

Received 6 August 2014

Accepted 18 December 2014

Edited by J. L. Smith, University of Michigan, USA

**Keywords:** crystallization; nucleation; molecular modelling; enhanced sampling; metadynamics

## Metadynamics studies of crystal nucleation

Federico Giberti,<sup>a</sup> Matteo Salvalaglio<sup>a,b,c</sup> and Michele Parrinello<sup>a,b\*</sup>

<sup>a</sup>Department of Chemistry and Applied Biosciences, ETH Zurich, CH-8092 Zurich, Switzerland, <sup>b</sup>Facoltà di informatica, Istituto di Scienze Computazionali, Università della Svizzera Italiana, CH-6900 Lugano, Switzerland, and <sup>c</sup>ETH Zurich, Institute of Process Engineering, Soneggstrasse 3, CH-8092 Zurich, Switzerland. \*Correspondence e-mail: parrinello@phys.chem.ethz.ch

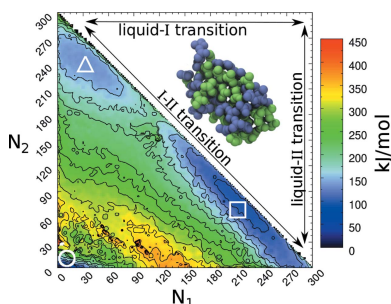
Crystallization processes are characterized by activated events and long timescales. These characteristics prevent standard molecular dynamics techniques from being efficiently used for the direct investigation of processes such as nucleation. This short review provides an overview on the use of metadynamics, a state-of-the-art enhanced sampling technique, for the simulation of phase transitions involving the production of a crystalline solid. In particular the principles of metadynamics are outlined, several order parameters are described that have been or could be used in conjunction with metadynamics to sample nucleation events and then an overview is given of recent metadynamics results in the field of crystal nucleation.

### 1. Introduction

Crystals are paradigmatic examples of long-range-ordered structures obtained through a self-assembly process *via* a first-order phase transition that starts from a disordered state, such as a liquid or a gas. The initial stage of a crystallization process, nucleation, is the fundamental and yet not well understood phenomenon leading to the formation of an embryonic structure with crystal-like properties. During nucleation, atoms or molecules gather together adopting a periodic pattern reproduced in space through crystal growth (Kashchiev, 2000; Debenedetti, 1996; Agarwal & Peters, 2013). Nucleation possesses a ubiquitous character that makes it one of the most important physical phenomena in nature and industry (Doherty, 2006; Agarwal & Peters, 2013). Liquid droplets in the upper atmosphere (Koop *et al.*, 2000; Li *et al.*, 2013), biomineralization processes such as the growth of gravity-sensing devices in insects and mammals (Gago-Duport *et al.*, 2008; Weiner *et al.*, 2003; Schüler, 2002; Schüler & Frankel, 1999), as well as the production of fine chemicals and drugs (Doherty, 2006) represent only a number of selected examples among the processes crucially depending on nucleation.

An elementary understanding of nucleation can be gained with classical nucleation theory (CNT) (Kashchiev, 2000; Debenedetti, 1996; Agarwal & Peters, 2013), initially developed by Gibbs for the formation of small liquid droplets in gas. It is based on the assumption that the free energy necessary to create a nucleus of  $n$  particles can be divided into a favourable term, proportional to the number of particles in the nucleus, and an unfavourable term, proportional to the dividing surface between the nucleus and the solution. The free energy difference can thus be analytically expressed as

$$\Delta G = -\Delta\mu n + \gamma S \quad (1)$$



where  $\Delta\mu$  is the difference in chemical potential between the crystal and the liquid phase,  $n$  is the number of molecules in the crystal phase,  $\gamma$  is the surface tension and  $S$  is the surface of the nucleus. In the above expression, the surface  $S$  can be expressed as a term proportional to  $n^\alpha$  in order to have an equation depending on the size of the cluster only, where  $\alpha$  relates to the scaling of the surface with respect to the volume of the nucleus. The simplicity of this expression is based on the assumptions that nuclei possess the same constant properties such as  $\Delta\mu$  and  $\gamma$  regardless of the cluster size  $n$ . This theory has proven to be very useful for many systems, and only in recent years has an in-depth investigation of the discrepancies between CNT predictions and experiments attracted increasing interest.

Among other mechanisms and theories that have been suggested and applied, an important role is played by the *two-step* nucleation mechanism, suggested by Vekilov (2004), Kuznetsov *et al.* (1998) and Ten Wolde & Frenkel (1997). According to the *two-step* mechanism, crystal nucleation is preceded by the formation of a dense liquid phase, in which the critical nucleus emerges and starts to grow. However, no single theory has been able to completely describe the process and a comprehensive theory is still missing. Consequently, despite its importance and intensive investigation, the true nature of the nucleation process remains elusive.

The main experimental difficulties in the study of the early stages of nucleation are the small dimension of the embryo nuclei and their transient nature. Normally, quantities that are indirectly correlated with nucleation events can be measured, such as the rate, or the crystal size and shape distribution. Only for some special cases has direct investigation on the nuclei been performed (Gasser *et al.*, 2001; Wu & Yang, 2001; Liang *et al.*, 2010; Kaiser *et al.*, 2002; Harano *et al.*, 2012). For these reasons, in recent years, computational techniques such as molecular dynamics (MD) and Monte Carlo (MC) have been applied to study nucleation (Auer & Frenkel, 2001; Lechner *et al.*, 2011; Sear, 2007; Zahn, 2004; Anwar & Zahn, 2011; Peters, 2009; Harding & Duffy, 2006; Duffy & Harding, 2004).

In MD, the Newton equations of motion are numerically solved for the atoms composing the system under study. The result of the simulation is a trajectory of the system, from which kinetic, dynamics and thermodynamics quantities can be calculated. In *ab initio* MD, electronic degrees of freedom are treated explicitly, limiting the number of approximations needed. However, the time required to run these simulations increases as the third power of the number of electrons; thus only systems composed of a few hundreds of atoms are typically accessible, for a timescale of the order of a few hundreds of picoseconds (Marx & Hutter, 2009). In contrast, in classical MD, where the interactions between nuclei are treated with simplified potentials, the number of atoms can be increased up to millions and timescales of microseconds can be reached. Consequently, nucleation events are typically simulated with classical MD (Frenkel & Smit, 2001). It is worth mentioning that simulations based on *ad hoc* neural network potentials can reach time and size scales comparable to those obtained in

standard MD, retaining an accuracy typical of quantum chemistry approaches, such as density functional theory (Behler & Parrinello, 2007; Sosso *et al.*, 2012, 2013).

An alternative approach to MD is to use MC simulations (Chandler, 1987; Frenkel & Smit, 2001). While MD is deterministic, MC is stochastic. In MC, a series of configurations is sequentially produced by performing random moves. The new configuration produced by a move is accepted or rejected so as to generate a desired statistical distribution. The ensemble of configurations generated in this way can be used to obtain static information on the system. In MC simulations the sequential generation of configurations does not relate to the time evolution of the system, consequently kinetic or dynamical information cannot be directly obtained from MC calculations.

The atomistic resolution of these methods makes them perfect candidates to investigate many-body problems such as crystal nucleation and growth. Unfortunately, the energy scale sampled by average fluctuations characteristic of MD or MC has an extent of a few  $k_B T$  while the energy barriers associated with nucleation can be much larger. The probability of observing a nucleation event is thus negligible on a typical molecular simulation timescale.

In the last two decades a number of enhanced sampling techniques have been developed and applied to address the nucleation problem without losing the characteristic atomistic resolution (Torrie & Valleau, 1977; Dellago *et al.*, 2002; Barducci *et al.*, 2008; Laio & Parrinello, 2002; Allen *et al.*, 2009; Huber *et al.*, 1994; Marsili *et al.*, 2006). A basic idea, common to several enhanced sampling methods, is to define a set of slow degrees of freedom, or collective variables (CVs), that identify all relevant states in the phase transition. The equilibrium distribution associated with these CVs is thus suitably altered to enhance the sampling of the rare event under investigation. Typically from the post-processing of enhanced sampling simulations the unperturbed equilibrium probability distribution in the space of the CVs can be recovered (Tiwary & Parrinello, 2015), thus enabling the calculation of thermodynamic and kinetic quantities associated with the rare event. In this review we shall restrict ourselves to the application of metadynamics (Barducci *et al.*, 2008; Laio & Parrinello, 2002), to obtain information on the mechanism and thermodynamics of homogeneous nucleation.

