Faraday
Discussions

Accepted Manuscript

Faraday Discussions

Royal Society of Chemistry

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/faraday_d

PAPER   www.rsc.org/faraday_d | Faraday Discussions

Urea homogeneous nucleation mechanism is solvent dependent

Matteo Salvalaglio,[*a,b] **Marco Mazzotti,**[a] **and Michele Parrinello**[b]

The composition of the mother phase plays a primary role in crystallization processes, affecting both crystal nucleation and growth. In this work the influence of solvents on urea nucleation has been investigated by means of enhanced sampling molecular dynamics simulations. We find that, depending on the solvent, the nucleation process can either follow a single-step or a two-step mechanism. While in methanol and ethanol a single-step nucleation process is favored, in acetonitrile a two-step process emerges as the most likely nucleation pathway. We also find that solvents have a minor impact on polymorphic transitions in the early stages of urea nucleation. The impact of finite sizes on the free energy surfaces are systematically considered and discussed in relation to the simulation setup.

1 Introduction

Fine chemicals and active pharmaceutical ingredients are typically produced in crystalline form. In such case it is of great importance to control both polymorphism and particles morphology. Both these characteristics in fact contribute crucially to the definition of key properties such as dissolution kinetics[1] and bioavailability[2]. Crystallization of organic molecules is typically carried out in solution. Addressing the role of solvents and predicting their impact on the process is therefore crucial to control and design crystallization. Solvents in fact can influence both the crystal morphology as well as the bulk crystal structure[3,4]. Crystallization proceeds in two subsequent stages: nucleation and growth. Nucleation is the process of formation of the first stable embryonic structure of a crystal, that eventually grows to become a macroscopic particle[5]. Understanding the effect of the mother phase composition on crystal growth might allow steering the crystallization process towards a predetermined outcome. As such, a large body of experimental and theoretical work deals with the impact of solvents and additives on crystal growth[6–8].

In this work we focus on nucleation with the aim of uncovering the effect of solvents on the nucleation mechanism by means of molecular simulations. Studying nucleation from solution represents a challenge for both experiments and simulations[9,10]. On the one hand nucleation is initiated at the molecular scale, thus making its direct experimental investigation extremely challenging. On the other hand, nucleation is a paradigmatic example of rare event, characterized by timescales that often more than not exceed those accessible to molecular

simulations. Therefore, while molecular modelling techniques display the clear advantage of providing an atomistic resolution of nucleation, timescale limitations require to be properly addressed[11]. To this aim a variety of enhanced sampling methods based on molecular dynamics have been proposed[12,12–21]. Taking advantage of these tools much work has been devoted to the investigation of homogeneous nucleation in simple model systems like Lennard-Jones particles or hard spheres[22–24]. Recently however the attention has been shifted to more complex and practically more relevant systems such as organic molecules in solution[11,25–34]. In this work we use Well Tempered (WT) metadynamics, a state-of-the-art enhanced sampling method that allows investigating nucleation events in explicit solution[12,35–38].

In particular, we focus our study on the effect of organic solvents such as acetonitrile (ACN), methanol (MeOH) and ethanol (EtOH) on the nucleation of urea crystals. Urea, despite being a simple organic molecule, exhibits a rich crystallization behaviour that has been extensively investigated through both experimental and computational studies[8,29,33,39–44]. It has emerged that different crystal faces grow following different mechanisms[8,41], and that solvents and additives may substantially affect urea crystal morphologies that can range from needle-like particles to compact tetrahedra[40,43,44]. A two-step nucleation mechanism has been identified as the most probable pathway for homogeneous nucleation of urea from aqueous solution[33]. Moreover, analysing both urea nucleation from the melt[29] and from aqueous solution[33] it has been found that two polymorphs compete in the early stages of nucleation: *form I*, the experimentally known urea crystal structure and *form II* a metastable polymorph. In this paper we investigate the effect of solvents on the nucleation mechanism of urea, focusing our attention on both the competition between polymorphs and on the dominant nucleation mechanisms. Similarly to the nucleation of liquid droplets from a vapour[45–47], the simulation of crystal nucleation from solution is heavily affected by finite-size effects[33,48]. In order to deal with these effects we shall apply the strategy discussed in detail in Ref.[33].

2 Theory and Methods

2.1 Reversible work of nucleation in a finite-sized system

Molecular dynamics (MD) simulations of crystallization processes in solution are conventionally performed in the isothermal-isobaric ensemble, in which temperature, pressure and number of atoms are kept constant. When nucleation takes place in a simulation box, the chemical potential of the mother solution inherently couples with the size of the nucleus, as described in depth in Ref.[33]. This effect is reflected in the free energy profile associated with the formation of a finite sized crystal nucleus of size N_c. In Ref.[33], the reversible work for the formation of a crystalline nucleus of size N_c in a confined system is thus derived as:

$$\Delta G_{\ell \to c} = -N_c k_\mathrm{B} T \ln\left(\frac{\gamma x}{\gamma^* x^*}\right) + \sigma' N_c^{\frac{2}{3}} + N_\mathrm{tot} k_\mathrm{B} T \ln\left(\frac{\gamma x}{\gamma_0 x_0}\right)$$
$$+ N_s k_\mathrm{B} T \ln\left(\frac{\gamma_s (1 - x)}{\gamma_{s,0} (1 - x_0)}\right) \tag{1}$$

where: N_c is the number of molecules belonging to the crystal phase, k_B the Boltzmann constant, T the temperature, γ_i are the activity coefficients, x is the actual molar fraction of the solute, x^* the molar fraction of the solute at equilibrium with the solid, x_0 its molar fraction in absence of any crystal-like nuclei, N_s the number of solvent molecules, N_{tot} the total number of solute molecules in the system, and σ' the effective surface energy that takes into account both the surface tension and the shape of the nucleus.

