Enhancing the sampling

How to save time, and time is money

Marcus Elstner and Tomáš Kubař

June 12, 2015

Problem

with normal nanosecond length MD simulations:

It is difficult to overcome barriers to conformational transitions,

only conformations around the initial structure may be sampled,

even if a different conformation is more likely – has lower ΔG

Special techniques are required to solve this problem.

Finding the global minimum of energy

MD may also be used for optimization

Assume a set of N atoms with many possible configurations
– this is truly the case with large (bio)molecules

The energy of these configurations is in general different,

- one of them will be the lowest;
- each of the configurations is a local minimum of energy
- separated from every other by an energy barrier

"A molecular dynamics primer" by Furio Ercolessi, University of Udine, Italy

Finding the global minimum of energy

- the most favorable structure
- tricky with traditional minimization techniques
  (steepest-descents, conjugate gradients, etc.)
- energy barriers cannot be overcome at all,
  the system falls into the nearest local minimum

Finding the global minimum of energy

- the most favorable structure
- tricky with traditional minimization techniques
  (steepest-descents, conjugate gradients, etc.)
- energy barriers cannot be overcome at all,
  the system falls into the nearest local minimum
- possible solution – try out several different starting points,
  hopefully in the neighborhood of different local minima,
  from which one would hopefully be the global

Finding the global minimum of energy

- the most favorable structure
- tricky with traditional minimization techniques
  (steepest-descents, conjugate gradients, etc.)
- energy barriers cannot be overcome at all,
  the system falls into the nearest local minimum
- possible solution – try out several different starting points,
  hopefully in the neighborhood of different local minima,
  from which one would hopefully be the global
- we cannot be really sure if we will find the global minimum

Simulated annealing

- key to overcome barriers in MD or MC – temperature
- state with energy E visited with probability (frequency)

$$\mathcal{P} \propto \exp\left[-\frac{E}{k_B T}\right]$$

Simulated annealing

- key to overcome barriers in MD or MC – temperature
- state with energy E visited with probability (frequency)

$$\mathcal{P} \propto \exp\left[-\frac{E}{k_B T}\right]$$

- if T large – many different minima populated
- what if we decrease T slowly to zero?

    system will be trapped in the deepest minimum possibly

Simulated annealing

- key to overcome barriers in MD or MC – temperature
- state with energy E visited with probability (frequency)

$$\mathcal{P} \propto \exp\left[-\frac{E}{k_B T}\right]$$

- if T large – many different minima populated
- what if we decrease T slowly to zero?
  system will be trapped in the deepest minimum possibly
- principle of simulated annealing:
- system is equilibrated at a certain temperature
- and then slowly cooled down to $T = 0$

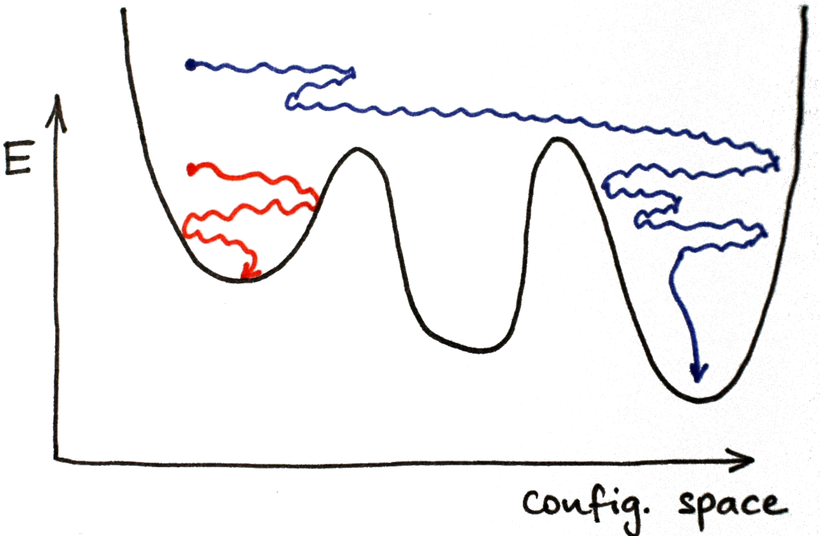
Simulated annealing

- key to overcome barriers in MD or MC – temperature
- state with energy E visited with probability (frequency)

$$\mathcal{P} \propto \exp\left[-\frac{E}{k_B T}\right]$$

- if T large – many different minima populated
- what if we decrease T slowly to zero?
  system will be trapped in the deepest minimum possibly
- principle of simulated annealing:
- system is equilibrated at a certain temperature
- and then slowly cooled down to $T = 0$
- no formal guarantee of success, but it often works
- no a priori assumptions / no intuition needed

Simulated annealing



Simulated annealing

– much more generally useful for optimization:

given an objective function $Z(\alpha_1, \ldots, \alpha_N)$ of N parameters, one can regard each of these parameters as a degree of freedom, assign it a "mass", and let the system evolve with MD or MC to perform simulated annealing.

an early application – problem of the traveling salesman

Kirkpatrick et al., Science 1983

Molecular dynamics with quenching

yet another possibility to make use of MD

- not only to get the minima of the energy,
- but even to approximate their relative free energies