The paper is divided into three parts. In the first part we briefly illustrate metadynamics, highlighting its features. Since this method is based on the definition of a set of order parameters to coarsen the phase space, in the second part of this paper we provide an overview of some order parameters proposed in the literature to study crystallization problems. After this methodological introduction we will illustrate a number of examples appearing in the literature. For practical reasons, we will focus on papers in which metadynamics has been applied to obtain nucleation of a crystalline material. We apologize if, in doing so, we do not mention a number of excellent papers based on other methods that have also substantially contributed to our understanding of the nucleation phenomenon and use other enhanced sampling techni-

ques, such as umbrella sampling (Ten Wolde *et al.*, 1995, 1996*a,b*; Ten Wolde & Frenkel, 1998; Valeriani *et al.*, 2005), transition path sampling (Bolhuis, 2003; Moroni *et al.*, 2005; Zahn, 2004) or forward flux sampling (Li *et al.*, 2011).

### 2. Enhanced sampling with metadynamics

Metadynamics belongs to the family of enhanced sampling techniques in which the probability of visiting high free energy states is increased by adding to the Hamiltonian an adaptive external potential. Such a potential, acting on some slow degrees of freedom, discourages the revisiting of states that have already been sampled and improves the exploration of the phase space. The potential is typically applied in a low-dimensional subspace defined as a function of a set of CVs, that are usually defined as continuous functions of its microscopic coordinates. The external repulsive potential is typically written as a series of Gaussian functions that are deposited during a normal MD simulation in the space of CVs as

$$V_M(S, t) = \int_0^t \omega \exp\left(-\sum_{i=1}^d \frac{\{S_i(r) - S_i[r(t')]\}^2}{2\sigma_i^2}\right) dt', \quad (2)$$

where  $V_M(S, t)$  is the total bias potential deposited at time  $t$  in the space of the CV  $S$ ,  $d$  is the dimensionality of the CVs space,  $S_i(r)$  is the  $i$ th CV,  $S_i[r(t')]$  is the instantaneous value of the  $i$ th CV, where the Gaussian contribution is centred, and  $\omega$  is an energy deposition rate.

We will illustrate the effect of this potential with a simple illustrative example, in which we have a transition between two minima, separated by a high free energy barrier. In normal conditions, an MD simulation would be trapped in one of the two states, and a complete sampling of conformational space would be impossible in a practical time. However, if metadynamics is applied, the repulsive potential will act on the system pushing it out of the starting minima, over the barrier, into the second one. If the simulation is extended, the potential will eventually compensate the free energy surface, and the system will freely diffuse above the minima and the barrier, completely sampling the CVs space. At this point, an estimation of the free energy surface (FES) as a function of the set of CVs  $S$  can be obtained as the negative of the repulsive bias deposited during the course of the simulation

$$V_M(S, t \rightarrow \infty) = -F(S) + C. \quad (3)$$

However, understanding when to terminate a standard metadynamics simulation is not trivial. In real applications, the metadynamics potential tends to overfill the underlying FES rather than compensating it. Consequently, the difference in free energy between two points does not converge to a defined value, but rather oscillates around it (Barducci *et al.*, 2008; Tiwary & Parrinello, 2015). A solution to this problem has been suggested by Barducci *et al.* (2008), by introducing a history dependence for the rate of bias deposition  $\omega(S, t)$  as

$$\omega(S, t) = \omega_0 \exp\left[\frac{-V(S, t)}{k_B \Delta T}\right] \quad (4)$$

where  $\omega_0$  is the initial bias deposition rate,  $V(S, t)$  is the total bias deposited prior to time  $t$  and  $\Delta T$  is a parameter dimensionally homogeneous to a temperature that determines the rate of decay of the Gaussian contributions. The meaning of equation (4) is that the Gaussian functions deposited along the MD trajectory are scaled with an exponential decay, that depends on the history of the system and on the position in the CV space through  $V(S, t)$ . In the long time limit, the height of the Gaussian functions became negligible, and instead of compensating the FES, the well tempered (WT) metadynamics potential converges to the following function of the free energy  $F(S)$

$$V_M(S, t \rightarrow \infty) = -\frac{\Delta T}{T + \Delta T} F(S) + C. \quad (5)$$

Since WT metadynamics allows one to converge the free energy and obtain a Boltzmann distribution in the space of the CV  $S$  (Bussi *et al.*, 2006; Dama *et al.*, 2014), this technique opens the possibility to account for the effect of the bias deposited as a function of  $S$  on other degrees of freedom through the formulation of reweighing algorithms (Bonomi, Barducci *et al.*, 2009; Tiwary & Parrinello, 2015). This is particularly useful, as often the degrees of freedom biased in the metadynamics calculations are not those that can be most easily connected to experiments.

It is important to highlight that the deposition of the WT metadynamics potential can be tuned in such a way that a desired probability distribution is sampled, leading to an enhancement of the fluctuations of the system along the CVs. An important application of this property is the parallel tempering (PT) well tempered ensemble simulation protocol (Bonomi & Parrinello, 2010), in which the probability of exchange between adjacent replicas, typically performed in PT (Earl & Deem, 2005), is enhanced through coupling with WT metadynamics, performed as a function of the potential energy of the system. It is important to recall that the external potential enhancing the sampling of the transition between the minima and the transition states is a function of the CVs. Consequently, the capability of metadynamics to produce a trajectory which is able to properly sample the transitions strongly depends on the choice of the CVs. Since nucleation is a many-body problem, identifying the correct CVs to generate the crystalline nuclei is far from being a simple task. Several order parameters that have been or could be used as CVs in metadynamics simulations have been proposed in the literature. Some of them are discussed in the following paragraphs.

We conclude this section by summarizing the strengths and weaknesses of the method. Metadynamics calculations are easy to set up and post-process. In contrast to other methods, it does not require an *a priori* knowledge of the CVs space, and it can be used to blindly explore the possible configurations that the system under study possesses. However, since the time that it takes to converge the FES scales exponentially with the dimension of CVs, usually no more than three CVs are used. In cases where the exploration is slow, multiple simulations or ‘walkers’ that move under the action of the same external biasing potential can be used to speed up the exploration of

the FES (Raiteri *et al.*, 2006). We conclude by highlighting that in a metadynamics simulation the unperturbed dynamics and kinetics are lost. However, a framework to recover unperturbed kinetic rates has recently been suggested (Tiwary & Parrinello, 2013), and this could be used to easily relate experimental rates with metadynamics simulations.

### 3. Collective variables

Detecting symmetry, and thus identifying molecular structures possessing a crystalline arrangement, is a task that humans can quickly and efficiently carry out based on innate pattern-recognition skills. Developing algorithms that allow one to discriminate between disordered and ordered states based on the sole Cartesian coordinates of atoms or molecules for which a molecular simulation is carried out is instead a challenging task.

The crystalline state of ensembles of molecules is typically defined on the basis of *order parameters* (OPs): functions of the Cartesian coordinates that relate a numerical value to a spatial configuration of an ensemble of atoms or molecules. An OP is typically zero for a disordered phase while it assumes characteristic nonzero values for specific spatial ordered arrangements.

Order parameters are key to the analysis of crystallization simulations as well as CVs to implement enhanced sampling protocols such as metadynamics. In the following an overview of OPs is given, with the aim of providing the flavour of the problem and some key ideas necessary to address it. Rather than compiling an exhaustive list of all the OPs developed in the literature and used in crystallization we choose to provide examples of OPs according to an incremental degree of complexity. Starting from OPs used to describe the crystal state of systems in which constituent entities possess a spherical symmetry (*i.e.* atoms or spherical particles), we will then proceed by introducing OPs suited to study crystallization in mono-component molecular systems, and then OPs used in multi-component systems such as crystals growing from solution. In this last section particular emphasis will be given to OPs used in metadynamics applications such as CVs.

