



Cite this: DOI: 10.1039/c4cp02817a

Nucleation free-energy barriers with Hybrid Monte-Carlo/Umbrella Sampling

M. A. Gonzalez,^a E. Sanz,^a C. McBride,^{†b} J. L. F. Abascal,^a C. Vega^a and C. Valeriani^{*a}

The aim of this work is to evaluate nucleation free-energy barriers using molecular dynamics (MD). More specifically, we use a combination of Hybrid Monte Carlo (HMC) and an Umbrella Sampling scheme, and compute the crystallisation barrier of NaCl from its melt. Firstly the convergence and performance of HMC for different time-steps and the number of MD steps within a HMC cycle are assessed. The calculated potential energies and densities converge regardless of the chosen time-step. However the acceptance ratio of the Metropolis step within the HMC scheme strongly depends on the time-step and affects the performance. It is shown that the acceptance ratio is close to 100% for time-steps of the order of those commonly used in molecular dynamics runs. We then explore the results obtained with a "non-Metropolised" version of HMC where the MD trajectories are always accepted (omitting the Metropolis criteria) and conclude that they are satisfactory for time-steps below 5 fs. Next, HMC is combined with Umbrella Sampling (HMC/US) to compute the nucleation free-energy for both the standard and the "non-Metropolised" HMC (using a small time-step) and in both cases find excellent agreement with the reported values. To conclude, we explore approximations to the HMC/US technique implementing HMC with isothermal–isobaric MD trajectories. The computed nucleation free-energy curve is coincident, within the statistical error, with previous calculations.

Received 26th June 2014,
Accepted 19th September 2014

DOI: 10.1039/c4cp02817a

www.rsc.org/pccp

1. Introduction

A meta-stable phase has to overcome a free-energy barrier in order to transform into a thermodynamically stable phase. The maximum of such a barrier corresponds to the emergence of a critical nucleus of the stable phase within the meta-stable parent phase. When the free-energy barrier is too high, the probability of observing a critical cluster during the course of a standard simulation becomes extremely low. Thus special rare-event simulation techniques such as Metadynamics,¹ Forward Flux Sampling² or Umbrella Sampling (US),^{3,4} are required to improve the sampling efficiency and study the nucleation process. US has been extensively used in combination with *NpT/NVT* Monte Carlo (MC) to calculate nucleation free-energy barriers for crystallisation,^{5–8} condensation⁹ and cavitation,¹⁰ to mention just a few examples. In simple fluids, such as those composed of Lennard-Jones or hard sphere sites, the use of US in combination with Monte Carlo (MC/US) has given excellent results and a general

consensus concerning the nucleation barrier has been achieved.^{5–9} However, when dealing with molecular liquids such as water, the situation is less well established (see ref. 11–13 and references therein). Therefore, the calculation of free-energy barriers for complex molecules is a matter of debate that deserves further attention.

The use of MC/US requires the implementation of a bespoke simulation code, which has somewhat restricted the number of researchers using US to investigate nucleation. This last decade has seen an upsurge in the use of highly optimised molecular dynamics (MD) packages, such as GROMACS¹⁴ and LAMMPS.¹⁵ Furthermore, MD is more suitable for parallelisation than MC. It seems then convenient to use MD to generate the input configurations to US. The Hybrid Monte Carlo^{16,17} (HMC) method was presented as a theoretically well founded scheme to that end. However, little attention has been paid thus far to the implementation of the HMC technique using popular MD codes. This is one of the goals of this work. It is our hope that this work will provide incentive for increased use of US as a technique for the calculation of free-energy barriers.

Once the implementation of the HMC technique is validated, one can combine it with Umbrella Sampling (HMC/US) to obtain free-energy barriers.^{11,13,18,19} In this paper we calculate the free-energy barrier for the crystallisation of NaCl from its melt and compare it with that obtained by MC/US.²⁰ Finally we also explore

^a Depto. Química-Física I, Fac. Ciencias Químicas, Univ. Complutense de Madrid, 28040 Madrid, Spain. E-mail: cvaleriani@ucm.es

^b Instituto de Química Física Rocasolano (IQFR), (CSIC), Serrano 119, 28006 Madrid, Spain

[†] Present address: Depto. Ciencias y Técnicas Fisicoquímicas, Facultad de Ciencias, UNED, Paseo Senda del Rey 9, 28040 Madrid, Spain.

approximations that can simplify the implementation of the HMC technique while retaining most of its theoretical robustness.

II. Theoretical background

A. Hybrid Monte Carlo

The Hybrid Monte Carlo method^{16,17,21–24} uses molecular dynamics micro-canonical trajectories (*i.e.* NVE trajectories) to sample the canonical *NVT* ensemble. It consists of the following steps:

(1) Perform a micro-canonical (NVE) MD simulation starting with a configuration whose momenta have been taken randomly from a Maxwell–Boltzmann distribution at the temperature of interest. The integration of the equations of motion must be performed with a symplectic and time-reversible algorithm.

