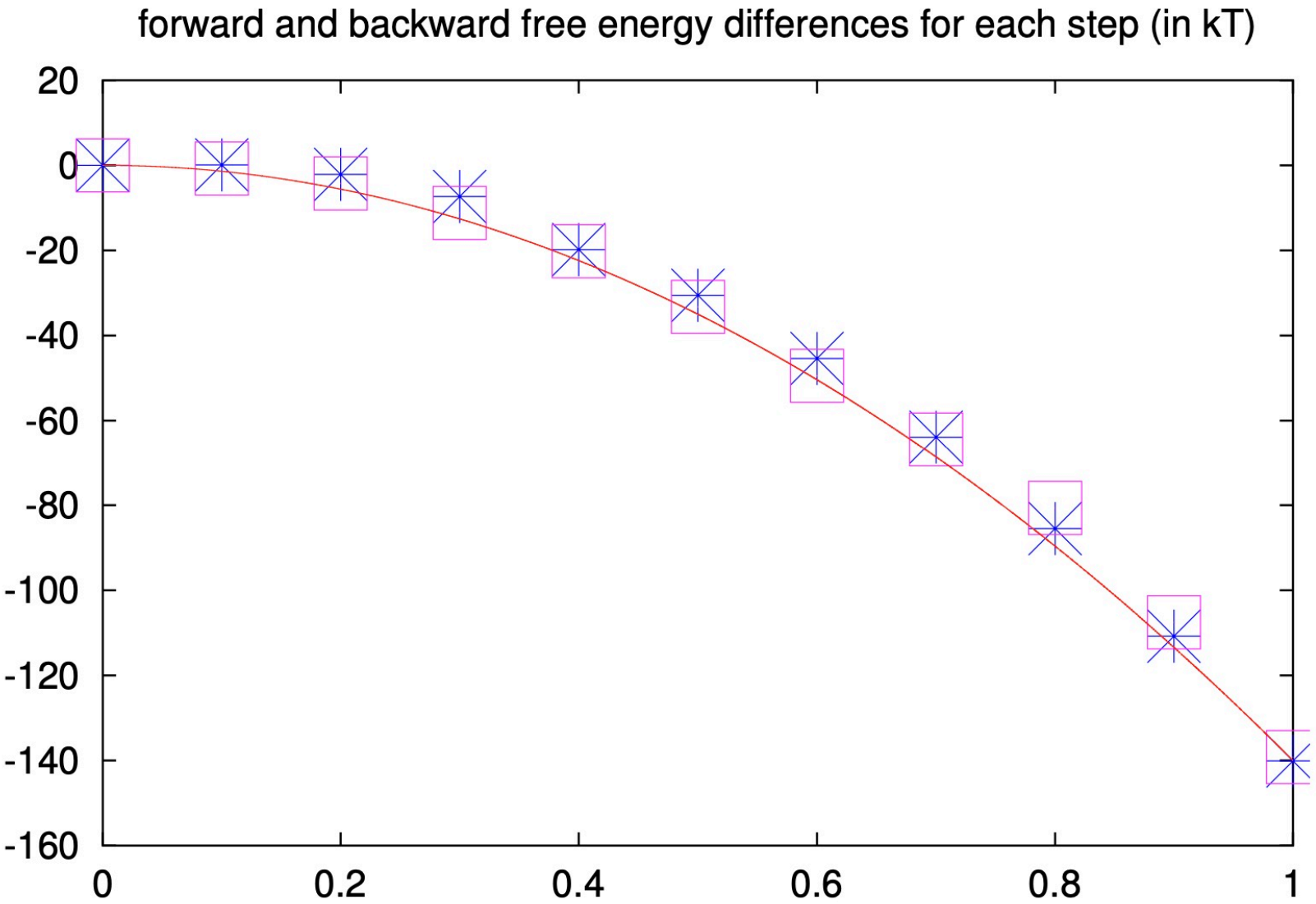
still a few configurations dominate the average!