For macroscopic systems, where $x \simeq x_0 = const$, this expression reduces to the free energy profile derived within the context of Classical Nucleation Theory (CNT) for a system at fixed composition:

$$\Delta G_{CNT}(N_c) = -N_c k_B T \ln \left(\frac{\gamma_0 x_0}{\gamma^* x^*} \right) + \sigma' N_c^{\frac{2}{3}} \tag{2}$$

The free energy profiles reported in Eq. 1 and 2, allow to determine the thermodynamic stability of crystal nuclei as a function of their size N_c in a finite-sized or an infinite system, respectively[5].

The most relevant consequence of the confinement effect is that the stability of the solution does not depend only on supersaturation $S = \gamma_0 x_0 / \gamma^* x^*$, but also on the system size[33,48]. Three regimes (labelled A, B, and C) can be identified as a function of the simulation box volume and of the initial composition of the liquid phase for $S > 1$ (see Fig. 1):

> A: the solution is supersaturated, however depletion effects render the formation of a crystal nucleus unfavourable. In this case the free energy profile described by Eq. 1 does not display stationary points, a critical nucleus thus cannot be identified and any state comprising a crystal-like particle is unstable with respect to dissolution.

> B: the absolute minimum is still the homogeneous solution, however the ΔG curve is not monotonous, exhibiting both a local maximum and a minimum. The latter is a state in which a crystallite is in metastable equilibrium with the solution. The maximum instead identifies the size of the critical nucleus that leads to the formation of the metastable state.

> C: here the curve has a shape similar to that of case B but the local minimum corresponding to the small crystallite in equilibrium with the solution becomes the stable state, towards which the homogeneous solution can evolve. Also in this case the critical nucleus size is defined by the location of the maximum in the free energy profile.

Examples of typical free energy profiles in the three cases, together with the domains in the space of the initial composition x_0 and the system volume are reported in Fig. 1. A detailed description and derivation of the nucleation free energy profile in a confined system is reported in Ref.[33].

From the free energy profiles defined in Eq. 2 and Eq. 1 we introduce the correction term:

$$\Delta G_{corr}(N_c) = \Delta G_{CNT} - \Delta G_{\ell \to c}$$
$$= -(N_{tot} - N_c) k_B T \ln \left(\frac{\gamma x}{\gamma_0 x_0} \right) - N_s k_B T \ln \left(\frac{\gamma_s (1 - x)}{\gamma_{s,0} (1 - x_0)} \right) \tag{3}$$
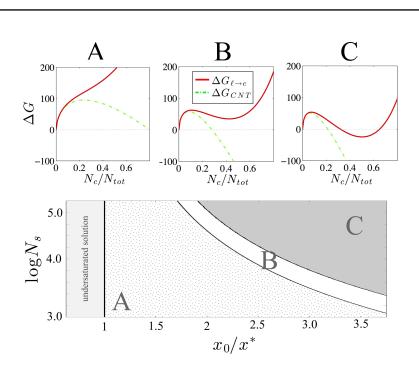
Fig. 1 *Upper row* Representation of characteristic free energy profiles calculated in regimes A, B, and C. *Lower row* existence domains for regimes A, B, and C as a function of supersaturation x_0/x^* and the number of solvent molecules N_s. Once x_0 is fixed the number of solvent molecules N_s determines the system volume, as $V \propto (N_{\text{tot}} + N_s) = (1 - x_0)^{-1} N_s$

Table 1 WT metadynamics simulations of urea nucleation in MeOH, EtOH, and ACN. Simulation conditions and parameters obtained from the fitting of the WT metadynamics FES.

	MeOH	EtOH	ACN
T [K]	300	300	300
x_0	0.0978	0.1348	0.1245
N_s	2767	1926	2110
time [ns]	1×10^3	1×10^3	1×10^3
x^*	0.0284	0.062	0.0019
σ' [kJ/mol]	54.90	29.05	75.05
N^* at $x = x_0$	1665	991	108

That expresses the difference in free energy between a finite size system and an infinitely large one at constant supersaturation[33,46]. In order to extract from finite sized simulations information in the limit of a macroscopic system we shall thus weight each configuration considering the correction term ΔG_{corr}.

The previous discussion has important consequences on simulations, in fact for computationally accessible system sizes we typically fall in case A, where neither stable nor metastable states other than the solution are possible. The states

in which the crystal has nucleated are to be found in the tail of the probability distribution and practically cannot be observed with unbiased simulations. Thus if we want to explore these states we need enhanced sampling. In this work we use WT metadynamics, an enhanced sampling method based on the deposition of a history dependent bias potential that enhances the fluctuations in a coarsened phase space defined by collective variables (CVs). In the following paragraphs details on the CVs definition as well as simulation details are provided. The interested reader can refer to a recent review[49] for a detailed description of the WT metadynamics method details and to Refs.[29,38] for examples of WT metadynamics simulations applied to nucleation processes.