(2) Accept or reject the final configuration according to the Metropolis acceptance criteria

$$\min(1, \exp^{-\beta\Delta H}) \quad (1)$$

where $\beta = 1/k_B T$ and ΔH is the change in the system's Hamiltonian [$H = \sum (p_i^2/2m_i) + U\{r_N\}$] after and before the MD run, *i.e.*, $\Delta H = H_{\text{new}} - H_{\text{old}}$.

Using the terminology of ref. 22 the HMC technique may be described as “Metropolised” thermostatted molecular dynamics. In what follows, skipping step 2 will be denoted “non-Metropolised” Hybrid Monte Carlo, HMC(nM). Sampling in the isothermal–isobaric *NpT* ensemble can be performed with an additional MC step involving a change in the system's volume.

B. Calculation of the free-energy barrier

The free-energy for the formation of a cluster of the stable phase containing n molecules can be approximated by:⁹

$$\beta\Delta G(n) = -\ln(N_n/N), \quad (2)$$

where N_n is the number of clusters with n molecules and N is the total number of molecules. From long “brute-force” simulations it is possible to estimate the first part of the barrier $\Delta G(n)$ up to a maximum size n_{ref} . Beyond that size one needs to resort to US to compute $\Delta G(n)$ because the likelihood of having clusters containing more than n_{ref} molecules is too low to be directly sampled in a standard simulation. If a cluster of size $n > n_{\text{ref}}$ appears in the system it will most likely be the largest cluster, n^* . Within the US scheme the simulation is biased to sample configurations containing a cluster of size n^* around n_0 . This is achieved by adding a harmonic term of the form

$$\omega[n^*(\mathbf{r}^N), n_0] = \frac{1}{2}k[n^*(\mathbf{r}^N) - n_0]^2 \quad (3)$$

to the potential energy, where $n^*(\mathbf{r}^N)$ is the number of particles in the biggest cluster as detected by a given order parameter that depends on the particles' coordinates \mathbf{r}^N , and k and n_0 control the width and the centre of the “umbrella” respectively. The free-energy that follows from a US run is

$$\beta\Delta G(n^*) = -\ln \left\langle \frac{\delta(n^*(\mathbf{r}^N) - n^*)}{\exp^{-\beta\omega(n^*(\mathbf{r}^N), n_0)}} \right\rangle_w + \text{constant}. \quad (4)$$

In this way a US simulation gives $\Delta G(n^*)$, plus a constant, for a certain n^* interval. Several consecutive US simulations must

be performed to provide overlapping umbrellas whose n^* ranges from n_{ref} to the critical cluster size. Finally, all $\Delta G(n^*)$ sections are “glued” together and the resulting second part of the barrier, for $n > n_{\text{ref}}$ (which is lacking a constant), is in turn glued to the first part of the barrier (for $n < n_{\text{ref}}$) obtained, as above, in an unbiased simulation.

Umbrella Sampling simulations require the calculation of n^* by means of some order parameter. Such calculations can be time consuming and, in practise, instead of computing n^* after each trial MC move, one usually runs an unbiased trajectory consisting of the order of 10^1 – 10^2 MC sweeps and accepts the configuration resulting from such a trajectory according to:

$$\text{acc}(\text{old} \rightarrow \text{new}) = \min \left(1, e^{-\beta[\omega(n^*(\mathbf{r}_{\text{new}}^N), n_0) - \omega(n^*(\mathbf{r}_{\text{old}}^N), n_0)]} \right). \quad (5)$$

This procedure ensures the generation of configurations according to $e^{-\beta[U(\mathbf{r}^N) + \omega(n^*(\mathbf{r}^N), n_0)]}$ where U is the intermolecular potential energy.^{9,25–27}

III. Simulation details

We calculate the nucleation free-energy barrier of NaCl from its melt at $T = 800$ K and 825 K. For the ion–ion interactions we use the Tosi–Fumi rigid-ion potential, which has the following form:^{28,29}

$$U_{ij}(r) = A_{ij}e^{[B(\sigma_{ij}-r)]} - \frac{C_{ij}}{r^6} - \frac{D_{ij}}{r^8} + \frac{q_i q_j}{r}. \quad (6)$$

This pair potential is written as the sum of a Born–Mayer repulsion, two attractive van der Waals contributions and a Coulombic interaction term.

We perform molecular dynamics simulations with the GRO-MACS 4.5 package^{14,30} using the time reversible and symplectic velocity Verlet algorithm. The particle mesh Ewald method³¹ has been used to calculate the long-range electrostatic forces. We truncate the van der Waals part of the potential at 9 Å and add long range corrections by assuming $g(r) = 1$ beyond this cutoff.