#### 3.1. Spherical constituent particles

*Steinhardt parameters.* Steinhardt parameters are bond-based OPs that have been widely used to describe the crystal packing of atoms, highly symmetric molecules, or more generally objects with a geometry that can be approximated to a sphere, such as colloidal particles (Steinhardt *et al.*, 1983). Steinhardt parameters have been applied in a variety of problems in material science, spanning from the phase transition of colloidal particles to the nucleation of ice (Trudu *et al.*, 2006; Matsumoto *et al.*, 2002; Gasser *et al.*, 2001).

First of all, the definition of the Steinhardt OPs is based on a concept of bonds that does not necessarily coincide with that of a *chemical* bond. In this context the bond between two constituent particles is defined as when their distance is below a certain threshold.

$\hat{r}_{ij}$  being the unitary vector defining the direction of the bond between particles  $i$  and  $j$ , the local environment around particle  $i$  can be characterized by the calculation of a set of  $\bar{q}_{lm}(i)$

$$\bar{q}_{lm}(i) = \frac{1}{N_b} \sum_{j=1}^{N_b} Y_{lm}(\hat{r}_{ij}), \quad (6)$$

where  $N_b$  is the number of bonds in which particle  $i$  is involved and  $Y_{lm}(\hat{r}_{ij})$  are spherical harmonics. The values of  $\bar{q}_{lm}(i)$  are strictly local, and thus to characterize the overall structure of an ensemble of  $N$  particles and obtain a global OP an average quantity has to be defined

$$\bar{Q}_{lm}(i) = \frac{\sum_{i=1}^N N_b(i) \bar{q}_{lm}(i)}{\sum_{i=1}^N N_b(i)}. \quad (7)$$

The Steinhardt OPs are defined as the first- and second-order invariant combinations of the  $\bar{Q}_{lm}(i)$

$$Q_l = \left( \frac{4\pi}{2l+1} \sum_{m=-l}^l |\bar{Q}_{lm}|^2 \right)^{1/2} \quad (8)$$

and

$$\hat{W}_l = \frac{W_l}{\left( \sum_{m=-l}^l |\bar{Q}_{lm}|^2 \right)^{3/2}} \quad (9)$$

where  $W_l$  is defined as

$$W_l = \sum_{\substack{m_1, m_2, m_3 \\ m_1 + m_2 + m_3 = 0}} \begin{pmatrix} l & l & l \\ m_1 & m_2 & m_3 \end{pmatrix} \bar{Q}_{lm_1} \bar{Q}_{lm_2} \bar{Q}_{lm_3} \quad (10)$$

and the term in parentheses is a Wigner  $3-j$  symbol.

Typical Steinhardt OPs used to characterize the crystal packing of Lennard-Jones spheres use four and six angular momenta channels. Third OPs are instead often used to characterize tetrahedral crystal environments such as the O atoms' sublattice in ice. In cases in which rotational invariance is not relevant or it is important to promote crystallization in a specific orientation, polynomial expressions inspired by spherical harmonics but computationally cheaper can also be used as OPs, as shown by Angioletti-Uberti *et al.* (2010).

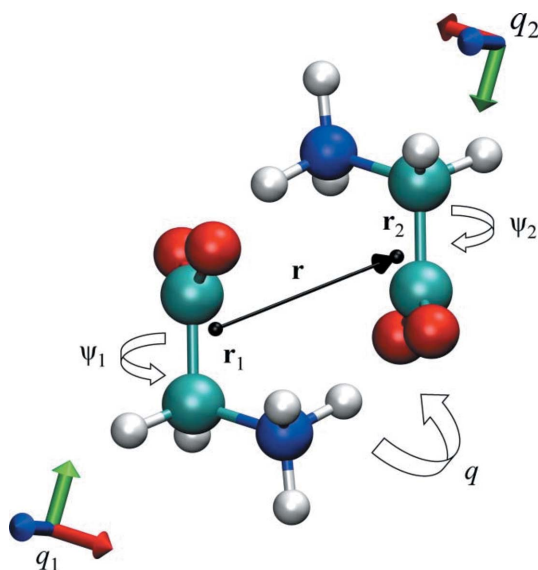
#### 3.2. Molecular crystals

In molecular crystals the simple relative position of neighbouring constituent particles is not sufficient to completely characterize the crystalline environment. In this case the complete description of the crystalline order has to account for the conformational flexibility of molecules and for the potentially wide variety of ordered arrangements. In this section we report the general framework for the construction of OPs in molecular crystals proposed by Santiso and Trout as well as some examples of OPs in which their ideas are implemented in a simplified manner. It is important to highlight that, rather than compiling an exhaustive list of all the OPs, we report some contributions that highlight different takes on the problem.

Describing crystal packing with an extended pair distribution function. A generalized approach to the construction of OPs able to distinguish between the ordered and disordered states of an ensemble of molecules has recently been proposed by Santiso & Trout (2011). Santiso OPs are constructed on the notion of an extended pair correlation function, describing not only the spatial correlation of molecules' positions, but also the spatial correlation of a set of axes describing the molecules' relative orientation and that of a set of relevant internal degrees of freedom, characterizing the structure of the molecule. By defining  $\psi$  as a given internal configuration of a molecule,  $\psi'$  the internal configuration of its neighbour,  $r$  their distance and  $q$  their relative orientation, the generalized pair distribution function  $G(r, q, \psi, \psi')$  represents the probability of finding, at distance  $r + dr$ , with mutual orientations  $q + dq$  from a molecule in configuration  $\psi$ , a second molecule in configuration  $\psi'$ . A graphical representation that helps in the rationalization of the extended pair distribution function  $G(r, q, \psi, \psi')$  is provided in Fig. 1. Such a function, for a perfect crystal at 0 K, can be modelled as a sum over the products of  $\delta$  functions as

$$G(r, q, \psi, \psi') = \sum_{i=1}^{\infty} \delta(r - r_i) \delta(q - q_i) \delta(\psi - \psi_i) \delta(\psi' - \psi'_i). \quad (11)$$

Such a sum is extended to  $\infty$ , but the relevant, non-redundant terms are those characterizing the unit cell. At finite temperature the convolution of delta functions becomes a characteristic probability distribution



**Figure 1** Construction of the extended pair distribution function. The vector  $\mathbf{r}$  is the vector joining the centres-of-mass of molecules 1 and 2, projected onto the molecule-centred frame of molecule 1 (represented by  $q_1$ ). The relative orientation  $q$  is the quaternion that rotates the frame of molecule 1 ( $q_1$ ) onto the frame of molecule 2 ( $q_2$ ), as seen from molecule 1. The internal degrees of freedom ( $\psi_1$  and  $\psi_2$ ) both are included in the pair distribution function. Reproduced with permission from Santiso & Trout (2011). Copyright 2011, AIP Publishing LLC.

$$G(r, q, \psi, \psi') = \sum_{i=1}^{\infty} f_i(r, q, \psi, \psi') \quad (12)$$

that can be approximated as the convolution of independent distributions

$$G(r, q, \psi, \psi') = \sum_{i=1}^{\infty} f_i(r) f_i(q) f_i(\psi) f_i(\psi'). \quad (13)$$

By defining specific probability distribution functions that describe  $f_i(r)$ ,  $f_i(q)$ ,  $f_i(\psi)$ ,  $f_i(\psi')$  specific OPs can be built. Such OPs can be constructed in order to account for relative distances, orientations and internal configurations of organic molecules or any subset of those properties. Several examples of specific OPs are provided in the Santiso work as well as in an application to benzene melt crystallization (Santiso & Trout, 2011; Shah *et al.*, 2011), an exhaustive review of which goes beyond the scope of this paper. However, we show below how the general ideas developed by Santiso can also be found in other OPs.

*Crystalline packing of polymeric chains.* Prototypical examples of systems in which conformational flexibility dominates in the definition of the crystalline order are polymer crystals. The nucleation from polymer melts is a complex problem requiring an important computational effort. The intra-chain degree of order is the crucial aspect that has to be captured by an OP aimed at the description of the crystal state of a polymer chain. A notable example of an OP successfully addressing this issue can be found in the works of the Rutledge group (Yi *et al.*, 2013). In this work the nucleation of crystals from polymer melts is typically characterized with the  $p_2$  OP. The local  $p_2$  OP is computed as

$$p_2(i) = \frac{1}{n_c} \sum_{j=1}^{n_c} \frac{3 \cos^2 \theta_{ij} - 1}{2}, \quad (14)$$

where  $\theta_{ij}$  is the angle between the vector oriented as the segment joining the  $(i-1)$ -th bead in a given polymer chain with the  $(i+1)$ -th bead, and the vector joining the  $(j-1)$ -th and the  $(j+1)$ -th beads. The quantity is then averaged over the  $n_c$  neighbouring beads lying within a predefined cutoff distance. A global OP can be constructed by averaging over all the  $i-j$  couples in the simulation box regardless of their distance.