2.2 Collective variables.

A key ingredient of metadynamics is the definition of collective variables (CVs), the coordinates of the coarsened space in which the metadynamics bias potential is adaptively constructed[49]. Collective variables should be able to capture the slow degrees of freedom associated to the rare transformation that is investigated. In this work, WT metadynamics simulations were carried out using a set of collective variables inspired by the study of urea growth from solution[8,44] and described in detail in a recent work on urea crystal nucleation from the melt[29]. The CVs were implemented in PLUMED-1.3[50]. The CVs are formulated with the aim of accounting for both local density and the mutual orientation of urea molecules. The general expression for the CVs, discussed in detail in Ref.[29] follows:

$$S = \frac{1}{N} \sum_{i=1}^{N} \Gamma_i \tag{4}$$

where N is the total number of solute molecules in the system and Γ_i is an estimate of the degree of order for the single solute molecule, defined with respect to a specific orientation axis of the urea molecule. In analogy to Ref.[29,33] we use two CVs: $S_1 = \frac{1}{N_{tot}} \sum_{i=1}^{N} \Gamma_{CO}(i)$ and $S_2 = \frac{1}{N_{tot}} \sum_{i=1}^{N} \Gamma_{NN}(i)$. Where $\Gamma_{CO}(i)$ and $\Gamma_{NN}(i)$ account for the degrees of local order with respect to the CO and NN axes within the urea molecule. $\Gamma_{CO}(i)$ and $\Gamma_{NN}(i)$ are defined as:

$$\Gamma_{CO}(i) = \frac{\rho_i}{n_i} \sum_{j=1}^{N} f_{ij} \left(e^{-\frac{(\vartheta_{ij}-\theta_1)^2}{2\delta^2}} + e^{-\frac{(\vartheta_{ij}-\theta_2)^2}{2\delta^2}} \right) \tag{5}$$

$$\Gamma_{NN}(i) = \frac{\rho_i}{n_i} \sum_{j=1}^{N} f_{ij} \left(e^{-\frac{(\vartheta_{ij}-\theta_1)^2}{2\delta^2}} + e^{-\frac{(\vartheta_{ij}-\theta_2)^2}{2\delta^2}} \right) \tag{6}$$

where:
$-$ n_i is the coordination number of molecule i.
$-$ ρ_i is a switching function defined in the space of the coordination number that decreases from one to zero when the number of neighbours of the ith urea

molecule is less than a threshold value. As in Ref. [29] a threshold of four neighbours within a cut-off distance of 0.6 nm has been used.

− ϑ_{ij} is the angle between the CO or the NN axis of the i and j urea molecules.

− f_{ij} is a function weighing the contribution of the relative orientation of neighbouring urea molecules with their Cartesian distance.

− θ_1, θ_2 represent characteristic orientations for the CO and NN vectors [29]. In Eq. 5 $\theta_1 = 0°$, $\theta_2 = 180°$, while in Eq. 6 $\theta_1 = 90°$, $\theta_2 = 270°$.

− δ is the standard deviation of the Gaussian functions centered in θ_1 and θ_2. In both Eq. 5 and Eq. 6 $\delta = 27.5°$.

A detailed description of the terms ρ_i and f_{ij} can be found in Ref. [29].

2.3 Analysis tools.

The addition of a bias potential with WT metadynamics [12], alters the statistical weight with which configurations are sampled. In order to restore canonical sampling a reweighing procedure needs to be put in place [36,37]. This allows to analyse the nucleation mechanisms by plotting the free energy surface (FES) relative to variables different from S_1 and S_2. In particular we found the introduction the variables (n, n_o) and (N_I, N_{II}) illuminating.

The first pair considers the number of molecules in the largest connected cluster. The difference is that to compute n we link molecules on the basis of their distance and on the density of their local environment, irrespective to their relative orientation, whereas in n_o the relative orientation is also taken into account; thus n_o represents the size of the largest cluster that is crystal-like, and $n > n_o$. In order to calculate n we cluster molecules that have at least four neighbours within a distance of 0.6 nm. Instead the connectivity criterion for n_o has the additional condition that the order parameter Γ_{CO} is greater than 0.5.

The variables N_I and N_{II} give the number of molecules in the largest crystalline cluster, with the structure of *form I*, and *form II*, respectively. To discriminate between *form I* and *form II* molecules, both molecular order parameters Γ_{CO} and Γ_{NN} were considered. Molecules were assigned to a *form I* configuration if $\Gamma_{CO} > 0.5$ and $\Gamma_{NN} > 0.3$, and to a *form II* configuration if $\Gamma_{CO} > 0.5$ and $\Gamma_{NN} \leq 0.3$. The choice of these threshold values is based on the relative position of basins identifying *form I* and *form II* configurations in the FES obtained for urea nucleation from its melt in Fig. 3 of Ref. [29].

Having defined a connectivity criterion, we identify the clusters using a *depth-first* algorithm, similar to the approach used in Ref. [51].

2.4 Simulation Details

Forcefield. In order to meet the requirement of an extensive sampling we have chosen to simulate our systems with a classical forcefield. Similarly to what has been done in previous works [8,29,33,44], both urea and solvent molecules have been explicitly represented using the Generalized Amber Force Field [52–54].