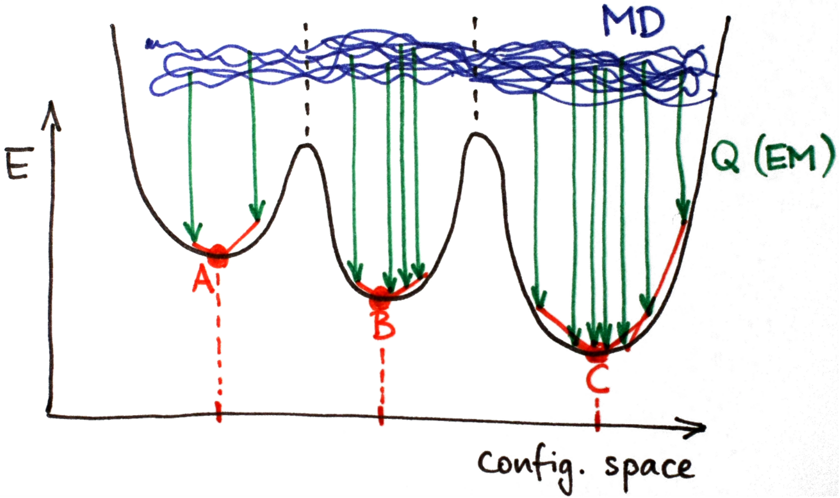
Molecular dynamics with quenching

yet another possibility to make use of MD
not only to get the minima of the energy,
but even to approximate their relative free energies

MD/quenching simulation

- make a usual MD simulation
- in regular intervals, energy-minimize from current structure
- the MD takes care of starting structures for minimizations

Molecular dynamics with quenching



Molecular dynamics with quenching

The obtained (possibly many) minimized structures
can be processed e.g. by a cluster analysis
to determine the set of unique optimal structures,
their total energies and number of hits.

For a small molecular system, we would observe
few unique structures, each occuring many times.
For larger systems, the number of unique structures grows rapidly.

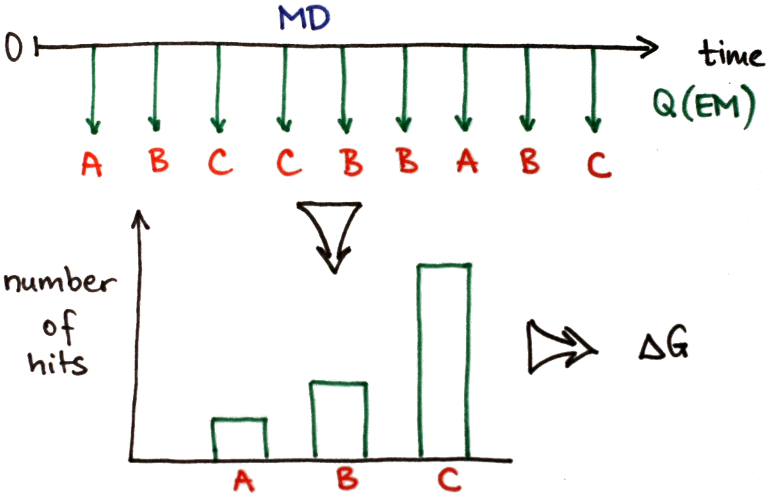
Free energies with MD/Q

If the MD simulation is long enough
(i.e. the sampling of configuration space is sufficient):

the ratio of occurrence of the individual minimized structures (n_i) determines the equilibrium constant K and the free energy ΔG:

$$\begin{aligned} K &= \frac{n_2}{n_1} \\ \Delta G &= -k_B T \log K = k_B T \log \frac{n_2}{n_1} \end{aligned}$$

Free energies with MD/Q



Note on free energies

We consider whole regions of configuration space rather than points to be the individual structures.

Therefore, we obtain no curves of free energy as a function of coordinate(s) but rather single values of free energy differences for certain pairs of “structures”.

Nearly philosophical question:
Is there something like “free energy surface” at all?
Or, is it only meaningful to ask
for discrete values of free energy differences?

Energy barriers in simulations

Energy landscapes in large (bio)molecular systems
– multitude of almost iso-energetic minima,
separated from each other by energy barriers of various heights
Each of these minima $\equiv$ one particular structure (conformation);
neighboring minima correspond to similar structures
Structural transitions are barrier crossings, and
the transition rate is determined by the height of the barrier.

Energy barriers in simulations

Energy landscapes in large (bio)molecular systems
– multitude of almost iso-energetic minima,
separated from each other by energy barriers of various heights
Each of these minima $\equiv$ one particular structure (conformation);
neighboring minima correspond to similar structures
Structural transitions are barrier crossings, and
the transition rate is determined by the height of the barrier.

Normal MD – only nanosecond time scales are accessible,
so only the smallest barriers are overcome in simulations,
and only small structural changes occur. $k = \exp\left[-E_\mathrm{A}/kT\right]$
The larger barriers are traversed more rarely
(although the transition process itself may well be fast),
and thus are not observed in MD simulations.

Using quotations by Helmut Grubmüller

Note – do not be afraid of Arrhenius

How often does something happen in a simulation?