*Asymmetry-based order parameters.* An intriguing alternative to conventional OPs used as CVs in enhanced sampling simulations or as post-processing tools comes from the biased Monte Carlo schemes employed in the works of Gavezzotti (2011, 2013). The degree of crystalline order in these works is quantified by the definition of translation and inversion asymmetry indexes  $s_{km}$ . For the couple of molecules  $k$  and  $m$  these indexes are, respectively, defined as

$$t_{km} = \frac{1}{N} \sum_{i=1}^{n_a} [(x_{ki} - x_{mi})^2]^{1/2} \quad (15)$$

and

$$i_{km} = \frac{1}{N} \sum_{i=1}^{n_a} [(x_{ki} + x_{mi})^2]^{1/2}, \quad (16)$$

where the index  $i$  runs over all the  $n_a$  atoms in a given molecule, and  $x_i$  is the position of atom  $i$  defined with respect to a reference atom. A global, single-molecule, symmetry index is then defined for molecule  $k$  as  $S_k = (N_{\text{mol}} - 1)^{-1} \sum_{m=1}^{N_{\text{mol}}} s_{km}$  while the total symmetry index for the whole simulation box can be obtained as  $S_{\text{all}} = (N_{\text{mol}}^{-1}) \sum_{k=1}^{N_{\text{mol}}} S_k$ . In this case, a perfect crystal has  $S_{\text{all}} = 0$  while disordered systems possess large  $S_{\text{all}}$  values.

### 3.3. Crystallization from solution

When crystallization takes place from a multi-component system, such as a binary solution, local density fluctuations of solute molecules are required to nucleate a crystal. As demonstrated by Giberti *et al.* (2013), producing solution domains locally concentrated is one of the limiting steps of the process; thus using OPs that account for local density is required to describe nucleation from solution. In their work Giberti *et al.* proposed using the concentration gradient as a CV to promote NaCl nucleation in water solution. It is clear that in the case of organic molecules the concentration gradient does not completely describe the crystal order observed in a molecular system as clearly discussed and systematically assessed by Santiso & Trout (2011). To account for this effect in our recent works on urea (Salvalaglio *et al.*, 2012, 2013; Giberti *et al.*, 2015; Salvalaglio *et al.*, 2015) and 1,3,5-tris(4-bromophenyl)benzene (3BrY) (Salvalaglio *et al.*, 2014) we have developed and used an OP that accounts for both the local density of solute molecules and their relative orientation. The global OPs can be written as the sum of individual molecular contributions ( $\Gamma_i$ ) over the whole ensemble of  $N$  solute molecules contained in the system of interest

$$S = \frac{1}{N} \sum_{i=1}^N \Gamma_i. \quad (17)$$

Each  $\Gamma_i$  contribution depends upon the local properties of the molecular environment that embeds the molecule  $i$ , namely its local density, measured using the coordination number, and its orientation with respect to its neighbours. The coordination number of molecule  $n_i$  is calculated using a sum of switching functions  $f_{ij}$

$$n_i = \sum_{j=1, j \neq i}^N f_{ij} = \sum_{j=1, j \neq i}^N \frac{1}{\{1 + \exp[a(r_{ij} - r_{\text{cut}})]\}}, \quad (18)$$

where  $r_{ij}$  is the distance between the centre of mass of molecules  $i$  and  $j$ . The state of molecule  $i$  is then determined by defining another switching function,  $\rho_i$ , in the domain of the coordination number  $n_i$  as a function of the lower coordination threshold for the crystal state  $n_{\text{cut}}$

$$\rho_i = \frac{1}{\{1 + \exp[-b(n_i - n_{\text{cut}})]\}}. \quad (19)$$

The slope of the switching functions can be tuned by choosing the parameters  $a$  and  $b$  in equations (18) and (19). To express the local order around  $i$ , rather than introducing a model of the generalized pair distribution function characteristic of a specific crystal structure as suggested by Santiso, an auxiliary function  $\theta_{ij}$  is used. This is defined in the domain of the angles between a given molecular vector describing the orientation of molecules  $i$  and  $j$  such that it is maximum for the characteristic angles observed in the crystal. We express  $\theta_{ij}$  as a sum of Gaussian functions centred at specific orientation angles  $\vartheta_k$  as

$$\theta_{ij} = \sum_{k=1}^{k_{\text{max}}} \exp\left[-\frac{(\vartheta_{ij} - \bar{\vartheta}_k)^2}{2\sigma_k^2}\right], \quad (20)$$

where  $k_{\text{max}}$  is the number of characteristic angles defining the local orientations of adjacent molecules. The choice of such a formulation introduces some flexibility in the sampling performed, *e.g.* with metadynamics. In fact, in addition to known structures, alternative local arrangements can be taken into account simply by expanding the series of  $\vartheta_k$  or by coupling more variables defined choosing different  $\vartheta_k$  values. The local OP  $\Gamma_i$  is then defined as

$$\Gamma_i = \frac{\rho_i}{n_i} \sum_{j=1}^N f_{ij} \theta_{ij}, \quad (21)$$

where also  $\theta_{ij}$  is weighted as a function of the distance between neighbours by multiplication by  $f_{ij}$  and normalized to one dividing by the coordination number  $n_i$ . Therefore, the global OP  $S$  becomes

$$S = \frac{1}{N} \sum_{i=1}^N \left( \frac{\rho_i}{n_i} \sum_{j=1}^N f_{ij} \left\{ \sum_{k=1}^{k_{\text{max}}} \exp\left[-\frac{(\vartheta_{ij} - \bar{\vartheta}_k)^2}{2\sigma_k^2}\right] \right\} \right), \quad (22)$$

where  $S$  approximates the molecular fraction of constituents possessing a crystal-like coordination number that are locally ordered according to the series of reference angles  $\vartheta_k$ .

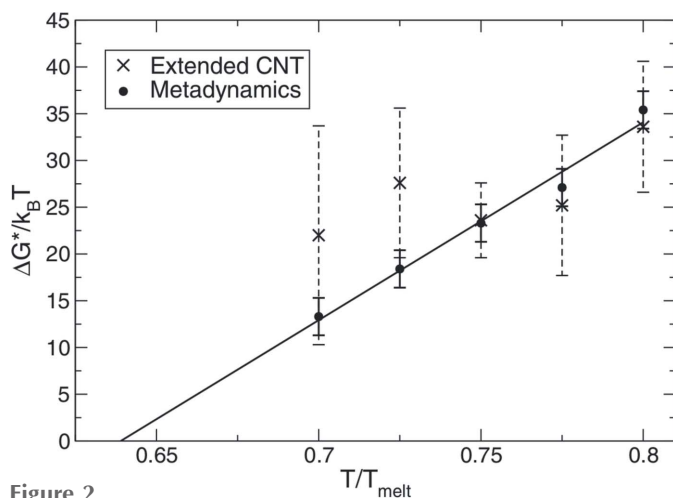
In a recent paper, Salvalaglio *et al.* (2015) investigated the nucleation process of urea from aqueous solution using the number of crystalline nuclei and their number of constituents. These order parameters are not used frequently to study the nucleation process with enhanced sampling techniques, since they do not possess a continuous definition. However, with metadynamics it is possible to obtain an FES by reweighing the bias deposited along the MD simulation (Bonomi, Barducci *et al.*, 2009; Tiwary & Parrinello, 2015). These order parameters could be used to link more easily with experiments where the rate of nucleation, and thus the number of nuclei per unit of time, can be directly observed.

## 4. Applications

In this part of the paper we will illustrate some examples of how metadynamics has been successfully applied to study problems related to crystallization.

