


Molecular simulation approaches to study crystal nucleation from solutions: Theoretical considerations and computational challenges

Aaron R. Finney^{1,2} | Matteo Salvalaglio¹ 

¹Department of Chemical Engineering, University College London, London, UK

²Department of Materials, Design and Manufacturing Engineering, The University of Liverpool, Liverpool, UK

Correspondence

Matteo Salvalaglio, Department of Chemical Engineering, University College London, London, UK.

Email: m.salvalaglio@ucl.ac.uk

Funding information

Engineering and Physical Sciences Research Council, Grant/Award Numbers: EP/R018820/1, EP/X033139/1

Edited by: Modesto Orozco, Associate Editor and Peter Schreiner, Editor-in-Chief

Abstract

Nucleation is the initial step in the formation of crystalline materials from solutions. Various factors, such as environmental conditions, composition, and external fields, can influence its outcomes and rates. Indeed, controlling this rate-determining step toward phase separation is critical, as it can significantly impact the resulting material's structure and properties. Atomistic simulations can be exploited to gain insight into nucleation mechanisms—an aspect difficult to ascertain in experiments—and estimate nucleation rates. However, the microscopic nature of simulations can influence the phase behavior of nucleating solutions when compared to macroscale counterparts. An additional challenge arises from the inadequate timescales accessible to standard molecular simulations to simulate nucleation directly; this is due to the inherent rareness of nucleation events, which may be apparent in silico at even high supersaturations. In recent decades, molecular simulation methods have emerged to circumvent length- and timescale limitations. However, it is not always clear which simulation method is most suitable to study crystal nucleation from solution. This review surveys recent advances in this field, shedding light on typical nucleation mechanisms and the appropriateness of various simulation techniques for their study. Our goal is to provide a deeper understanding of the complexities associated with modeling crystal nucleation from solution and identify areas for further research. This review targets researchers across various scientific domains, including materials science, chemistry, physics and engineering, and aims to foster collaborative efforts to develop new strategies to understand and control nucleation.

Abbreviations: API, active pharmaceutical ingredient; BF, brute force (MD simulation); CNT, classical nucleation theory; CV, collective variable; C_{μ} MD, constant chemical potential molecular dynamics; FFS, forward flux sampling; HEN, heterogeneous nucleation; HON, homogeneous nucleation; MD, molecular dynamics; MFPT, mean first-passage time; MLD, modified liquid droplet; PNC, prenucleation clusters; TPS, transition path sampling; US, umbrella sampling; WTMetaD, well-tempered metadynamics.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *WIREs Computational Molecular Science* published by Wiley Periodicals LLC.

This article is categorized under:

Molecular and Statistical Mechanics > Molecular Dynamics and Monte-Carlo Methods

Molecular and Statistical Mechanics > Free Energy Methods

Theoretical and Physical Chemistry > Statistical Mechanics

KEYWORDS

crystallization, enhanced sampling, molecular dynamics, molecular simulations, nucleation, rare events

1 | INTRODUCTION: WHY SIMULATE CRYSTAL NUCLEATION FROM SOLUTION?

Crystalline materials are solids composed of microscopic repeating units arranged in a regular, periodic array. At the smallest scales, the repeat units are formed from just a handful of atomic/molecular “building blocks” arranged with a well-defined symmetry. Like construction bricks, these building blocks can assemble into a plurality of structures (polymorphs), providing materials with different physical and chemical properties. A key characteristic of molecular and ionic crystalline solids is that their building blocks are held together by non-covalent interactions, resulting in a decoupling of material chemical properties from their physical and mechanical properties. Additionally, the reversible self-assembly of these solids can often be carried out under mild conditions in solution. This flexibility facilitates endless possibilities towards material design with applications in, for example, construction, pharmaceutical manufacturing, separations, catalysis, and organic electronics. For instance, most active pharmaceutical ingredients (APIs) are formulated as crystalline solids.^{1–3}

A significant focus of the computational design of crystalline materials is dedicated to predicting a rigorous thermodynamic stability ranking among all possible bulk phases that may result from crystallization. To this aim, increasingly sophisticated approaches have emerged in recent years, including accurate dispersion models in electronic structure calculations⁴ and the introduction of quantum effects⁵ to calculate lattice energies, the application of machine learning methods to accurately estimate the thermodynamic stability of crystal polymorphs⁶, and the development of novel approaches for the calculation of relative lattice free energies.^{7–12} Establishing the relative stability of crystal polymorphs using computational calculations is extremely valuable in identifying those that are thermodynamically plausible. This has proven particularly useful for the development of APIs, and is finding increasing applications outside academia.^{13–19} These methods, however, fail to capture the role played by components in a preceding crystal-forming fluid phase in directing phase separation toward a specific outcome. In the case of crystallization from solution, the parent liquid often has a composition radically different from the crystal itself, where solvent and other solution additives are often excluded. Even in the simplest example of a single-component molecular crystal emerging from a two-component solution—comprising the solute building blocks and solvent—varying the solute concentration can change the mechanism, rate and polymorphic outcome of precipitation. Computational material discovery/prediction methods based solely on thermodynamic assumptions, therefore, cannot identify *how*, or even *if* a thermodynamically favorable polymorph can be obtained by crystallization from solution.

The key to understanding why thermodynamics alone cannot determine crystallization outcomes is related to the fact that crystallization from an out-of-equilibrium solution is dominated by kinetic factors that are sensitive to changes in the reaction environment.^{20–24} Both of the mechanistic steps necessary for crystallization to occur in metastable solutions—namely, the formation of a crystal embryo (nucleation) and its subsequent growth into a bulk phase—are determined by the dynamics and frequency of transfer of building blocks from the solution to the crystal. As such, the choice of solvent, temperature and solute concentration, in particular, can change the polymorphic outcome of precipitation.^{25–27} Because the dynamics of the building blocks assembly discussed above are so important, molecular dynamics (MD) is typically the simulation tool of choice to investigate the crystallization of molecular and ionic solids *in silico*. Through the lens of statistical mechanics, MD simulations unlock both thermodynamic and kinetic information by tracing the motion of

up to 10^9 atoms over simulation times ranging from 10^{-9} to 10^{-6} s. Nevertheless, understanding and predicting how crystals with well-defined composition, structure and properties assemble from a supersaturated solution remains a formidable task using molecular simulations. The small time scales probed by MD simulations are generally inadequate to study the complex microscopic steps involved in crystal nucleation and growth. As we highlight below, however, advanced sampling schemes and theoretical treatments are often invoked to circumvent these limitations. In doing so, MD simulations can help to determine crystal nucleation and growth rates, polymorphism, crystal habit and defect density as a function of solution composition, temperature and the presence of additive/impurities and external forces. Indeed, a wealth of information can be gained by combining complementary computational tools to probe different crystallization steps from solution, but it is not always apparent which tools are most appropriate for the specific task at hand.

In this review, we provide an extended summary of the key theoretical and methodological features associated with atomic-scale modeling and simulation of crystallization triggered by homogeneous nucleation from *solution*. In particular, we focus on the key differences and complementarities between simulation methods implying that nucleation follows a classical mechanism and methods that enable us to ask questions about the mechanism itself. In this process, we contrast simulation methods based on theoretical prior knowledge of nucleation mechanisms (e.g., seeding) with rare events simulation methods based on the introduction of a bias potential (e.g., metadynamics) and methods based on sampling nucleation trajectories in path space (e.g., forward flux sampling). This review focuses on methods to simulate crystal nucleation, and we also note that significant progress was made in recent years to simulate crystal growth from solution by combining MD simulations, Kinetic Monte Carlo, and enhanced sampling.^{28–31}

Our review complements an overview of the literature provided by Agarwal et al.,³² which delves into the theoretical background of rate theories applied to nucleation problems, as well as the review by Sosso et al.³³ that provides a wider appraisal of simulation methods to investigate crystal nucleation from liquids, without a specific focus on crystal precipitation from solution. We believe solutions deserve particular attention because, even if crystallization from multicomponent liquids is extremely common, it presents peculiar challenges associated with the fact that the product phase is usually characterized by a different composition than the parent phase and that the driving force for the process depends on solute concentration. After defining the remit of applicability of atomistic simulations to nucleation problems and highlighting the theoretical basis associated with the main features of nucleation processes investigated by molecular simulations, we briefly review the literature describing the methods to model nucleation from solution with atomistic detail. Finally, we provide a critical comparison of insights obtained from works that investigate the nucleation of NaCl(s) from aqueous solutions: a problem that has been tackled using approaches covering the entire spectrum of methods reviewed in this work.

2 | CLASSIFYING NUCLEATION FROM SOLUTION

Before delving into an overview of simulation methods to study crystal nucleation from multicomponent liquids, it is important to discuss the additional adjectives that qualify the nucleation process in the scientific literature and their meaning. This is essential to define the applicability domain of molecular simulation methods and to formulate appropriate research questions that could emerge when different nucleation processes and mechanistic hypotheses are being investigated. Figure 1 summarizes the different classes of nucleation discussed below and provides a sequence of considerations in order to determine an appropriate classification.

2.1 | Primary versus secondary nucleation

Primary nucleation is the spontaneous formation of new crystalline particles from a metastable solution phase without any interplay with pre-existing crystalline particles. This happens due to rare fluctuations in local solute order. Primary nucleation rates are usually low and can be influenced by factors such as solute concentration and temperature. In contrast, *secondary* nucleation occurs when pre-existing crystals or crystal surfaces (of the same nucleating substance) promote the formation of new crystals by attracting and attaching crystal growth units. Secondary nucleation is often much faster than primary nucleation and can be influenced by external factors such as agitation and shear forces that lead to particle attrition.

2.2 | Homogeneous versus heterogeneous nucleation

Homogeneous nucleation (HON) occurs in the bulk of a supersaturated solution and can be controlled by changing the solution environment. As well as temperature and pressure control, solution additives may affect the nucleation behavior. While the rates for HON can be predicted using a suitable theory/model, the location where nucleation occurs cannot be determined a priori. HON requires specific conditions for it to occur and is much less prevalent in nature than *heterogeneous* nucleation (HEN). In HEN, crystal nucleation is facilitated by interfaces—usually a solid submerged in the solution phase. Surfaces can act as nucleants to direct the site-specific crystallization of particular crystal polymorphs or, more generally, enhance the rates for crystallization. The direct simulation of heterogeneous nucleation is a daunting task, as, in general, information on the local structure of the surfaces promoting nucleation is not available. Hence, while the methods discussed in this review are, in principle, applicable to the study of heterogeneous nucleation, as shown by the recent works on heterogeneous ice nucleation,^{35–39} examples of their application to HEN from solution are currently lacking. As such, in Figure 1, we report HEN as being *on the edge* of the applicability domain for molecular simulations of nucleation from solution.

2.3 | Single-step versus multi-step nucleation

Single-step nucleation refers to the direct formation of crystals from their primary building blocks in solution via a single energy barrier in the free energy landscape, giving rise to an induction time associated with a single bottleneck toward crystallization. In contrast, multi-step nucleation encompasses a variety of crystallization pathways, which can involve the formation of intermediate precipitate phases, such as liquids, amorphous solids or other crystalline phases. The pathway to the most stable crystal form from solutions may involve multiple intermediates occurring in sequence reproducibly. If the relative stability of the intermediates increases with each step, this type of mechanism is consistent with the Ostwald–Lussac empirical rule of stages.^{40,41}

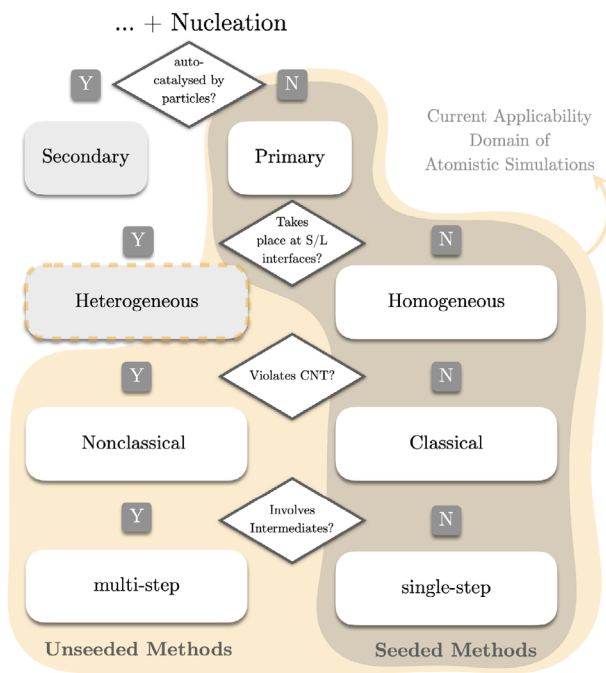


FIGURE 1 To gain insight into nucleation mechanisms and evaluate rates, an appropriate simulation method must be chosen to investigate the system at hand. This flow diagram identifies where theory-based simulation methods (based on CNT and its extensions) are most suitable. In particular, we highlight that secondary nucleation typically involves interactions of crystalline particles with the fluid at a scale inaccessible to molecular simulations and, with some notable exceptions,³⁴ it is typically outside the scope of atomistic simulation studies. Heterogeneous nucleation is only partly included in the applicability domain as, while most methods discussed in this review can in principle be employed to study nucleation at solid/liquid (S/L) interfaces, appropriate examples exist only for crystal nucleation from the melt.³³

We specify that intermediates must be demonstrably involved in the formation of crystals to classify crystallization as multi-step. The mere presence of intermediate structures before or even during crystallization does not exclude the possibility that single-step crystal nucleation occurs (e.g., by a dissolution/re-precipitation reaction).