Simulations setup. Three simulations were carried out in order to study urea nucleation in explicit ACN, MeOH, and EtOH. All the systems have been prepared distributing 300 urea molecules in an appropriate number of solvent molecules (see Tab. 1). Urea molecules were dispersed in solvent using the *genbox* utility of

the gromacs simulation package, three dimensional periodic boundary conditions were applied. Production runs were performed at 300 K and 1 bar. The meta-dynamics production runs were carried out for approximately one microsecond each. Each system was at first minimized with the conjugate gradient algorithm with a tolerance on the maximum force of 200 kJ mol^{-1} nm^{-1}. A 20 ns unbiased NPT equilibration of the systems was carried out prior performing the meta-dynamics simulations. All the simulations were carried out in the isothermal-isobaric ensemble at 1 bar using the Bussi-Donadio-Parrinello stochastic thermostat[55] and the isotropic Parrinello-Rahman barostat[56] as implemented in GRO-MACS 4.5.4[57]. The particle-mesh Ewald approach was used to calculate long range electrostatic interactions, with a cutoff of 1 nm[58]. The LINCS algorithm was applied at each step to preserve the bond lengths with 1×10^{-5} tolerance and 4 iterations[59]. The time step used in the production run was 2×10^{-3} ps. In our simulations the bias deposition stride was set to be 1 ps; the height of the Gaussian bias rate was set to ≈ 2 k$_b$T. The width of the Gaussian bias rate was set to $2 \cdot 10^{-2}$. In all simulations the WT algorithm was applied with a bias factor of 200[12,29].

3 Results

3.1 Reversible work of crystal nucleation

The expression of the free energy of nucleation in a closed, confined system formulated in Eq. 1, has been derived under the hypotheses that: *(i)* only one nucleus is formed, *(ii)* surface tension can be defined also for small crystal-like clusters, and *(iii)* the nucleus possesses a well defined crystal structure. As discussed at length in Ref.[33], these hypotheses are considered only to derive an analytical expression for the reversible work of formation of a nucleus, and do not reflect the complexity of the system evolution at the molecular scale.

In this first paragraph we report the free energy profiles calculated from WT metadynamics for the nucleation of a single cluster of urea molecules in a crystal-like configuration. We compare these free energy profiles with those derived from the analytical expression reported in Eq. 1. In Tab. 1, the fitted values of σ' and x^* are reported.

It can be seen in Fig. 2, that the finite size expression of equation 2 fits rather well the data for all three solvents. Together with our previous results[33], this provides further validation of this equation. In the right column of Fig. 2, we report the solvent-specific location in the parameter space of domains A, B, and C described in the previous section. Their relative position and extension are clearly solvent-dependent, reflecting the differences in x^* and σ'.

The fitted values of x^* display a solubility for urea in ACN at least 20 times lower than in alcohols, in agreement with the experimental observations (Tab. 2). The relative solubility of urea in MeOH and EtOH, appears instead to rank in the opposite order with respect to the experimental values. This discrepancy reflects the typical energy accuracy of the molecular modelling approach used in our simulations.

Consider for instance the thermodynamic driving force to crystallization $\Delta\mu_{ref} = k_{\mathrm{B}}T \ln(x_{ref}/x^*)$ calculated with respect to a common reference state at
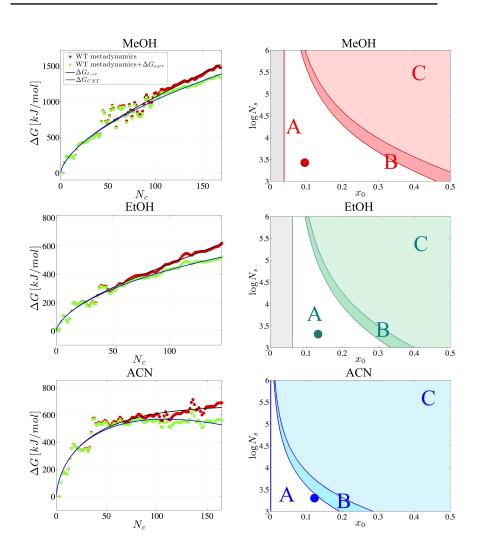
Fig. 2 *Left column* Comparison between the analytical expression of the free energy of nucleation of a crystal-like cluster calculated with parameters reported in Tab. 1, and the free energy profile computed from WT metadynamics simulations for the nucleation of a single urea cluster in solution. The analytical expression describes consistently the free energy profile obtained from simulations and the correction term reported in Eq. 3 allows to obtain the free energy profile in the limit of a macroscopic system at constant composition $x = x_0$. *Right column* Regimes A, B, and C in the domain of x_0 and N_s. The point corresponding to the simulation conditions is highlighted with a circle. Systems prepared with similar initial conditions in the x_0, N_s space may be affected by drastically different finite size effects, depending on the choice of the solvent.

composition x_{ref} (Tab. 2). It can be noted that the discrepancy between the calculated and the experimental $\Delta\mu$ values, as well as the difference in $\Delta\mu$ between MeOH and EtOH is of the order of $k_B T$, the typical level of accuracy expected from classical forcefields. The difference between $\Delta\mu_{ref}$ between alcohols and ACN is instead much larger than $k_B T$, leading to significant differences in sol-

Table 2 Solubility values obtained from simulation compared with their experimental counterpart. To estimate the error in the $\Delta\mu$ a common reference state at $x_{ref} = 0.3$ has been chosen. The value of x_{ref} has been such that the reference state is supersaturated with respect to all solvents, *i.e.* $x_{ref} > \max x^*$. The particular choice of the reference state does not influence the absolute error, reported in the last column.

	x^* calc.	x^* exp.	$\Delta\mu_{ref}$ calc. (kJ/mol)	$\Delta\mu_{ref}$ exp. (kJ/mol)	$\left\|\frac{\Delta\mu_{ref}^{exp} - \Delta\mu_{ref}^{calc}}{k_{\mathrm{B}}T}\right\|$
MeOH	0.039	0.107^a	-5.13	-2.57	1.02
EtOH	0.062	0.045^{60}	-3.94	-4.76	0.33
ACN	0.0019	0.0035^a	-12.65	-11.13	0.61

^{a} Data obtained from the *Open Notebook Science Solubility Challenge as of November 25, 2014*[61].

ubility. It is interesting to notice that the highest discrepancy with respect to experimental data occurs in the case of the most polar solvent, pointing towards the lack of polarizability as a possible source of error.

