Supporting Information

Umbrella sampling with path integrals

In the path-integral method, trace averages over the quantum density matrix are obtained from discretising a closed-loop path in which each of the discretised segments can be thought of as a replica of the classical system, connected to its neighbours by a harmonic spring term which acts between each atom and its neighbour in next and previous replica. The total system comprising all the replicas may then be described by an effective classical Hamiltonian, which in the PIMD algorithm is sampled via standard constant temperature molecular dynamics.

Fortunately, it is relatively simple to adapt PIMD to umbrella sampling. In our implementation, the reaction coordinate becomes a function of the *centroid* of each atom's path, where the centroid is defined to be the geometrical centre of that atom's coordinates across all replicas. The resulting free-energy profiles then correspond to the free energy of moving the centre of the two H centroids in a tagged H_2 molecule between two cages. For those who require it, this will be described in more mathematical detail in the following section.

To gauge the accuracy of the path-integral calculations we examined the convergence of the free energy curves with respect to the number of replicas used. Fig. S1 shows path-integral calculations of the free energy curve for the quadruple occupation case, with the barrier heights plotted in Fig. S2. Results are shown for 1, 2, 4 and 8 replicas, where 1 replicas is identical to the classical calculation.

The free energy curve converges quite rapidly with the number of replicas, with there being no statistically significant difference in the barrier height between 4 and 8 replicas. This may seem surprising given that typically, path-integral simulations of water require in the regions of tens of replicas for accurate energy convergence at around 200K. However, there is no contradiction; the barrier energy simply converges faster than the bulk energy.

Path-integral calculations

In this section, we describe the path-integral calculations in more mathematical detail.

In the path-integral method, the *P*-replica approximation to the quantum partition function is given by (e.g. see the treatment of Smith 1)

$$Z^{PIMD} = \Pi_i \left(\frac{m_i}{2\pi\hbar^2\beta_0} \right)^{\frac{3P}{2}} \int e^{-\beta_0 \Phi^{PIMD}(\mathbf{r})} d^{3N}\mathbf{r}_0 d^{3N}\mathbf{r}_1 .. d^{3N}\mathbf{r}_{P-1}$$

where *P* is the number of replicas, $\beta_0 = \beta / P$ is an effective inverse temperature and Φ^{PIMD} is the path-integral effective potential-energy surface. Φ^{PIMD} is a function of *3NP* coordinates, *P* being the number of replicas of the system and *N* the number of nuclei in each replica. The above integral is carried out over all *3NP* coordinates with r_p being the coordinates of the pth replica, having *3N* components.

The path-integral effective potential-energy surface is given by the sum of the potential energies of each replica and a sum over quadratic spring terms which connect the same atom in neighbouring replicas along a path.

$$\Phi^{PIMD}(\mathbf{r}, \dot{\mathbf{r}}) = \sum_{p=0}^{P-1} \left[V(\mathbf{r}_p) + \sum_i \frac{\kappa_i}{2} \left(\dot{\mathbf{r}}_{p,i} - \dot{\mathbf{r}}_{p-1,i} \right)^2 \right]$$

where $\mathbf{r}_{p,i}$ is the coordinate of the ith atom in the pth replica, $\mathbf{r}_{-1,i} = \mathbf{r}_{p-1,i}$ such that the path forms a closed loop, and κ_i is a mass and temperature dependent spring constant given by

$$\kappa_i = \frac{m_i}{\beta_0^2 \hbar^2}$$

Formulation in normal-mode coordinates

It Is convenient to convert the spring term to normal-mode coordinates. Working with one particle moving in the *x* direction and setting $\kappa = 1$, we first recast the quadratic term in terms of a matrix product,

$$S = \frac{1}{2} \sum_{p=1}^{P} (x_p - x_{p-1})^2 = \mathbf{x}^T \mathbf{M} \mathbf{x}$$

where

$$\mathbf{x}^T = (x_1, x_2, ..., x_P)$$

and **M** is a matrix of dimension *P×P* elements, with non-zero elements given by

$$M_{i,j} = \delta_{i,j} - \frac{1}{2}\left(\delta_{i,j+1} + \delta_{i+1,j}\right); \quad 0 \leq i,j < P$$

where it is assumed that $M_{0,j} = M_{P,j}$ and $M_{i,0} = M_{i,P}$.