An archetypal example of multi-step nucleation is two-step nucleation, widely described as a process where crystalline order emerges in dense liquid precursors that form during the first phase transformation in solution.^{42,43} Figure 2b shows how a two-step pathway deviates from a single-step (one-step) pathway on a two-dimensional reaction coordinate defining cluster size and order. It is assumed that the rate determining step along a two-step pathway is the emergence of order within dense liquid domains (see the snapshot in Figure 2b), which form readily when the supersaturation of the parent phase is sufficiently high. In this case, there are often two bottlenecks to crystallization^{44–46}; however, it may be that a single energy barrier is involved in the formation of crystals from solution and that under certain conditions, traversing the lowest energy pathway leads to amorphous solute cluster intermediates.^{47–49} In other words, multi-step nucleation does not necessarily refer to a cascade of single-step nucleation events.

The process of nucleation in complex solutions may be dependent upon chemical reactions and changes in species stoichiometry occurring locally. Several studies^{50–52} have attributed the multi-step nature of nucleation to these factors. Simulating reactive crystallization events is often beyond the scope of classical molecular simulations using semi-empirical, non-reactive force fields. As such, this review focuses on simulating nucleation from building blocks which are already present in the parent solution phase. These examples, however, highlight the importance of simulation practitioner's understanding of the chemical speciation and valency of molecular and ionic species involved in crystal formation.

2.4 | Classical versus nonclassical nucleation

In Section 3.1, we briefly describe classical nucleation theory (CNT), the thermodynamic and kinetic framework for *classical* nucleation. For the purposes of classifying nucleation pathways, we stress that classical nucleation is a

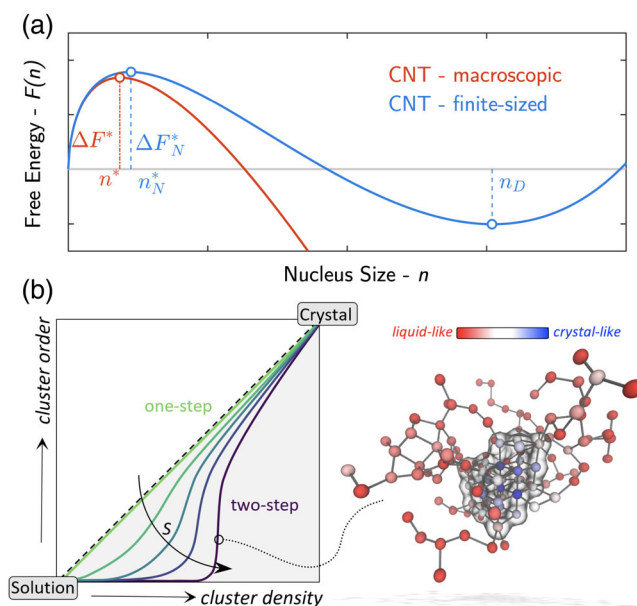


FIGURE 2 (a) The finite size effects of closed simulations manifest in a free energy minimum for the stable thermodynamic phase resulting from nucleation. This is different to the macroscopic case due to a bounded partition function and a depletion of crystal building blocks that transfer from the parent phase to the nucleating phase. (b) Crystal nucleation pathways from solution to crystal represented on a two-dimensional reaction coordinate characterizing cluster density and order. The diagonal marks the case for the concomitant increase in solute cluster order with density, indicative of the capillary approximation adopted in CNT. Pathways that deviate from this limiting case include two-step nucleation, where, for example, crystalline order is established in the core of liquid-like precursors, as shown on the right of B. In the case of NaCl nucleation (discussed below), simulations demonstrate a transition from one-step to two-step crystal nucleation that occurs when the supersaturation ratio, $S = \exp(\Delta\mu_{\ell \rightarrow \text{xtal}}/k_B T)$, is increased far into the metastable zone for phase separation.

single-step process that takes place through the attachment of monomers from solution to a cluster with an internal structure matching the bulk crystal. Within the capillary approximation, the surface tension of the clusters of a new phase is independent of cluster size and also matches the bulk; we should expect a sharp solid–solution interface and the same crystal faceting observed at equilibrium and ignore any curvature effects on the surface tension for small clusters.⁵³ In terms of kinetics, we assume an abundance of monomers in the solution phase and that the out-of-equilibrium growth of nuclei occurs at a steady state.

It is highly unlikely that any single-step, homogeneous crystal nucleation occurs according to the above mechanism. The capillary approximation is particularly problematic for crystal nucleation because the smallest clusters are unlikely to display an interface structure with surface tension that matches highly faceted bulk crystals with infinitely large planar surfaces. Furthermore, it may not be the case that the density and structure in the emerging crystal are homogeneous throughout. The CNT framework also fails to account for the role that growth units beyond monomers might play in the formation of crystal nuclei.

Some of the effects described above can be accounted for in frameworks which we label as *extended classical* nucleation. For example, models that depart from CNT include a surface tension term that is dependent on the cluster size.^{54–56} Other theoretical and simulation studies have identified that the structure, particularly the density, of the smallest nuclei differs from the bulk and stated that this should be accounted for in extensions of CNT.⁵⁷ Formulations based on classical density functional theory provide corrections to the CNT nucleation rate derived from the excess system free energy that accounts for varying cluster density.⁵⁸

Given the description of two-step nucleation above, it may be perceived that this type of phase separation mechanism is inconsistent with classical nucleation. However, theoretical studies have demonstrated that two-step nucleation can be described using classical concepts adopting a composite cluster model.^{48,59} Changes to the relative supersaturation of the system with respect to a dense liquid and crystal phase lead to changes in the pathways to crystals from solutions⁴⁸: a common observation in experimental studies. Some nucleation theories predict mechanisms that are clearly different from those described by CNT and its extensions and are thus termed *nonclassical*. Because this classification encompasses a large family of different nucleation frameworks, it is usually not a useful or informative description. A comprehensive review of nonclassical nucleation is beyond the scope of this review, and we refer the reader to perspectives and reviews on the topic.^{60–64}

A paradigmatic example of nonclassical nucleation worth mentioning is the *prenucleation cluster pathway*. The prenucleation cluster (PNC) pathway was first proposed for CaCO_3 ⁶⁵ but has since been attributed to phase separation in a diverse range of systems.⁶⁶ The PNC pathway suggests that the parent solution phase is comprised of hydrated solute clusters in pseudo-equilibrium with solvated monomers. The population of PNC sizes is determined by the equilibrium constant (K) for the reaction $(\text{monomer})_x \xrightleftharpoons[\text{-monomer}]{\text{+monomer}} (\text{monomer})_{x+1}$, which is assumed to be constant and independent of the value of x . PNCs are highly dynamic, evolving their structure and topology over very short time-scales (typically picoseconds).⁶⁷ The nucleation step in the PNC pathway involves changes to the order and dynamics of monomers within PNC assemblies that renders them a new thermodynamic phase and gives rise to an interface with the solution.⁶⁸ This rate-determining step typically produces a dense liquid phase that contains a large amount of solvent. Subsequent transformations of this liquid are necessary to produce crystals, which *may* include the further dehydration of dense liquids to form amorphous solid-like phases.

3 | DYNAMIC SIMULATIONS OF CRYSTALLIZATION FROM SOLUTION: TIME- AND LENGTH-SCALE CHALLENGES

Obtaining information on molecular-scale crystallization events, particularly nucleation, requires a high degree of time and space resolution. This poses significant challenges for in situ experiments, often leading to only speculative interpretations of molecular mechanisms. In contrast, modeling techniques based on MD simulations inherently provide insight into the time evolution of systems with atomistic detail,⁶⁹ making them a powerful tool for understanding complex processes such as nucleation. However, the small system sizes and limited simulation times available to the simulator mean that advanced simulation methods must often be used to simulate nucleation.

In MD, Newton's equations of motion are adopted to evolve atom positions in time.⁷⁰ Here, chemical bonds are typically approximated as classical “springs” whose displacement from an equilibrium bond distance is modeled using Hooke's law. Similar simple functions can be used to approximate bond and dihedral angle rotation to capture the forces associated with intramolecular degrees of freedom. Intermolecular interactions can be approximated using

simple, classical interpretations of Van der Waals forces and Coulombic forces for point-charged atoms. Importantly, all molecular species are assumed to be in their electronic ground state. As such, MD adopting classical, semi-empirical force fields are not well-suited to simulate chemical reactions and dynamical changes to chemical speciation, though reactive force fields⁷¹ and the advent of machine learning methods to force field design^{72,73} make this a possibility. Particularly in the case of ionic systems, polarizability can be simulated using a relatively cheap treatment of charge displacement from atomic nuclei, for example, using springs. By assigning a velocity (sampled from a Maxwell–Boltzmann distribution at a defined temperature) to atoms at the beginning of a simulation, the Hamiltonian dynamics of the system can be propagated according to a fixed temperature and the force field.^{70,74} This time integration is performed iteratively using a small time step, typically on the order of 1 fs, to capture the fastest atomic displacements in the system, usually molecular bond vibration.

The choice of force field can have important consequences for simulation observations. In terms of simulating crystallization, the force field should reproduce the structure, density and stability of the crystal phase as a minimum requirement. Furthermore, the solubility of the crystal phase should be reasonably close to that determined experimentally if a comparison to experiments is intended. The properties of the solution should also be modeled accurately. For example, the free energy of solvation of solutes, their activity in solution as a function of concentration and their mobilities in the solvent are all important properties necessary to accurately capture the thermodynamics and kinetics of crystal nucleation.

Assuming a suitable force field is available (see Section 6), simulating crystallization from multicomponent liquids using MD typically requires addressing two main system-specific limitations: *timescale* and *finite size* effects. Despite the high space and time resolution capability of MD, which makes it suitable for understanding molecular-scale processes, crystal nucleation in microscopic MD volumes can be very slow to realize. Therefore, many novel simulation methods have been developed to efficiently overcome one or both of these limitations to modeling crystallization.

3.1 | Timescale limitations

Simulating crystallization using MD, especially crystal nucleation, presents a significant challenge due to the infrequency of many of the atomic and molecular scale elementary steps required for these processes to occur. For instance, depending upon the thermodynamic state of the parent phase, crystal nucleation can occur on time scales typically orders of magnitude larger than those accessible to brute-force simulations.^{75,76} The separation of these timescales has made stimulating crystallization an ideal playground for the development of enhanced sampling methods based on MD. Thanks to enhanced sampling, significant early progress was made to analyze the early stages of crystallization in simple systems of uniform particles^{47,77–80} and toward the in-depth investigation of nucleation in monocomponent molecular systems, such as pure water.⁸¹ In recent years, an evolution toward systems with increased complexity and practical relevance has begun, enabling for example, the study of solute precipitate nucleation. These tend to focus on model systems,^{32,82,83} such as two-component Potts-lattice models, or inorganic systems such as NaCl(aq).^{33,84–89} Attempts to simulate organic solids nucleating from solution have been successful in highlighting the structural features of early-stage crystallization precursors^{90,91} and to extract qualitative information on nucleation mechanisms.^{92,93}

3.1.1 | Origin of the timescale separation in nucleation: Nucleation free energy barriers and CNT

CNT provides a quasi-mechanistic description of the nucleation process, connecting crystallization equilibrium thermodynamics and nucleation kinetics.⁵³ CNT foundations were laid down by Gibbs 150 years ago to describe the formation of liquid droplets from saturated vapors, and the theory was developed over the following century, now often being invoked to explain the emergence of ordered materials from liquids. In CNT, there are two contributions to the free energy of a metastable system: a volume and a surface free energy both associated with the new, emerging phase embedded in an out-of-equilibrium parent phase. Forming a stable phase is thermodynamically favorable, but the resulting interface, which delimits the new phase from the parent one, carries an energy penalty. While the free energy gain associated with the formation of a stable phase scales linearly with the volume of a nucleus, the free energy cost of forming an interface scales with its surface area. This simple but general argument provides a rationale for the barrier to nucleation and a definition for the *transition state* associated with the nucleation process.^{27,94}

The size of the nucleus was originally described using a linear descriptor, that is, its radius r . However, the number of constituent monomers belonging to the nucleus, n , provides a variable which makes fewer assumptions about the nucleus shape. In this context, the volume and surface free energy contributions lead to the following expression for the nucleation free energy:

$$F(n) = -n\Delta\mu_{\ell \rightarrow \text{xtal}} + \sigma' n^{2/3}, \quad (1)$$

where $\Delta\mu_{\ell \rightarrow \text{xtal}} = \mu_{\ell} - \mu_{\text{xtal}}$ is the difference in chemical potential between the metastable liquid and the stable crystal phase and is strictly positive in conditions where nucleation is thermodynamically favorable and the system is *supersaturated*. The surface term scales as $n^{2/3}$, where σ' is the product of a shape factor, area per monomer adsorption site and the surface tension, σ . A typical free energy profile for nucleation in the CNT framework is shown by the red curve in Figure 2a.

By solving the derivative of the equation above with respect to n , we find that the free energy barrier associated with the nucleation process ΔF^* is

$$F(n^*) = \frac{4\sigma'^3}{27\Delta\mu_{\ell \rightarrow \text{xtal}}^2} = \frac{1}{2}\Delta\mu_{\ell \rightarrow \text{xtal}}n^*, \quad (2)$$

where the *critical* size of the crystal nucleus is

$$n^* = \left(\frac{2\sigma'}{3\Delta\mu_{\ell \rightarrow \text{xtal}}} \right)^3 \quad (3)$$

and CNT provides analytical solutions for the minimum work required to form a crystal nucleus. The addition of monomers to the nucleus beyond n^* reduces the free energy of the system.