Inserting these values in Eq. 2 allows reconstructing the theoretical free energy profile for an infinite system at constant composition x_0 (See Tab. 1). Under these conditions we find critical nuclei sizes (N^*) that far exceed the size of the simulation box for MeOH and EtOH (see Tab. 1), thus implying that in the simulations of these two cases we are observing only the first steps of the nucleation process. For ACN instead the critical nucleus for an infinite system at $x = x_0$ is estimated to be around 108 molecules. The differences in nucleus size estimated for ACN, with respect to MeOH and EtOH can be rationalised analysing the dependence of N^* on concentration as predicted by CNT. Under the hypothesis of spherical nucleus the critical size can be written as[5] $N^* = 32\pi v_0^2 \sigma^3 / \Delta\mu^3$, where v_0 is the molecular volume, σ is the surface tension, and $\Delta\mu$ the difference in chemical potential between the solution and the crystal. This means that the critical nucleus size for crystals of the same substance scales as $N^* \propto (\sigma/\Delta\mu)^3$. The term at the denominator can be straightforwardly expressed as: $\Delta\mu = k_{\mathrm{B}}T \ln(x_0/x^*)$, assuming unitary activity coefficients. The estimate of surface tension σ requires instead the formulation of a simplified model. Typically theoretical expressions adopt the general functional form[62–64]: $\sigma \propto \ln(C_s/C_l)$, where C_s is the solute concentration in the solid and C_l its concentration in solution. As C_s is constant by definition and C_l is very similar in the three simulations, σ is within the same order of magnitude for all three solvents. In agreement with the simulation results, the dependence of N^* on the solvent results is dominated by $\Delta\mu$, and therefore by the supersaturation ratio x_0/x^*.

3.2 The nucleation mechanism is solvent-dependent

Single step vs. two step nucleation. As mentioned in the previous paragraph, explicit solvent molecular simulations allow to obtain mechanistic information on the nucleation process, helping to decrypt its inherent complexity. In this paragraph we analyse the WT metadynamics simulations in order to estimate whether urea homogeneous nucleation is a single-step or a multiple step process

in each of the solvents investigated. To this aim we proceed as done in Ref.[33] for the nucleation of urea in aqueous solution, *i.e.* projecting the free energy surface obtained from metadynamics simulations in the space defined by variables n and n_o (see the Methods section). By definition $n \geq n_o$, thus the free energy surface assumes a triangular shape with finite values only above the $n = n_o$ line. In order to account for finite size effects the free energy obtained from WT metadynamics has been corrected with the term reported in Eq. 3, calculated solely as a function of n_o. In other words this means that we are considering only the crystal phase as a separate phase that depletes the mother solution. Thus molecules belonging to locally dense, disordered clusters are still considered as belonging to the liquid phase. This assumption is equivalent to considering disordered clusters as the result of local solute density fluctuations rather than a separate metastable phase. In the (n, n_o) space a perfectly homogeneous solution is at the origin, while in the upper right corner states characterised by a large ordered clusters can be identified, and in the upper left corner large disordered clusters are represented. In this space a typical single step nucleation mechanism corresponds to an ideal path along the diagonal, where every constituent in the forming nucleus is crystal-like. In Fig. 3, the free energy surfaces $F(n_o, n)$ computed from WT metadynamics simulations of urea nucleation in methanol, ethanol, and acetonitrile are reported. For each of these surfaces the most probable pathway connecting the homogeneous solution and the largest crystal-like cluster sampled during the simulation is reported. The pathway is constructed as the locus of the minima in $F(n)|_{n_o}$ the free energy profile in n at fixed n_o and plotted as a function of n_o. Each of the points belonging to such pathway represents the most probable state with respect to n, at n_o fixed.