Running Averages



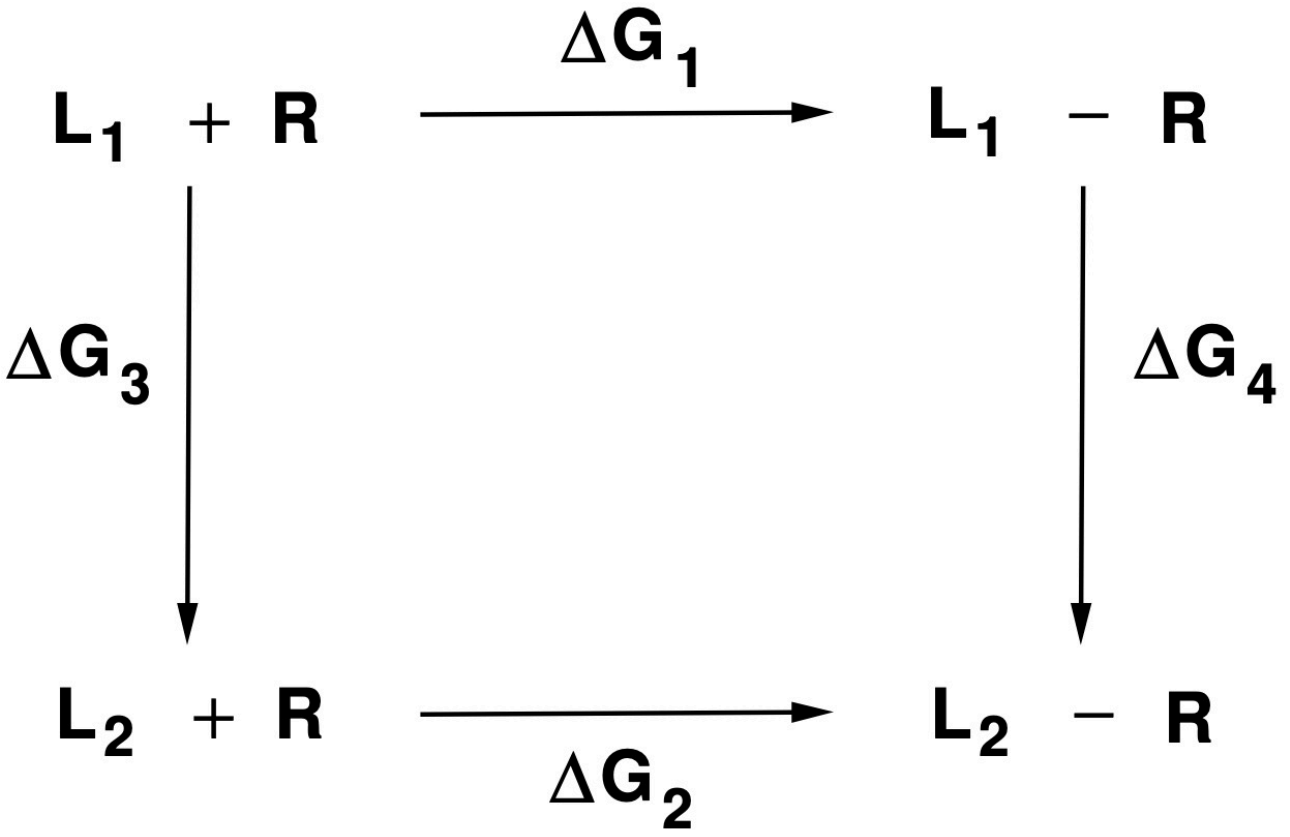
many configurations determine average

Running Averages



- realisation of the Born model

Mutation Simulations



$$\Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3$$

frequently used technique in drug design studies

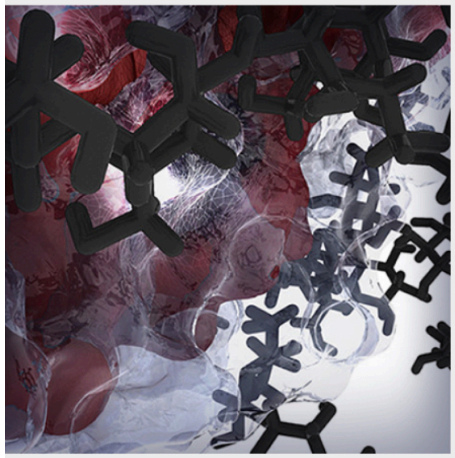
Atomistic Simulation?

- modeling of real physical (chemical / biological) systems
- a subdiscipline of statistical mechanics
- calculates thermodynamic properties
- follows the 'real' trajectory of a system
- studies structure, evolution thereof
- Molecular Dynamics: studies the **dynamics** of
  - individual atoms
  - molecules
  - clusters
  - structural elements
  - ….
  - parts of living cells
  - materials
- chemical reaction dynamics

Protein Agglomeration ('Crowding')

VIRTUAL ISSUE

Protein Crowding and Stability



Protein structure determines function, but prote
molecules are shaped *in vivo* where biology tak
environments: a significant molecular diversity
macromolecules. This leads to a distracting mo
against which biological function must be main
crowding effects and may also facilitate aggreg
how biological environments affect protein stab
Read more

- mechanisms of protein aggregation
- RDFs
- specific interactions
- classical MD

