# Supporting Information

## Boxed molecular dynamics: a simple and general technique for accelerating rare event kinetics and mapping free energy

David R. Glowacki, Emanuele Paci, and Dmitrii V. Shalashilin

The Supporting Information includes:
1. A detailed derivation of the AXD formalism
2. Further details of the trajectory inversion procedure

**Derivation of the AXD formalism**

The residence probability in the phase space volume $\Delta\Gamma$ (see Scheme I of this article) may be obtained using classical MD, Langevin dynamics, or a Monte-Carlo random walk as

$$P(\rho \in \Delta\Gamma) = \int\limits_{\rho \in \Delta\Gamma} p(\rho)d\rho \qquad (AI.1)$$

$P(\rho \in \Delta\Gamma)$ is also defined as

$$P(\rho \in \Delta\Gamma) = \frac{\tau_{\Delta\Gamma}}{\tau_{total}} \qquad (AI.2)$$

where $\tau_{\Delta\Gamma} = \tau(\rho \in \Delta\Gamma)$ is the time that the system spends within the $\Delta\Gamma$ region, and $\tau_{total}$ is the total MD simulation time. If the dynamics is statistical, then $P(\rho \in \Delta\Gamma)$ corresponds to the ratio of two phase volumes (or classical partition functions)

$$P(\rho \in \Delta\Gamma) = \frac{\Delta\Gamma}{\Gamma_{tot}} \qquad (AI.3)$$

where $\Gamma_{tot}$ is the total phase volume. For the canonical case of a thermal ensemble averaged over all energies, $E(\mathbf{p},\mathbf{q})$, the terms in AI.3 may be rewritten as

$$\Gamma_{tot} \propto \iint e^{-\frac{E(\mathbf{p},\mathbf{q})}{kT}} d\mathbf{p}\,d\mathbf{q}$$

$$\Delta\Gamma \propto \iint\limits_{\rho \in \Delta\Gamma} e^{-\frac{E(\mathbf{p},\mathbf{q})}{kT}} d\mathbf{p}\,d\mathbf{q} \qquad (AI.4)$$

For the microcanonical case, where the ensemble has a fixed energy, the terms are

$$\Gamma_{tot} = \iint \delta(E(\mathbf{p,q})) - E) \, d\mathbf{p} \, d\mathbf{q}$$

$$\Delta\Gamma = \iint_{\rho \in \Delta\Gamma} \delta(E(\mathbf{p,q})) - E) \, d\mathbf{p} \, d\mathbf{q} \qquad\qquad \text{(AI.5)}$$

For large molecules, the results for $P(\rho \in \Delta\Gamma)$ calculated for canonical or microcanonical ensembles are indistinguishable if the microcanonical energy is equal to the canonical average energy – i.e., when $E(\mathbf{p,q}) = <E(\mathbf{p,q})>$.

It is often difficult to converge both $P(\rho)$ and $p(\rho)$ over the full phase space volume because the probability of finding the system at high free energy regions of the phase space is very small. Following on from the ideas in refs 27 and 28, and using the definitions in Scheme I, we may accelerate the convergence by rewriting Eq (AI.3) as follows:

$$P(\rho \in \Delta\Gamma) = \frac{\Delta\Gamma}{\Gamma_1 + \Delta\Gamma} \times \frac{\Gamma_1 + \Delta\Gamma}{\Gamma_{tot}} = P^{AXD} \times P^{CORR} \qquad\qquad \text{(AI.6)}$$

where $\Gamma_1$ is bounded by $\rho_{lock}$, and therefore adjacent to $\Delta\Gamma$. In Eq (AI.6), $P(\rho \in \Delta\Gamma)$ is a product of two factors. The first, $P^{AXD}$, is an accelerated integrated probability

$$P^{AXD} = \frac{\Delta\Gamma}{\Gamma_1 + \Delta\Gamma} \qquad\qquad \text{(AI.7)}$$

and the second, $P^{CORR}$, is a correction factor for the accelerated probability

$$P^{CORR} = \frac{\Gamma_1 + \Delta\Gamma}{\Gamma_{tot}} \qquad\qquad \text{(AI.8)}$$

$P^{CORR}$ may be obtained by running MD and using Eq (AI.2) with $\tau_{\Delta\Gamma}$ replaced by $\tau_{lock}$ where $\tau_{lock}$ is the time the trajectory spends in $\Gamma_1 + \Delta\Gamma$, and $P^{AXD}$ may be obtained using Eq (AI.2) with $\tau_{total}$ replaced by $\tau_{lock}$ – i.e., the MD constrained so that it only occurs in the phase space region $\Gamma_1 + \Delta\Gamma$, bounded by $\rho_{lock}$. The rare event acceleration of these equations derives from the fact that *it is less expensive to*

*converge $P^{AXD}$ and $P^{CORR}$ separately than their small product in Eq (AI.6) using brute force MD.*

        Similar arguments can be also applied to the calculation of rate coefficients because a TST thermal rate coefficient, $k(T)$, is inversely proportional to $\Gamma_{react}$ – i.e.,

$$k(T) \propto \frac{1}{\Gamma_{react}} \qquad (AI.9)$$

where $\Gamma_{react} = \Gamma_1 + \Gamma_2$ is the total reactant phase volume. Therefore, $k(T)$ may also be written in an analogous fashion as $P(\rho \in \Delta\Gamma)$ in Eq. (AI.6):

$$k(T) = k^{AXD} \frac{\Gamma_1}{\Gamma_1 + \Gamma_2} = k^{AXD} P^{CORR} \qquad (AI.10)$$

where $k^{AXD}$ is the accelerated rate coefficient obtained from the dynamics restricted within $\Gamma_1$, restricted by $\rho_{lock}$, and $P^{CORR}$ is slightly different from that given in AI.8 when correcting probabilities – i.e.,

$$P^{CORR} = \frac{\Gamma_1}{\Gamma_1 + \Gamma_2} \qquad (AI.11)$$

As has been discussed in Ref. 27, $k_{AXD}$ may account for some non-statistical effects such as direct trajectories or fast recrossing, and therefore may differ from the TST rate coefficient. Strictly speaking, $k_{AXD}$ should contain a transmission coefficient that accounts for these effects; however, we note that irrespective of such corrections, $k(T)$ remains inversely proportional to $\Gamma_{react}$ as written in Eq (AI.9). Thus, the transmission coefficient is simply folded into $k^{AXD}$, and the definition of $P_{CORR}$ remains unchanged – i.e., it may be calculated statistically. For the sake of simplicity, the equations written above do not include transmission factor corrections.