From a comparison of the most probable pathways in the (n_o, n) space in the case of MeOH and ACN it can be noted that, while in the former the evolution towards larger nucleus sizes proceeds following the diagonal, in the latter the most probable pathway proceeds at first across states characterised by a low degree of order and then evolves towards more ordered states. The mechanism of crystal nucleation in MeOH can thus be described as a single-step process during which the crystal phase emerges from solution without intermediates: in this case in fact small crystal-like clusters are self-assembled from solution. It is important to note that in this case the formation of disordered clusters is still possible, however the nucleation of crystal-like phase within disordered clusters does not appear to be the most favourable pathway towards the formation of a cluster with crystal-like structure. On the contrary, in acetonitrile, the formation of an ordered phase within disordered clusters represents the most favourable pathway. In analogy with the mechanism of nucleation observed in water[33], the most probable nucleation mechanism involves two steps: the formation of locally dense disordered urea clusters and a subsequent nucleation of a crystal phase within the dense clusters. In Fig. 3 these two alternative mechanisms are illustrated with configurations extracted from the simulations in methanol and acetonitrile. It is interesting to notice that the FES obtained in ethanol exhibits a competition between the two mechanism, as while the single step process still seems to dominate, the most favourable pathway coincides less precisely with the diagonal. Moreover a secondary transition pathway associated to the formation of crystalline domains within disordered clusters is also present.
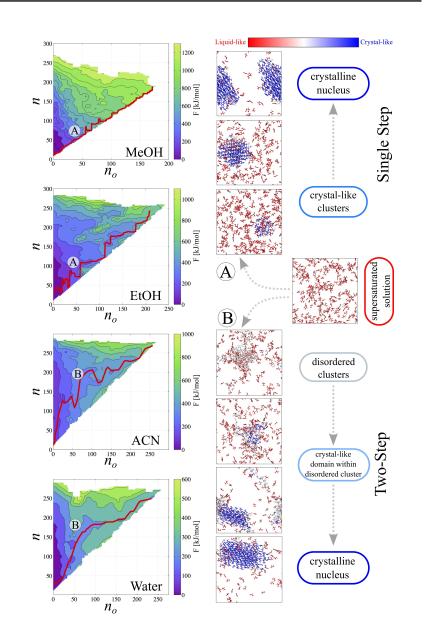
Fig. 3 $F(n_o, n)$ computed from the nucleation simulations in methanol, ethanol, acetonitrile and water (reported from Ref.[33]). It can be observed that the most favourable process follows a single-step mechanism in the case of methanol and ethanol (A), and a two step process in the case of acetonitrile and water (B). Differences in the two mechanisms are shown on the right with configurations extracted from the trajectories obtained in methanol and acetonitrile, where urea molecules were coloured according to their respective degree of crystallinity expressed through their Γ_{CO} value, ranging from red, for liquid-like molecules, to blue, for crystal-like molecules. In the snapshots periodic boundary conditions apply.

Crystal structure of the nuclei. In our analysis of urea nucleation both from aqueous solution and from the melt we have identified two polymorphic structures appearing in the early stages of the nucleation process[29,33]. The structures observed are the result of a replication in the three dimensional space of the intermolecular interaction motifs displayed by *head-to-tail* and *cyclic* dimers[65]. The former gives rise to the experimental crystal structure (*form I*), while the latter produces a distorted structure (*form II*) in which the relative orientation between carbonyl bonds of neighbouring molecules is preserved, while the relative orientation of the axis joining the two nitrogen atoms is changed[29,33]. In the previous paragraph we have focused our analysis on the nucleation pathway in relation to the development of crystalline clusters. In doing so we have considered nuclei belonging to both polymorphs as crystalline structure contributing to the value of n_o.

Now we make a step further and analyse the relative stability of the two polymorphs as a function of the cluster size. As shown in our previous work on urea nucleation from the melt[29], in the limit of an infinite crystal, *form I* is the most stable polymorph in the temperature range extending from 300 K to the melting temperature. This means that the polymorphic transformation between *form I* and *form II* can be considered monotropic. However, for small finite-sized crystal nuclei surface effects play an important role, and the relative stability between *form I* and *form II* clusters becomes a size dependent property.

In Fig. 4 on the left column the FES in the (N_I, N_{II}) space is reported for urea nucleation in methanol, ethanol, acetonitrile, and water[33]. On the right column of Fig. 4 the free energy profiles associated to the conversion of a cluster of constant size from *form I* to *form II* are displayed. It can be observed that the nucleation of small *form II* clusters appears to be favourite with respect to *form I* clusters. The difference in the relative stability of the two forms decreases with increasing the cluster size. This indicates that for larger clusters the surface energy contributions become gradually less important and the relative stability of the two forms tends to the limit of an infinite crystal. This effect of size-dependent stability of different polymorphs has been recently hypothesised also for the nucleation of NaCl in water solution[30] and is compatible with the Ostwald empirical rule of stages[66] whereby the least stable polymorph is the first appearing during the nucleation process.

Let us now analyse the features of the free energy profile associated with the conversion between *form I* and *form II* clusters. In Fig. 4 it can be seen that for large cluster sizes the transition between *form I* and *form II* becomes a barrier-less process. We interpret this feature as a consequence of the mechanism of conversion between the two forms. Such a transition involves in fact a distortion of the crystal lattice, taking place without relevant conformational rearrangements of the solute molecules nor a dissolution step, as shown in Fig.5. The *form I* to *form II* transition is thus not a solvent-mediated process, and can take place due to characteristics of the urea crystal bulk[26,29,33]. In fact such a transition is observed consistently across all solvents considered in this study. The barrier less polymorphic transformation of crystal-like clusters from *form II* to *form I* is thus confirmed to be dominated by the bulk properties of the crystal
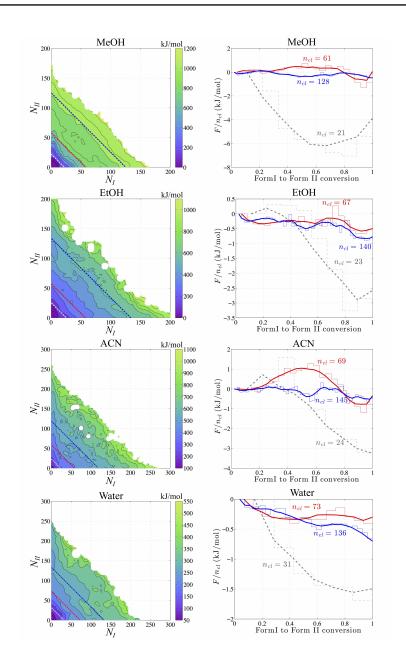
Fig. 4 *Left column* Contour plots of the free energy surfaces obtained for urea nucleation in methanol, ethanol, acetonitrile, and water (reported from Ref. [33]) in the space defined by N_I and N_{II}, the number of molecules in the largest *form I* or *form II* cluster (see the Methods section for further details). *Right column* Free energy profiles associated with the conversion of a crystal-like cluster from *form I* to *form II* for three increasing cluster sizes. In each curve the number of constituents n_{cl} is constant. Curves are extracted as cuts of the FES reported on the right and normalized by the number of constituents.
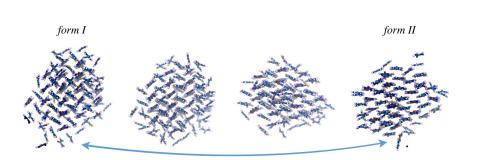
form I *form II*