Protein Agglomeration

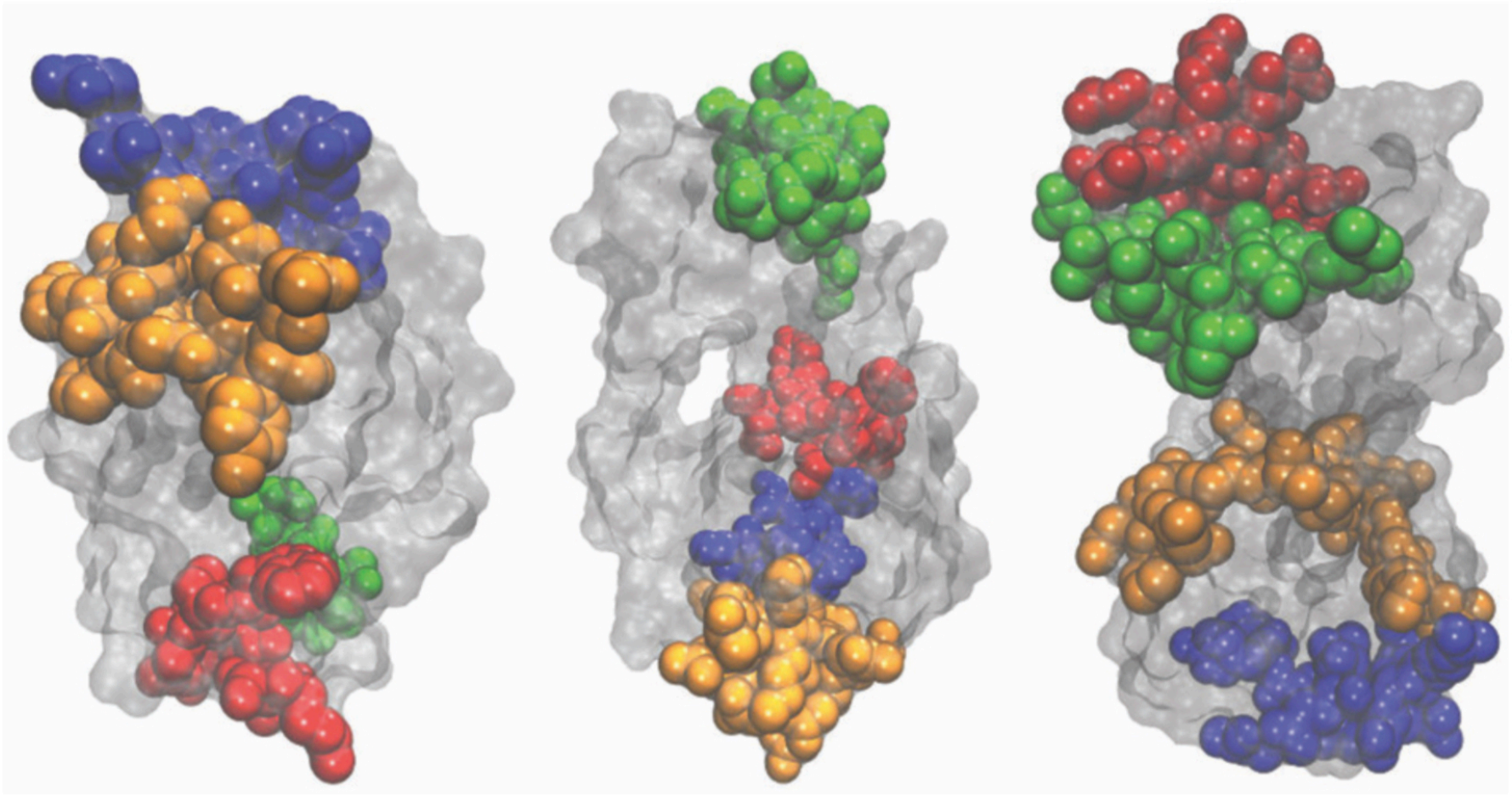
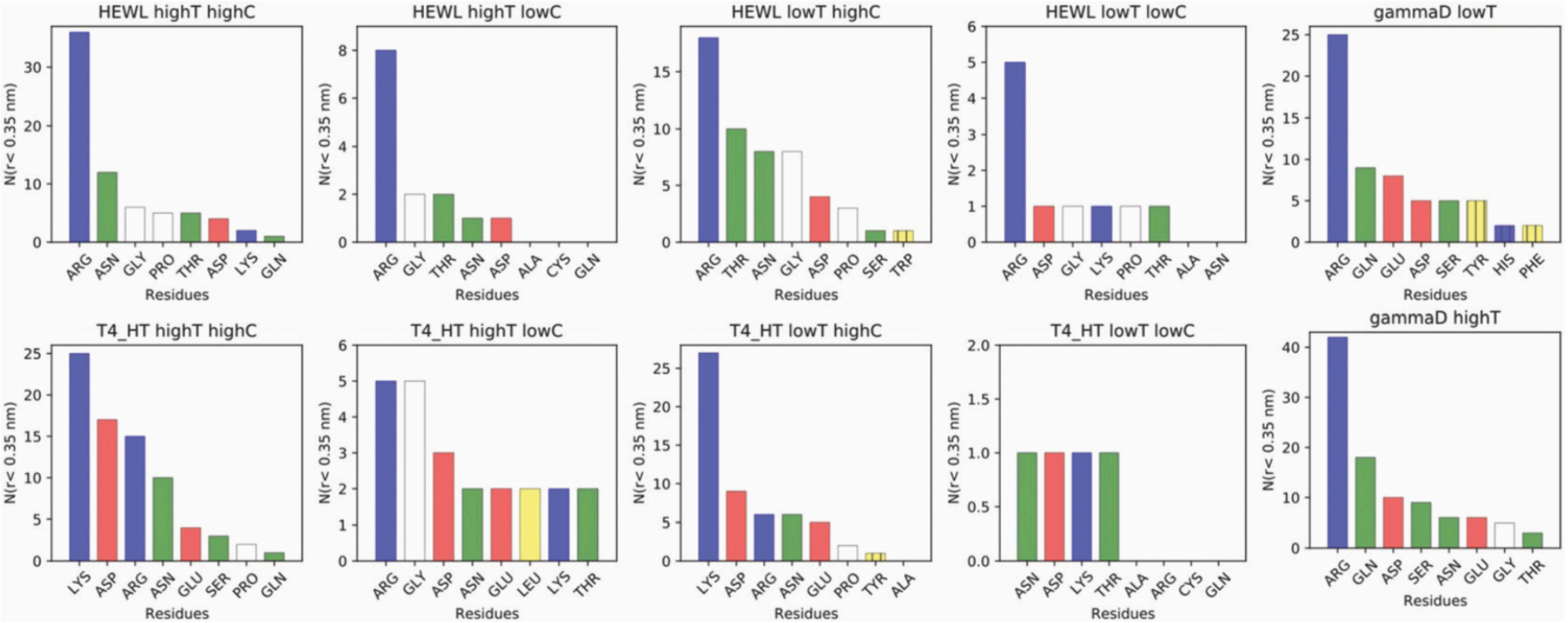