$k = A \times \exp\left[-E_\mathrm{A}/kT\right]$, e.g. $A = 1 \times 10^9\ \mathrm{s}^{-1}$

Note – do not be afraid of Arrhenius

How often does something happen in a simulation?

$$k = A \times \exp\left[-E_\mathrm{A}/kT\right], \text{ e.g. } A = 1 \times 10^9 \text{ s}^{-1}$$

E_A kcal/mol	k 1/s	$1/k$ μs
1	0.19×10^9	0.005
3	6.7×10^6	0.15
5	0.24×10^6	4.2
7	8.6×10^3	120

Note – do not be afraid of Arrhenius

How often does something happen in a simulation?

$$k = A \times \exp\left[-E_\mathrm{A}/kT\right], \text{ e.g. } A = 1 \times 10^9 \text{ s}^{-1}$$

E_A kcal/mol	k 1/s	$1/k$ μs
1	0.19×10^9	0.005
3	6.7×10^6	0.15
5	0.24×10^6	4.2
7	8.6×10^3	120

If the process has to overcome a barrier of 5 kcal/mol,
we have to simulate for 4 μs to see it happen once on average.

Conformational flooding

- to accelerate conformational transitions in MD simulations by several orders of magnitude
- makes it possible to simulate slow conformational transitions

Conformational flooding

- to accelerate conformational transitions in MD simulations by several orders of magnitude
- makes it possible to simulate slow conformational transitions

1. generate a trajectory with a normal MD simulation
2. using the ensemble of structures from that trajectory, construct a localized artificial flooding potential V_{fl}:

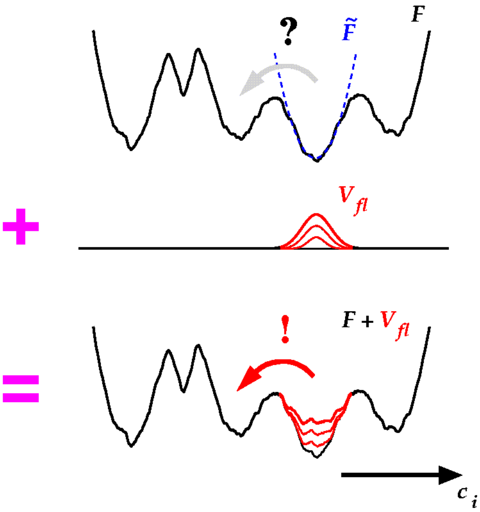
Conformational flooding

- to accelerate conformational transitions in MD simulations by several orders of magnitude
- makes it possible to simulate slow conformational transitions

1. generate a trajectory with a normal MD simulation
2. using the ensemble of structures from that trajectory, construct a localized artificial flooding potential V_{fl}:

- V_{fl} shall affect only the initial conformation and vanish everywhere outside of this region of conf. space
- V_{fl} shall be well-behaved (smooth) and 'flood' the entire initial potential-energy well

Conformational flooding



from the website of Helmut Grubmüller

Flooding potential

a multivariate (n-dimensional) Gaussian function is good:

$$V_{\mathrm{fl}} = E_{\mathrm{fl}} \cdot \exp\left[-\frac{E_{\mathrm{fl}}}{2k_{\mathrm{B}}T} \cdot \sum_{i=1}^{n} q_i^2 \lambda_i\right]$$

E_{fl} – strength of the flooding potential (constant)
q_i – coordinates along the first n essential dynamics modes (PCA)

the first n PCA modes with eigenvalues λ_i will be flooded

The course of flooding simulation

The flooding potential is added to the force field,
and 'flooding' (biased) simulations are performed.

The course of flooding simulation

The flooding potential is added to the force field,
and 'flooding' (biased) simulations are performed.

The energy minimum of the initial conformation is elevated
$\rightarrow$ the height of barriers is reduced
$\rightarrow$ the transitions are accelerated (TS theory)

The course of flooding simulation

The flooding potential is added to the force field,
and 'flooding' (biased) simulations are performed.

The energy minimum of the initial conformation is elevated
$\rightarrow$ the height of barriers is reduced
$\rightarrow$ the transitions are accelerated (TS theory)

Only the energy landscape within the minimum was modified $\rightarrow$

- the dynamics is already known there $\rightarrow$ uninteresting
- the barriers and all the other minima are unbiased
  – may be studied (are usually of interest)
- CF is expected to induce unbiased transitions
  – those that would occur without flooding (but slower)

Metadynamics

- ‘to reconstruct multidimensional ΔG of complex systems’
- artificial dynamics (metadynamics) performed
  in the space defined by a few collective variables S,
  assumed to give a coarse-grained description of the system
- history-dependent biasing potential constructed as a sum
  of Gaussians centered at points visited in the simulation

Laio & Parrinello, Proc. Natl. Acad. Sci. 2002

using quotations by Alessandro Laio

Metadynamics – how it works

- a new Gaussian is added at every time interval t_G
- the biasing potential at time t is given by

$$V_G(S(x),t) = \sum_{t'=t_G,2t_G,3t_G,\ldots} w \cdot \exp\left[\frac{(S(x)-s_{t'})^2}{2\cdot\delta s^2}\right]$$

w and δs – height and width of the Gaussians (preset)
$s_t = S(x(t))$ – value of the collective variable at time t