*Lennard–Jones nucleation.* Metadynamics has been applied for the first time in the field of crystallization to study the freezing of an undercooled Lennard–Jones liquid. This system

has been intensely studied over the years with brute-force MD and with other biasing methods, providing useful insight into homogeneous nucleation processes (Ten Wolde *et al.*, 1995, 1996b; Ten Wolde & Frenkel, 1998; Bolhuis, 2003; Moroni *et al.*, 2005). The first advantage of using an enhanced sampling technique is that it is possible to study nucleation even at moderate and low undercooling, while severe undercooling has to be employed in brute-force MD or MC. This allows a better comparison with experimental data, normally sampled at low undercooling. Despite the simple structure and the absence of internal degrees of freedom, finding the correct set of CVs for this system has been a difficult task. Many authors have pointed out that the local Steinhardt parameters are the correct OPs to study the phase transition of this system, being able to identify both the f.c.c. (face-centred cubic) and the b.c.c. (body-centred cubic) arrangements emerging during the nucleation. Trudu *et al.* (2006) used the  $Q_6$  of a subset of atoms together with the potential energy of the system with plain metadynamics to drive the formation of the critical nucleus as a function of temperature. The nuclei obtained at low supercooling were for the most part non-spherical. In this regime the nucleation can be described with a modified CNT formalism that accounts for the asphericity of the nuclei. However, the mechanism changes substantially when the temperature is lowered to severe supercooling. In these conditions nucleation proceeds with the formation of more than one nucleus, that are rarely spherical and often present a fractal-like shape. The calculation of the rates as a function of temperature illustrate how at  $T/T_{\text{melt}} = 0.64$  the nucleation rate was approaching zero. An analysis of the potential energy of the system illustrated how, in the second case, the liquid was unstable, rather than metastable. These shows support the thesis that the barrier vanishes and the surface of the critical nucleus becomes very broad and eventually diverges. The trend of the nucleation

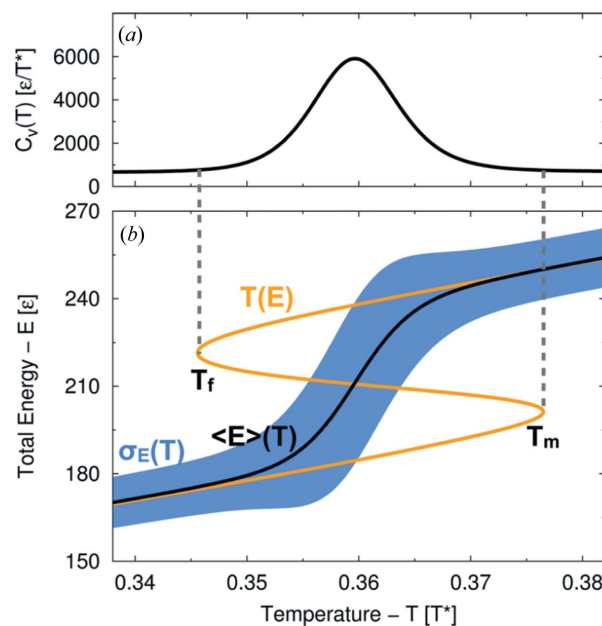


**Figure 2** The free energy barriers as computed from metadynamics at different temperatures (circles) are compared to the work of formation of the critical nuclei and as predicted within our extended CNT model (crosses with dashed error bars). A linear fit to the metadynamics data is also shown. Reproduced with permission from Trudu *et al.* (2006). Copyright (2006) The American Physical Society.

barriers as a function of the temperature, computed with metadynamics, is reported in Fig. 2.

Obtaining the FES associated with the nucleation profile from enhanced sampling techniques is of great relevance from a practical point of view. The direct calculation of the CNT FES from equation (1) with molecular simulations requires the evaluation of parameters such as the difference in chemical potential and the surface tension at the solid–liquid interface. This quantity is usually hard to evaluate, and normally only related quantities can be monitored. However, Angioletti-Uberti *et al.* developed a metadynamics framework from which surface tensions can be calculated in a simple and robust way (Angioletti-Uberti *et al.*, 2010). The key to the method is the use of a CV that is not invariant with respect to rotation, and thus enhances the formation of a solid/liquid interface oriented in a specific direction. By dividing the simulation box into two subcells and using a non-rotationally invariant degree of crystallinity in the two sub-domains as CVs, it is possible to enhance the formation of a crystal phase consistent with periodic boundary conditions. At the end of the calculation, the surface tension can thus be easily obtained as the difference between the free energy in the presence and absence of the solid/liquid interface, divided by the surface area.

Of particular interest for its thermodynamics analysis is the paper of Valsson & Parrinello (2013). In this paper, a parallel tempering well tempered ensemble (PT-WTE) (Bonomi & Parrinello, 2010) simulation was employed to achieve the nucleation of small droplets of 147 Lennard–Jones atoms. At



**Figure 3** Phase-coexistence region (LJ147-1 results): (a) canonical specific heat  $C_v(T)$ ; (b) canonical caloric curve  $\langle E \rangle(T)$  (black line), standard deviation of the total energy in the canonical ensemble  $\sigma_E(T)$  (blue shaded area) and microcanonical caloric curve  $T(E)$  (orange line, note that for this curve the independent variable, *i.e.* the total energy  $E$ , is on the vertical axis). Reproduced with permission from Valsson & Parrinello (2013). Copyright (2013) American Chemical Society.

this small size, the process is a quasi-first-order phase transition, since first-order phase transitions are only defined for samples with an infinite number of particles. Taking advantage of the Boltzmann distribution and combining the probability density function at different temperatures, the authors were able to construct the microcanonical caloric curve  $T(E)$  and the canonical caloric curve  $\langle E \rangle(T)$  (see Fig. 3). The two curves present a difference at the phase-coexistence temperature, that should vanish when an infinite amount of particles is considered. From the canonical curve, the variance and thus the specific heat were calculated. Although the calculation was performed on a toy system, the possibility of extracting this kind of information is valuable, in particular for nanoparticles and nanorods.

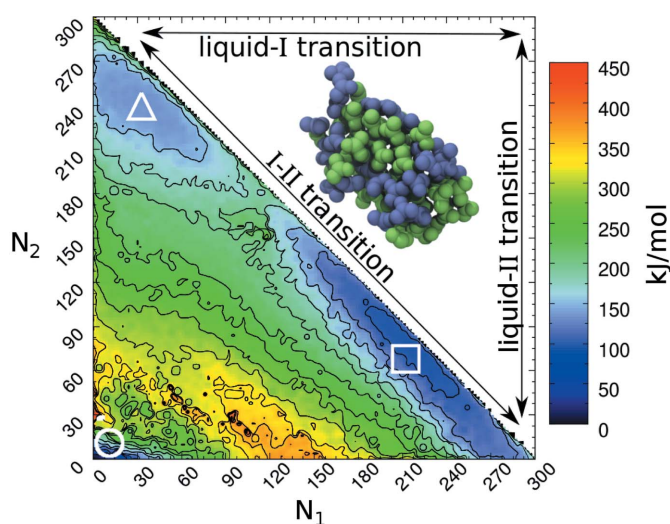
As illustrated in the previous section, molecular crystals possess a higher complexity if compared to the atomic crystals discussed so far. However, there are some interesting and promising results in the literature.

**Ice nucleation.** A prime example of nucleation is the formation of ice from water. The sampling limitations associated with the activated nature of nucleation have been recently elegantly addressed by Molinero *et al.* by introducing a simple model of water (Molinero & Moore, 2009), that allows the unbiased sampling of long timescales and the characterization of rare events without the need to introduce a bias (Knott *et al.*, 2012; Moore & Molinero, 2011). Contributions in understanding ice nucleation have also been obtained using metadynamics. Quigley *et al.* enhance nucleation in ice at moderate supercooling using plain metadynamics and four OPs,  $Q_6$ ,  $Q_4$ , a tetrahedral OP  $\chi$  (Radhakrishnan & Trout, 2003) and the potential energy of the system  $U$  (Quigley &

Rodger, 2008b). The CVs were chosen in order to sample the formation of both the  $I_c$  and  $I_h$  polymorphs. The nucleation of the solid phases was obtained through non-spherical embryos, typically stabilized by periodic boundary conditions. Quite often interstitial defects were included in the nucleating embryos, thus leading to an overestimation of the nucleation barrier. Despite the parameter choice, only a minor number of  $I_h$  water molecules were obtained, and the dominant crystal structure was found to be ice  $I_c$ , probably due to an effect of the periodic boundary conditions.