Fig. 5 Four regions of protein molecules that most frequently participate in protein–protein contacts for HEWL (left), T4 WT* (middle), and γ-D crystallin (right).

S. Brudar, J. Gujt, E. Spohr, and B. Hribar-Lee, PCCP 23, 415, (2021)

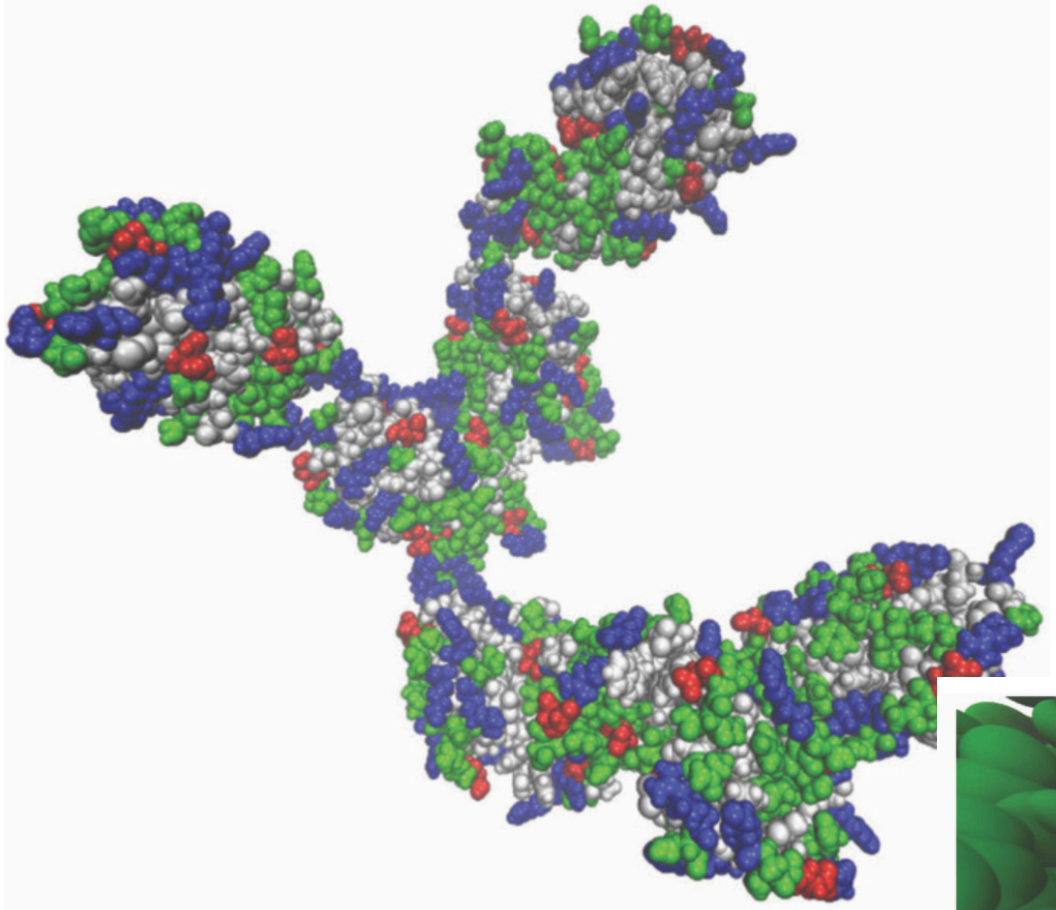
Protein Agglomeration

statistics on cluster-'initializing' polar amino acid groups



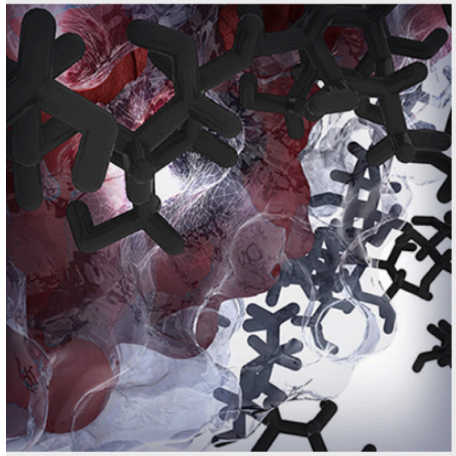
S. Brudar, J. Gujt, E. Spohr, and B. Hribar-Lee, PCCP 23, 415, (2021)