The computed number density of ions in the bulk solid at 800 K and 825 K and 10^5 Pa is 0.041 Å^{-3} , in agreement with experiments.³² The density in the super-cooled liquid at the same temperature and pressure is 0.034 Å^{-3} .

All simulations are performed within a cubic simulation box in conjunction with periodic boundary conditions. A super-cooled system of the liquid at ambient pressure is obtained by cooling the fluid phase to below the melting temperature. For the present model, Aragones *et al.*³³ have recently computed its melting temperature to be $T_m = (1082 \pm 13) \text{ K}$, which is very close to the experimental melting temperature of $T_m^{\text{exp}} = 1074 \text{ K}$. Thus, using constant pressure Monte Carlo simulations, we cool the system down to $T_1 = 800 \text{ K}$ and $T_2 = 825 \text{ K}$, (corresponding to 26% and 23.8% under-cooling respectively), at which point we study crystal nucleation.

To distinguish between solid-like and liquid-like particles, and to identify the particles belonging to a solid cluster, we

used the same local bond order parameter as in ref. 20. First we compute a normalised complex vector q_4 for every ion i , whose m components are:

$$q_{4,m}(i) = \frac{\frac{1}{N_b(i)} \sum_j^{N_b(i)} Y_{4,m}(\theta_{i,j}, \phi_{i,j})}{\left(\sum_{m=-4}^4 |q_{4,m}(i)|^2 \right)^{1/2}}, \quad m = [-4, 4], \quad (7)$$

where $N_b(i)$ is the number of neighbours of particle i within a cut-off radius of 4 Å (the first minimum in the Na–Cl radial distribution function). Then we compute the scalar product $\left(\sum_{m=-4}^4 q_{4,m}(i) \cdot q_{4,m}^*(j) \right)$ for every particle i with each of its neighbouring particles j . Particle i would then be considered as ‘solid-like’ if at least 6 of its scalar products are greater than 0.35. Finally two ‘solid-like’ particles are considered to be neighbours in the same cluster if they are closer than 3.4 Å.

IV. Results

Initially we simulate molten NaCl at 800 K and 1 bar using HMC and discuss the dependence of the results on the chosen time-step or the number of MD steps within a HMC cycle, considering both the ‘Metropolised’ and ‘non-Metropolised’ versions of HMC. Next we combine HMC with Umbrella Sampling (HMC/US) to obtain free-energy barriers and compute the nucleation free-energy of NaCl from its melt, again at 800 K and 1 bar, for both the ‘Metropolised’ and ‘non-Metropolised’ versions of HMC, and compare it with that obtained by MC/US in ref. 20. To conclude, we explore approximations of the HMC/US technique, implementing HMC with the molecular dynamics trajectories now sampling the NpT ensemble, and compute the nucleation free-energy of NaCl from its melt at 800 K and 825 K.

A. Hybrid Monte Carlo

Fig. 1 represents the potential energy of molten NaCl at 800 K and 1 bar as a function of the time-step Δt used in the molecular dynamic runs. In all cases the number of HMC cycles is 30 000, where a HMC cycle consists of 50 steps of the MD NVE trajectory, whose final configuration is accepted or rejected based on the Metropolis criteria eqn (1), along with an attempt to change the system's volume (based on the usual accepting prescription in NpT MC runs). Results were obtained for a system composed of 600 NaCl pairs.

It can be seen that the results for the internal energy are independent of the time-step for the Metropolised HMC, even though the acceptance ratio of the Metropolis criteria is strongly dependent on Δt . For low values of Δt the acceptance ratio is close to 100% (notice that this acceptance ratio refers to that of eqn (1)). This suggests that, for small time-steps, a simple approximation could be used: the Metropolis criteria test may be omitted altogether given that the configurations obtained after the MD runs are always accepted. The results obtained with this ‘non-Metropolised’ scheme²² are also shown in Fig. 1. It can be seen that for Δt up to approximately

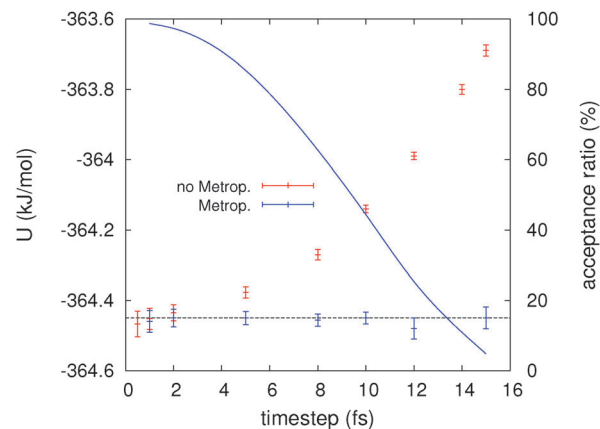


Fig. 1 Potential energy (per mole of ions) of NaCl at 800 K and 1 bar calculated via HMC for different MD time-steps with the MD trajectory set to 50 steps. Blue and red points refer to data calculated with the full HMC (‘Metropolised’) and the ‘non-Metropolised’ schemes, respectively (see text). The dashed line is the mean value of the potential energy of different full HMC runs. The continuous line represents the acceptance ratio of the Metropolised–HMC (eqn (1), y-axis on the right side). The volume move acceptance ratio was $\approx 60\%$.

