



UvA-DARE (Digital Academic Repository)

Path finding on high-dimensional free energy landscapes

Díaz Leines, G.; Ensing, B.

DOI

[10.1103/PhysRevLett.109.020601](https://doi.org/10.1103/PhysRevLett.109.020601)

Publication date

2012

Published in

Physical Review Letters

[Link to publication](#)

Citation for published version (APA):

Díaz Leines, G., & Ensing, B. (2012). Path finding on high-dimensional free energy landscapes. *Physical Review Letters*, *109*(2), 020601. <https://doi.org/10.1103/PhysRevLett.109.020601>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Supplementary Information, Path Finding on High-Dimensional Free Energy Landscapes

Grisell Diaz Leines and Bernd Ensing*

*Van 't Hoff Institute for Molecular Sciences, Universiteit van Amsterdam,
Science Park 904, 1098 XH Amsterdam, The Netherlands*

(Dated: April 18, 2012)

Derivation of the path collective variable

To define a reaction coordinate as a function of collective variables $\sigma(\mathbf{z})$, we use a projection of \mathbf{z} onto a discretized path of $M + 1$ equidistant points, or nodes, $\mathbf{s}_0, \mathbf{s}_1, \dots, \mathbf{s}_M$. The path of nodes forms a smooth curve that connects the reactant and product free energy minima and the projection provides a foliation of the \mathbf{z} -space such that it attributes to each configuration of the system a measure of the progress along the reaction coordinate, $\sigma(\mathbf{z})$.

The left panel in figure 1 illustrates the geometric projection of a point \mathbf{z} in a three-dimensional collective variable space onto a path of nodes, yielding the point $\mathbf{s}(\sigma(\mathbf{z}))$. Here, $\sigma(\mathbf{z})$ is the desired reaction coordinate value, which we wish to be equal to zero for points \mathbf{z} that are in the plane S_0 perpendicular to the curve through the reactant state minimum and equal to one for \mathbf{z} -configurations that lie in the perpendicular plane S_1 through the product state minimum. The geometric expression (eq. 4 in the paper) that we use for this projection is derived in the following way.

First, we locate the path node, \mathbf{s}_m , that is closest to the point \mathbf{z} that we wish to project onto the path. Next, we make use of the fact that, irrespective of the dimensionality, we can draw a circle with center \mathbf{c} through the triplet of nodes \mathbf{s}_m and its neighboring nodes \mathbf{s}_{m-1} and \mathbf{s}_{m+1} . In figure 1 (left), point \mathbf{z} happens to lie in front of the plane through the nodes. A gray triangle depicts the plane S_σ through \mathbf{c} and \mathbf{z} that is perpendicular to the plane of the circle. As the circle is a smooth curve through the nodes, and the plane S_σ is normal to this curve, we choose S_σ to be the plane that connects all points that are projected to the same point $\mathbf{s}(\sigma(\mathbf{z}))$ on the path and thus have the same σ value.

The white slice indicates the range of all possible planes S_σ that are closest to node \mathbf{s}_m . This range runs from $\sigma(\mathbf{z}) = \frac{m-1/2}{M}$, where the plane bisects the line segment between \mathbf{s}_{m-1} and \mathbf{s}_m , to $\sigma(\mathbf{z}) = \frac{m+1/2}{M}$, where the plane is equidistant from \mathbf{s}_m and \mathbf{s}_{m+1} . As at these limits the plane is also perpendicular to the line segment, it is easy to see that the plane at one end coincides with the plane at the corresponding end of the next triplet of nodes, resulting a smooth foliation along the entire path.

To find the plane through \mathbf{z} and thus the reaction coordinate value $\sigma(\mathbf{z})$, we set out to locate the point \mathbf{x} on

the circle arc between \mathbf{s}_m and \mathbf{s}_{m+1} , for which

$$|\mathbf{z} - \mathbf{s}_{m-1}| = |\mathbf{x} - \mathbf{z}|. \quad (1)$$

However, to avoid computation of arc lengths, we make an approximation by assuming \mathbf{x} to lie on the linear segment between \mathbf{s}_m and \mathbf{s}_{m+1} (see figure 1) and then define the fraction of the arc length from \mathbf{s}_m to \mathbf{x} as:

$$f = \frac{|\mathbf{x} - \mathbf{s}_m|}{|\mathbf{s}_{m+1} - \mathbf{s}_m|}. \quad (2)$$

This fraction is directly related to the projection of \mathbf{z} to σ as:

$$f \in [0, 1] \Rightarrow \sigma \in \left[\frac{m-1/2}{M}, \frac{m}{M} \right] \quad (3)$$