Metadynamics – how it works

- a new Gaussian is added at every time interval t_G
- the biasing potential at time t is given by

$$V_G(S(x), t) = \sum_{t'=t_G, 2t_G, 3t_G, \ldots} w \cdot \exp\left[\frac{(S(x) - s_{t'})^2}{2 \cdot \delta s^2}\right]$$

  w and δs – height and width of the Gaussians (preset)
  $s_t = S(x(t))$ – value of the collective variable at time t

- this potential is filling the minima on the free energy surface that the system visits during the MD
- energy surface $\equiv$ true free energy + sum of biasing Gaussians – is becoming constant (as function of col. vars S)

Metadynamics – how it works

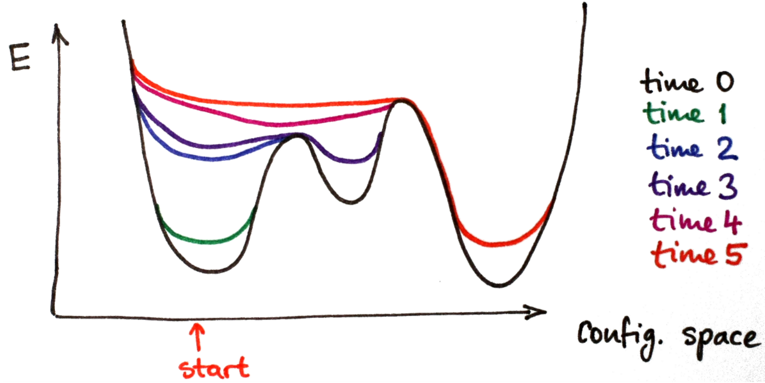
- a new Gaussian is added at every time interval t_G
- the biasing potential at time t is given by

$$V_G(S(x),t) = \sum_{t'=t_G,2t_G,3t_G,\ldots} w \cdot \exp\left[\frac{(S(x)-s_{t'})^2}{2\cdot\delta s^2}\right]$$

  w and δs – height and width of the Gaussians (preset)
  $s_t = S(x(t))$ – value of the collective variable at time t

- this potential is filling the minima on the free energy surface that the system visits during the MD
- energy surface $\equiv$ true free energy $+$ sum of biasing Gaussians – is becoming constant (as function of col. vars S)
- the MD has a kind of memory via the biasing potential

Metadynamics – what it looks like



https://www.youtube.com/watch?v=IzEBpQ0c8TA

https://www.youtube.com/watch?v=iu2GtQAyoj0

Properties of metadynamics

- explores new reaction pathways
- accelerate rare events
- estimates free energies efficiently

Properties of metadynamics

- explores new reaction pathways
- accelerate rare events
- estimates free energies efficiently
- the system escapes a local free energy minimum
  through the lowest free-energy saddle point.
- the free-energy profile is filled with the biasing Gaussians
- the sum of the Gaussians $\rightarrow$ negative of the free energy
  (if the dynamics along S is much slower than
  the dynamics along the remaining degrees of freedom)

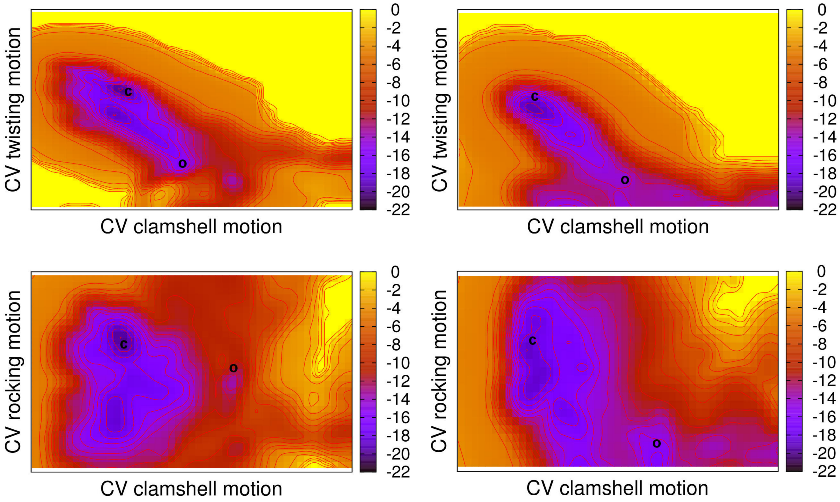
Properties of metadynamics

Crucial point – identify the variables that are of interest
and are difficult to sample because of barriers
that cannot be cleared in the available simulation time.

These variables $S(x)$ are functions of the coordinates of the system;
practical applications – up to 3 such variables,
and the choice depend on the process being studied.

Typical choices – principal modes of motion obtained with PCA
Still, the choice of S may be far from trivial.