2 fs, the results of the non-Metropolised scheme coincide with those for the full (Metropolised) HMC procedure. The same conclusions may be reached from the HMC calculations of Holzgräfe *et al.*²³ for a Lennard-Jones system, and Allen and Quigley²² for a soft-repulsive potential. Naturally, for larger values of Δt the non-Metropolised results increasingly depart from the Metropolised ones, although the difference is still not too large (0.2%) even for a time-step of 10 fs.

It could be assumed that a large time-step in conjunction with a small number of MD steps would be the optimal choice in order to reduce computational time. However, due to the reduced acceptance ratios encountered using large time-steps this may not be the case. One may wonder which is the best combination leading to the highest computational efficiency. By computational efficiency we mean the time required to obtain the potential energy (or any other thermodynamic quantity) within a given uncertainty. To evaluate this we split the simulations into 25 blocks and increase the number of MD cycles until the standard deviation of the sub-averages is below a given threshold. The results are shown in Fig. 2.

All the curves for a given time-step exhibit a minimum, indicating that for each time-step there is an optimum value of the number of MD steps. The CPU time needed to obtain the best accuracy is similar for all values of Δt . We conclude that the optimal length of the MD runs is around 50 steps where large values of the time-step are coupled to short MD trajectories.

B. Calculation of the free-energy barrier

An interesting application that can benefit from the use of HMC is the calculation of free-energy barriers using Umbrella Sampling.^{11,13,18,19} We undertake a similar study to that in the previous subsection, this time to test the calculation of the free-energy barrier of NaCl combining HMC and US. From the

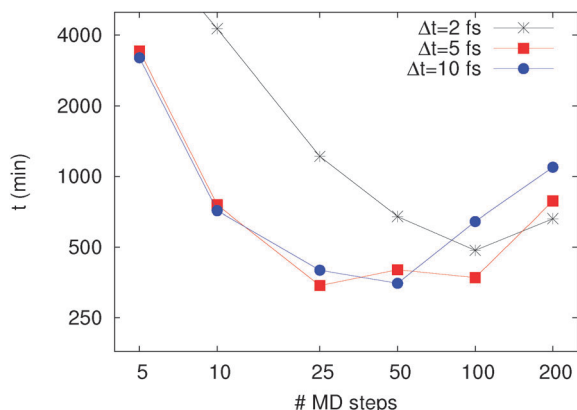


Fig. 2 Computer time (t) to obtain the potential energy within a given uncertainty as a function of the length of the MD trajectories for several time-steps Δt . Lines are a guide to the eye. Results were obtained at $T = 800$ K and $p = 1$ bar.

previous subsection it seems that the use of a large time-step results in only a modest reduction of the computational cost; from about 400 min for $\Delta t = 10$ fs to about 500 min for $\Delta t = 2$ fs (see Fig. 2). In view of this, to compute the free-energy barrier we shall use a time-step of 2 fs and NVE/MD runs of 100 steps. A typical run consists of $30\text{--}70 \times 10^3$ MD trajectories. The value of the biasing constant, k , was chosen to be $0.15\text{--}0.20k_{\text{B}}T$ which produces an acceptance (eqn (5)) close to 20%. The values of n_0 used in the Umbrella Sampling runs were 20, 30, ..., 120. For the Umbrella Sampling calculations we used a system composed of 1500 NaCl pairs.

Fig. 3 shows the results for the second part of the calculation of $\beta\Delta G$ (see eqn (4)) as a function of the cluster size n^* . In the figure we include calculations performed with both the full (Metropolised) HMC/US and the non-Metropolised (HMC(nM)/US) scheme. The uncertainty of each $\beta\Delta G$ section is about 0.1, thus the overall uncertainty of the second part of the barrier is of the order