Determining $\sigma(z)$ thus requires finding the fraction f . By combining equations 2 and 3, we can write:

$$|\mathbf{z} - \mathbf{s}_{m-1}| = |f(\mathbf{s}_{m+1} - \mathbf{s}_m) + \mathbf{s}_m - \mathbf{z}| \quad (4)$$

Introducing $\mathbf{v}_1 = \mathbf{s}_m - \mathbf{z}$, $\mathbf{v}_2 = \mathbf{z} - \mathbf{s}_{m-1}$ and $\mathbf{v}_3 = \mathbf{s}_{m+1} - \mathbf{s}_m$, this equation is simplified to

$$|\mathbf{v}_2| = |f\mathbf{v}_3 + \mathbf{v}_1|. \quad (5)$$

Rewriting to solve for f results in a quadratic equation,

$$|\mathbf{v}_3|^2 f^2 + 2(\mathbf{v}_1 \cdot \mathbf{v}_3)f + (|\mathbf{v}_1|^2 - |\mathbf{v}_2|^2) = 0 \quad (6)$$

with (relevant) root:

$$f = \frac{\sqrt{(\mathbf{v}_1 \cdot \mathbf{v}_3)^2 - |\mathbf{v}_3|^2(|\mathbf{v}_1|^2 - |\mathbf{v}_2|^2)} - (\mathbf{v}_1 \cdot \mathbf{v}_3)}{|\mathbf{v}_3|^2}. \quad (7)$$

The reaction coordinate value is related to f by:

$$\sigma = \frac{m}{M} \pm \left(\frac{f-1}{2M} \right), \quad (8)$$

where the \pm sign depends on whether $\sigma(\mathbf{z})$ is on the right or the left side of node \mathbf{s}_m . Combining the expression for f (eq. 7) with equation 8 results in the full expression for the projection:

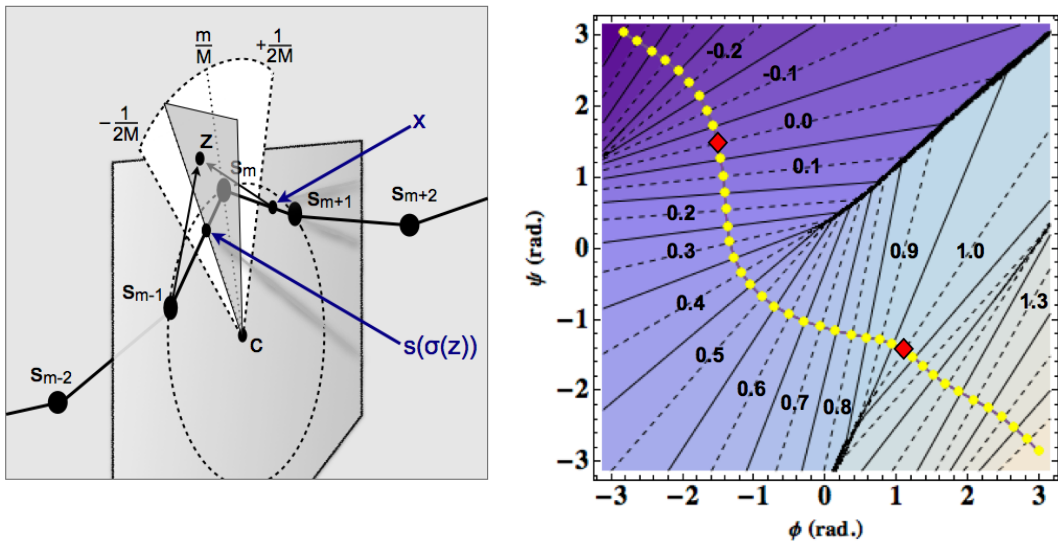


FIG. 1: Left: a point \mathbf{z} in a, here, three-dimensional, collective variable space is projected onto a path of nodes, $\mathbf{s}(\sigma)$, leading to point $\mathbf{s}(\sigma(\mathbf{z}))$. Right: the reaction coordinate value σ as a function of the two torsion angle collective variables used for the alanine dipeptide.

$$\sigma = \frac{m}{M} \pm \frac{\sqrt{(\mathbf{v}_1 \cdot \mathbf{v}_3)^2 - |\mathbf{v}_3|^2(|\mathbf{v}_1|^2 - |\mathbf{v}_2|^2)}}{2M|\mathbf{v}_3|^2} - \frac{(\mathbf{v}_1 \cdot \mathbf{v}_3) - |\mathbf{v}_3|^2}{2M|\mathbf{v}_3|^2} \quad (9)$$

This measure of the progress along the path collective variable is robust and very easy to compute, also for large sets of collective variables. Moreover, it provides in the neighborhood of the path a smooth foliation in \mathbf{z} -space as illustrated in the right panel of figure 1 for the case of the alanine dipeptide.