Protein Agglomeration

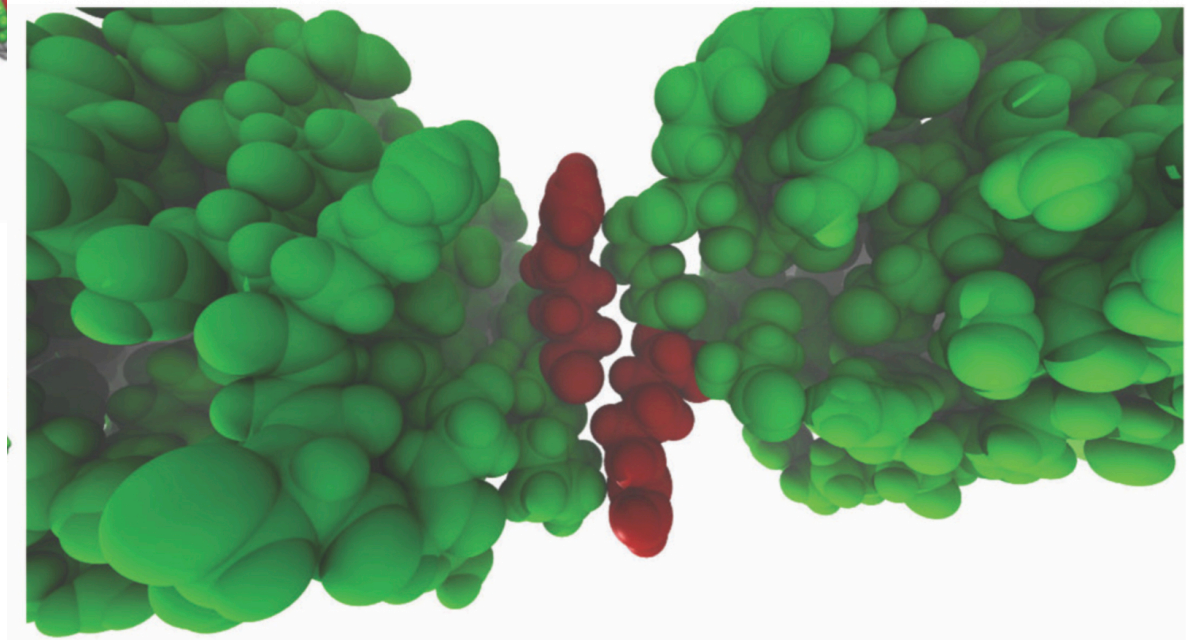


VIRTUAL ISSUE

Protein Crowding and Stability



Protein structure determines function, but prote
molecules are shaped *in vivo* where biology tak
environments: a significant molecular diversity
macromolecules. This leads to a distracting mo
against which biological function must be main
crowding effects and may also facilitate aggreg
how biological environments affect protein stab
Read more