Fig. 5 *Form I* to *form II* polymorphic transition observed in clusters obtained from simulations of urea nucleation in aqueous solution[33]. Intermediate structures visited during the polymorphic transition are shown in order to highlight that the lattice distortion does not involve conformational changes of urea molecules, but simply a reorientation of urea chains within the crystal lattice. This transition mechanism is observed for all solvents investigated.

4   Conclusions

In this work, taking advantage of the sampling of nucleation events made possible by WT metadynamics, we have investigated urea nucleation from methanol, ethanol and acetonitrile solutions. The main finding of our analysis regards the effect of solvent on the nucleation mechanism. While in acetonitrile a two-step nucleation process emerges as most likely mechanism, in methanol and ethanol a classical, single-step, process dominates. While in acetonitrile the crystal phase is nucleated within amorphous urea clusters, in methanol small crystal-like clusters are formed directly from solution.

On the contrary the competition between urea polymorphs is confirmed to be present with similar features in all three solvents. In all three cases metastable form II clusters are favored over form I clusters for cluster sizes reaching up to approximately 50 molecules. For larger cluster sizes the free energy difference between the two forms tends to progressively level off, with the transition between the two crystal forms becoming essentially barrier-less. These findings indicate that, unlike the nucleation mechanism, heavily affected by the solvent, the polymorph interconversion is dominated by the bulk structure of the urea polymorphs.

To complete our analysis we show that finite size effects associated with solution depletion are not solely a function of the initial composition (x_0) and system size ($\propto N_s$) but also crucially depend on solubility and surface tension and thus on solvent. These effects show lead to significant differences in the shape of the free energy profiles calculated for systems prepared in similar x_0 and N_s conditions but characterized by different solvents.

[a] *ETH Zurich, Institute of Process Engineering, Soneggstrasse 3, CH-8092 Zurich, Switzerland*
[b] *Facoltà di informatica, Istituto di Scienze Computazionali, Università della Svizzera Italiana, CH-6900 Lugano, Switzerland, and Department of Chemistry and Applied Biosciences, ETH Zurich,CH-8092 Zurich, Switzerland*
[*] *E-mail: salvalaglio@ipe.mavt.ethz.ch*