Convergence of the path and the free energy

The path-metadynamics simulation of the $C7_{\text{eq}} \rightarrow C7_{\text{ax}}$ transition for the alanine dipeptide started from a configuration that belongs to the $C7_{\text{eq}}$ reactant state. The (biased) dynamics of the collective variables of the first 1.5 ns simulation is shown in figure 2. The top and middle panels show the evolution of the torsion angles ψ and ϕ . The bottom panel shows the corresponding σ -value along the (moving) path of nodes. The first barrier crossing occurs after about 175 ps and after about 280 ps the system re-crosses back to the reactant state. The time between successive barrier crossings does not reduce here because after each recrossing the size of the repulsive Gaussian potentials that are continuously added to the metadynamics biasing potential is reduced in order to further refine the estimation of the free energy profile.

Figure 3 shows the convergence behavior of the path-metadynamics method for the case of the alanine dipep-

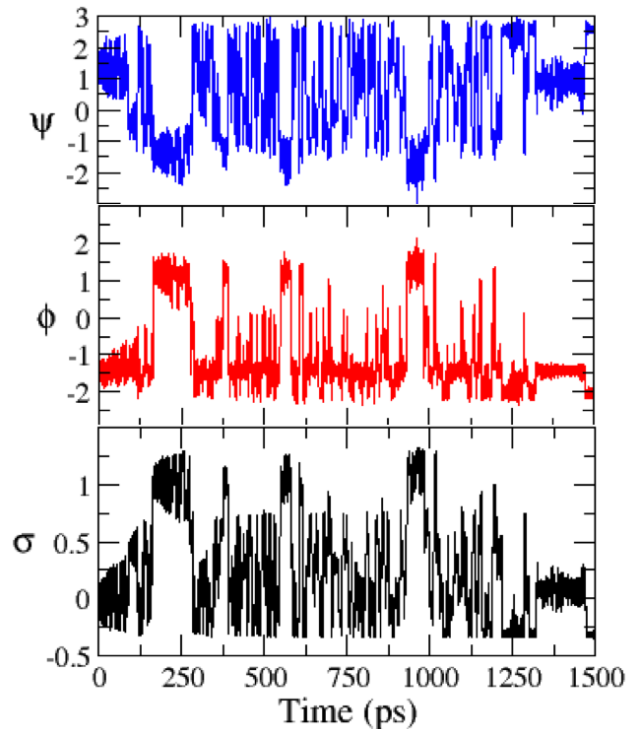


FIG. 2: Evolution of the path collective variable, σ (bottom panel) and the torsion angles ψ and ϕ in radians of the alanine dipeptide during the path-metadynamics simulation (top and middle panel).

ptide conformational transition. The convergence of the path is computed by summing the root mean square distance between the path nodes at time t and at time zero,

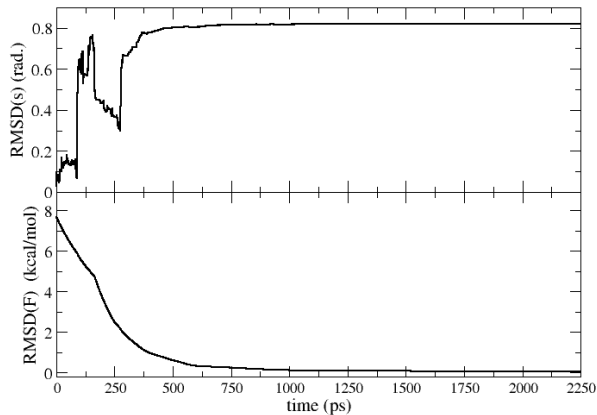


FIG. 3: Convergence of the path shown as the mean square deviation of the node positions with respect to those of the initial guess path as a function of time (top) and convergence of the free energy profile shown as the mean square deviation of the free energy at each node with respect to the final value (bottom).

as follows:

$$\text{RMSD}(s) = \sqrt{\frac{1}{M} \sum_{i=1}^M \left(|s_i(t) - s_i(0)| \right)^2}, \quad (10)$$

with M the number of nodes. The top panel in figure 3 shows that the path is converged after about 500 ps path-metadynamics simulation. The dip in the curve between 175 and 280 ps matches with the time that the system visits and reconstructs the product state for the first time. Due to algorithm used for the reparametrization of the curve to maintain equidistant nodes, the positions of the nodes in the reactant state deteriorate somewhat while the system is in the product state. After the next barrier crossing this is fixed again and the path converges quickly.