Arginine Solution

OPLS force field

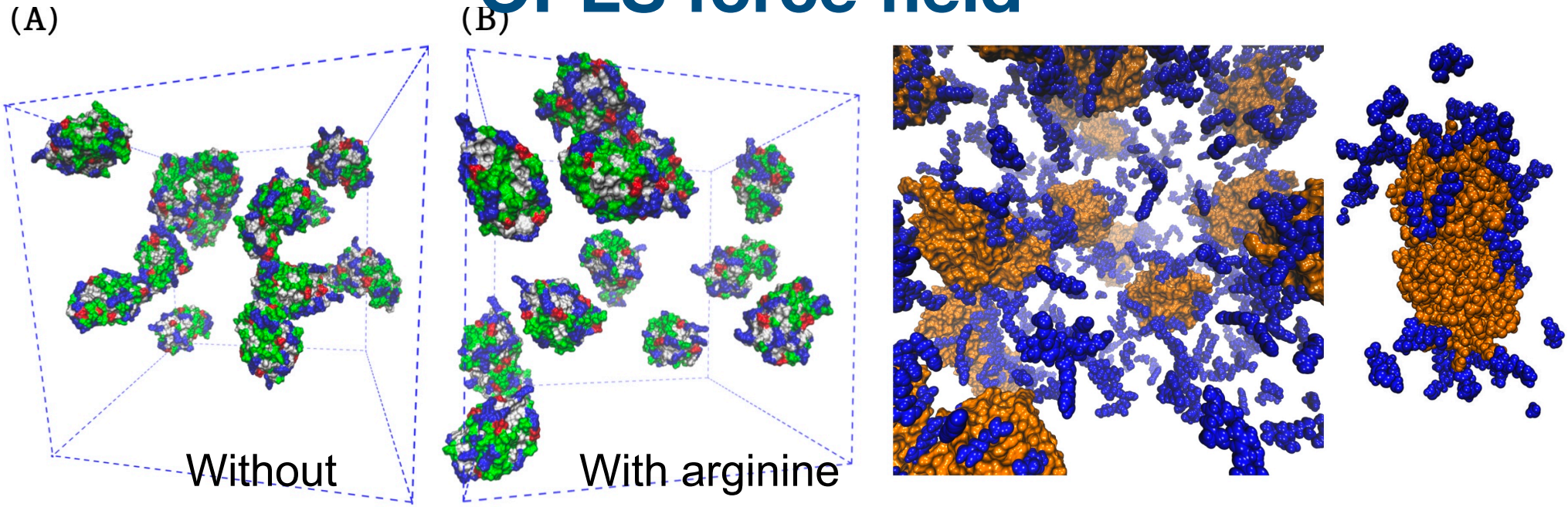


Figure 4. Snapshots from the computer simulations showing just HEWL molecules (**A**) without the presence of free arginine and (**B**) with added 0.3 M arginine.

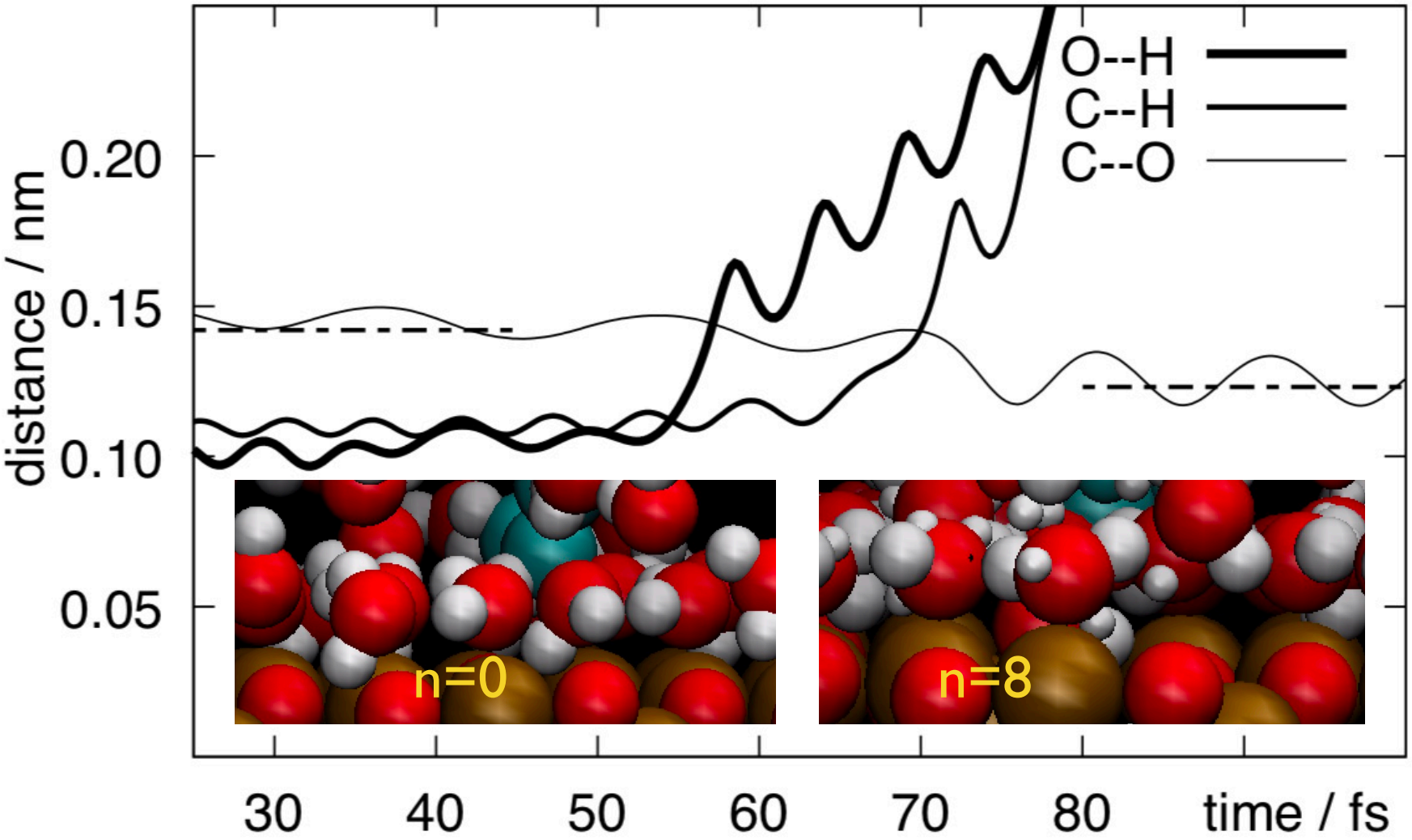
- Experiments show that the addition of arginine increases the stability of the HEWL solutions.
- The computer simulation results indicate that arginine molecules tend to self-associate
- If arginine residues are located on the protein surface, the free arginine molecules stay in their vicinity and prevent protein molecules from “connecting” through them to form clusters.