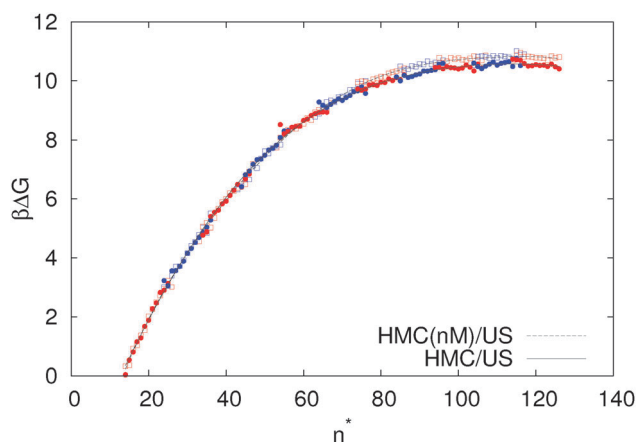


Fig. 3 Second part of the calculation of the free-energy barrier (eqn (4)) as a function of the cluster size n^* . HMC/US refers to calculation with the HMC coupled to Umbrella Sampling. For the results labelled as HMC(nM)/US the non-Metropolised approximation is used. Lines are a guide to the eye. Results were obtained at $T = 800$ K and $p = 1$ bar.

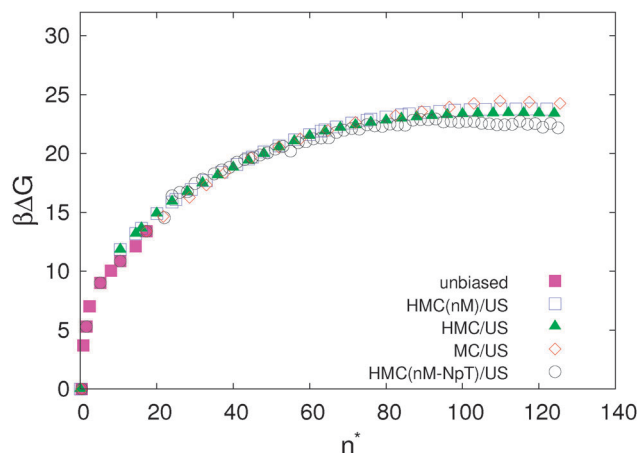


Fig. 4 Free energy barrier obtained merging the first (unbiased) and second parts (eqn (2) and (4)). For the results labelled as HMC(nM)/US the non-Metropolised approximation is used in HMC. HMC/US refers to the calculation with the HMC method coupled to an Umbrella Sampling. For comparison we have also shown the reported results using a MC/US scheme.²⁰ The HMC(nM-NpT)/US corresponds to a non-Metropolised approximation coupled with short NpT-MD runs. Results were obtained at $T = 800$ K and $p = 1$ bar.

of 0.5, significantly larger than the difference between the HMC/US and HMC(nM)/US curves.

To calculate the entire free-energy barrier we also evaluate its initial part, which has been computed according to eqn (2) with $n_{\text{ref}} = 17$. The second part shown in Fig. 3 is then “glued” to the first, thus obtaining the whole free-energy barrier. This procedure introduces some uncertainty in the complete calculation, which we estimate to be slightly larger than $\pm 1k_{\text{B}}T$. Fig. 4 shows the whole free-energy barrier as a function of the number of particles in the crystalline nucleus. Given that the first part is common to all schemes the differences between the results are all well within the estimated uncertainty of the calculations. Fig. 4 also shows the results reported for the same system using a MC/US scheme.²⁰ The difference at the top of the barrier is about $1k_{\text{B}}T$, indicative of an excellent agreement among all schemes. A summary of the numerical results obtained with the different approaches for the top of the barrier, and for the size of the critical cluster can be found in Table 1. As shown in Fig. 4 and Table 1, HMC(nM)/US, HMC/US and MC/US²⁰ predict the same shape of the free-energy barrier, the same height (within the combined error bars) and the same critical cluster size (within 10 particles).

The agreement between MC/US and HMC/US was to be expected, but the fact that HMC(nM)/US predicts the features of

Table 1 Top of the nucleation free-energy barrier ($\beta\Delta G_{\text{crit}}$) and the corresponding value of the critical cluster (n_{crit}^*) for each barrier reported in Fig. 4. Results were obtained at $T = 800$ K and $p = 1$ bar

	$\beta\Delta G_{\text{crit}}$	n_{crit}^*
HMC(nM)/US	23.8	115
HMC/US	23.5	116
MC/US ²⁰	24	120
HMC(nM-NpT)/US	22.6	114

the free-energy barrier is a relatively surprising conclusion, though understandable when considering the results shown in Fig. 1. In summary, for a small time step (*i.e.*, 2 fs) it is possible to accurately describe the free-energy barrier within the HMC(nM)/US approximation.

Our HMC implementation is performed in combination with the GROMACS MD package. With a simple wrapper script we iterate over a number of cycles, each consisting of a call to an NVE MD trajectory with moments chosen from a Gaussian distribution, followed by a Metropolis acceptance check of the final configuration (skipped in the approximate HMC(nM) scheme). To maintain the pressure constant we also perform a trial volume-change and accept it using the standard MC prescription of the NpT ensemble. This is entirely correct, but also very inefficient since one needs to make a call to the MD code from the wrapper script just to obtain the energy of the system for the trial volume.