It is then possible to perform a unitary transformation which diagonalises M, i.e. a unitary matrix U can be chosen such that $UMU^\dagger$ is diagonal. Then

$$S = \left(\mathbf{x}^T\mathbf{U}^\dagger\right)\left(\mathbf{UMU}^\dagger\right)\left(\mathbf{Ux}\right) = \sum_{m=0}^{P-1} \lambda_m \left|q_m\right|^2$$

where $\mathbf{q} = \mathbf{Ux}$, and λ_p are the eigenvalues forming the (real) diagonal elements of $\mathbf{UMU}^\dagger$.

It can be shown that the particular unitary matrix which diagonalises M is given by

$$U_{jk} = \frac{1}{\sqrt{P}} \exp\left(2\pi i \frac{jk}{P}\right); \quad 0 \leq j,k \leq P-1$$

with eigenvalues given by

$$\lambda_m = 1 - \cos\left(2\pi m/P\right); \quad 0 \leq m \leq P-1$$

so the transform is equivalent to a Fourier expansion of $\mathbf{x}$, with $\mathbf{q}$ being the normal modes of the ring.

It is somewhat inconvenient to work with a complex valued unitary transform. Fortunately, it is not too difficult to convert to a real valued orthogonal transformation. We begin by noting that if $\mathbf{e}$ is an eigenvector of M, then if $\mathbf{e}^* \neq \mathbf{e}$, then $\mathbf{e}^*$ must also be a separate linearly independent eigenvector with the same eigenvalue. Indeed, it is simple to show that the eigenvectors and their complex conjugates are paired according to

$$U_{j,k} = U^*_{P-j,k}; \quad 0 < j < P$$

with associated eigenvalues

$$\lambda_m = \lambda_{P-m}; \quad 0 < m < P$$

Taking the sum and difference of each pair gives us a new set of this time real-valued linearly independent eigenvectors. For P being odd:

$$O_{0,k} = \frac{1}{\sqrt{P}}$$

$$O_{2j-1,k} = \sqrt{\frac{2}{P}} \cos\left(2\pi jk/N\right); \quad 0 < j < P/2$$

$$O_{2j,k} = \sqrt{\frac{2}{P}} \sin\left(2\pi jk/N\right); \quad 0 < j < P/2$$

with associated eigenvalues

$$\gamma_{2m} = \gamma_{2m-1} = 1 - \cos\left(2\pi m/P\right); \quad 0 \le m \le P/2$$

When P is even, the above is slightly modified to include

$$O_{P-1,k} = \sqrt{\frac{1}{P}} \cos\left(2\pi Pk/2N\right)$$

The resulting matrix is orthogonal, obeying $\mathbf{OO}^\mathsf{T} = \mathbf{I}$, and $\mathbf{OMO}^\mathsf{T}$ is diagonal and defining real-valued normal mode coordinates $\mathbf{s}$ from $\mathbf{s} = \mathbf{O\,x}$, we have

$$S = \left(\mathbf{x}^T\mathbf{O}^T\right)\left(\mathbf{OMO}^T\right)\left(\mathbf{Ox}\right) = \sum_{m=0}^{P-1} \gamma_m \left|s_m\right|^2$$

The effective potential-energy surface in normal-mode coordinates is then given by

$$\Phi^{PIMD}(\mathbf{s}) = \sum_{m=0}^{P-1}\left[V(\mathbf{r}_p[\mathbf{s}]) + \sum_i \kappa_i \gamma_p s_{i,m}^2\right]$$

where $s_{i,k}$ is the coordinate of the kth normal mode on the ith particle, being a Cartesian vector with x, y and z components, and $m_{i,k}$ is the mass of that mode.