The convergence of the free energy profile is monitored using the root mean square deviation of the free energy at each node with respect to the final values:

$$\text{RMSD}(F) = \sqrt{\frac{1}{M} \sum_{i=1}^M \left(F_i(t) - F_i^{\text{final}} \right)^2}. \quad (11)$$

As expected, the free energy profile converges soon after the path nodes are converged, as shown in the bottom panel of figure 3.

Using the gradient of the free energy landscape

Here we briefly discuss the effects of two important features of the path-metadynamics method. The first is

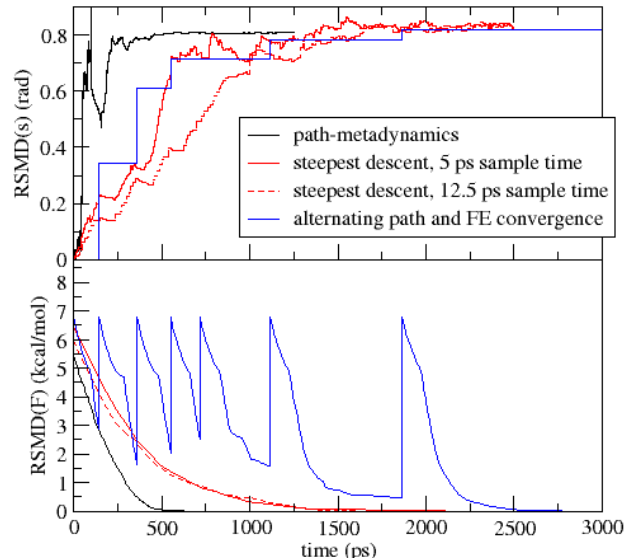


FIG. 4: Convergence of the path (top panel) and the free energy (bottom panel) of the path-metadynamics as presented in the Letter (black lines) compared to that of a variant in which the path follows the steepest descent along the gradient of the free energy landscape (red lines) and a third variant in which the free energy and path are optimized in an alternating fashion until convergence (blue lines).

the approach to advance the path of nodes to the highest transition density. This is a new alternative to the approach of evolving the nodes along the gradient of the free energy. Secondly, the method reconstructs the free energy profile along the path at the same time as it evolves the guess path, rather than converging the free energy and the path in an alternating fashion.

We performed a series of simulations, using as a simplified model of the alanine dipeptide a single particle on the two-dimensional Ramachandran potential energy surface coupled to a Langevin thermostat with a friction coefficient of 0.5, such that it mimics to a high degree the dynamics of the alanine dipeptide molecule. The black graphs in Figure 4 show the convergence of the path and the free energy for this system using path-metadynamics, in very good agreement with figure 3.

By using a very stiff harmonic “tube” potential that confines the sampling to stay very close to the guess path, we can effectively accumulate the average force due to the tube potential at each node. This average force is an estimator of the gradient of the free energy perpendicular to the gradient. We evolved the guess path following the steepest descent toward the minimum free energy path, which is where the perpendicular gradient is zero. The step-size for the evolution of the nodes was taken as a compromise between good efficiency and stability of the integrator for the evolution of the nodes. The metric tensor was taken to be the 2x2 unit ma-

trix. Figure 4 shows the convergence of the path and the free energy for this model system using two different sampling lengths between each update of the path: 5 ps (solid red lines) and 12.5 ps (dashed lines). Clearly, for this system, the approach of following stepwise the free energy gradient is significantly slower than the path-metadynamics technique of allowing the dynamics of the nodes to move unhindered to the highest transition density. Note that a shorter sampling length leads initially to a faster convergence, however, the less accurate estimates of the average force results in larger fluctuations in path motion on the long run, without really converging. We also tried shorter sampling lengths of 1.0 and 0.5 ps (data not shown), which do not converge at all. The algorithm can be improved by increasing the sampling length during the simulation. A factor that would further decrease the performance of the gradient approach for an actual molecular system is that each update of the nodes would require an unproductive equilibration simulation to accommodate the system to its new confining path configuration.

Finally, we show the advantage of optimizing simul-

taneously the path and the free energy profile by comparing to the alternative approach of optimizing them in an alternating fashion. That is, first the free energy profile is converged along a fixed guess path. Then the path is moved to the sampled average transition density, after which the second metadynamics simulation reconstructs the free energy profile along the new guess path, and so forth. Not surprisingly this approach is much less efficient than simultaneous optimization. We tried to improve the alternating approach by, in the first four cycles, only performing a rough free energy reconstruction, by only sampling one barrier crossing and recrossing. Although this renders a 3 to 5 times speedup compared to converging the free energy profile in every cycle, this is still much less efficient compared to the simultaneous optimization as done in the path-metadynamics method. The results are shown by the blue lines in Figure 4.

* b.ensing@uva.nl; <http://molstim.chem.uva.nl>