To avoid the expensive volume move, we tried a new approach that, though approximate, may provide reasonable results. It consists of generating new configurations using short MD runs in the NpT ensemble, and accepting/rejecting the final configuration using the bias potential with the prescription provided in eqn (5). In this way, both the particles' positions and system's volume are modified in one go, thus avoiding the need for a specific trial volume move (and the increase in computational efficiency that this entails). Notice that this is an approximate route since we are not applying a Metropolis scheme to the final configuration generated with the short MD NpT runs. A correct implementation of this algorithm would require the use of a Metropolis scheme similar to that used in eqn (1), where the system's Hamiltonian, H , is now replaced by the conserved energy of the MD NpT run (denoted \tilde{H}).³⁴ The expression of this conserved quantity depends on the particular thermostat and barostat employed. Bussi *et al.*³⁵ and Leimkuhler and Reich³⁶ have expressed the conserved quantity \tilde{H} for different thermostats, indicating that correct results in the NVT ensemble can be obtained by generating configurations which are accepted or rejected according to $\min(1, \exp^{-\beta\Delta\tilde{H}})$ (thus correcting for the error introduced by a finite time step). In accordance with the results presented above, a way to avoid the implementation of such a Metropolis test is to use a small time step in the MD runs. We denote this scheme HMC(nM-NpT).

Let us now analyse whether this approximate route produces reasonable values for the free-energy barrier. We make use of the velocity-rescaling thermostat³⁵ and the Parrinello–Rahman barostat³⁷ with a relaxation time of 0.2 ps. The time step is set to 1 fs and the MD NpT runs consisted of 200 steps. Notice that the acceptance ratio for HMC depends more on the time step than on the length of the MD runs (see Fig. 4 of ref. 23). A simple bespoke code is used to evaluate n^* at the end of each trajectory in order to decide, *via* eqn (5), whether it is accepted or not. If a trajectory is accepted, the last configuration of the trajectory is used as a starting point for the subsequent one. In the case of rejection, the first configuration of the trajectory is used to start the next one. In both cases, upon continuing the simulation, we re-assign the momenta at

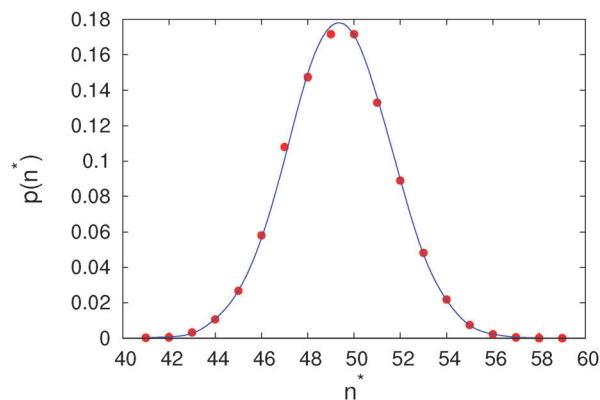


Fig. 5 Normalised probability distribution at 800 K and 1 bar of the largest cluster size obtained from Umbrella Sampling runs using $n_0 = 50$ and $k = 0.20k_B T$. Results obtained with HMC/US are represented with points. Results obtained with HMC(nM-NpT)/US are represented by a continuous line. In both cases 240×10^3 MD trajectories (2 fs and 100 MD steps each) were used.

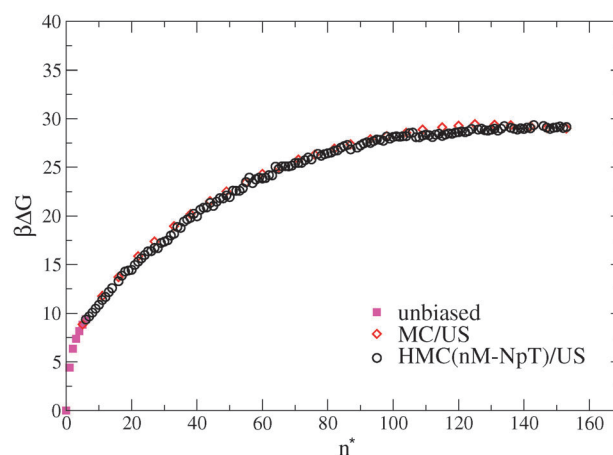


Fig. 6 Free energy barrier at 1 bar and 825 K computed using HMC(nM-NpT)/US, obtained merging the first (unbiased) and second parts (eqn (2) and (4)). For comparison we have also shown the reported results using a MC/US scheme under the same thermodynamic conditions.²⁰

random from a Boltzmann distribution consistent with the temperature of interest.

A first check is to compare the distribution of the largest clusters obtained using the HMC(nM-NpT) approximation with that provided by the “exact” (to within the statistical error) HMC/US scheme.