**Urea nucleation from the melt.** Giberti *et al.* (2015) recently provided the formulation of a simple OP, reported in the previous paragraphs, to be employed in the description of the phase transitions in molecular systems characterized by directional orientations and few internal degrees of freedom. The OP has been applied to nucleate urea from its melt, illustrating how even simple organic molecules can exhibit complex behaviour. Systems composed of 128, 300 and 1000 urea molecules were investigated just above the melting temperature. In the 128 case, a simulation at a deep quench of  $\simeq 0.28 T/T_{\text{melt}}$  and one above the melting temperature were also carried out. Several solid/liquid phase transitions were obtained for all the systems simulated. Information on the nucleation mechanisms was obtained from the 300 and 1000 molecules, revealing a behaviour that can hardly be captured by CNT. Nuclei obtained from WT metadynamics simulations were in fact composed of molecules arranged in two different crystallographic structures. The first one is the well known urea crystal structure, composed of *head-to-tail* dimers arranged in anti-parallel chains. The second is composed of cyclic dimers. An interesting finding of this work is that a variation in the ratio between the molecules arranged as one of the two polymorphs within the nucleus corresponds to a barrierless transition, as long as the number of solid particles in the nucleus is conserved. It is important to emphasize that the process observed does not correspond to the definition of the Ostwald step rule, as the two polymorphs are not mutually exclusive. An FES obtained from the reweighing of a WT metadynamics simulation showing minima that correspond to the melt, polymorph I and polymorph II is shown in Fig. 4.

**Sodium chloride nucleation from water solution.** So far, only studies of nucleation from the molten state have been discussed. However, metadynamics has been applied in homogeneous nucleation from solution too. Giberti *et al.* (2013) investigated the homogeneous nucleation of NaCl from aqueous solution close to saturation conditions. The formation of a nucleus was driven using as CV the gradient of the concentration in the simulation box. This OP directly correlates with the presence of an interface between the NaCl ions and the water molecules. Surprisingly, in addition to the well known rock salt structure, NaCl was nucleating with another structure similar to that of wurtzite. The latter, even if less stable than the rock salt, possesses a smaller surface tension due to more favourable interactions with the solvent. This suggests a possible Ostwald step rule, where the wurtzite phase transforms into rock salt after the nucleus reaches a certain size, where the two CNT profiles cross each other.



**Figure 4** FES reweighted as a function of  $N_1$  and  $N_2$ , indicating the number of molecules in a form I and form II structure, respectively. Landmarks highlight the position in the CVs space of the melt (circle), polymorph I (square) and polymorph II (triangle) configurations. The transitions between the three basins of the FES described in the text are reported in the upper triangle. A nucleus in which polymorph I (blue molecules) and II (green molecules) sub-domains coexist is also shown. Reproduced with permission from Giberti *et al.* (2015). Copyright (2015) Elsevier.



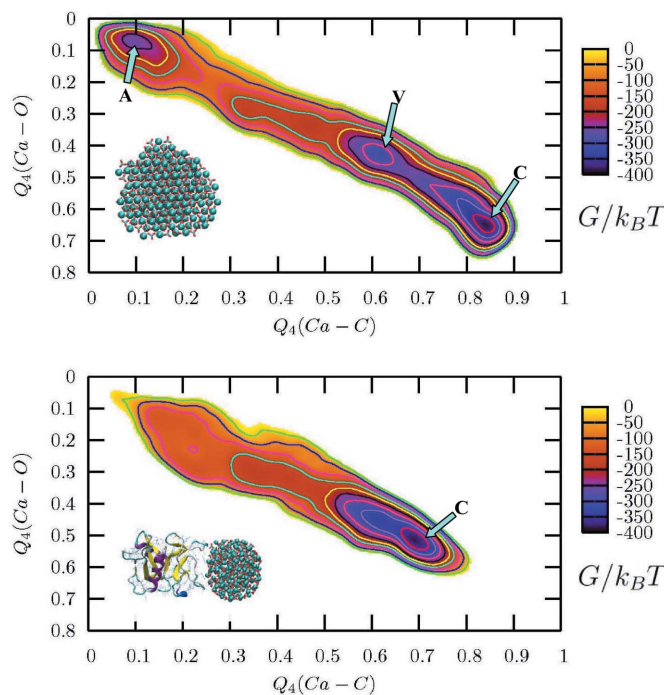
**Calcium carbonate nucleation from water solution.** In recent years, several studies aimed at clarifying calcium carbonate ( $\text{CaCO}_3$ ) nucleation in water have been carried out using metadynamics. Biased and unbiased MD simulations have predicted a non-classical mechanism for this inorganic compound (Quigley & Rodger, 2008a; Quigley *et al.*, 2011; Freeman *et al.*, 2010; Raiteri & Gale, 2010). In particular, several different amorphous and crystalline states, such as vaterite, aragonite and hydrated or anhydrous amorphous calcium carbonate (ACC), can be generated before forming the most stable crystalline solid (calcite). A comprehensive nucleation mechanism has still to be presented, and the relative stability of the polymorphs as well as the different nucleation rates are still debated. Consequently, metadynamics simulations provided interesting insight into the early stages of nucleation of  $\text{CaCO}_3$ , Quigley & Rodger (2008a), using as CV the local OP  $Q_4$  drive ACC particles of 300, 192 and 75 formula units to calcite in aqueous solution. A fine tuning of the OP allows them to identify all the possible polymorphs involved in the phase transition, ACC, vaterite, aragonite and calcite. While the larger and medium-size systems were simulated at constant pressure, the smallest was investigated at constant density too. In the 75 formula-units system, the relative stability of the ACC and the calcite structure was found to depend on the simulation set-up. In particular, at constant density the stable structure was found to be ACC, while at constant pressure calcite was the most

stable form. Also the barrier between the two structures changes remarkably, from several hundreds of  $k_B T$  at constant density to a few tens of  $k_B T$  at constant pressure. For all the other particle sizes, the ACC state was found to be metastable. The two states in all the reported cases were separated by a barrier of  $\simeq 350 k_B T$ , illustrating that the process needs a long time to occur, as highlighted by experimental results (Freeman *et al.*, 2010).

The authors were able to identify a shallow metastable minimum between the ACC and calcite basins that could have been identified probably with vaterite, but the number of vaterite-like particles in this state was found to be too small to unambiguously assign the minimum. Aragonite was not found in the phase-transition process. The information provided in these two papers has been used by Freeman *et al.* (2010) to simulate the effect that the ovocleidin-17 (OC-17) protein would have during the nucleation of a carbonate particle. Unbiased MD simulations of the protein in contact with nanoparticles composed of 192 and 300 formula units were used to select the most relevant binding configurations. Because of the different curvature, the protein was tightly bound to the 192 formula-units particles, with two clusters of arginine, while the interaction with the 300 formula-units particle was weaker. Then, using the same set-up as Quigley & Rodger (2008a), four metadynamics simulations for each nanoparticle were conducted. When OC-17 is present and bound to the ACC, the FES changes drastically, as displayed in Fig. 5. The barriers separating the ACC from the calcite minimum, as the intermediate vaterite-like basin, disappear. While for the small case, the protein remains bound to the  $\text{CaCO}_3$  particle, in the bigger one, after the calcite crystallization, the protein desorbs from the surface. This is an intriguing fact, suggesting that this protein could in fact be part of a loop in which ACC particles are bound and transformed in calcite. The OC-17 can desorb and restart the cycle with a new ACC particle. The early stages of carbonate aggregation were studied by Tribello *et al.* (2009), using umbrella sampling and metadynamics coupled with simple MD. As observed in experiments, their simulations suggested that the formation of a hydrated phase of amorphous  $\text{CaCO}_3$  is preferred rather than the nucleation of a calcite nucleus. They illustrated how water molecules could have been kinetically trapped inside an amorphous pre-critical cluster, preventing the formation of an anhydrous crystal.