References

1 R. C. Snyder, S. Veesler and M. F. Doherty, *Cryst. Growth Des.*, 2008, **8**, 1100–1101.

2 N. Variankaval, A. S. Cote and M. F. Doherty, *AIChE J.*, 2008, **54**, 1682–1688.

3 I. Weissbuch, L. Addadi, M. Lahav and L. Leiserowitz, *Science*, 1991, **253**, 637–645.

4 I. Weissbuch, R. Popovitz-Biro, M. Lahav, L. Leiserowitz and Rehovot, *Acta Crystallogr. B*, 1995, **51**, 115–148.

5 D. Kashchiev, *Nucleation, Basic Theory with Applications*, Elsevier, 2000.

6 J. P. Sizemore and M. F. Doherty, *Cryst. Growth Des.*, 2009, **9**, 2637–2645.

7 T. Vetter, M. Mazzotti and J. Brozio, *Cryst. Growth Des.*, 2011, **11**, 3813–3821.

8 M. Salvalaglio, T. Vetter, F. Giberti, M. Mazzotti and M. Parrinello, *J. Am. Chem. Soc.*, 2012, **134**, 17221–17233.

9 R. J. Davey, S. L. M. Schroeder and J. H. TerHorst, *Angew. Chem. Int. Ed.*, 2013, **52**, 2166–2179.

10 A. S. Myerson and B. L. Trout, *Science*, 2013, **341**, 855–856.

11 J. Anwar and D. Zahn, *Angew. Chem. Int. Edit.*, 2011, **50**, 1996–2013.

12 A. Barducci, G. Bussi and M. Parrinello, *Phys. Rev. Lett.*, 2008, **100**, 020603.

13 A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 12562–12566.

14 P. G. Bolhuis, C. Dellago and D. Chandler, *Proc. Nat. Acad. Sci.*, 2000, **97**, 5877–5882.

15 A. F. Voter, *J. Chem. Phys.*, 1997, **106**, 4665–4677.

16 H. Grubmüller, *Phys. Rev. E*, 1995, **52**, 2893–2906.

17 T. Huber, A. E. Torda and W. F. van Gunsteren, *J. Comp. Mol. Des.*, 1994, **8**, 695–708.

18 R. J. Allen, C. Valeriani and P. R. ten Wolde, *J Phys-Condens. Mat.*, 2009, **21**, 463102.

19 C. Snow, H. Nguyen, V. Pande and M. Gruebele, *Nature*, 2002, **420**, 102–106.

20 D. Hamelberg, C. A. F. de Oliveira and J. A. McCammon, *J. Chem. Phys.*, 2007, **127**, 155102.

21 A. Kawska, J. Brickmann, R. Kniep, O. Hochrein and D. Zahn, *J. Chem. Phys.*, 2006, **124**, 024513.

22 P. R. ten Wolde and D. Frenkel, *Science*, 1997, **277**, 1975–1978.

23 S. Auer and D. Frenkel, *Nature*, 2001, **409**, 1020–1023.

24 F. Trudu, D. Donadio and M. Parrinello, *Phys. Rev. Lett.*, 2006, **97**, 105701.

25 E. E. Santiso and B. L. Trout, *J. Chem. Phys.*, 2011, **134**, 064109.

26 B. Peters, *J. Chem. Phys.*, 2009, **131**, 244103.

27 G. A. Tribello, F. Bruneval, C. C. Liew and M. Parrinello, *J. Phys. Chem. B*, 2009, **113**, 11680–11687.

28 D. Zahn, *Phys. Rev. Lett.*, 2004, **92**, 040801.

29 F. Giberti, M. Salvalaglio, M. Mazzotti and M. Parrinello, *Chem. Eng. Sci.*, 2015, **121**, 51 – 59.

30 F. Giberti, G. Tribello and M. Parrinello, *J. Chem. Theory Comput.*, 2013, **9**, 2526–2530.

31 B. Knott, V. Molinero, M. Doherty and B. Peters, *J. Am. Chem. Soc.*, 2012, **134**, 19544–19547.

32 M. Salvalaglio, F. Giberti and M. Parrinello, *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.*, 2014, **70**, 132–136.

33 M. Salvalaglio, C. Perego, F. Giberti, M. Mazzotti and M. Parrinello, *Proc. Natl. Acad. Sci. USA*, 2014.

34 P. Ectors, P. Duchstein and D. Zahn, *Cryst. Growth Des.*, 2014.

35 J. F. Dama, M. Parrinello and G. A. Voth, *Phys. Rev. Lett.*, 2014, **112**, 240602.

36 M. Bonomi, A. Barducci and M. Parrinello, *J. Comput. Chem.*, 2009, **30**, 1615.

37 P. Tiwary and M. Parrinello, *J. Phys. Chem. B*, 2014, in press.

38 F. Giberti, M. Salvalaglio and M. Parrinello, *IucJ, in press*, 2014.

39 R. Docherty, K. Roberts, V. Saunders, S. Black and R. Davey, *Faraday Discussions*, 1993, **95**, 11–25.

40 V. Bisker-Leib and M. Doherty, *Cryst. Growth Des.*, 2001, **1**, 455–461.

41 S. Piana, M. Reyhani and J. D. Gale, *Nature*, 2005, **438**, 70–73.

42 S. Piana and J. D. Gale, *J. Am. Chem. Soc.*, 2005, **127**, 1975–1982.

43 X. Y. Liu, E. S. Boek, W. J. Briels and B. P., *Nature*, 1995, **374**, 342–345.

44 M. Salvalaglio, T. Vetter, M. Mazzotti and M. Parrinello, *Angew. Chem. Int. Ed.*, 2013, **52**, 13369–13372.

45 D. Reguera, R. K. Bowles, Y. Djikaev and H. Reiss, *J. Chem. Phys.*, 2003, **118**, 340–353.

46 J. Wedekind, D. Reguera and R. Strey, *J. Chem. Phys.*, 2006, **125**, –.

47 J. W. W.P. Schmelzer and A. S. Abyzov, *J. Non-Cryst. Solids*, 2014, **384**, 2.

48 R. Grossier and S. Veesler, *Cryst. Growth Des.*, 2009, **9**, 1917–1922.

49 A. Barducci, M. Bonomi and M. Parrinello, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2011, **1**, 826–843.

50 M. Bonomi, D. Branduardi and P. M., *Comput. Phys. Commun.*, 2009, **180**, 1961–1972.

51 G. C. Sosso, G. Miceli, S. Caravati, F. Giberti, J. Behler and M. Bernasconi, *J. Phys. Chem. Lett.*, 2013, **4**, 4241–4246.

52 W. Cornell, P. Cieplak, C. Bayly, I. Gould, K. Merz, D. Ferguson, D. Spellmeyer, T. Fox, J. Caldwell and P. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.

53 J. Wang, R. Wolf, J. Caldwell, P. Kollman and D. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

54 D. Van der Spoel, P. van Maaren and C. Caleman, *Bioinformatics*, 2012.

55 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.

56 P. M. and R. A., *J. Appl. Phys.*, 1981, **52**, 7182–7190.

57 B. Hess, C. Kutzner and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.

58 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 1008910092.

59 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comp. Chem.*, 1997, **18**, 14631472.

60 F.-M. Lee and L. E. Lahti, *J. Chem. Eng. Data*, 1972, **17**, 304–306.

61 J.-C. Bradley, C. Neylon, A. Williams, R. Guha, B. Hooker, A. S. Lang, B. Freisen, T. Bohinski, D. Bulger, M. Federici *et al.*, *Open notebook science challenge: Solubilities of organic compounds in organic solvents*, ONS Books, 2009.

62 A. E. Nielsen and O. Söhnel, *J. Cryst. Growth*, 1971, **11**, 233–242.

63 A. Mersmann, *J. Cryst. Growth*, 1990, **102**, 841–847.

64 J. Christoffersen, E. Rostrup and M. R. Christoffersen, *J. Cryst. Growth*, 1991, **113**, 599–605.

65 G. ÖzpInar, W. Peukert and T. Clark, *J. Mol. Model.*, 2010, **16**, 1427–1440.

66 R. A. Van Santen, *J. Phys. Chem.*, 1984, **88**, 5768–5769.