In the path-integral molecular-dynamics method, the effective potential-energy surface is augmented by a kinetic energy term to form an effective, or *de facto*, Hamiltonian. The kinetic term does not have any physical meaning *per se*, but it is necessary in order to derive an equation of motion for the particles so that the averages can be evaluated via constant-temperature molecular dynamics. In normal-mode coordinates, the effective Hamiltonian is given by

$$H^{PIMD}(\mathbf{s}) = \sum_{m=0}^{P-1}\left[V(\mathbf{r}_p[\mathbf{s}]) + \sum_i \kappa_i \gamma_p s_{i,m}^2 + \sum_i \frac{p_{i,m}^2}{2m_i}\right]$$

where $p_{i,m}$ and $m_{,m}$ are the momentum and mass of the mth normal mode of the ith particle. The masses may be chosen freely, but it is usual to give the 0th mode the full particle mass

and the higher-modes masses which oscillate at frequencies in the range of the normal modes of the classical system.

Free-energy calculations using the centroid coordinate

The s_0 normal mode coordinate is proportional to the *centroid* of the path, being the average position of all the replicas of a particle. We have

$$C = \frac{1}{P} \sum_{m=0}^{P-1} x_m$$

and so

$$s_0 = \frac{1}{\sqrt{P}} \sum_{m=0}^{P-1} x_m = \sqrt{P} C$$

Thus, the zeroth normal mode which has a zero eigenvalue corresponds to the centroid coordinate, whereas the other normal modes describe fluctuations in the path.

Returning now to the problem of calculating the free-energy profile of moving a hydrogen molecule when quantum nuclear effects are included, we observe that since each particle no longer has a definite position, it is no longer possible to attach the umbrella spring to a particular nuclear coordinate, or any function of nuclear coordinates. However, we can use the *centroids,* and so, for our calculation we attach the spring to the midpoint of the centroids of the two hydrogen nuclei. Although the path-integral centroids do not have an obvious physical interpretation, the resulting free-energy barriers will give the total free-energy barrier, including quantal corrections, for moving the hydrogen molecule between the cages.

This free-energy is given by

$$A e^{-\beta F(\xi')} = P(\xi') = \frac{1}{\mathscr{Z}^{PIMD}} \int e^{-\beta_0 H^{PIMD}(\mathbf{s},\mathbf{p})} \delta\{\xi' - \xi(\mathbf{s}_0)\} d^{3PN}\mathbf{s}\, d^{3PN}\mathbf{p}$$

where $\mathscr{Z}^{PIMD}$ is given by

$$\mathscr{Z}^{PIMD} = \int e^{-\beta_0 H^{PIMD}(\mathbf{s},\mathbf{p})} d^{3PN}\mathbf{s}\, d^{3PN}\mathbf{p}$$

Adapting the path-integral technique to our umbrella-sampling method is then achieved by simply augmenting the path-integral effective Hamiltonian by the umbrella-biasing potential

$$H^{PIMD'}(\mathbf{s}) = H^{PIMD}(\mathbf{s}, \mathbf{p}) + \frac{\kappa}{2}\left(\xi(\mathbf{s_0}) - \xi\right)^2$$

where, as in the classical case, ξ is the reaction coordinate, which is given by

$$\xi = \frac{l_1}{l_1 + l_2}$$

with l_1 and l_2 being the distances of the hydrogen-molecule centre to each of the two cage-centres in a cage-pair, with the H-molecule centre being taken as the centre of the two centroid coordinates of the nuclei in the tagged H-molecule. Note that this means the reaction coordinate a function of the $\mathbf{s_0}$ path coordinates only.

Table S1.