**Details of the velocity inversion procedure**

Below, we describe the procedure we have used for inverting velocities at a box boundary in a fashion that conserves momentum and angular momentum. Inverting the velocities is a simple procedure that may be applied to other reaction coordinates involving an arbitrary number of atoms. In general, if the reaction coordinate is a function of a number of system coordinates and/or momenta, then the inversion is undertaken as follows:

(1) Transform atomic velocities from the Cartesian frame to the center-of-mass (CM) frame

(2) Calculate the unit vector along the gradient of the reaction coordinate $\rho$ as
$$\vec{n} = \nabla\rho / |\nabla\rho| \qquad\qquad\qquad\qquad (AII.1)$$

(3) Calculate the parallel projections of the center of mass atomic velocities on the reaction coordinate unit vector, and then invert these velocities

(4) Transform back to the Cartesian frame

Below, we illustrate the velocity inversion procedure for the peptide extension reaction coordinate discussed in the article. If the trajectory crossed the boundary at the $t+1$ timestep, then we go back to the positions, $r$, and velocities, $v$, at the $t$ timestep, and invert the velocities in the centre of mass frame to yield $v'$. For the peptide extension, we start with the positions and velocities of the peptide C terminus ($\vec{r}_1$, $\vec{v}_1$) and N terminus ($\vec{r}_2$, $\vec{v}_2$), and work out the velocity of their CM

$$\vec{v}_{cm} = \frac{m_1\vec{v}_1 + m_2\vec{v}_2}{m_1 + m_2} \qquad\qquad\qquad\qquad (AII.2)$$

Their velocities in the CM frame are

$$\vec{v}_{1cm} = \vec{v}_1 - \vec{v}_{cm}, \qquad\qquad \vec{v}_{2cm} = \vec{v}_2 - \vec{v}_{cm} \qquad\qquad (AII.3)$$

The peptide extension reaction coordinate is the distance between atoms 1 and 2 – i.e., $\rho = \sqrt{r_x^2 + r_y^2 + r_z^2}$ where $r_x$, $r_y$, and $r_z$ are elements of the vector $\vec{r}_2 - \vec{r}_1$. The unit vector along the reaction coordinate, $\vec{n}_{12}$, is defined using eq. (AII.1) as:

$$\vec{n}_{12} = \frac{\nabla\rho}{|\nabla\rho|} = \nabla\rho = \frac{\vec{r}_2 - \vec{r}_1}{|\vec{r}_2 - \vec{r}_1|} \qquad\qquad\qquad (AII.4)$$

The projection of the vector $\vec{v}_{1cm}$ onto $\vec{n}_{12}$ is defined as

$$\vec{v}_{1cm}{}^{\parallel} = \vec{n}_{12}(\vec{v}_{1cm}\vec{n}_{12}) \tag{AII.5}$$

Inverting the velocity simply means inverting the sign of the $\vec{v}_{1cm}{}^{\parallel}$ contribution to the vector $\vec{v}_{1cm} = \vec{v}_{1cm}{}^{\perp} + \vec{v}_{1cm}{}^{\parallel}$, which means replacing $\vec{v}_{1cm}$ by $\vec{v}_{1cm}{}'$ where

$$\vec{v}_{1cm}{}' = \vec{v}_{1cm} - 2\vec{v}_{1cm}{}^{\parallel} = \vec{v}_{1cm}{}^{\perp} - \vec{v}_{1cm}{}^{\parallel} \tag{AII.6}$$

Now we go back to the original laboratory frame and replace $\vec{v}_1$ by $\vec{v}_1{}'$

$$\vec{v}_1{}' = \vec{v}_{1cm}{}' + \vec{v}_{cm} \tag{AII.7}$$

We undertake the same procedure for atom 2: $\vec{v}_2$ is replaced by $\vec{v}_2{}'$, where the latter is calculated as follows:

$$\vec{v}_{2cm}{}^{\parallel} = \vec{n}_{12}(\vec{v}_{2cm}\vec{n}_{12}) \tag{AII.8.1}$$

$$\vec{v}_{2cm}{}' = \vec{v}_{2cm} - 2\vec{v}_{2cm}{}^{\parallel} \tag{AII.8.2}$$

$$\vec{v}_2{}' = \vec{v}_{2cm}{}' + \vec{v}_{cm} \tag{AII.8.3}$$

The rest of the velocities remain unchanged. So long as the numerical integration algorithm calculates velocities and positions at the same time, the velocity inversion procedure described above conserves both the linear and angular momentum of the system, and thus conserves energy perfectly for *in vacuo* peptide simulations using a microcanonical *NVE* ensemble. The requirement that both velocities and positions are calculated at the same time means that this inversion procedure does not conserve energy for *in vacuo* simulations using integrators such as Leapfrog Verlet. In the case of Langevin dynamics, where energy is not conserved, then the procedure described above may be implemented within the Leapfrog Verlet algorithm, and the numerical integration is stable.