The results, shown in Fig. 5 for the umbrella around $n_0 = 50$, indicate that both methods lead to nearly identical results. We may then compute the second part of the free-energy barrier using the HMC(nM-NpT) approach. The free energy barrier calculated in this manner is shown in Fig. 4 and Table 1. The HMC(nM-NpT) results at 800 K agree with those of the “exact” schemes MC/US and HMC/US within the statistical noise.

Moreover, in order to ensure that the agreement observed at 800 K is not fortuitous, we have repeated the free-energy barrier calculations at 825 K.

Fig. 6 shows that, again, the HMC(nM-NpT) scheme yields results comparable to the “exact” ones.

V. Conclusions and discussion

The aim of this work was to investigate the evaluation of nucleation free-energy barriers using widely used molecular dynamic packages such as GROMACS or LAMMPS. Our present work is based on GROMACS, but most of the procedure we have used can be equally applied to LAMMPS. From the variety of numerical schemes developed to evaluate the nucleation free-energy barrier, we focus on the combination of the Hybrid Monte Carlo with the Umbrella Sampling method because it can be easily implemented. We have chosen to compute the crystallisation barrier of NaCl and compare it to the previously reported data using MC/US.

The first step we undertake is to examine whether the implementation of these packages is able to produce consistent results for the thermodynamic properties of NaCl in the context of HMC. We have shown that the procedure converges to the same value for the potential energy, irrespective of the selected time-step. As expected, the acceptance ratio of the Metropolis step strongly depends on the time-step. This implies that the computational efficiency of the overall method must be carefully checked. The optimal performance of the method is obtained for runs of between 25 and 100 steps, depending on the time-step. Interestingly, for the smallest time-steps, close to and below those usually employed in MD simulations, the acceptance ratio is only slightly below 100%. This suggests that when using a small time-step the Metropolis acceptance criteria may be omitted. We then analyse to which extent this “non-Metropolised” HMC(nM) approximation is able to account for the thermodynamic properties of the system. We conclude that for time-steps below 5 fs the results are acceptable (depending, of course, on the desired accuracy).

Once we have the optimised parameters of the HMC, we can tackle the calculation of the free-energy barrier by coupling the HMC to the Umbrella Sampling method. We demonstrate that the barriers obtained with the full HMC/US and the approximate HMC(nM)/US (using a small time-step) schemes are coincident to within statistical uncertainty. Moreover, these results are in excellent agreement with previous calculations using a combination of Monte Carlo and Umbrella Sampling.²⁰

The implementation of HMC is extremely simple and can be performed using MD packages such as GROMACS or LAMMPS. It requires the writing of a simple wrapper script iterating over a number of cycles, each consisting of a call to the NVE MD trajectory with moments chosen from a Gaussian distribution, and to the Metropolis acceptance check of the final configuration (skipped in the approximate HMC(nM) scheme). The calculation of the molecular interactions can be performed entirely with a well tested and extremely efficient external MD package. Thus, the effort of implementing sophisticated and efficient simulation techniques can be completely avoided. Moreover, the calculation of free-energy barriers incorporating the US sampling into the

HMC method does not involve the evaluation of the molecular interactions but that of the biasing potential, which is much simpler. Even though MD is more suitable for parallelization than MC, it is not so clear a priori whether launching short MD trajectories with a script wrapper is computationally more efficient than using an integrated MC/US scheme. The relative efficiency will strongly depend on the system size, on the number of steps of the short trajectories, and on the specific implementation of the MD and MC codes. In any case for the system considered in this work the combination of MD and US was significantly more efficient (from a computational point of view) than the combination of MC and US.

If one wishes to perform a study at constant pressure a volume move must be added to each cycle. This may be easily implemented in the wrapper script by two calls to the MD package in which the volume of the system is modified *via* trajectories consisting of a single step. Since MD programs usually require an initialisation period before starting calculations, calling the MD code for a single step may somewhat degrade the computational efficiency of the overall process. For this reason it is convenient to substitute the HMC methods based on the microcanonical ensemble for a method based on the isothermal-isobaric ensemble, thus avoiding any volume moves. This HMC-NpT must be based on a magnitude conserved throughout the trajectory (the idea of conserved magnitudes associated with a given barostat/thermostat has been already reported in the literature by other authors). Since the general features of the HMC-NpT method are similar to those outlined in this work for the HMC-NVE scheme, we may assume that, for small time-steps, the Metropolis acceptance check of each cycle may be skipped. In accordance with this idea, we have also calculated the free-energy barrier of NaCl crystallisation using the approximate HMC(nM-NpT) combined with the US method. The results are indistinguishable (to within the statistical noise) to those obtained in this work using the rigorous HMC/US procedure and to previous MC/US calculations.