Example – opening of a protein binding pocket



courtesy Tino Wolter

Replica-exchange molecular dynamics

REMD / parallel tempering

- method to accelerate the sampling of configuration space in case of high barriers between relevant configurations
- several (identical) replicas of the system are simulated simultaneously, at different temperatures
- coordinates+velocities of the replicas may be switched (exchanged) between two temperatures

Probability of replica exchange

- probability of exchange between $T_1 < T_2$
- determined in regular time intervals

Probability of replica exchange

- probability of exchange between $T_1 < T_2$
- determined in regular time intervals
- instantaneous potential energies U_1 and U_2 in the two simulations needed

$$\mathcal{P}(1 \leftrightarrow 2) = \begin{cases} 1 & \text{if } U_2 < U_1, \\ \exp\left[\left(\frac{1}{k_B T_1} - \frac{1}{k_B T_2}\right) \cdot (U_1 - U_2)\right] & \text{otherwise.} \end{cases}$$

Probability of replica exchange

- probability of exchange between $T_1 < T_2$
- determined in regular time intervals
- instantaneous potential energies U_1 and U_2 in the two simulations needed

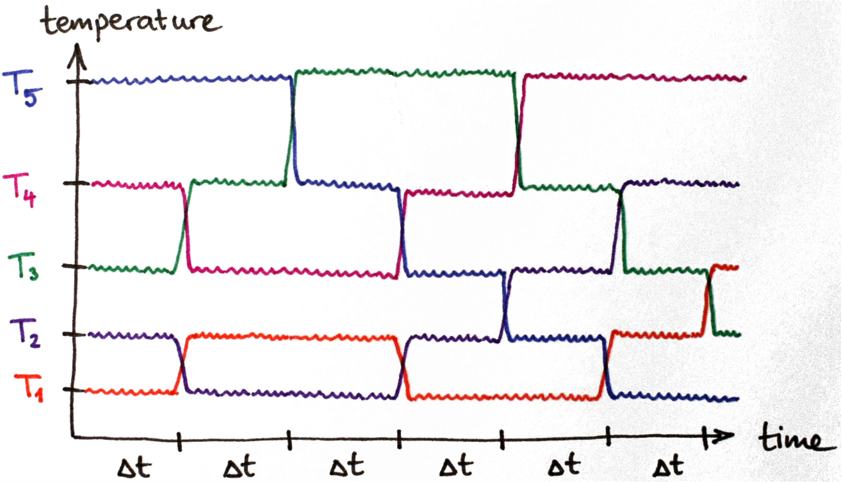
$$\mathcal{P}(1 \leftrightarrow 2) = \begin{cases} 1 & \text{if } U_2 < U_1, \\ \exp\left[\left(\frac{1}{k_B T_1} - \frac{1}{k_B T_2}\right) \cdot (U_1 - U_2)\right] & \text{otherwise.} \end{cases}$$

- if $\mathcal{P}(1 \leftrightarrow 2) >$ random number from $(0, 1)$, then replicas in simulations at T_1 and T_2 are exchanged
- a flavor of Metropolis' Monte Carlo

Setup of the simulation of replicas

- one replica at the temperature of interest ($T_1 = 300$ K)
- several others at higher temperatures ($T_1 < T_2 < T_3 < \ldots$)
- after 1 ps, attempt exchanges $1 \leftrightarrow 2$, $3 \leftrightarrow 4$ etc.
- after another 1 ps, do the same for $2 \leftrightarrow 3$, $4 \leftrightarrow 5$ etc.
- so, try to exchange replicas at "neighboring" temperatures

Setup of the simulation of replicas



Advantages of REMD

- due to the simulations at high temperatures
- faster sampling and more frequent crossing of energy barriers
- correct sampling at all temperatures obtained,
  above all at the (lowest) temperature of interest

Advantages of REMD

- due to the simulations at high temperatures
- faster sampling and more frequent crossing of energy barriers
- correct sampling at all temperatures obtained,
  above all at the (lowest) temperature of interest
- increased computational cost (multiple simulations)
  pays off with largely accelerated sampling
- simulations running at different temperatures are independent
  except at attempted exchanges → easy parallelization

Advantages of REMD

- due to the simulations at high temperatures
- faster sampling and more frequent crossing of energy barriers
- correct sampling at all temperatures obtained,
  above all at the (lowest) temperature of interest

- increased computational cost (multiple simulations)
  pays off with largely accelerated sampling
- simulations running at different temperatures are independent
  except at attempted exchanges → easy parallelization

- first application – protein folding (Sugita & Okamoto, Chem. Phys. Lett. 1999)

Choice of temperatures to simulate

Important – suitable choice of temperatures T_i – criteria:

- how frequent exchanges we wish (average prob. $\mathcal{P}(1 \leftrightarrow 2)$)
- the size of the system (degrees of freedom N_{dof})
- the number of temperatures/simulations

Choice of temperatures to simulate

Important – suitable choice of temperatures T_i – criteria:

- how frequent exchanges we wish (average prob. $\mathcal{P}(1 \leftrightarrow 2)$)
- the size of the system (degrees of freedom N_{dof})
- the number of temperatures/simulations

For protein/water systems with all bond lengths constrained:

- $N_{\mathrm{dof}} \approx 2N$ (N – number of atoms)
- average probability is related to $T_2 - T_1 = \varepsilon T_1$ as

$$\overline{\mathcal{P}(1 \leftrightarrow 2)} \approx \exp\left[-2\varepsilon^2 N\right]$$

- set of temperatures may be designed to suit the problem

REMD generalized

- multiple different simulation parameters...
- different temperatures and different (e.g. biasing) potentials
- great flexibility

REMD generalized

- multiple different simulation parameters. . .
- different temperatures and different (e.g. biasing) potentials
- great flexibility

Simulations 1 and 2 performed

- at different temperatures T_1 and T_2
- with different potentials U_1 and U_2 (umbrella or other)

REMD generalized

- multiple different simulation parameters…
- different temperatures and different (e.g. biasing) potentials
- great flexibility

Simulations 1 and 2 performed

- at different temperatures T_1 and T_2
- with different potentials U_1 and U_2 (umbrella or other)

$$
\begin{aligned}
\Delta &= \frac{1}{kT_1}\left(U_1(q_2) - U_1(q_1)\right) - \frac{1}{kT_2}\left(U_2(q_1) - U_2(q_2)\right) \\
\mathcal{P}(1 \leftrightarrow 2) &= \begin{cases} 1 & \text{if } \Delta \leq 0, \\ \exp\left[-\Delta\right] & \text{otherwise.} \end{cases}
\end{aligned}
$$

REMD generalized

Barostat

- common problem of REMD simulations
- our experience – NVT is reliable, NPT is not
- box scaling $\rightarrow$ scaling of atom coordinates necessary
  – not (always) performed in the RE protocol
- in Gromacs: 'LINCS' warnings before crash etc.
- $\mathcal{P}$ also affected (for REST2: much smaller than in NVT)
- conclusion: do NVT

Extended sampling methods

Biasing potential methods – US, METAD

- required: a priori choice of reaction coordinate(s) to be biased
- problem – success depends on that choice, possibly non-trivial

Extended sampling methods

Biasing potential methods – US, METAD

- required: a priori choice of reaction coordinate(s) to be biased
- problem – success depends on that choice, possibly non-trivial

REMD (parallel tempering)

- $+$ no such required, can be used rather blindly
- $-$ all of the system heated $\rightarrow$ may destroy something
- $-$ no knowledge of the system may be embedded
- $-$ poor efficiency for big systems: $\overline{\mathcal{P}(1 \leftrightarrow 2)} \approx \exp\left[-2\varepsilon^2 N\right]$
  $\rightarrow$ critical problem

Extended sampling methods

Hamiltonian replica exchange (HREX)

- in intermediate position between US/METAD and REMD/PT
- simpler to use than US/METAD
  - results depend not so strongly on the choices to be made
- efficiency does not depend on the overall system size
- many possibilities; our choice: REST2

REST1: Berne et al., Proc. Natl. Acad. Sci. USA 2005
modif: Ceulemans et al., J. Chem. Theory Comput. 2011
modif: Takada et al., J. Comput. Chem. 2011
REST2: Berne et al., J. Phys. Chem. B 2011
review and Gromacs implementation: Bussi, Mol. Phys. 2014

Replica-exchange with solute tempering

$$P = \exp\left[-\frac{U}{kT}\right] = \exp\left[-\beta U\right]$$

Replica-exchange with solute tempering

$$P = \exp\left[-\frac{U}{kT}\right] = \exp\left[-\beta U\right]$$

- note: $\frac{1}{2}U$ would be the same as $2T$

Replica-exchange with solute tempering

$$P = \exp\left[-\frac{U}{kT}\right] = \exp\left[-\beta U\right]$$

- note: $\frac{1}{2}U$ would be the same as $2T$
- U is combined from terms that we can scale individually
  - is not possible for T
  - 'heating' of a portion of the system
  - a group of atoms, or just a group of interaction terms

REST2

- divide the system into two parts:
- hot – small, will be subject to extended sampling
- cold – all of the rest

REST2

- divide the system into two parts:
- hot – small, will be subject to extended sampling
- cold – all of the rest

Generate replicas with different $\lambda_m < 1$, modify parameters in hot:

- scale the charges by $\sqrt{\lambda_m}$
- scale the LJ depths ε by λ_m
- scale the amplitudes of dihedrals within hot by λ_m
- scale dihedrals partly within hot by $\sqrt{\lambda_m}$

REST2

- divide the system into two parts:
- hot – small, will be subject to extended sampling
- cold – all of the rest

Generate replicas with different $\lambda_m < 1$, modify parameters in hot:

- scale the charges by $\sqrt{\lambda_m}$
- scale the LJ depths ε by λ_m
- scale the amplitudes of dihedrals within hot by λ_m
- scale dihedrals partly within hot by $\sqrt{\lambda_m}$

Then, the 'effective' temperatures are

- inside hot: $T/\lambda_m > T$
- interactions between hot and cold: $T/\sqrt{\lambda_m}$
- inside cold: T is retained

REST2

Meaning of temperature

- kinetic energy $\leftarrow$ velocities
  - does not change, is the same in hot and cold (300 K)
  - simulation settings need not be adjusted (time step!)
  - unlike in parallel tempering
- factor affecting the population of states
  - we play with this