CNT assumes that a single energy barrier separates the parent liquid from a bulk crystal in equilibrium with a solution. It also assumes that the surface tension and the shape factor comprising σ' are unchanging as a function of n . This is the so-called capillary approximation that likely fails for the smallest crystals whose shape and faceting are expected to deviate from that of the bulk crystal.⁴¹ Extensions of the theory can be made to account for varying surface tension.^{54–56}

The crystal embryo growth is assumed to occur by the transfer of *monomers* from the solution to the crystal.⁹⁵ This is justified by considering the probability of finding clusters of size n : $p(n) \propto \exp(-\Delta F(n)/k_B T)$, where k_B is Boltzmann's constant and T is temperature. If $\Delta F(n^*) \gg k_B T$ and n^* is relatively small; it is reasonable to expect the transfer of growth units from the solution to the crystal to be dominated by solute monomers. The rate, J , of formation of clusters of size n per unit volume of the solution and per unit time takes an Arrhenius-type form^{96,97}:

$$J = \rho f^+ Z \exp\left(\frac{-\Delta F^*}{k_B T}\right); \quad Z = \frac{1}{n^*} \left(\frac{\Delta F^*}{3\pi k_B T}\right)^{\frac{1}{2}}, \quad (4)$$

where ρ is the volume density of solute monomers and f^+ is the rate of attachment of monomers to the nucleus. Z is the Zeldovich factor for homogeneous nucleation, which accounts for the low probability of observing a large population of supercritical clusters that progress to bulk crystals (and is determined by the shape of the free energy barrier in n).^{53,98} Due to the exponential term in Equation (4), small changes to $\Delta F(n^*)$, as determined by $\Delta\mu_{\ell \rightarrow \text{xtal}}$ and σ' , can result in orders of magnitude changes to crystal nucleation rates.

Following units for the rate as $\text{m}^{-3} \text{s}^{-1}$, the characteristic time for a single nucleation event scales according to V^{-1} , the reciprocal volume of the solution. As described above, computational costs limit the total volume (number of atoms) that can be simulated using standard MD. With this in mind, consider a hypothetical aqueous solution system undergoing relatively fast nucleation with $J = 10^{20} \text{ m}^{-3} \text{ s}^{-1}$. In an MD simulation of this solution containing around 10^5 water molecules, we can expect the mean simulation time required to observe one nucleation event to be roughly 1 h. This is

far beyond the capabilities of MD, which typically achieves simulation times up to 10^{-6} s on a powerful CPU/GPU. From this example, it is easy to appreciate why enhanced sampling simulations based on MD are often essential to study phase separation in some solutions.

3.2 | Finite-size dependence of the crystallization driving force

Simulations of crystallization from solution are typically carried out in the canonical or isothermal-isobaric ensembles, where the total number of atoms/molecules is constant. When dealing with out-of-equilibrium, multicomponent liquid phases undergoing a phase transition, this constraint introduces a coupling between the number of solute monomers available to a growing crystal nucleus and the time-dependent crystallization driving force, $\Delta\mu_{\ell \rightarrow \text{xtal}}$.^{32,75,92,99,100} Indeed, nucleation can be completely inhibited in a microscopic, closed system, if the transfer of monomers from the liquid to a critical cluster of the new phase renders the solution undersaturated.^{75,92,101} In a dense system, such as a liquid solution, this coupling cannot be efficiently removed by simulating in the grand canonical ensemble (where the solute is replenished from an artificial, external reservoir to maintain a constant solution chemical potential), due to a low probability for insertion of solute in the liquid phase. Instead, molecular simulations using a constant number of molecules require either the application of theoretical corrections to account for the change in $\Delta\mu_{\ell \rightarrow \text{xtal}}$ or the development of specialized methods to mimic open boundary conditions,^{102–106} unless sufficiently large systems can be simulated to minimize the finite-size effects.

3.2.1 | Nucleation free energy in small systems

A rationale for the effect of system size on the phase behavior of metastable liquids is gained by developing a model for crystal nucleation in confined volumes analogous to the modified liquid droplet (MLD) model developed by Reguera et al.^{99,107} to describe depletion effects on the thermodynamics of nucleation of liquid droplets.^{86,92} Consider a two-component supersaturated solution, ℓ . The chemical potential of solute i in solution can be written as,

$$\mu_{\ell} = \mu_0 + k_{\text{B}}T \ln a^i \approx \mu_0 + k_{\text{B}}T \ln x^i, \quad (5)$$

where μ_0 is a reference chemical potential and a^i is the activity of i , which is approximately equal to the mole fraction, x^i , assuming close-to ideal solution behavior. When a crystal forms, the chemical potential of this phase equals the chemical potential of the solution at equilibrium, such that,

$$\mu_{\text{xtal}} = \mu_0 + k_{\text{B}}T \ln a^{i,*} \approx \mu_0 + k_{\text{B}}T \ln x^{i,*} \quad (6)$$

and asterisks indicate the activity and mole fraction of the solute under coexistence conditions.

The transfer of monomers from the solution to the crystal during nucleation depletes the surrounding solution. However, this effect is negligible in a macroscopic system, as the abundance of monomers in the bulk liquid quickly replenishes the solution surrounding the nucleus (assuming that crystallization is not diffusion limited). In a simulation, however, where the total number of monomers is fixed (N), the depletion changes the driving force for crystallization and must be accounted for in an additional term to Equation (1):

$$\Delta F(n) = N(\mu_{\ell}(n) - \mu_{\ell}(n=0)) - n(\mu_{\ell}(n) - \mu_{\text{xtal}}) + \sigma' n^{2/3}. \quad (7)$$

Depending on the volume of the simulated system, and the supersaturation of the mother phase, the deviation between a macroscopic and a finite-size nucleation free energy profile can be responsible for significant changes to the outcome of nucleation simulations. In supersaturated solutions, confinement leads to a minimum in the free energy profile shown in Figure 2a, the depth of which is determined by the initial concentration and number of monomers in solution. Furthermore, the height and position of the barrier is perturbed compared to the macroscopic case. If association of monomers leads to the solution becoming undersaturated below the critical size of the nucleus in the macroscopic limit, then nucleation can be inhibited altogether.

One could take advantage of the limitations imposed by finite size by performing simulations initiated at different N and V and analyzing the steady-state properties of the system obtained from simulations to evaluate $\Delta F(n)$. For example, Li et al.¹⁰⁸ performed a series of unbiased MD simulations where they varied the concentration and total number of coarse-grained, intrinsically disordered peptides in a continuum solvent. Starting from a homogeneous phase, spontaneous separation occurred when simulations were sufficiently large and concentrated, resulting in dense liquid peptide droplets in equilibrium with lean solutions. Fitting the steady-state simulation data to a model analogous to Equation (7) allowed the authors to evaluate x^{i*} , σ and n^* over a range of saturation levels and identify the conditions where nucleation is completely inhibited by the system size.^{92,93}

Practically, computational challenges mean this approach has yet to find applications to study crystal nucleation. This is because the spontaneous decomposition of a highly supersaturated solution is unlikely to lead to crystals due to the slow decay times for monomer relaxation to a lattice structure in a dense amorphous precipitate. Even if sampling this monomer ordering were feasible over MD timescales, many more crystal geometries could result from a rapid crystallization step, making precise estimates of σ difficult. Based on the local density, alternative formulations of CNT have also been constructed,¹⁰⁹ and the finite size effects can be included here to account for the changing thermodynamic driving force.¹⁰¹

3.2.2 | Simulating condensed matter systems with pseudo-open boundaries

A limited number of strategies have been developed to avoid the effects of solute depletion that shift the thermodynamic driving force for steady-state crystallization in simulations. The constant chemical potential MD method ($C\mu$ MD)^{102,104,110} employs a closed system that is segregated into a transition region (TR), housing the process of interest; a control region (CR), representing a bulk fluid with fixed composition; and an internal reservoir that supplies the CR with solute monomers as they are removed from the liquid phase toward the growing crystal in the TR. In doing so, steady-state crystal growth is maintained for relatively long simulation times, as was shown in the case of, for example, urea,¹⁰² isoniazid¹¹¹ and naphthalene,¹¹² which demonstrate crystal facet and solvent dependent growth rates. In addition, crystal nucleation can be studied using a spherical variant of $C\mu$ MD that was applied to simulate NaCl crystallization.¹⁰⁴

Combining multiple, carefully prepared closed system simulations (representing steps on a nucleation reaction coordinate) to maintain the solution concentration as solute molecules are transferred to a crystal nucleus, out-of-equilibrium sampling in the so-called Osmotic ensemble can be performed to avoid solute depletion effects.¹⁰⁵ This was successfully applied to understand the nucleation and polymorph selection of sulfamerazine.¹⁰⁶ Unlike $C\mu$ MD, the reaction coordinate for crystallization must be known a priori here to set up the initial configurations, and dynamical information cannot be obtained using this approach.

Another promising simulation strategy to avoid solute depletion is the adaptive resolution scheme.¹¹³ Here, a closed system contains a high-resolution region of interest coupled to a low-resolution, coarse-grained representation of solute/solvent. In the case of nucleation, for example, the crystal nucleus and its immediate environment can be modeled with atomic detail. As the nucleus grows, a smooth transformation of the model for solute and solvent molecules occurs by extending the region of the system modeled with high resolution. There are no known examples of applying such approaches to study crystallization. Both the coarse-grained model and the method by which the molecule representations are transformed must be prepared/performed carefully to ensure the properties of the system are maintained.

4 | MODELING APPROACHES TO SIMULATE NUCLEATION FROM SOLUTION

As highlighted in Sections 2 and 3, modeling nucleation from solution introduces unique theoretical and practical challenges. As such, a range of different approaches have been applied to this problem, populating a variety of methods that span epistemic interpretations of the role of simulations. On the one hand, there is the application of molecular simulation methods to estimate physical parameters appearing within independently established theoretical frameworks. On the other hand, molecular simulations can be utilized as computational experiments, yielding a direct observation of the fundamental steps underpinning crystal nucleation from solution. These types of simulations contribute to the assessment of existing theoretical frameworks and, if necessary, to the development of new ones. The space between these two simulation extremes is populated by a plethora of different approaches that vary not just in their technical

implementation but also in the degree to which they rely on reference theoretical frameworks to yield estimates of nucleation kinetics and mechanisms. In Figure 3, we provide a graphical representation of a variety of such methods in relation to their reliance on CNT and its underpinning assumptions. At one extreme, if computational resources were abundant, then one could perform many, large-scale MD simulations for sufficiently long enough times to observe crystal nucleation and growth, and, following validation of an appropriate reaction coordinate, extract kinetics for nucleation using Poisson statistics (see right of Figure 3). This approach requires no a priori assessment of the mechanism and rate equation for nucleation. Unfortunately, the computational cost here means that it is unlikely to succeed for most cases. At the other extreme, if one assumes a mechanism following CNT, equilibrium MD simulations or experimental data can be leveraged to plug into the theory master equation (see left of Figure 3). This approach can be very cheap; however, estimates of crystallization rates here are likely to be erroneous (given the shortcomings of CNT and observations from experiments discussed in Section 2) and no insight into the mechanism is gained, as well as this approach being largely unsatisfying for the simulator. The simulation methods we discuss below fall between these extremes and were either developed to simulate nucleation directly or to simulate rare events more generally.

A somewhat arbitrary classification of nucleation from solution guides the structure of the core sections in our review. In the following, we summarize the basics of different simulation approaches by grouping them into two main categories: theory-based (Section 4.1.1) and exploration (Section 4.2) simulation methods. The latter rely on the vast literature covering rare event sampling methods. We report on both biased and unbiased simulation methods, prioritizing those that have been applied to study crystal nucleation from multicomponent solutions. Finally, we propose a global overview of the type of insight available from nucleation simulations by reporting results that have accumulated in the last decade on the homogeneous nucleation of NaCl in an aqueous solution. For this system, we have examples of many, if not all, possible implementations of nucleation simulation strategies. This *unicum* in the literature provides an opportunity to compare different approaches and, at the same time, offers an overview of the field. Numerous studies of NaCl crystal nucleation using different simulation methods facilitate an assessment of the suitability of these approaches to investigate crystal nucleation in silico more generally.

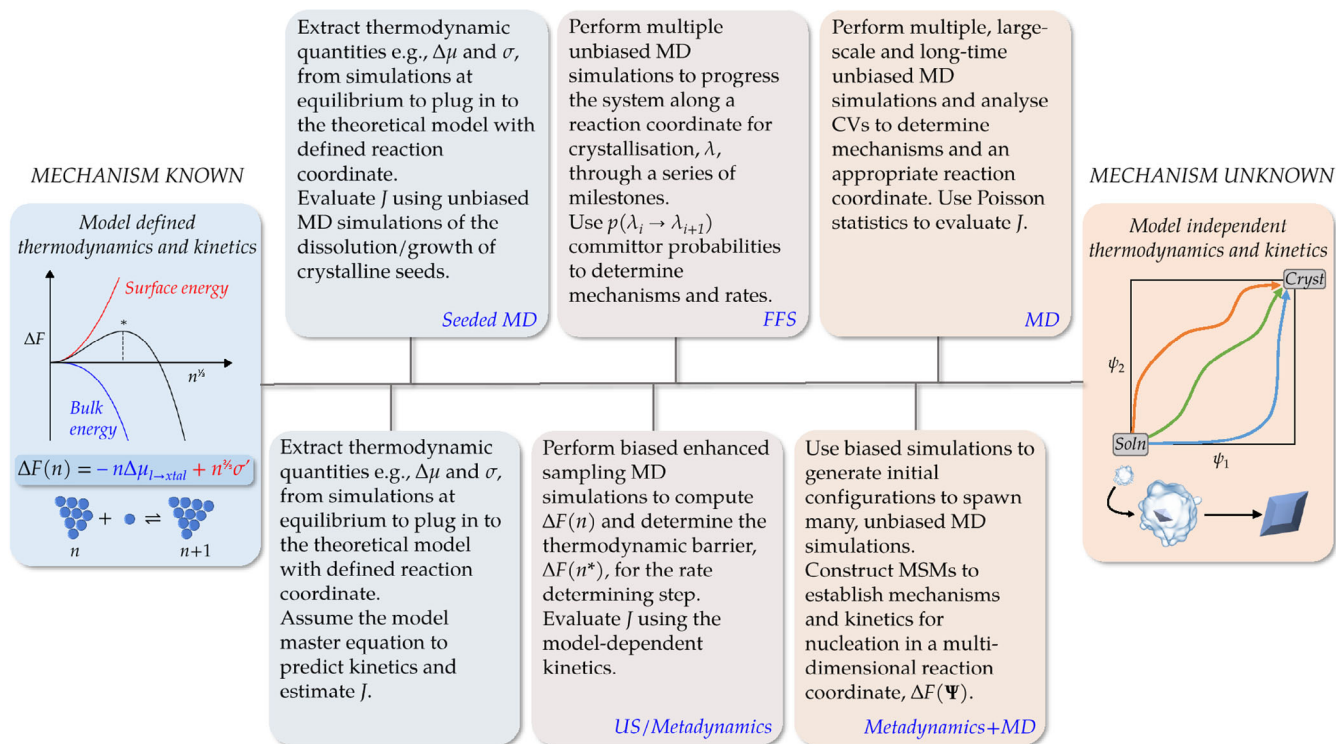


FIGURE 3 A variety of molecular simulation techniques that can be applied to study nucleation from solution ranges from completely theory-based (left side) to exploration-driven computational experiments (right side). The insight available and its associated computational cost vary significantly across this range of methods, highlighting how an a-priori formulation of the research question to be addressed by simulation is essential to guide the selection of an appropriate simulation method.