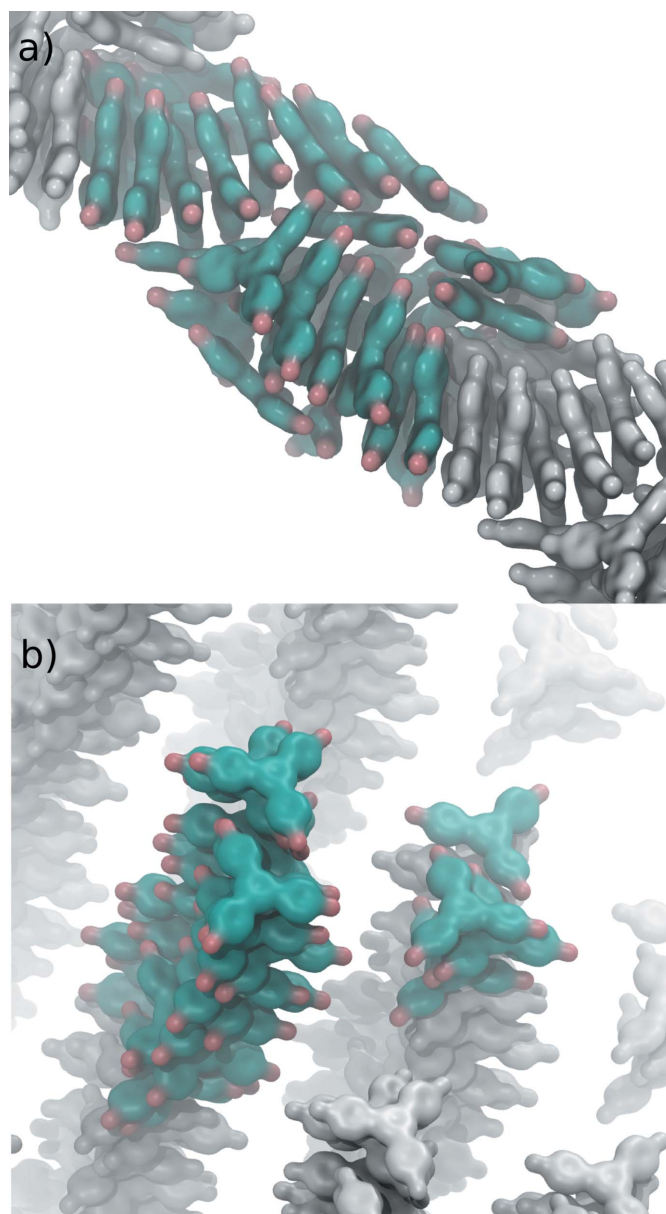
**1,3,5-Tris(4-bromophenyl)benzene nucleation from solution.** More recently, Salvalaglio *et al.* (2015) investigated with WT metadynamics the early stages of nucleation of 1,3,5-tris(4-bromophenyl)benzene (3BrY).

The nucleation mechanism proposed by Harano *et al.* (2012) for this system is a two-step mechanism, where the first stage is the formation of a dense droplet of 3BrY, in which the crystal nucleates (Vekilov, 2012). The structure of the clusters obtained from WT metadynamics simulation of the 3BrY nucleation was revealed to be strongly influenced by the solvent used. In particular, long columnar structures were obtained in ethanol, while in water, where the driving force for the nucleation was stronger, 3BrY clusters are characterized



**Figure 5** Projections of Gibbs free energy maps for a nanoparticle containing 192 units of  $\text{CaCO}_3$ . Top: nanoparticle in water. Bottom: nanoparticle with OC-17 bound in water. The OPs used for the axes measure symmetry in the arrangement of the C or O atoms about Ca ions. The letters label minima where local order is associated with a macroscopic polymorph: A is ACC; C is calcite; V is vaterite-like. Reproduced from Freeman *et al.* (2010) with permission from Wiley.

by a disordered structure. Examples of such structures are reported in Fig. 6. Regardless of the solvent used, at the basis of the formation of amorphous clusters there is a strong  $\pi$ -stacking interaction of the 3BrY molecules that drives them to align in a crystal-like configuration. Even in small oligomers, individual molecules are arranged with the same relative orientation displayed in the crystal. However, even if similar, the structure observed in solution and the structure of the crystalline bulk are characterized by a slightly different orientation of 3BrY molecules belonging to adjacent columns, suggesting that the small size of the crystal could actually play a role in its internal structure.



**Figure 6**  
Examples of cluster structures extracted from the metadynamics trajectories in water (a) and methanol (b). Reproduced with permission from Salvalaglio *et al.* (2014).

## 5. Conclusions

To conclude, in this brief review we have given a description of the ingredients necessary to investigate crystallization processes with metadynamics. First of all it is important to remark that using an enhanced sampling approach such as metadynamics is essential to compute the FES associated with the phase transition and to provide an efficient sampling of nucleation events. We highlight that the choice of the OPs to be used as collective variables is of paramount importance; we have therefore provided examples of OPs that have been or could be used in conjunction with metadynamics to explore crystallization in systems at increasing levels of complexity. Finally, a series of recent examples of crystallization problems addressed using metadynamics have been reported.

## Acknowledgements

For some of the figures included in this review, the authors would like to thank Erik Santiso and Bernard Trout (Fig. 1), Federica Trudu and Davide Donadio (Fig. 2), Omar Valsson (Fig. 3), Colin L. Freeman, John H. Harding, David Quigley and P. Mark Rodger (Fig. 5). The authors also acknowledge the work of the *PLUMED* (<http://www.plumed-code.org>; Bonomi, Branduardi *et al.*, 2009; Tribello *et al.*, 2014) development team. *PLUMED* is an external plugin that can be interfaced with some of the most popular MD codes to perform free energy calculations as a function of many CVs, using state-of-the-art methods including metadynamics, umbrella sampling and steered MD.

## References

- Agarwal, V. & Peters, B. (2013). In *Advances in Chemical Physics*, Vol. 155, edited by S. A. Rice & A. R. Dinner. New Jersey: John Wiley.
- Allen, R. J., Valeriani, C. & Ten Wolde, P. R. (2009). *J. Phys. Condens. Matter*, **21**, 463102.
- Angioletti-Uberti, S., Ceriotti, M., Lee, P. D. & Finnis, M. W. (2010). *Phys. Rev. B*, **81**, 125416.
- Anwar, J. & Zahn, D. (2011). *Angew. Chem. Int. Ed.* **50**, 1996–2013.
- Auer, S. & Frenkel, D. (2001). *Nature (London)*, **409**, 1020–1023.
- Barducci, A., Bussi, G. & Parrinello, M. (2008). *Phys. Rev. Lett.* **100**, 020603.
- Behler, J. & Parrinello, M. (2007). *Phys. Rev. Lett.* **98**, 146401.
- Bolhuis, P. (2003). *J. Phys. Condens. Matter*, **15**, S113–S120.
- Bonomi, M., Barducci, A. & Parrinello, M. (2009). *J. Comput. Chem.* **30**, 1615–1621.
- Bonomi, M., Branduardi, D., Bussi, G., Camilloni, C., Provasi, D., Raiteri, P., Donadio, D., Marinelli, F., Pietrucci, F., Broglia, R. A. & Parrinello, M. (2009). *Comput. Phys. Commun.* **180**, 1961–1972.
- Bonomi, M. & Parrinello, M. (2010). *Phys. Rev. Lett.* **104**, 190601.
- Bussi, G., Laio, A. & Parrinello, M. (2006). *Phys. Rev. Lett.* **96**, 090601.
- Chandler, D. (1987). *Introduction to Modern Statistical Mechanics*. Oxford University Press.
- Dama, J., Parrinello, M. & Voth, G. (2014). *Phys. Rev. Lett.* **112**, 240602.
- Debenedetti, P. G. (1996). *Metastable Liquids: Concepts and Principles*. Princeton University Press.
- Dellago, C., Bolhuis, P. & Geissler, P. L. (2002). *Adv. Chem. Phys.* **123**, 1–78.
- Doherty, M. (2006). *Chem. Eng. Educ.* **40**, 116–125.
- Duffy, D. M. & Harding, J. H. (2004). *Langmuir*, **20**, 7630–7636.