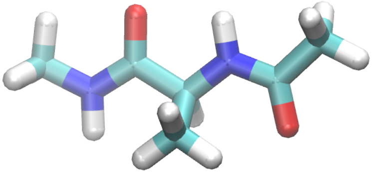
REST2 – technical

- implemented in Gromacs+Plumed
- independent topology files may be used – great flexibility
- scripts for topology modification available
- $\mathcal{P}$ computed from the general expression
- low overhead – extra computational cost up to 10 %
- also possible with Gromacs' free energy code (slower)

REST2 – example

Solute tempering – dialanine

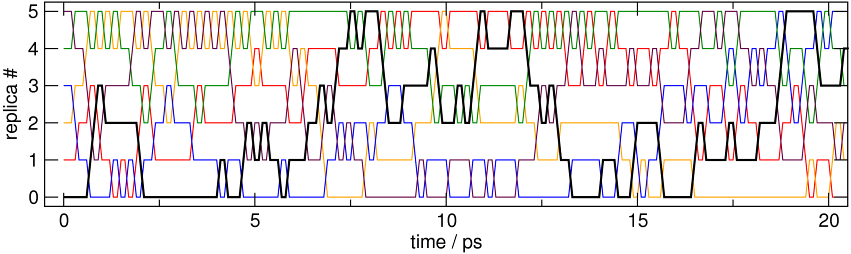
- alanine dipeptide – 22 atoms, 1 pair of $\varphi - \psi$
- Amber99SB + TIP3P
- 5 replicas, $\lambda = 1 \ldots 0.18$ i.e. $T_m = 300 \ldots 1700$ K
- exchange every 0.1 ps, observed $\overline{\mathcal{P}}$ =0.25–0.50



REST2 – example

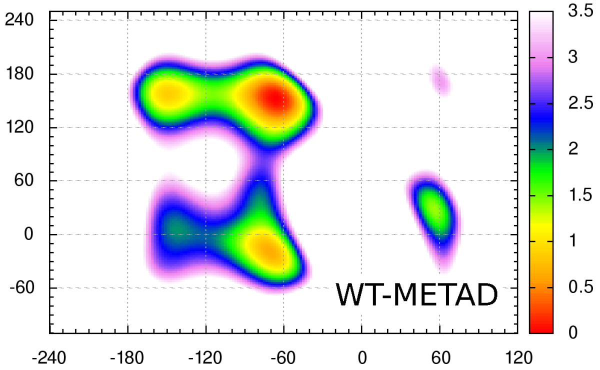
Solute tempering – dialanine

- alanine dipeptide – 22 atoms, 1 pair of $\varphi - \psi$
- Amber99SB + TIP3P
- 5 replicas, $\lambda = 1 \ldots 0.18$ i.e. $T_m = 300 \ldots 1700$ K
- exchange every 0.1 ps, observed $\overline{\mathcal{P}}$ =0.25–0.50



REST2 – example

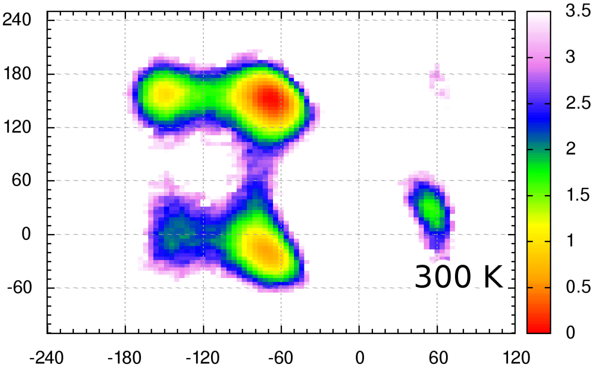
Solute tempering – dialanine – reference result from WT metadyn.



$\varphi - \psi$ in degrees, ΔF in kcal/mol

REST2 – example

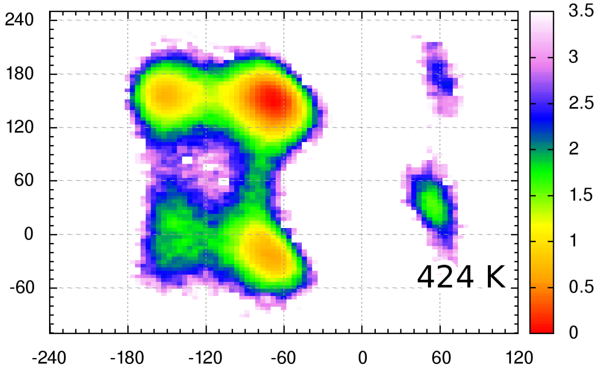
Solute tempering – dialanine – replica #0



$\varphi - \psi$ in degrees, ΔF in kcal/mol

REST2 – example

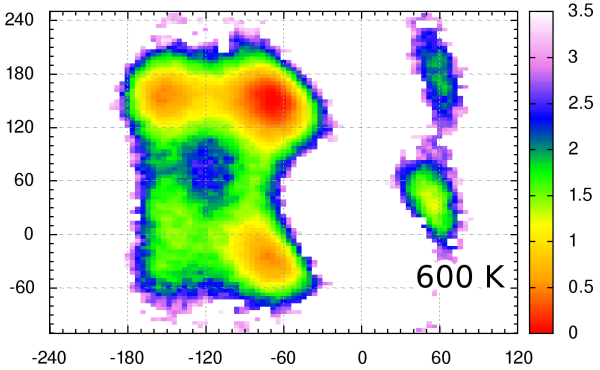
Solute tempering – dialanine – replica #1