4.1 | Informing theory with molecular simulations

4.1.1 | The seeding method

In the seeding method, popularized by the Espinosa, Vega, Valeriani, Sanz,^{114,115} Quigley,⁸³ and Peters^{89,116} groups, molecular dynamics simulations are used to inform nucleation rate expressions based on CNT. By construction, the seeding method relies on the a priori acceptance of a quasi-classical nucleation pathway and, within this context, enables the calculation of nucleation rates. Importantly, as extensively discussed by Zimmerman et al.,¹¹⁶ the seeding method is applicable when the size of the nucleus is the reaction coordinate for nucleation, and the capillary approximation can be safely applied (see Section 2). Because analysis of seeded simulations is based on CNT, the approach is only applicable to systems that nucleate classically (see Figure 1).

The key expression in the seeding method is the CNT nucleation rate, that is Equation (4). In this expression, CNT is used to obtain estimates of both the free energy barrier ΔF^* appearing in the exponent and the prefactor $\rho f^+ Z$, where ρ is the density of growth units in solution, f^+ is the attachment frequency, and Z the Zeldovich factor, expressed as a function of the crystallization driving force, as shown in Equation (4).^{114–116}

To solve the rate equations, the parameters that should be estimated at a given T are the solute density ρ , the thermodynamic driving force $\Delta\mu_{\ell \rightarrow \text{xtal}}$, the critical nucleus size n^* and the attachment frequency f^+ . The seeding method leverages unbiased MD simulations to estimate n^* and f^+ by averaging the dynamics of growth or dissolution of nuclei that are prepared with initial size n_0 and equilibrated in a solution with composition ρ and temperature T . The rate is then evaluated according to an estimate of $\Delta\mu_{\ell \rightarrow \text{xtal}}$, which has to be independently obtained at the composition of interest. Along with others discussed below, the method has only been applied to study nucleation from homogeneous solutions thus far. As well as requiring a supply of sufficient monomers to a post-critical seed, the initial seed configuration should be mechanically relaxed which is more challenging in the presence of surfaces. Algorithmic frameworks were designed to combat this latter challenge in simulations of ice nucleation and these tools may also prove useful to study nucleation from solutions.¹¹⁷

The critical nucleus size n^* , for a specific value of ρ and T is obtained by performing ensembles of simulations varying n_0 and, by computing for each ensemble of trajectories initialized at the same n_0 , the ensemble average of the initial drift velocity $\langle \dot{n}(0) \rangle_{n_0}$. The initial size n_0 associated with $\langle \dot{n}(0) \rangle_{n_0} = 0$ is by definition the critical nucleus size n^* . As such, seeding simulations can be performed using any MD engine that can handle the force field. Typically, for a solute concentration, one performs at least five simulations at different values of n_0 (where each seed is carefully relaxed in the solution), each of which involves running a handful of independent simulations to obtain the mean trajectory that indicates the relative stability of the cluster. Seeding methods can, therefore, be relatively cheap and easy to perform in systems where the attachment of monomers from the solution to crystal occurs readily over simulation timescales.

The attachment frequency f^+ can be obtained by considering that the dynamics of the nuclei obey the overdamped Langevin dynamics; Zimmerman et al.^{86,116} have shown that the attachment frequency is limited by the desolvation process. Under this condition $f^+ = f_{des}^+$:

$$f_{des}^+ = 4\pi k_s \sigma_s \rho, \quad (8)$$

where k_s is a second-order rate constant and σ_s is the surface concentration of attachment sites. A more simulation-driven approach to estimate the attachment frequency is instead proposed by Auer and Frenkel, as discussed by Espinosa et al.¹¹⁸:

$$f_{AF}^+ = \frac{\langle (n(t) - n(0))^2 \rangle}{2t}. \quad (9)$$

These two approaches have been shown to lead to the same order-of-magnitude estimates of the attachment frequency f^+ for NaCl nucleation from aqueous solution,¹¹⁵ a case study that will be further explored and discussed in Section 5. A thorough discussion on the estimate of the attachment frequency is reported in Lifanov et al.⁸³

It should be noted that, in the literature, we can find examples of applications where seeding-inspired approaches are used to compute only the prefactor of Equation (4), while enhanced sampling methods are used to independently compute ΔF^* .¹⁰⁴

Notable applications of the seeding method to the investigation of nucleation from solutions include, for example, the study of the nucleation of methane hydrates,¹¹⁹ and the study of urea nucleation from aqueous solution.¹²⁰

4.1.2 | Prenucleation species

Many simulation studies seek to gain an understanding of nucleation without directly simulating the process. These types of simulations are usually the cheapest as the steady-state solution behavior can be achieved readily in standard MD simulations. Often, though, enhanced sampling techniques (see Section 4.2) are also used to determine the thermodynamic stability of associated species. In complex solutions, simulations have provided information on the species potentially involved in nucleation; careful analysis of these species and their assembly can aid the classification of nucleation pathways.^{49,67,121}

Demichelis et al.¹²² performed metadynamics simulations (described in Section 4.2) to confirm the relative stability of calcium phosphate complexes as the first associates to form in solution and thought to be directly involved in mineral nucleation from experiments.¹²³ Simulations have also shed light on structural building units that act as precursors in the nucleation of metal–organic frameworks^{124–126} and the complexes that assemble to form inorganic functional materials.

It was hypothesized that the structural motifs of API dimers in solution encode the polymorphic outcomes of crystal nucleation. As such, MD simulations of organic molecules in solutions have focused on the association of monomer building blocks to support experimental studies.^{127–131} In simple 2D models of flexible chiral molecules, Carpenter and Grünwald¹³² recently demonstrated how bulk crystal structures are related to the organization of building blocks prior to nucleation and the importance of kinetics in predicting polymorphism.

Investigating prenucleation species is particularly useful to identify systems that may undergo phase separation following the PNC pathway and other nonclassical crystallization routes. In particular, many studies have considered the structure and stability of CaCO₃ assemblies that emerge in solution and the effects of additives and solution environment on their properties.^{67,133–139} These studies have demonstrated that entropy drives the formation of PNCs¹³⁵ and questioned the relationship between PNCs and microscopic precursors to dense liquids.^{136,139} Generally, it was observed that organic additives stabilize liquid-like assemblies that form prior to crystallization.^{133,137,138}

Predicting the phase behavior in systems that follow the PNC pathway to phase separation requires evaluation of the equilibrium constant for monomer association, as discussed in Section 2.⁶⁸ Enhanced sampling simulations, particularly umbrella sampling and metadynamics,^{134,138,140} have been successfully applied to evaluate these constants, which corroborate experimental measurements.¹³⁵ Both biased and unbiased MD simulations were informative in predicting the structure and dynamic properties of the PNCs and determining the thermodynamic driving forces for their formation.^{67,137,139} In this regard, the application of simulations has been critical to understanding and evolving theories for the PNC pathway.

4.2 | Molecular simulations as computational experiments

As discussed in Section 4.1.1, the seeding method provides information on the nucleation kinetics, implying a nucleation process that closely follows the nucleation pathway postulated in CNT. By construction, seeding methods cannot answer research questions pertaining to the nucleation mechanism itself. In order to discover and investigate mechanisms deviating from the one postulated in CNT and its extensions, unseeded nucleation simulations are necessary. In unseeded nucleation simulations, the assembly of crystalline nuclei is explicitly sampled, starting from a solution where a critical nucleus is absent. As such, the nucleation mechanism is not a priori defined and emerges from the collective evolution of the system, thus allowing for an open-ended exploration of the nucleation process. This can be useful for studying systems that exhibit complex behavior involving intermediates along the nucleation pathway or that can yield different crystal structures upon nucleation.

In order to sample nucleation events in unseeded simulations, enhanced sampling methods are key. Under conditions of supersaturation of interest for practical applications, spontaneous fluctuations across the nucleation-free energy barriers are too rare to be observed over timescales accessible with standard MD simulations (see Section 3.1). Enhanced sampling methods can thus be used to overcome this limitation. Broadly speaking, enhanced sampling methods used to investigate nucleation from multicomponent liquid phases can be classified depending on whether they introduce a bias potential as a perturbation of the system's Hamiltonian—such as

metadynamics and umbrella sampling—or are based on efficiently sampling the space of reactive paths using techniques such as Transition Path, Transition Interface, and Forward Flux Sampling, or the construction of Markov State Models from unbiased simulation trajectories. Biased enhanced sampling methods typically aim to estimate the free energy barrier associated with the nucleation and to discover nucleation pathways. Nucleation rates are usually obtained by complementing evaluation of the free energy barrier with estimates of the rate prefactor, often carried out using approaches similar to those adopted in the seeding method. A direct calculation of nucleation rates here is only practical in simple cases.⁷⁵

4.2.1 | Biased enhanced sampling approaches

Biased enhanced sampling methods in the context of crystal nucleation are typically deployed to achieve two aims: the calculation of the free energy barrier associated with the nucleation process and the enhanced exploration of the nucleation mechanism. Typically, the former objective can be achieved by either static or history-independent biasing strategies while the second objective is pursued by the deployment of adaptive, history-dependent biasing methods. In the following, we briefly recap the methodological bases of two biased enhanced sampling methods representative of the static and adaptive categories: Umbrella sampling (US) and metadynamics (MetaD). Here we do not report on sampling methods based on constrained dynamics such as the string method used by Santiso et al. to model crystallization both from the melt¹⁴¹ and from solution,¹⁰⁵ and we refer the interested reader to the original publications for an overview of this method, related to US and MetaD. Biased enhanced sampling methods depend crucially on the choice of low-dimensional descriptors of the system configuration—that is, the collective variables (CVs)—that in biased sampling are used to define the bias potential.^{46,142} We briefly discuss this point at the end of this section and for a comprehensive overview, we refer the interested reader to reviews on this topic from Giberti et al.⁸⁷ and Neha et al.¹⁴³ These types of simulations are typically expensive, especially if multiple CVs are biased. Very many (US) and/or very long (MetaD) simulations may be necessary to obtain convergent thermodynamics and kinetics. However, the computational cost is much lower than if one were to observe crystal nucleation spontaneously in such systems. US can be performed in the most commonly available MD engines that allow for implementation of harmonic restraints; though the CVs typically used to simulate nucleation are unlikely to be available in standard MD codes. A useful and noteworthy plugin is the PLUMED software¹⁴⁴ which interfaces with most MD engines and offers a wide range of CVs, including ones that are useful to study crystal nucleation and discussed in this review. PLUMED also allows the practitioner to perform US and MetaD, as well as other types of biased enhanced sampling, in favorite MD engine that has been patched with PLUMED.

Umbrella sampling

Umbrella sampling is a computational method used in MD simulations to calculate the free energy profile associated with an activated transition.^{145–147} The method involves performing a number of independent simulations, or *windows*, where a harmonic bias potential, defined as a function of a CV \mathbf{s} as $V_i = k_i(\mathbf{s} - \mathbf{s}_i)^2$ is added to the Hamiltonian of the system. In the i th window, the bias potential is designed to sample configurations that, in \mathbf{s} , are projected in the vicinity of \mathbf{s}_i . Performing simulations for \mathbf{s}_i values describing a pathway between reactants and products allows collecting configurations distributed between a supersaturated solution and a crystalline nucleus. Within each window, the biased probability density can then be reweighted,^{145,148} and a global free energy profile is then obtained by employing either the weighted histogram analysis (WHAM)^{146,149} the multiple Bennett Acceptance Ratio (mBAR) method,¹⁵⁰ or umbrella integration (UI).¹⁵¹ US is routinely used to compute free energy surfaces associated with activated processes and its most common application in nucleation studies is the calculation of the free energy barrier to nucleation, that is ΔF^* , without resorting to a theoretical formulation of the barrier dependence on thermodynamic parameters such as the solubility and surface tension.¹⁵² For the calculation of nucleation rates, US calculations are complemented by estimates of the nucleation rate prefactor. This can be obtained by using CNT-inspired expressions¹¹⁸ following an approach similar to seeding simulations or, more generally, by drawing from the *Bennett–Chandler* formulation of the kinetic prefactor associated with a rare-event transition.^{86,153}