- Earl, D. J. & Deem, M. W. (2005). *Phys. Chem. Chem. Phys.* **7**, 3910–3916.
- Freeman, C. L., Harding, J. H., Quigley, D. & Rodger, P. M. (2010). *Angew. Chem. Int. Ed.* **49**, 5135–5137.
- Frenkel, D. & Smit, B. (2001). *Understanding Molecular Simulation: from Algorithms to Applications*, Vol. 1. London: Academic Press.
- Gago-Duport, L., Briones, M., Rodriguez, J. & Covelo, B. (2008). *J. Struct. Biol.* **162**, 422–435.
- Gasser, U., Weeks, E., Schofield, A., Pusey, P. & Weitz, D. (2001). *Science*, **292**, 258–262.
- Gavezzotti, A. (2011). *CrystEngComm*, **13**, 3573–3579.
- Gavezzotti, A. (2013). *Cryst. Growth Des.* **13**, 3801–3815.
- Giberti, F., Salvalaglio, M., Mazzotti, M. & Parrinello, M. (2015). *Chem. Eng. Sci.* **121**, 51–59.
- Giberti, F., Tribello, G. A. & Parrinello, M. (2013). *J. Chem. Theory Comput.* **9**, 2526–2530.
- Harano, K., Homma, T., Niimi, Y., Koshino, M., Suenaga, K., Leibler, L. & Nakamura, E. (2012). *Nature Mater.* **11**, 877–881.
- Harding, J. & Duffy, A. D. (2006). *J. Mater. Chem.* **16**, 1105–1112.
- Huber, T., Torda, A. E. & van Gunsteren, W. F. (1994). *J. Comput. Mol. Des.* **8**, 695–708.
- Kaiser, U., Muller, D., Grazul, J., Chuvilin, A. & Kawasaki, M. (2002). *Nature Mater.* **1**, 102–105.
- Kashchiev, D. (2000). *Nucleation, Basic Theory with Applications*. Oxford: Elsevier.
- Knott, B., Molinero, V., Doherty, M. & Peters, B. (2012). *J. Am. Chem. Soc.* **134**, 19544–19547.
- Koop, T., Luo, B., Tsias, A. & Peter, T. (2000). *Nature (London)*, **406**, 611–614.
- Kuznetsov, Y. G., Malkin, A. & McPherson, A. (1998). *Phys. Rev. B*, **58**, 6097.
- Laio, A. & Parrinello, M. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 12562–12566.
- Lechner, W., Dellago, C. & Bolhuis, P. G. (2011). *Phys. Rev. Lett.* **106**, 085701.
- Li, T., Donadio, D. & Galli, G. (2013). *Nature Commun.* **4**, 1887.
- Li, T., Donadio, D., Russo, G. & Galli, G. (2011). *Phys. Chem. Chem. Phys.* **13**, 19807–19813.
- Liang, Y., Lynn, D. G. & Berland, K. M. (2010). *J. Am. Chem. Soc.* **132**, 6306–6308.
- Marsili, S., Barducci, A., Chelli, R., Procacci, P. & Schettino, V. (2006). *J. Phys. Chem. B*, **110**, 14011–14013.
- Marx, D. & Hutter, J. (2009). *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge University Press.
- Matsumoto, M., Saito, S. & Ohmine, I. (2002). *Nature (London)*, **416**, 409–413.
- Molinero, V. & Moore, E. (2009). *J. Phys. Chem. B*, **113**, 4008–4016.
- Moore, E. & Molinero, V. (2011). *Nature (London)*, **479**, 506–508.
- Moroni, D., Ten Wolde, P. & Bolhuis, P. (2005). *Phys. Rev. Lett.* **94**, 235703.
- Peters, B. (2009). *J. Chem. Phys.* **131**, 244103.
- Quigley, D., Freeman, C., Harding, J. & Rodger, P. M. (2011). *J. Chem. Phys.* **134**, 044703.
- Quigley, D. & Rodger, P. M. (2008a). *J. Chem. Phys.* **128**, 221101.
- Quigley, D. & Rodger, P. M. (2008b). *J. Chem. Phys.* **128**, 154518.
- Radhakrishnan, R. & Trout, B. L. (2003). *J. Am. Chem. Soc.* **125**, 7743–7747.
- Raiteri, P. & Gale, J. D. (2010). *J. Am. Chem. Soc.* **132**, 17623–17634.
- Raiteri, P., Laio, A., Gervasio, F. L., Micheletti, C. & Parrinello, M. (2006). *J. Phys. Chem. B*, **110**, 3533–3539.
- Salvalaglio, M., Giberti, F. & Parrinello, M. (2014). *Acta Cryst.* **C70**, 132–136.
- Salvalaglio, M., Perego, C., Giberti, F., Mazzotti, M. & Parrinello, M. (2015). *Proc. Natl Acad. Sci. USA*, **112**, E6–E14.
- Salvalaglio, M., Vetter, T., Giberti, F., Mazzotti, M. & Parrinello, M. (2012). *J. Am. Chem. Soc.* **134**, 17221–17233.
- Salvalaglio, M., Vetter, T., Mazzotti, M. & Parrinello, M. (2013). *Angew. Chem. Int. Ed.* **52**, 13369–13372.
- Santiso, E. & Trout, B. (2011). *J. Chem. Phys.* **134**, 064109.
- Schüler, D. (2002). *Int. Microbiol.* **5**, 209–214.
- Schüler, D. & Frankel, R. B. (1999). *Appl. Microbiol. Biotechnol.* **52**, 464–473.
- Sear, R. P. (2007). *J. Phys. Condens. Matter*, **19**, 033101.
- Shah, M., Santiso, E. & Trout, B. (2011). *J. Phys. Chem. B*, **115**, 10400–10412.
- Sosso, G. C., Miceli, G., Caravati, S., Behler, J. & Bernasconi, M. (2012). *Phys. Rev. B*, **85**, 174103.
- Sosso, G. C., Miceli, G., Caravati, S., Giberti, F., Behler, J. & Bernasconi, M. (2013). *J. Phys. Chem. Lett.* **4**, 4241–4246.
- Steinhardt, P., Nelson, D. & Ronchetti, M. (1983). *Phys. Rev. B*, **28**, 784–805.
- Ten Wolde, P. R. & Frenkel, D. (1997). *Science*, **277**, 1975–1978.
- Ten Wolde, P. & Frenkel, D. (1998). *J. Chem. Phys.* **109**, 9901–9918.
- Ten Wolde, P., Ruiz-Montero, M. & Frenkel, D. (1995). *Phys. Rev. Lett.* **75**, 2714–2717.
- Ten Wolde, P., Ruiz-Montero, M. & Frenkel, D. (1996a). *J. Chem. Phys.* **104**, 9932–9947.
- Ten Wolde, P., Ruiz-Montero, M. & Frenkel, D. (1996b). *J. Chem. Phys.* **104**, 9932–9947.
- Tiwary, P. & Parrinello, M. (2013). *Phys. Rev. Lett.* **111**, 230602.
- Tiwary, P. & Parrinello, M. (2015). *J. Phys. Chem. B*, **119**, 736–742.
- Torrie, G. M. & Valleau, J. P. (1977). *J. Comput. Phys.* **23**, 187–199.
- Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. (2014). *Comput. Phys. Commun.* **185**, 604–613.
- Tribello, G. A., Bruneval, F., Liew, C. & Parrinello, M. (2009). *J. Phys. Chem. B*, **113**, 11680–11687.
- Trudu, F., Donadio, D. & Parrinello, M. (2006). *Phys. Rev. Lett.* **97**, 105701.
- Valeriani, C., Sanz, E. & Frenkel, D. (2005). *J. Chem. Phys.* **122**, 194501.
- Valsson, O. & Parrinello, M. (2013). *J. Chem. Theory Comput.* **9**, 5267–5276.
- Vekilov, P. G. (2004). *Cryst. Growth Des.* **4**, 671–685.
- Vekilov, P. G. (2012). *Nature Mater.* **11**, 838–840.
- Weiner, S., Levi-Kalisman, Y., Raz, S. & Addadi, L. (2003). *Connect. Tissue Res.* **44**, 214–218.
- Wu, Y. & Yang, P. (2001). *J. Am. Chem. Soc.* **123**, 3165–3166.
- Yi, P., Locker, C. & Rutledge, G. (2013). *Macromolecules*, **46**, 4723–4733.
- Zahn, D. (2004). *Phys. Rev. Lett.* **92**, 040801.