$\varphi - \psi$ in degrees, ΔF in kcal/mol

REST2 – example

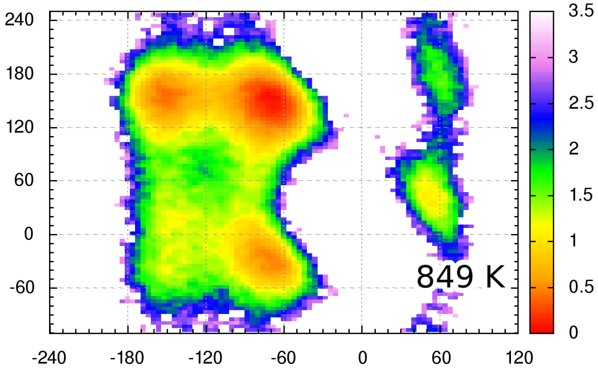
Solute tempering – dialanine – replica #2



$\varphi - \psi$ in degrees, ΔF in kcal/mol

REST2 – example

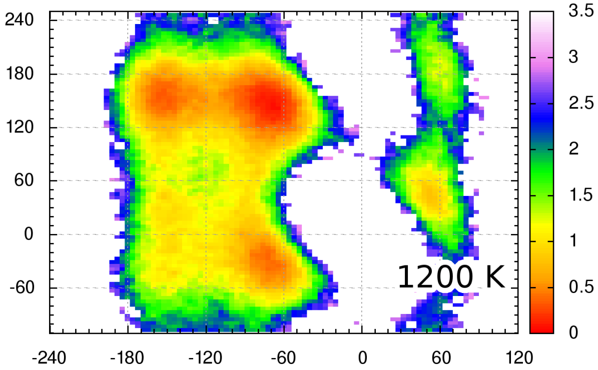
Solute tempering – dialanine – replica #3



$\varphi - \psi$ in degrees, ΔF in kcal/mol

REST2 – example

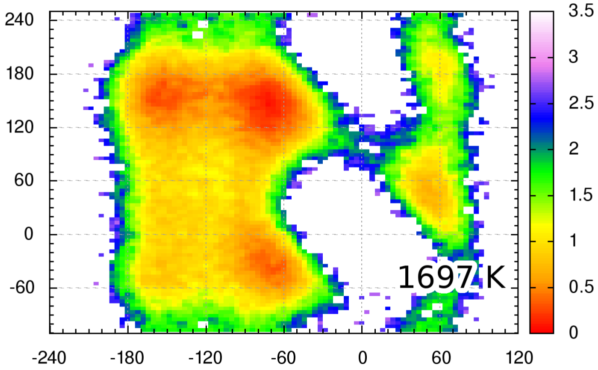
Solute tempering – dialanine – replica #4



$\varphi - \psi$ in degrees, ΔF in kcal/mol

REST2 – example

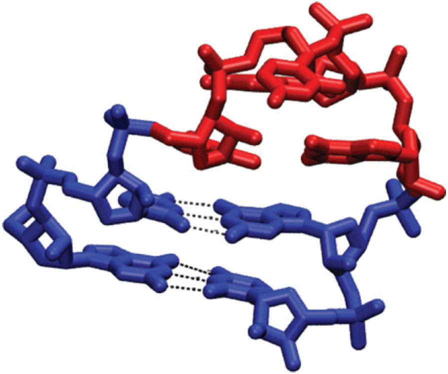
Solute tempering – dialanine – replica #5



$\varphi - \psi$ in degrees, ΔF in kcal/mol

REST2 – example

Partial tempering – RNA tetraloop



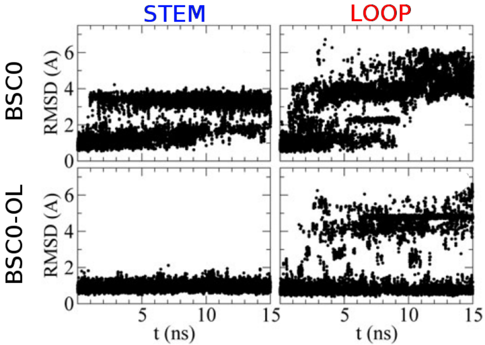
REST2 – example

Partial tempering – RNA tetraloop

- GC-UUCG-GC
- difficult – slow sampling, force field issues – Olomouc FF
- stem – WC HB restrained, kept 'cold'
- loop – 'hot', 16 replicas, $\lambda = 1 \ldots 0.3 \rightarrow \mathcal{P}$ =0.3–0.5
- 4600 TIP3P waters, 14 Na^{+}, 7 Cl^{-}

REST2 – example

Partial tempering – RNA tetraloop



defficiency of BSC0 manifests quickly: ladder-like structure of stem