Metadynamics

Metadynamics (MetaD) is a molecular simulation technique to study the thermodynamics and mechanisms of rare events.^{154,155} It involves introducing a time-dependent bias potential to the system being simulated, which acts to push

the system out of local energy minima and explore a wider range of possible configurations. Poorly explored regions of configuration space, that might not be easily accessed using traditional molecular dynamics simulations, can be visited using this approach. MetaD is one among many methods in which sampling is enhanced by adaptively perturbing the original Hamiltonian of the system by introducing a bias potential, a seminal idea historically introduced with umbrella sampling.^{145,154,156,157} In MetaD such a potential is adaptively constructed as a sum of Gaussian kernels defined in a low-dimensional space of collective variables (CVs), usually indicated as \mathbf{s} . CVs are formulated as continuous and differentiable functions of the microscopic coordinates of a system.^{87,155,158,159} In recent years, several adaptations of MetaD have been proposed, the most relevant being Well-Tempered metadynamics (WTmetaD) introduced by Barducci et al.¹⁵⁵ In WTmetaD, the external repulsive potential is iteratively updated as:

$$V_n(\mathbf{s}) = V_{n-1}(\mathbf{s}) + G(\mathbf{s}, \mathbf{s}_n) \exp \left[-\frac{V_{n-1}(\mathbf{s}_n)}{k_B \Delta T} \right], \quad (10)$$

where $n-1$ and n refer to consecutive iterations of bias deposition, \mathbf{s}_n defines the position of the system in CV space \mathbf{s} at iteration n , V_{n-1} is the total bias potential at iteration $n-1$, k_B is the Boltzmann constant and ΔT is a parameter akin to a temperature. A key feature of the WTmetaD algorithm is the fact that the scaling factor $\exp \left[-\frac{V_{n-1}(\mathbf{s}_n)}{k_B \Delta T} \right]$ decays as $1/n$ leading to a convergent behavior. WTmetaD convergence was demonstrated initially for infinitesimally narrow kernel functions and more recently for any $G(\mathbf{s}, \mathbf{s}_n)$. By changing the parameter ΔT , a controlled enhancement of the fluctuations in \mathbf{s} can be achieved, leading to the asymptotic convergence of the bias in the long-time limit: $V(\mathbf{s}, t \rightarrow \infty) = -\frac{\Delta T}{T + \Delta T} F(\mathbf{s}) + c(t)$, where $F(\mathbf{s})$ is the free energy in the set collective variables \mathbf{s} , and $c(t)$ is a time-dependent constant.¹⁶⁰

In the context of the study of nucleation processes *from solution*, metadynamics was applied to explore the configurational landscapes associated with the nucleation of small organic molecules from liquid solutions,^{91-93,161,162} the nucleation of salts from aqueous solution,^{49,104,163} the nucleation of methane clathrates,¹⁶⁴ and the assembly of perovskites.¹⁶⁵ Figure 4 highlights some of these examples. In all of these cases, metadynamics enables the calculation of continuous reactive trajectories for crystal nucleation and often enables estimates of the free energy landscape associated with such pathways.

4.2.2 | Unbiased enhanced sampling approaches

In this review, we indicate with the adjective *unbiased* those rare event sampling methods that are not based on the perturbation of a system's Hamiltonian by introducing any biasing potential or any artificial force, yet achieve an enhancement of the sampling of activated events by efficiently sampling trajectories obtained using standard MD. Such techniques are based on efficiently sampling the Transition Path Ensemble (TPE) and have been spearheaded by Transition Path Sampling (TPS). After briefly introducing TPS, hereafter, we focus on reporting only techniques that have recently been used to model nucleation from multicomponent liquids, such as Transition Interface Sampling^{166,167} and Forward Flux Sampling^{153,168} that was recently applied to simulate NaCl nucleation from aqueous solution at moderate supersaturations¹⁶⁹ (see Section 5).

Transition path and transition interface sampling

Transition path sampling (TPS) leverages a Monte Carlo algorithm to sample the TPE starting from a single reactive trajectory that connects reactants and products. In the context of homogeneous nucleation, this trajectory connects a supersaturated solution to a solution containing a crystal particle. This initial reactive trajectory is often generated using biased sampling techniques. New trajectories are then sampled by implementing a *shooting* algorithm. While multiple shooting algorithms exist, a typical approach consists of perturbing a configuration sampled by the initial reactive trajectory that is, by slightly modifying the momenta, thus generating a new trajectory via backward and forward propagation in time. If the new trajectory, proceeding through the set of configurations x in n steps, connects reactants (A) and products (B), it is accepted with a path weight that depends on the equilibrium phase space probability density of the initial point in the trajectory $\rho(x_0)$ as:

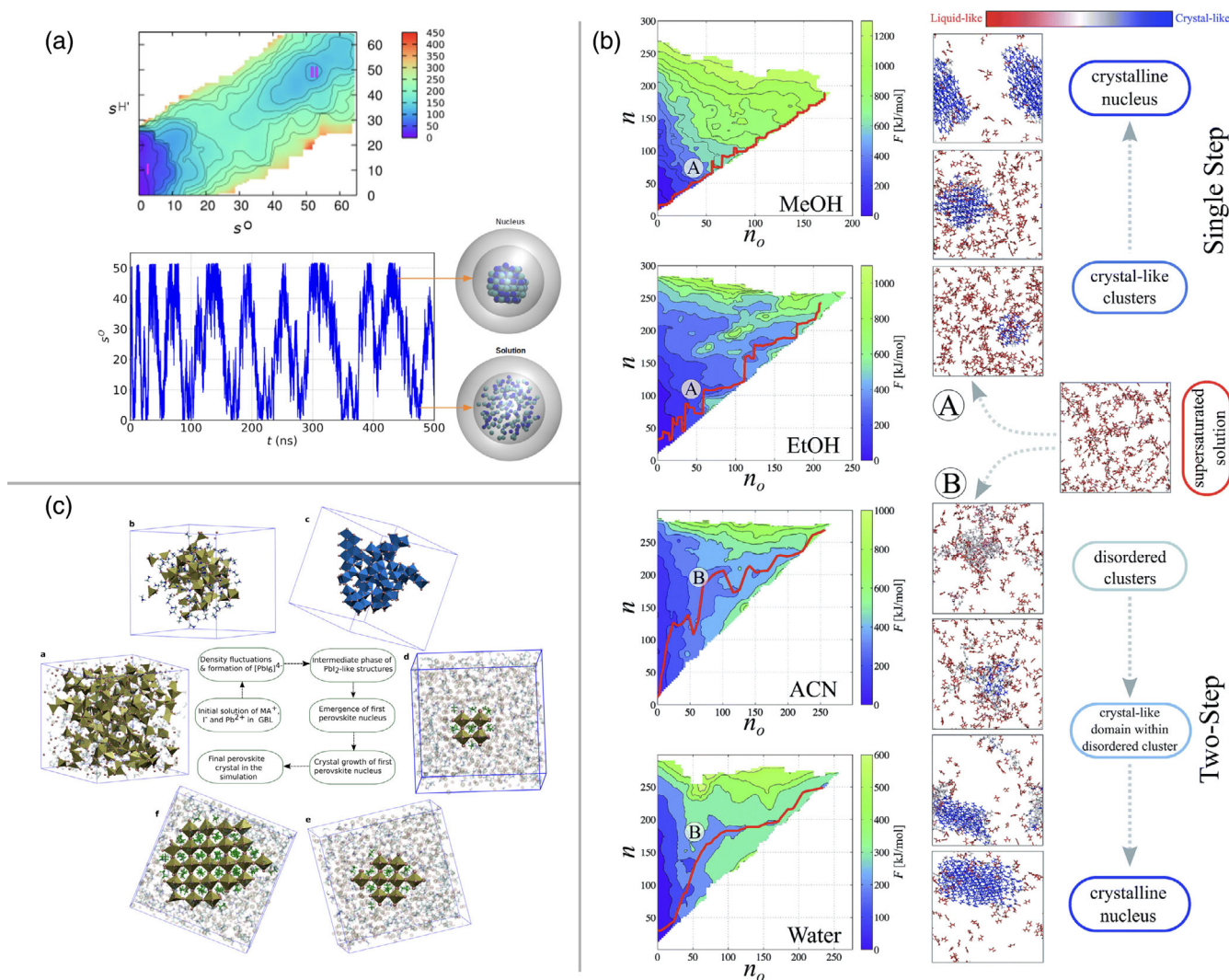


FIGURE 4 MetaD simulations of crystal nucleation from solutions. (a) Combining $C\mu MD$ ^{104,110} and WTMetaD Karmakar et al. compute the free energy surface for NaCl nucleation from solution, limiting the finite-size effects due to confinement. Here the CV enhanced the sampling of crystal structures, and the method was focused on the calculation of the free energy barrier to nucleation. Adapted with permission from Ref.¹⁰⁴ copyright 2019 American Chemical Society. (b) Nucleation of urea from solution. Different solvents induce different nucleation mechanisms: while MeOH and EtOH promote a single-step classical-like process, water and ACN lead to a two-step process. Reproduced from Ref.⁹³ with permission from the Royal Society of Chemistry. (c) Metadynamics enables the discovery of a two-step nucleation mechanism for the synthesis of Methylammonium Lead Iodide Perovskites from solution. Reproduced with permission from Ref.¹⁶⁵ copyright 2020 American Chemical Society.

$$P[x] = \rho(x_0) \prod_{i=0}^{n-1} p(x_i \rightarrow x_{i+1}), \quad (11)$$

where $p(x_i \rightarrow x_{i+1})$ is the probability to transition from configuration x_i to configuration x_{i+1} . The calculation of the rate constant associated with the AB transition is based on the estimate of the correlation function $C(t) = \frac{\langle h_A(x_0) h_B(x_t) \rangle}{\langle h_A(x_0) \rangle}$, where h_i is a characteristic function that is equal to one in state i , and null everywhere else. As discussed in detail in references 153, 170, 171 the calculation of the kinetic constant requires the generation of trajectories targeting intermediate states between A and B , typically generated using an order parameter, or CV. It should be noted that, while the sampling of trajectories is independent from the choice of CVs, the rate calculation does depend on the CV choice.

Transition interface sampling (TIS) was introduced to improve the efficiency of the rate calculation, associated with TPS. In TIS, configuration space is sectioned into non-intersecting interfaces based on a CV (usually indicated with λ in the TPS/TIS literature). The order parameter lambda should enable a mutually exclusive partitioning of the

configuration space between a reactant basin, where $\lambda < \lambda_A = \lambda_0$ and a product basin where $\lambda > \lambda_B = \lambda_n$. The reaction rate constant k_{AB} is then obtained as¹⁶⁷:

$$k_{AB} = \frac{\overline{\Phi}_{A,0}}{\overline{h}_A} \prod_{i=0}^{n-1} P(\lambda_{i+1}|\lambda_i), \quad (12)$$

where $\overline{\Phi}_{A,0}$ is the steady state flux of trajectories leaving the reactant state, which can be easily evaluated by a brute-force MD simulation of the system in the reactant state; \overline{h}_A is a function that is equal to unity if a trajectory was more recently in the reactant state than in the product state; while $\prod_{i=0}^{n-1} P(\lambda_{i+1}|\lambda_i) = P(\lambda_n|\lambda_0)$ is the probability that a trajectory crosses the product interface λ_n , when starting from reactant interface λ_0 . In order to improve the efficiency of the computational evaluation of $P(\lambda_n|\lambda_0)$, further developments of the TIS algorithm, such as replica exchange TIS (RETIS^{171,172}), were proposed, introducing exchange moves between interface ensembles in order to enhance the ergodicity of the sampling. As for TPS, in TIS the CV λ is used to conveniently partition configuration space, and does not affect the sampling of reactive trajectories.

Forward flux sampling

Forward flux sampling (FFS) is a computational method used to study rare events in complex systems. Unlike TPS and TIS, FFS does not require the system to be at equilibrium, and can thus be applied to out-of-equilibrium processes. Similarly to TIS, FFS is based on sampling transitions between non-intersecting interfaces defined in a low-dimensional order parameter space describing the transition from a reactant state, A , to a product state, B . Analogously to TIS, the initial interface λ_A marks the boundary in CV space between the reactants and all other configurations, and λ_B indicates the boundary between the products and all other configurations. The pathway from A to B is described by crossing a series of intermediate interfaces $\lambda : \lambda_i, \dots, \lambda_{n-1}$, with $\lambda_{i+1} > \lambda_i$ for every value of i .

FFS uses the same expression proposed by TIS for the calculation of the transition rate between A and B , namely Equation (12). However, it differs in the computational approach adopted for the calculation of the term $P(\lambda_{i+1}|\lambda_i)$, that provides the attribute *forward* to the method's name. In FFS the probability of crossing interface λ_{i+1} starting from interface λ_i is obtained from the forward-only integration of the system's dynamics. This term is obtained as the fraction of trial runs initiated in λ_i that reach λ_{i+1} .⁷⁹ Implementations of the FFS algorithm differ in the specifics of the algorithm adopted to generate configurations at interfaces and thus in the details associated with the calculation of $P(\lambda_{i+1}|\lambda_i)$. Additional information on FFS can be found in original publications and a number of reviews describing specific of the method. Particularly important in the context of nucleation, where the size of the largest nucleus is typically a good choice of order parameter (see Section 4.2.3), is the fact that *jumpy* order parameters require a specialized treatment in order to consistently yield estimates $P(\lambda_{i+1}|\lambda_i)$ as discussed in detail by Haji-Akbari.¹⁷³ Furthermore, Hall et al.¹⁶⁸ provide a practical guide and comparison between the RETIS and FFS methods.

The need to spawn many MD trajectories (typically, more than 100 crossings are required to achieve good statistical accuracy in the probabilities) along many points (the density of interfaces must be high if the free energy barrier to nucleation is large) in a reaction coordinate make FFS particularly expensive. As FFS requires unbiased MD, it can be performed with all MD packages and using scripts to automate the spawning and analysis of trajectories. Of course, the CV must be implemented in order to use the method. In this regard, SSAGES¹⁷⁴ is software that interfaces with popular MD engines and facilitates FFS simulations with several FFS protocols implemented. Though several CVs are available, they are not typically used to study nucleation; however, a guide is provided in the software documentation on how to add CVs to the code.