To summarise, we have shown for the first time that free-energy barriers for nucleation obtained from different HMC schemes are the same (within the uncertainty) as those previously obtained using the standard MC/US scheme. It is also shown that those calculations can be performed easily with a molecular dynamic package such as GROMACS. Thus HMC techniques can indeed be useful in the future to obtain nucleation free-energy barriers for complex/molecular systems.

Acknowledgements

The authors thank the financial support from the MCINN Grant FIS2013-43209-P. CV and ES acknowledge the financial support from a Marie Curie Career Integration Grant (322326-COSAACFP7-PEOPLE-CIG-2012 and 303941-ANISOKINEQ-FP7-PEOPLE-CIG-2011 respectively), together with the Juan de La Cierva (JCI-2010-06602) and Ramon y Cajal (RYC-2010-06098). CM wishes to acknowledge the financial support from the

S2009/ESP-1691 (MODELICO-CM) project in the form of a post-doctoral contract.

References

- 1 A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12562.
- 2 R. J. Allen, P. B. Warren and P. R. ten Wolde, *Phys. Rev. Lett.*, 2005, **94**, 018104.
- 3 G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187.
- 4 J. S. van Duijneveldt and D. Frenkel, *J. Chem. Phys.*, 1992, **96**, 4655.
- 5 L. Filion, M. Hermes, R. Ni and M. Dijkstra, *J. Chem. Phys.*, 2010, **133**, 244115.
- 6 S. Auer and D. Frenkel, *Nature*, 2001, **409**, 1020.
- 7 I. Saika-Voivod, F. Romano and F. Sciortino, *J. Chem. Phys.*, 2011, **135**, 124506.
- 8 R. P. Sear, *Int. Mater. Rev.*, 2012, **157**, 328.
- 9 P. R. ten Wolde and D. Frenkel, *J. Chem. Phys.*, 1998, **109**, 9901.
- 10 V. K. Shen and P. G. Debenedetti, *J. Chem. Phys.*, 1999, **111**, 3581.
- 11 A. Reinhardt and J. P. K. Doye, *J. Chem. Phys.*, 2012, **136**, 054501.
- 12 E. Sanz, C. Vega, J. R. Espinosa, R. Caballero-Bernal, J. L. F. Abascal and C. Valeriani, *J. Am. Chem. Soc.*, 2013, **135**, 15008.
- 13 J. Palmer, F. Martelli, Y. Liu, R. Car, A. Panagiotopoulos and P. Debenedetti, *Nature*, 2014, **510**, 385.
- 14 B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435.
- 15 S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1.
- 16 S. Duane, A. D. Kennedy, B. J. Pendleton and D. Roweth, *Phys. Lett. B*, 1987, **195**, 216.
- 17 B. Mehlig, D. W. Heermann and B. M. Forrest, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1992, **45**, 679.
- 18 D. T. Limmer and D. Chandler, *J. Chem. Phys.*, 2011, **135**, 134503.
- 19 P. Geiger and C. Dellago, *J. Chem. Phys.*, 2013, **139**, 164105.
- 20 C. Valeriani, E. Sanz and D. Frenkel, *J. Chem. Phys.*, 2005, **122**, 194501.
- 21 M. E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*, Oxford University Press, Oxford, 2010.
- 22 M. P. Allen and D. Quigley, *Mol. Phys.*, 2013, **111**, 3442.
- 23 C. Holzgräfe, A. Bhattacharjee and A. Irback, *J. Chem. Phys.*, 2014, **140**, 044105.
- 24 A. Irback, *J. Chem. Phys.*, 1994, **101**, 1661.
- 25 L. Gelb, *J. Chem. Phys.*, 2003, **118**, 7747.
- 26 M. Schönherr, B. Slater, J. Hutter and J. VandeVondele, *J. Phys. Chem. B*, 2014, **118**, 590.
- 27 R. Iftimie and D. Salahub, *Chem. Phys.*, 2000, **113**, 4852.
- 28 F. Fumi and M. Tosi, *J. Phys. Chem. Solids*, 1964, **25**, 31.
- 29 M. Tosi and F. Fumi, *J. Phys. Chem. Solids*, 1964, **25**, 45.
- 30 S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess and E. Lindahl, *Bioinformatics*, 2013, **29**, 845.
- 31 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577.
- 32 G. J. Janz, *Molten Salts Handbook*, Academic Press, New York, 1987.
- 33 J. L. Aragones, E. Sanz, C. Valeriani and C. Vega, *J. Chem. Phys.*, 2012, **137**, 104507.
- 34 M. E. Tuckerman and G. J. Martyna, *J. Phys. Chem. B*, 2000, **104**, 159.
- 35 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.
- 36 B. Leimkuhler and S. Reich, *ESAIM: Math. Modell. Numer. Anal.*, 2009, **43**, 743.
- 37 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182.