σ (H_2O), Å	3.1589
ε (H_2O), kJ/mol	0.7749
σ (H_2), Å	3.038
ε (H_2), kJ/mol	0.2852
q_H (H_2O), Coulombs	0.5564 e
q_M (H_2O), Coulombs	-1.1128 e
q_H (H_2), Coulombs	0.4932 e
q_M (H_2), Coulombs	-0.9864 e
k_{HH}, (H_2), kJ / mol Å^{-2}	3475.0
r_{HH} (H_2), Å	0.7417
γ (H_2O)	0.264

The Lennard Jones interactions are of the form:

$$U^{LJ} = 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r}\right)^{12} - \left(\frac{\sigma_{ij}}{r}\right)^{6} \right)$$

where r is the inter-nuclear separation, and the following combining rules are used:

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$$
$$\sigma_{ij} = \sqrt{\sigma_i \sigma_j}$$

The H_2O LJ interaction site is positioned on each water molecules' O site, whereas the H_2 LJ molecule interaction site is positioned on the midpoint between the two H nuclei on each molecule.

Each water molecule has a positive charge, q_H, on each H atom, with a charge of -2q_H positioned on a massless 'M-site', given by

$$\mathbf{r}_M = \mathbf{r}_O + \frac{\gamma}{2}(\mathbf{r}_1 + \mathbf{r}_2)$$

Similarly, there is a negative charge placed on the midpoint of each neutral hydrogen molecule.

The intramolecular H_2 potential energy surface is given by

$$U^{HHintra} = \frac{1}{2}k_{HH}(r - r_{HH})^2$$

Fig S1.