4.2.3 | Collective variables

Crystal nucleation from solution is a collective process of assembly that involves, by definition, an ensemble of growth units (molecules, ions, atoms, particles, etc.) that are inherently equivalent and that come together to form a nucleus of a new phase characterized by a well-defined structural arrangement. In order to describe and ultimately understand the salient features of the assembly process, it is necessary to develop low-dimensional descriptors of the system characterizing the transformation. In the theoretical and computational literature on phase transitions, such descriptors take the name of order parameters (OPs). In the context of enhanced sampling, OPs fall into the broader category of Collective Variables (CVs, often indicated as $\mathbf{s}(\vec{\mathbf{r}})$). These are typically functions of the atomic coordinates that are used to define

the bias potential added to the Hamiltonian in biased enhanced sampling or that mark the progress between reactants and products in unbiased enhanced sampling and MD. Very simple CVs to monitor the emergence of crystalline order include average monomer coordination number, system potential energy and local density, though the correct choice of CVs will depend upon the simulated system. In addition, these simple examples may not be able to distinguish between different crystal polymorphs, in which case some characterization of the local symmetry between coordinated monomers in crystal structures should be captured by the (combination of) CVs.

It is important that CVs approximate the reaction coordinate associated with the nucleation process and therefore distinguish important states along the reaction path, such as reactant, product and, ideally, configurations belonging to the transition state ensemble.¹⁷⁵ Moreover, to be used in biased enhanced sampling simulations, CVs should be continuous and differentiable functions of the atomic coordinates. It was demonstrated by Peters et al. that to study single-step nucleation mechanisms and to determine rates, CVs should characterize the size of the largest crystalline cluster in solution; indeed, the application of CVs to describe phase separation can be tested to ensure they are good ones to determine mechanisms and rates and to compare to reaction coordinates adopted in established nucleation theories.^{86,175} In two-step processes, a two-dimensional CV space representing the extent of the largest cluster and of the largest *ordered* domain in the nucleating phase have also emerged as good descriptors of the reaction coordinate,^{46,49,92,93} which also lend themselves to a theoretical description of two-step nucleation.⁴⁸ More recently, the application of Machine Learning methods and the data-driven identification of low-dimensional reaction coordinates for nucleation has emerged as a viable strategy to identify combinations of CVs that enable an effective, low-dimensional description of nucleation processes,^{46,143,176–178} and that allow for the application of biased enhanced sampling to drive polymorph-specific crystal nucleation.¹⁶²

The definition of effective CVs to describe and enhance the sampling of complex nucleation processes in solution also hinges on our ability to define order parameters that can resolve well the atomic environments that are characteristic of specific crystalline structures. While this is routinely done for crystals constructed from atomic or (spherically uniform) single particle monomers—simple colloids being an example—using bond-orientational OPs such as the Steinhardt order parameters^{87,143,179,180} it remains a challenge for molecular crystals. Approaches based on the calculation of generalized pair distribution functions¹⁸¹ or on the calculation of properties of the distributions of order parameters^{182–184} appear promising options but still require significant improvements in terms of computational efficiency, generalizability to molecular systems with a significant degree of conformational complexity¹⁸⁵ and with hundreds of putative polymorphs.^{186,187} All these aspects limit their current applicability to study nucleation from solution and represent one of the most limiting bottlenecks in the current applicability of systematic nucleation studies in molecular systems.

5 | COMPARING APPROACHES: NaCl NUCLEATION FROM AQUEOUS SOLUTION

As anticipated in Section 4, here we report on the results obtained for the nucleation of NaCl in aqueous solution by different researchers over the previous decade. NaCl(aq) is arguably one of the simplest mineralizing solutions and represents one of the earliest case studies for crystallization,^{88,188} instrumental for our understanding of crystal nucleation from solution and for assessing simulation methods applied to this problem. The examples we report are representative of the methods that we briefly introduced in the previous sections and show how different approaches can yield complementary information on NaCl nucleation kinetics and mechanisms from aqueous solutions spanning a wide range of concentrations. Many of the studies conducted on this system employ the same solute and solvent forcefield combination and can thus be quantitatively compared, such as in Figure 5. The force field of reference in the studies described in this section, unless otherwise stated, is the NaCl Joung-Cheatham (JC)¹⁸⁹ force field that makes use of the SPC/E water model.¹⁹⁰ The force field is very simple, using Na⁺ and Cl⁻ ions in water modeled using a three-point water model with a constrained geometry. This force field performs reasonably well to capture the solution properties, including ion solvation energies.¹⁸⁹ This is perhaps surprising for such a simple model that offers a cheap solution to simulate NaCl(aq). It was demonstrated that scaling the charges on ions can improve the model of the solution and several schemes are available in order to do this; see references 191, 192. However, these improvements to modeling the solution phase do not always lead to improved thermodynamic and mechanical properties for the solid NaCl phase when compared to experimental measurements. Recent parameterization strategies reduce some of these errors and provide force fields that predict solubilities close to those determined experimentally.¹⁹³

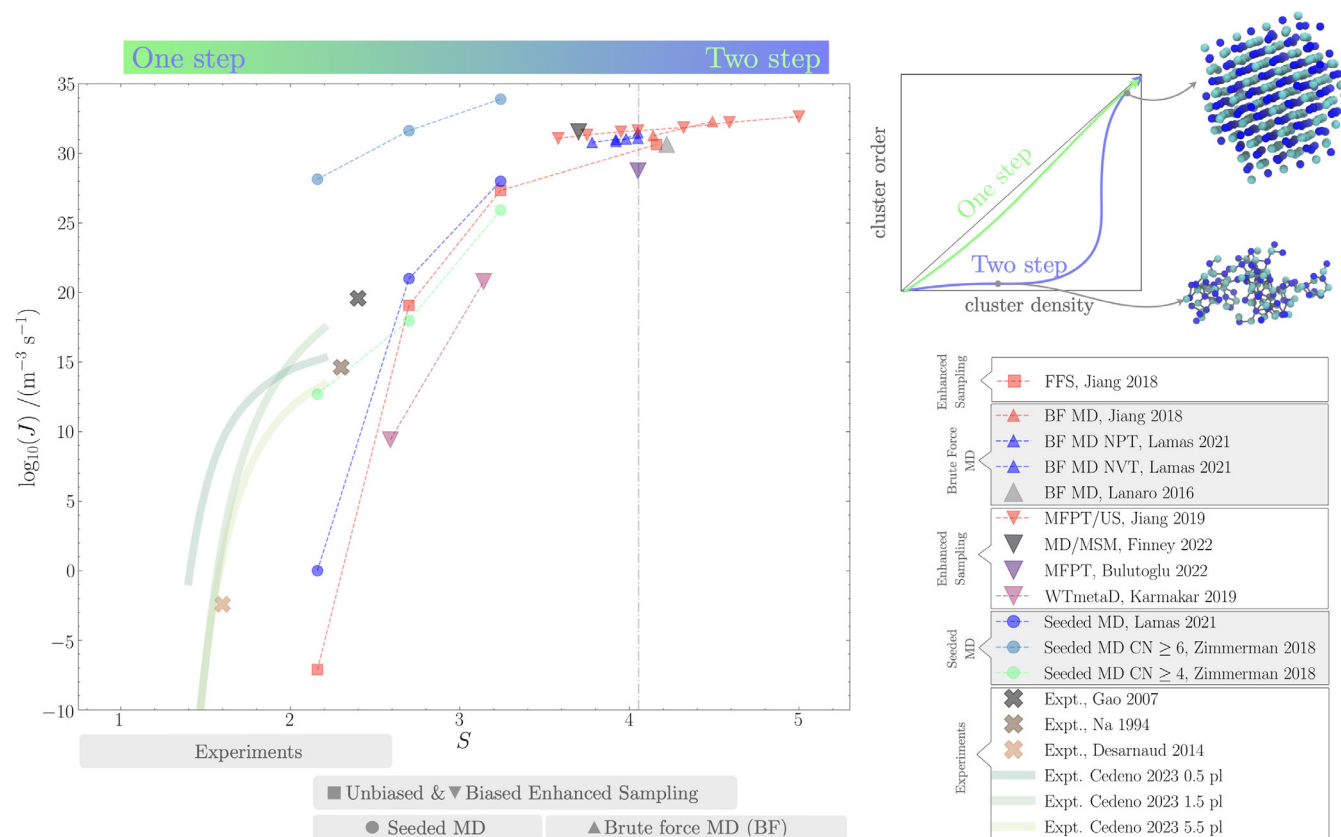


FIGURE 5 Rates for NaCl crystal nucleation from aqueous solution taken from the literature. Here, BF refers to brute force MD simulations in which crystal nuclei emerge spontaneously. The limit of solution stability is indicated by the vertical dashed line at $S = 4.05$. The rate for Bulutoglu was evaluated using their literature rate value of $4 \times 10^3 \text{ s}^{-1}$ and with an ionic density for NaCl(aq) of 6.5 nm^{-3} at 15 mol/kg. Crosses indicate experimental literature values taken from reference 115. All simulations are performed at 298 K, except in the case of Karmakar et al.,¹⁰⁴ where simulations were performed at 350 K. The data reported in this plot are available at https://github.com/mme-ucl/NaCl_water_Nucleation_Rates.git.

Despite more accurate force fields being available, the JC/SPC/E force field is, arguably, the most characterized model to simulate NaCl(aq) and has been adopted in most studies of nucleation where both mechanisms and nucleation rates have been extracted from simulations. Importantly, for the JC/SPC/E force field, consensus on the room temperature solubility has been reached by multiple groups, using a range of different approaches based both on free energy calculations and direct coexistence methods.^{118,192,194–200} An accurate determination of the solubility enables the correct assessment of $\Delta\mu$ that is used to evaluate rates in the seeding method (see Section 4.1.1),^{89,115,116} and to consistently attribute a supersaturation level to simulations performed with theory-agnostic methods such as FFS, and Markov State Models (MSMs).^{49,152,169,201} The accepted solubility for JC NaCl in SPC/E water at 298.15 K and 1 bar is 3.7 m , which we label b^{sat} .²⁰² As discussed in detail by Zimmermann et al.,¹¹⁶ this estimate is based on the values independently estimated by Moucka et al. ($3.64 \pm 0.2 \text{ m}$), Mester et al. ($3.71 \pm 0.04 \text{ m}$), Benavides et al. ($3.71 \pm 0.20 \text{ m}$), Kolafa et al. ($3.6 \pm 0.4 \text{ m}$) and Espinosa et al. ($3.7 \pm 0.4 \text{ m}$), and is the product of a process that has seen the refinement of simulation approaches and correction of earlier inaccurate estimates.

5.1 | Nucleation kinetics and mechanisms at moderate supersaturation with forward flux sampling

Employing Forward flux Sampling (see Section 4.2), Jiang et al. have investigated the nucleation of NaCl from brine at supersaturation ratios ($S = b/b^{\text{sat}}$) ranging between 2.1 and 4.5.¹⁶⁹ FFS allows evaluation of both the nucleation mechanism and nucleation kinetics at these conditions. Concerning the former, the authors observe that nuclei tend to assemble directly into an FCC-like rocksalt structure, independently of their size, and that the level of crystallinity of the

nuclei has a strong influence on their lifetime and probability of growing. This suggests that the nucleation process at moderate supersaturations follows a mechanism that can be described by CNT. The nucleation kinetics obtained by FFS tend to underestimate experimental measurements, possibly due to an overestimation of the crystal/solution interfacial tension.¹⁶⁹ Nevertheless, this dataset provides an important benchmark for studying nucleation from solution as rates here are computed independently from any theoretical interpretation of the self-assembly process.

5.2 | Nucleation kinetics within the remit of CNT

The quality of the dataset provided by Jiang et al.¹⁶⁹ comes with a significant computational cost. Hence, there is a strong incentive to test this result against less computationally intensive methods such as seeding (see Section 4.1.1). Two papers^{115,116} have provided independent attempts at comparing nucleation kinetics of seeding with the FFS dataset. Zimmerman et al.^{89,116} exploit the fact that the ion attachment frequency f^+ is dominated by ion desolvation and adopt the theoretical expression of the nucleation rate prefactor reported in Equation (8). Moreover, in contrast with other seeding studies,¹¹⁸ the authors adopt the Girshick–Chiu correction²⁰³ $\Gamma = \exp(F(1)/k_B T)$ (where $F(1)$ corresponds to the free energy associated to the formation of a monomer in a crystalline configuration) in the prefactor to the rate expression. Adopting the correct value of $\Delta\mu_{e \rightarrow \text{xtal}}$ ¹¹⁶ allowed Zimmerman et al. to estimate nucleation rates in a range of supersaturations overlapping with the Jiang dataset. A similar simulation strategy was recently employed by Lamas et al.¹¹⁵ who adopted the Auer and Frenkel expression of the attachment frequency f^+ (see Equation 9) and did not explicitly introduce the Girshick–Chiu correction in their working expression for the calculation of the nucleation rates, yielding a set of results slightly less reliant on theoretical considerations than those of Zimmermann et al.

As shown in Figure 5, these two approaches yield substantial discrepancies in the estimate of the nucleation rates, which can only partly be attributed to subtle differences in the expressions adopted for the nucleation rate prefactor. The discrepancy appears to originate, instead, from the classification criteria used to estimate the number of crystalline particles in the evolving seeds. The classification problem is discussed at length by both Zimmerman et al.¹¹⁶ and Lamas et al.¹¹⁵ In particular, Zimmerman et al. show that nucleation rates are extremely sensitive to the number of crystal-like neighbors necessary to consider an ion part of the crystal nucleus. For example, Figure 5 demonstrates how the predicted rates change by around 10–15 orders of magnitude when crystalline ions are identified as those with a coordination number ≥ 4 or ≥ 6 . Lamas et al., instead, resort to developing a systematic approach for the classification of the ions as part of a crystalline particle based on the analysis of the overlap of distributions of the local q_4 order parameter for a bulk crystalline phase and a bulk solution. The identification of the optimal threshold yields nucleus size estimates that give rise to nucleation rates in good agreement with FFS results.