[S.Brudar,B. Hribar Lee, Int. J. Mol. Sci. **2023**,24,1197]

Atomistic Simulation?

- modeling of real physical (chemical / biological) systems
- a subdiscipline of statistical mechanics
- calculates thermodynamic properties
- follows the 'real' trajectory of a system
- studies structure, evolution thereof
- Molecular Dynamics: studies the dynamics
  - individual atoms
  - molecules
  - clusters
  - structural elements
  - ….
  - parts of living cells
  - materials
- chemical **reaction dynamics**

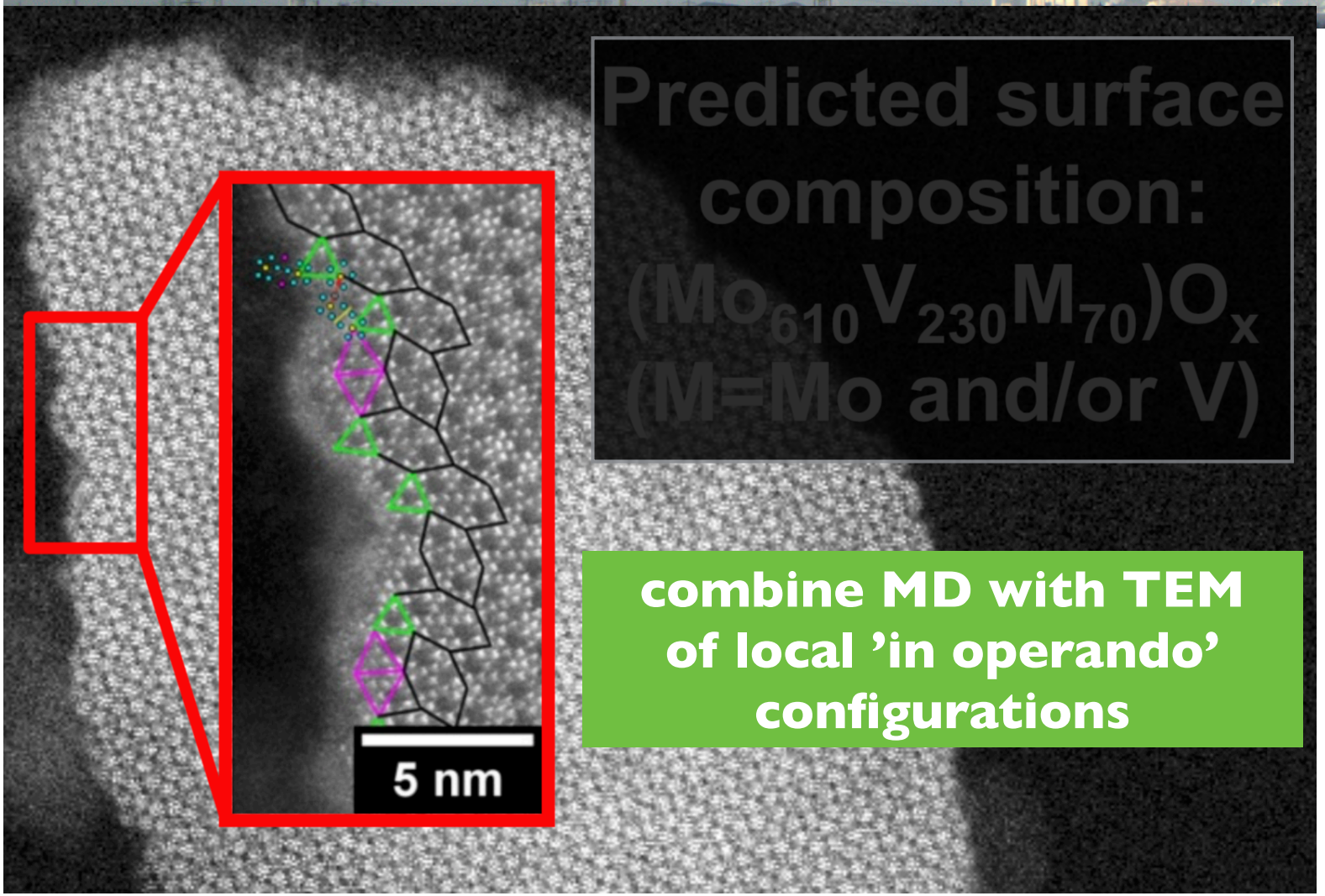
‘Electrochemical’ Approach



Computer Experiment

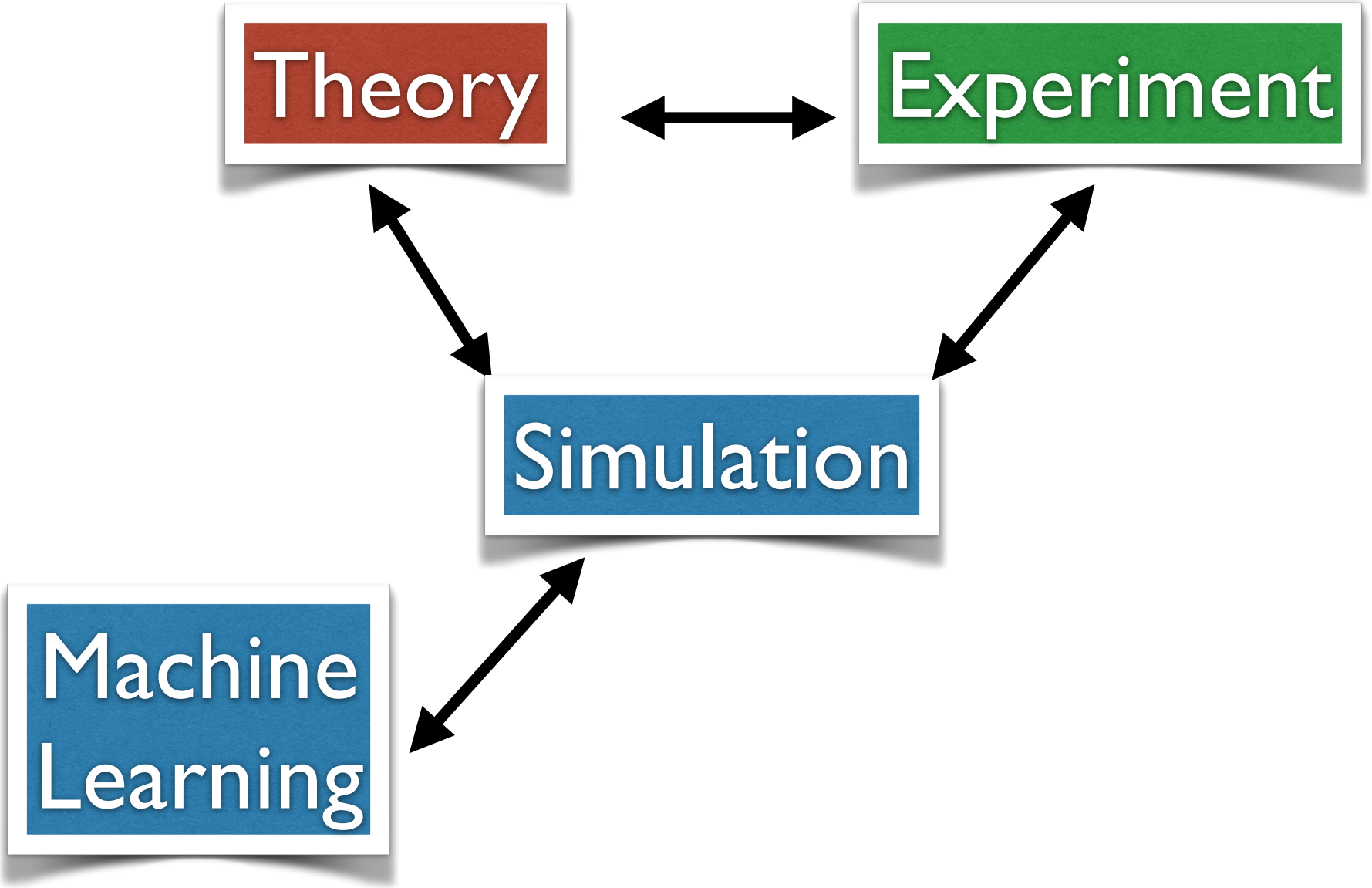
- remove n H atoms (from water molecules or from surface OH)

Perspective



- aqueous environment and proton transfer modulate (electro,photo)catalytic rxns
- local structure is essential
- transition metal atoms: 'buffer' for electrons
- aqueous phase: 'buffer' for protons

Perspective



Thank You!