Path-integral free-energy curves for the quadruple occupation case for 1, 2, 4 and 8 replicas, with 1 replica being the classical free-energy curve.
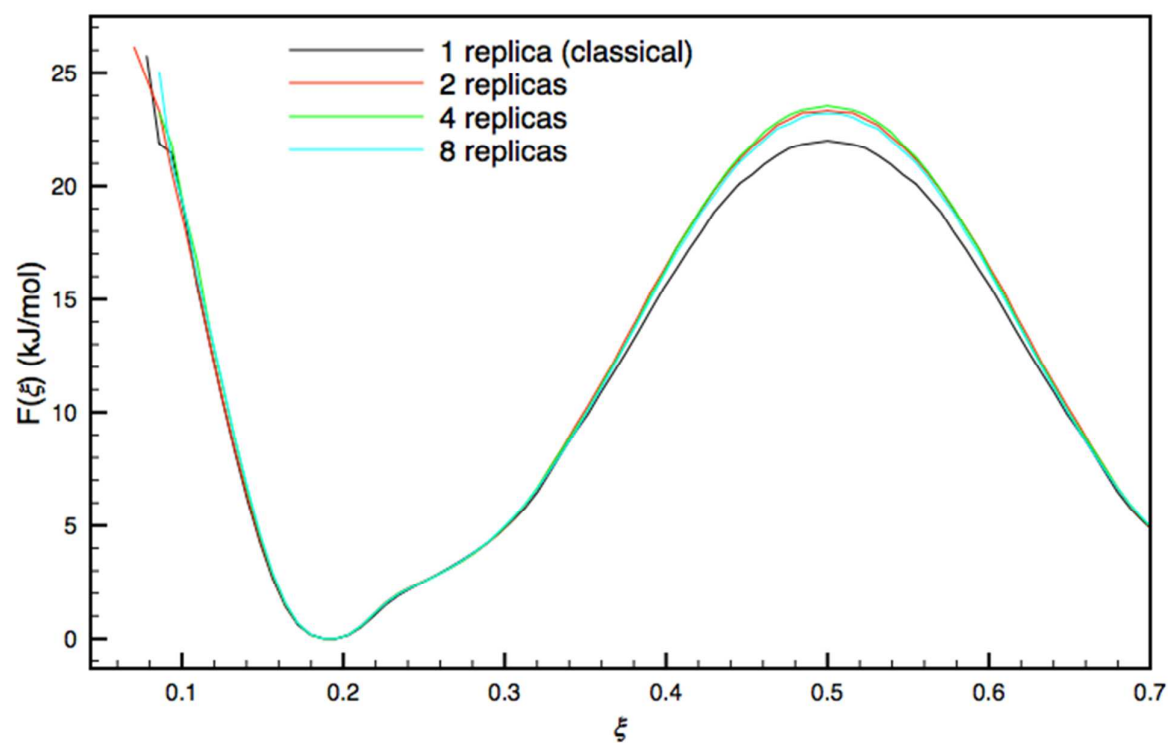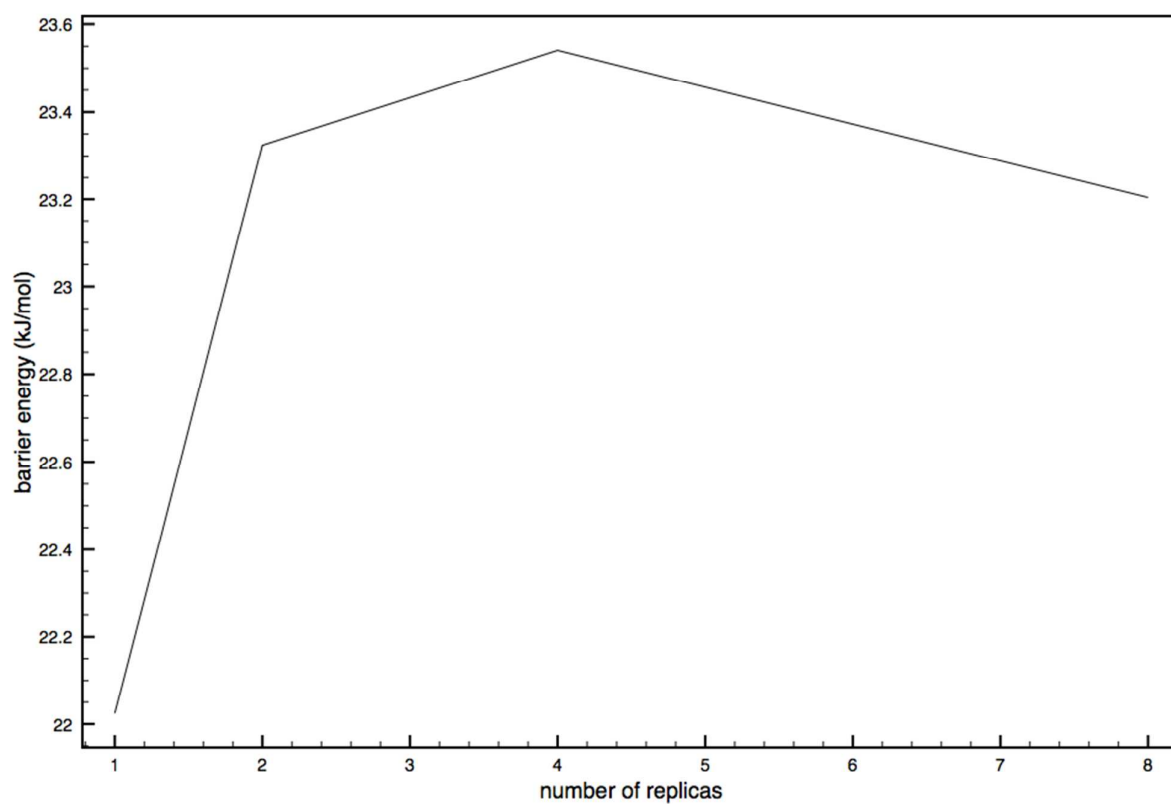
Fig S2.

Convergence of the path-integral barrier height for the quadruple occupation case as a function of the number of replicas. The calculated error in each value is ~0.1 kJ/mol.

Bibliography

(1)   Smith, W. Elements of Molecular Dynamics, 2015.