5.3 | Overcoming solution depletion

Karmakar et al.,¹⁰⁴ compute the nucleation barrier at two distinct supersaturation conditions from unseeded simulations (see Section 4.2 and Figure 4a) by combining WTmetaD¹⁵⁵ with $C\mu\text{MD}$.¹⁰² This approach allows a decoupling of the size of the nucleus from the chemical potential of the parent phase by mimicking an open boundary system at constant composition, therefore overcoming the *depletion* issues that typically affect both the qualitative and quantitative behavior of nucleating systems in small volumes (see Section 3.2).^{33,92} Depletion artifacts affect the shape of the nucleation free energy profile, the estimate of the critical nucleus size, and in severe cases, can even completely inhibit nucleation. Depletion effects are also present in regular seeding simulations. However, depletion only affects the estimate of nucleation rates when the critical nucleus size $n^* \ll V(c - c^{\text{sat}})$,⁸⁹ where c^{sat} is the concentration of a saturated solution. For systems where c^{sat} is large, satisfying this equation may require very large simulation volumes.

Karmakar et al. use the nucleation barriers estimated via WTmetaD and $C\mu\text{MD}$ to compute the exponential term in the nucleation rate expression (Equation 4). The prefactor is then estimated with the approach adopted by Espinosa et al.¹¹⁸ The values of the nucleation rates obtained by Karmakar et al. following this route are reported in Figure 5. It should be noted that while the forcefield used in this study is the JC/SPC/E model, the results are not directly comparable as calculations were performed at 350 K, rather than at room temperature.

5.4 | NaCl nucleation mechanism is supersaturation-dependent

As mentioned in the introductory section, an important feature of nucleation from solutions is the fact that both the rates and the mechanism (or pathway) for nucleation depends on the composition of the mother phase⁹³ and, in particular, its degree of supersaturation. Studies based on seeding cannot lead to the discovery of pathways departing from CNT-compliant ones by construction. In contrast, unseeded simulations can reveal pathways that depart from CNT predictions (see Section 2).

Using unseeded simulations, Panagiotopoulos and co-workers¹⁵² discovered the mechanism of NaCl nucleation from aqueous solution depends on supersaturation. By employing large-scale MD simulations and free energy calculations, they identified the limit of stability for aqueous NaCl solutions with respect to a liquid/amorphous phase separation (reported in Figure 5 as a vertical dashed line). The dense *amorphous salt* clusters observed beyond the limit of solution stability act as intermediates in the crystallization of NaCl, where crystalline order emerges within these disordered clusters following a two-step nucleation pathway¹⁵² (see Section 2). Nucleation rates in this region of the phase diagram were obtained by Jiang et al. by estimating Mean First Passage times (MFPTs) from brute force sampling of nucleating trajectories and by performing umbrella sampling simulations to estimate nucleation barriers. Lamas et al.¹¹⁵ corroborate these values of nucleation rates in the proximity of the limit of solution stability by constructing survival probability distributions from brute force simulations that yield nucleation rates within the same order of magnitude (see Figure 5). An overview of both the MFPT and survival probability methods for the calculation of nucleation rates from brute force simulation is provided by Chkonia et al.²⁰⁴

The fact that at the limit of solution stability, the nucleation of NaCl follows a nonclassical, multi-step pathway has been further corroborated by the work of Bulutoglu et al.,²⁰⁵ where the nucleation mechanism of NaCl is analyzed by performing free energy calculations and by developing a theoretical approach based on the composite-cluster model⁵⁹ able to interpret the atomistic simulation results. Most notably, fitting the composite-cluster model to simulation data revealed that beyond the limit of solution stability, the amorphous salt clusters are thermodynamically favored compared to the aqueous solution, thus further validating the existence of a two-step nucleation pathway for NaCl at high supersaturation.

5.5 | Discovering nucleation mechanisms by combining biased and unbiased simulations

Motivated by the observation of dense liquid-like ionic clusters emerging in the double layer in simulations of a solid-liquid interface²⁰⁶ at supersaturation levels significantly lower than the limit of solution stability discovered by Jiang et al.,¹⁵² and by recent experimental and simulation observations suggesting significant ion-ion correlations occurring in supersaturated NaCl¹²¹; we have recently investigated the emergent nucleation mechanism for NaCl(s) in supersaturated solutions *below* the limit of solution stability. To this aim, we performed multiple metadynamics simulations where we enhanced fluctuations in the local ion density to explore the configuration space of the nuclei forming in an aqueous solution. By mapping the nuclei configurations as a function of two order parameters indicative of the size of a dense cluster and of the size of a dense cluster with a crystalline structure, we collected configurations able to describe the emergence of crystalline NaCl nuclei without prescribing a specific pathway.⁴⁹

By initializing hundreds of brute force MD simulations from uniformly distributed configurations in the $n, n(q_6)$ space (q_6 being a sixth-order Steinhardt bond-order parameter), we then constructed a MSM able to yield model-free estimates of the nucleation rate of the committor surface in order parameter space. The MSM reveals that when $S = 3.7$, both one and two-step nucleation mechanisms are indeed accessible, with the two-step nucleation pathways being slightly more favorable. Interestingly, the analysis of the committor probability surface in the $n, n(q_6)$ space suggests that at these conditions, one may need to extend the attribute critical to an ensemble of clusters that, despite displaying a broad range of structures, include sizeable disordered domains and have an equal probability of evolving toward a macroscopic crystal or dissolving.

As well as characterizing the nucleation mechanisms far into the metastable solution zone, we also ruled out the PNC pathway for NaCl. Using umbrella sampling, we computed the equilibrium constant K for ion pair association in the dilute limit. Despite the significant ion association that occurs in solutions across all of the metastable solutions investigated $S = 1 - 4$, the result that $K < 1$ means that ion dissociation is thermodynamically favorable in dilute solutions and ion assembly into liquid-like clusters is due to non-idealities in the solution phase. These liquid-like entities

can reach significant sizes, containing up to hundreds of ions at the high end of concentration, and evolve their topology rapidly over simulation timescales.

5.6 | Uncovering multi-step processes involving crystal polymorphs

Using simulations as *computational experiments*, involves methods that enable the *discovery* of nucleation mechanisms as emergent, collective evolution of systems. The nucleation of NaCl from aqueous solution offers an interesting case study in this regard, showcasing the potential of simulation approaches while contextually providing a cautionary tale about the quality of the models used to explore nucleation.¹⁶³ Performing WMetaD simulations of NaCl nucleation from an aqueous solution, where the NaCl ions were modeled with the GROMOS forcefield,²⁰⁷ Giberti et al. discovered that small NaCl clusters might preferentially adopt structures that differ from that of bulk rock salt.¹⁶³ By employing a CV designed to enhance local density fluctuations without favoring a specific crystal structure,^{46,163} the authors discovered that for the model adopted there is a competition between hydrated amorphous NaCl, rocksalt nuclei, and nuclei of a new wurtzite-like phase. An analysis of the CNT nucleation free energies of the rocksalt and wurtzite phases revealed that indeed, *according to the molecular model*, wurtzite-like arrangements are more favorable than rocksalt at small sizes. While this result is not representative of the real NaCl nucleation mechanism, due to the fact that the GROMOS model for NaCl strongly underestimates its aqueous solubility¹⁶⁹ and overestimates the stability of the wurtzite structure at the supersaturation conditions sampled by Giberti, it nevertheless provides a very important observation pertaining to simulation methods. This study, in fact, demonstrates that unexpected pathways can be discovered from a direct sampling obtained with enhanced MD techniques.

6 | PERSPECTIVES AND CONCLUSIONS

Molecule and particle based simulation methods provide useful tools to gain mechanistic insight into the nucleation of crystals from solutions. With proper sampling, quantitative thermodynamics and kinetics can be obtained to compare with experiments and test theories for crystal nucleation. In this review, we have highlighted state-of-the-art techniques to gain such information. These can be broadly categorized as methods which rely on theory with a well-defined reaction coordinate—mainly CNT—and those which are truly explorative.

Simulating nucleation in multicomponent solutions has been facilitated by the advent of enhanced sampling techniques and novel approaches to simulate nucleation using standard MD, for example, seeding. Exemplary studies of NaCl crystal nucleation, which we have discussed, highlight the range of methods available to overcome time- and length-scale challenges associated with the direct simulation of crystal nucleation in microscopic, closed systems. Alongside the referenced simulations to study a wide range of nucleating systems, these works demonstrate the capability of molecular simulations to predict crystallization outcomes and determine rates.

Figure 5 presents NaCl crystal nucleation rates evaluated from simulations and measured in experiments. Accepting that the y-axis spans 45 orders of magnitude, and given the known intricacies associated with nucleation rate calculations,²⁰⁸ there appears to be reasonable consensus in the rates from unbiased and biased simulation strategies. System size effects (including the large critical nucleus sizes that are observed at low supersaturation, as well as solute depletion) mean that simulation studies typically adopt high values of S . In experiments, however, NaCl nucleation becomes so fast at the higher end of S that special experimental techniques must be used to determine rates even up to $S \approx 2$. In the small range of S where simulations overlap with experiments, a comparison of the rates leaves much to be desired. While it is clear from Zimmerman et al.¹¹⁶ that the vast difference in the rates from seeding simulations depends on the definition of the order parameter (OP), this issue is addressed by the mislabeling analysis of Lamas et al.¹¹⁵ Taking the latter seeding results along with the forward flux sampling (FFS) rates from Jiang et al.,¹⁶⁹ the force field predicts crystal nucleation rates that are approximately 20 orders of magnitude lower than experiments. This discrepancy may indicate inaccuracies in the molecular model; however, we note that there is even a discrepancy in the experimental data, spanning around five orders of magnitude. This example, simple system typifies the challenge of comparing simulation and experimental studies of crystal nucleation. The moonshot challenge, therefore, is to develop simulation *and* experimental strategies that consistently reproduce rates that can be confidently compared to one another. Mechanistically, the outlook is less bleak, as both simulations and experiments indicate one- and two-step crystal nucleation according to the solution supersaturation.^{49,121,152,205}

Impressive advances in simulation capabilities were made over the previous decade; however, it is undoubtedly evident to the reader that the studies we highlight where quantitative thermodynamics and kinetics were evaluated apply to very simple systems. Even so, the determination of nucleation rates requires exhaustive computational resources. Sampling crystal nucleation pathways in more complex systems, such as in solutions of molecules with conformational freedom that are so important to the pharmaceutical industry, is still a major challenge that is yet to be realized beyond small molecules. Nevertheless, increasing computational power and the progress made in machine learning techniques may make the routine simulation of crystal nucleation realizable in the near future.

Assuming that computational resources are abundant, we doubt that a one-size-fits-all simulation approach will be achieved any time soon. This is principally due to the many different crystallization pathways that are evident in different systems. Indeed, simulations are utilized to support perspectives regarding the invalidity of established theories to describe nucleation generally. Adopting the most suitable simulation strategy can therefore be of critical importance. When departing from CNT-compliant methods, it is far less obvious what reaction coordinate should be sampled. As we have discussed, even in the CNT framework, how the cluster size is determined can drastically affect the predicted nucleation rates. While methods based on transition path sampling are less susceptible to errors in rates due to inappropriately defined reaction coordinates, a rigorous mechanistic understanding requires appropriate variables to describe crystal nucleation and countless examples are found in the literature. Unfortunately, there is no silver bullet when it comes to quantifying the emergence of order, though perhaps, this is good for explorative purposes.

Simulating crystal nucleation in extremely large atomic systems is unlikely to be achieved in the near future. For instance, some systems, such as proteins, are just too big; others have such low solubility that simulating crystallization directly from monomers in solution at experimental concentrations requires system's sizes that are simply beyond our reach (hence, the necessity to develop increasingly accurate coarse-grained representations of these systems). However, with continued advances in simulation techniques and an increased understanding of the factors that govern nucleation from solution, we believe that the next step will be to systematically extend nucleation studies to systems where the growth units are conformationally flexible organic molecules and/or are polymorphic. In turn, this will enable accounting for crucial out-of-equilibrium effects in computationally-assisted material design.

AUTHOR CONTRIBUTIONS

Matteo Salvalaglio: Conceptualization (equal); data curation (supporting); funding acquisition (lead); project administration (lead); supervision (lead); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Aaron R. Finney:** Conceptualization (equal); data curation (lead); visualization (equal); writing – original draft (equal); writing – review and editing (equal).

ACKNOWLEDGMENTS

We thank Dr. Gabriele Cesare Sosso for his feedback on an early version of the manuscript. We acknowledge funding from the *Crystallisation in the Real World* EPSRC Programme Grant (Grant EP/R018820/1). Matteo Salvalaglio acknowledges support from the EPSRC via the UKRI Horizon Frontier Research Guarantee Grant (EP/X033139/1).

FUNDING INFORMATION

This study was supported by the EPSRC, UKRI Horizon Frontier Research Guarantee Grant, EP/X033139/1; EPSRC, *Crystallisation in the Real World* Programme Grant, EP/R018820/1.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

OPEN RESEARCH BADGES



This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at https://github.com/mme-ucl/NaCl_water_Nucleation_Rates.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Matteo Salvalaglio  <https://orcid.org/0000-0003-3371-2090>

RELATED WIREs ARTICLES

[Metadynamics](#)

[Umbrella Sampling](#)

REFERENCES

1. Price SL. Control and prediction of the organic solid state: a challenge to theory and experiment. *Proc R Soc A Math Phys Eng Sci.* 2018;474(2217):20180351.
2. Price SL. Predicting crystal structures of organic compounds. *Chem Soc Rev.* 2014;43(7):2098–111.
3. Day GM. Current approaches to predicting molecular organic crystal structures. *Crystallogr Rev.* 2011;17(1):3–52.
4. Hoja J, Ko HY, Neumann MA, Car R, DiStasio RA, Tkatchenko A. Reliable and practical computational description of molecular crystal polymorphs. *Sci Adv.* 2019;5(1):eaau3338.
5. Rossi M, Gasparotto P, Ceriotti M. Anharmonic and quantum fluctuations in molecular crystals: a first-principles study of the stability of paracetamol. *Phys Rev Lett.* 2016;117(11):115702.
6. Kapil V, Engel EA. A complete description of thermodynamic stabilities of molecular crystals. *Proc Natl Acad Sci.* 2022;119(6):e2111769119.
7. Li L, Totton T, Frenkel D. Computational methodology for solubility prediction: application to the sparingly soluble solutes. *J Chem Phys.* 2017;146(21):214110.
8. Cheng B, Ceriotti M. Computing the absolute Gibbs free energy in atomistic simulations: applications to defects in solids. *Phys Rev B.* 2018;97(5):054102.
9. Li L, Totton T, Frenkel D. Computational methodology for solubility prediction: application to sparingly soluble organic/inorganic materials. *J Chem Phys.* 2018;149(5):054102.
10. Kapil V, Engel E, Rossi M, Ceriotti M. Assessment of approximate methods for anharmonic free energies. *J Chem Theory Comput.* 2019;15(11):5845–57.
11. Abraham NS, Shirts MR. Adding anisotropy to the standard quasi-harmonic approximation still fails in several ways to capture organic crystal thermodynamics. *Cryst Growth des.* 2019;19(12):6911–24.
12. Abraham NS, Shirts MR. Statistical mechanical approximations to more efficiently determine polymorph free energy differences for small organic molecules. *J Chem Theory Comput.* 2020;16(10):6503–12.
13. Reilly AM, Cooper RI, Adjiman CS, Bhattacharya S, Boese AD, Brandenburg JG, et al. Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallogr Sect B Struct Sci Crystal Eng Mater.* 2016;72(4):439–59.
14. Zhang P, Wood GP, Ma J, Yang M, Liu Y, Sun G, et al. Harnessing cloud architecture for crystal structure prediction calculations. *Cryst Growth Des.* 2018;18(11):6891–900.
15. Mortazavi M, Hoja J, Aerts L, Quéré L, van de Streek J, Neumann MA, et al. Computational polymorph screening reveals late-appearing and poorly-soluble form of rotigotine. *Commun Chem.* 2019;2(1):1–7.
16. Yang M, Dybeck E, Sun G, Peng C, Samas B, Burger VM, et al. Prediction of the relative free energies of drug polymorphs above zero kelvin. *Cryst Growth Des.* 2020;20(8):5211–24.
17. Kendrick J, Leusen FJ, Neumann MA, van de Streek J. Progress in crystal structure prediction. *Chem A Eur J.* 2011;17(38):10736–44.
18. Cruz-Cabeza AJ, Reutzel-Edens SM, Bernstein J. Facts and fictions about polymorphism. *Chem Soc Rev.* 2015;44(23):8619–35.
19. Nyman J, Reutzel-Edens SM. Crystal structure prediction is changing from basic science to applied technology. *Faraday Discuss.* 2018;211:459–76.
20. Srinivasan K. Crystal growth of α and γ glycine polymorphs and their polymorphic phase transformations. *J Cryst Growth.* 2008;311(1):156–62.
21. Price SL. Why don't we find more polymorphs? *Acta Crystallogr Sect B.* 2013;69(4):313–28.
22. Gadewar SB, Doherty MF. A dynamic model for evolution of crystal shape. *J Cryst Growth.* 2004;267:239–50.
23. Gadewar SB, Hofmann HM, Doherty MF. Evolution of crystal shape. *Cryst Growth Des.* 2004;4:109–12.
24. Zhang Y, Sizemore JP, Doherty MF. Shape evolution of 3-dimensional faceted crystals. *AIChE J.* 2006;52:1906–15.
25. Davey RJ, Dent G, Mughal RK, Parveen S. Concerning the relationship between structural and growth synthons in crystal nucleation: solution and crystal chemistry of carboxylic acids as revealed through ir spectroscopy. *Cryst Growth Des.* 2006;6:1788–96.
26. Davey RJ, Schroeder SLM, Ter Horst JH. Nucleation of organic crystals—a molecular perspective. *Angew Chem Int Ed.* 2013;52(8):2167–79.
27. Sun CC, Sun W, Price S, Hughes C, Ter Horst J, Veessler S, et al. Solvent and additive interactions as determinants in the nucleation pathway: general discussion. *Faraday Discuss.* 2015;179:383–420.
28. Piana S, Reyhani M, Gale JD. Simulating micrometre-scale crystal growth from solution. *Nature.* 2005;438:70–3.
29. Piana S, Gale JD. Three-dimensional kinetic Monte Carlo simulation of crystal growth from solution. *J Cryst Growth.* 2006;294:46–52.
30. Piana S, Gale JD. Understanding the barriers to crystal growth: dynamical simulation of the dissolution and growth of urea from aqueous solution. *J Am Chem Soc.* 2005;127:1975–82.

31. Piana S, Jones F, Gale JD. Aspartic acid as a crystal growth catalyst. *CrstEngComm*. 2007;9:1187–91.
32. Agarwal V, Peters B. Nucleation near the eutectic point in a Potts-lattice gas model. *J Chem Phys*. 2014;140(8):084111.
33. Sosso GC, Chen J, Cox SJ, Fitzner M, Pedevilla P, Zen A, et al. Crystal nucleation in liquids: open questions and future challenges in molecular dynamics simulations. *Chem Rev*. 2016;116(12):7078–116.
34. Anwar J, Khan S, Lindfors L. Secondary crystal nucleation: nuclei breeding factory uncovered. *Angew Chem Int Ed*. 2015;54(49):14681–4.
35. Sosso GC, Tribello GA, Zen A, Pedevilla P, Michaelides A. Ice formation on kaolinite: insights from molecular dynamics simulations. *J Chem Phys*. 2016;145(21):211927.
36. Sosso GC, Sudera P, Backes AT, Whale TF, Fröhlich-Nowoisky J, Bonn M, et al. The role of structural order in heterogeneous ice nucleation. *Chem Sci*. 2022;13(17):5014–26.
37. Factorovich MH, Naullage PM, Molinero V. Can clathrates heterogeneously nucleate ice? *J Chem Phys*. 2019;151(11):114707.
38. Glatz B, Sarupria S. Heterogeneous ice nucleation: interplay of surface properties and their impact on water orientations. *Langmuir*. 2018;34(3):1190–8.
39. Lata NN, Zhou J, Hamilton P, Larsen M, Sarupria S, Cantrell W. Multivalent surface cations enhance heterogeneous freezing of water on muscovite mica. *J Phys Chem Lett*. 2020;11(20):8682–9.
40. Mullin JW. *Crystallization*. 4th ed. Oxford, UK: Butterworth-Heinemann; 2001.
41. De Yoreo JJ, Vekilov PG. Principles of crystal nucleation and growth. *Rev Mineral Geochem*. 2003;54(1):57–93.
42. Erdemir D, Lee AY, Myerson AS. Nucleation of crystals from solution: classical and two-step models. *Acc Chem Res*. 2009;42(5):621–9.
43. Vekilov PG. The two-step mechanism of nucleation of crystals in solution. *Nanoscale*. 2010;2(11):2346.
44. Schöpe HJ, Bryant G, van Meegen W. Two-step crystallization kinetics in colloidal hard-sphere systems. *Phys Rev Lett*. 2006;96(17):175701.
45. Vatamanu J, Kusalik PG. Observation of two-step nucleation in methane hydrates. *Phys Chem Chem Phys*. 2010;12(45):15065.
46. Finney AR, Salvalaglio M. A variational approach to assess reaction coordinates for two-step crystallization. *J Chem Phys*. 2023;158(9):094503.
47. ten Wolde PR, Frenkel D. Enhancement of protein crystal nucleation by critical density fluctuations. *Science*. 1997;277:1975–8.
48. Kashchiev D. Classical nucleation theory approach to two-step nucleation of crystals. *J Cryst Growth*. 2020;530:125300.
49. Finney AR, Salvalaglio M. Multiple pathways in NaCl homogeneous crystal nucleation. *Faraday Discuss*. 2022;235:56–80.
50. De Yoreo JJ, Biao J, Chen Y, Pyles H, Baer M, Legg B, et al. Formation, chemical evolution, and solidification of the calcium carbonate dense liquid phase. *Res Sq*. 2022.
51. Whitehead CB, Finke RG. Particle formation mechanisms supported by in situ synchrotron XAFS and SAXS studies: a review of metal, metal-oxide, semiconductor and selected other nanoparticle formation reactions. *Mater Adv*. 2021;2(20):6532–68.
52. Krautwurst N, Nicoleau L, Dietzsch M, Lieberwirth I, Labbez C, Fernandez-Martinez A, et al. Two-step nucleation process of calcium silicate hydrate, the Nanobrick of cement. *Chem Mater*. 2018;30(9):2895–904.
53. Kashchiev D. *Nucleation*. Oxford, UK: Elsevier; 2000.
54. Dillmann A, Meier GEA. A refined droplet approach to the problem of homogeneous nucleation from the vapor phase. *J Chem Phys*. 1991;94(5):3872–84.
55. Ford IJ, Laaksonen A, Kulmala M. Modification of the Dillmann–Meier theory of homogeneous nucleation. *J Chem Phys*. 1993;99(1):764–5.
56. Prestipino S, Laio A, Tosatti E. Systematic improvement of classical nucleation theory. *Phys Rev Lett*. 2012;108(22):225701.
57. Ten Wolde PR, Frenkel D. Computer simulation study of gas–liquid nucleation in a Lennard–Jones system. *J Chem Phys*. 1998;109(22):9901–18.
58. Lutsko JF, Durán-Olivencia MA. A two-parameter extension of classical nucleation theory. *J Phys Condens Matter*. 2015;27(23):235101.
59. Iwamatsu M. Free-energy landscape of nucleation with an intermediate metastable phase studied using capillarity approximation. *J Chem Phys*. 2011;134(16):164508.
60. Sear RP. The non-classical nucleation of crystals: microscopic mechanisms and applications to molecular crystals, ice and calcium carbonate. *Int Mater Rev*. 2012;57(6):328–56.
61. Lee J, Yang J, Kwon SG, Hyeon T. Nonclassical nucleation and growth of inorganic nanoparticles. *Nat Rev Mater*. 2016;1(8):16034.
62. Karthika S, Radhakrishnan TK, Kalaichelvi P. A review of classical and nonclassical nucleation theories. *Cryst Growth Des*. 2016;16(11):6663–81.
63. De Yoreo J. A perspective on multistep pathways of nucleation. In: Zhang X, editor. ACS symposium series. Volume 1358. Washington, DC: American Chemical Society; 2020. p. 1–17.
64. Jun YS, Zhu Y, Wang Y, Ghim D, Wu X, Kim D, et al. Classical and nonclassical nucleation and growth mechanisms for nanoparticle formation. *Annu Rev Phys Chem*. 2022;73(1):453–77.
65. Gebauer D, Volkel A, Colfen H. Stable prenucleation calcium carbonate clusters. *Science*. 2008;322(5909):1819–22.
66. Gebauer D, Kellermeier M, Gale JD, Bergström L, Cölfen H. Pre-nucleation clusters as solute precursors in crystallisation. *Chem Soc Rev*. 2014;43(7):2348–71.
67. Demichelis R, Raiteri P, Gale JD, Quigley D, Gebauer D. Stable prenucleation mineral clusters are liquid-like ionic polymers. *Nat Commun*. 2011;2(1):590.
68. Avaro JT, Wolf SLP, Hauser K, Gebauer D. Stable prenucleation calcium carbonate clusters define liquid–liquid phase separation. *Angew Chem Int Ed*. 2020;59(15):6155–9.

69. Palmer JC, Debenedetti PG. Recent advances in molecular simulation: a chemical engineering perspective. *AIChe J.* 2015;61(2):370–83.
70. Frenkel D, Smit B. *Understanding molecular simulation.* Cambridge, MA: Elsevier; 2002.
71. Senftle TP, Hong S, Islam MM, Kylasa SB, Zheng Y, Shin YK, et al. The ReaxFF reactive force-field: development, applications and future directions. *npj Comput Mater.* 2016;2(1):15011.
72. Meuwly M. Reactive molecular dynamics: from small molecules to proteins. *WIREs Comput Mol Sci.* 2019;9(1):1–22.
73. Meuwly M. Machine learning for chemical reactions. *Chem Rev.* 2021;121(16):10218–39.
74. Tuckerman ME. *Statistical mechanics: theory and molecular simulation.* 2nd ed. Oxford, UK: Oxford Graduate Texts; 2019.
75. Salvalaglio M, Tiwary P, Maggioni GM, Mazzotti M, Parrinello M. Overcoming timescale and finite-size limitations to compute condensation rates at physically relevant conditions. *J Chem Phys.* 2016;145:211925.
76. Diemand J, Angéilil R, Tanaka KK, Tanaka H. Large scale molecular dynamics simulations of homogeneous nucleation. *J Chem Phys.* 2013;139(7):074309.
77. Auer S, Frenkel D. Prediction of absolute crystal-nucleation rate in hard-sphere colloids. *Nature.* 2001;409:1020–3.
78. Valeriani C, Sanz E, Frenkel D. Rate of homogeneous crystal nucleation in molten NaCl. *J Chem Phys.* 2005;122:194501.
79. Allen RJ, Valeriani C, Tanase-Nicola S, ten Wolde PR, Frenkel D. Homogeneous nucleation under shear in a two-dimensional Ising model: cluster growth, coalescence, and breakup. *J Chem Phys.* 2008;129:134704.
80. Schilling T, Schöpe HJ, Oettel M, Opletal G, Snook I. Precursor-mediated crystallization process in suspensions of hard spheres. *Phys Rev Lett.* 2010;105:025701.
81. Haji-Akbari A, Debenedetti PG. Direct calculation of ice homogeneous nucleation rate for a molecular model of water. *Proc Natl Acad Sci.* 2015;112(34):10582–8.
82. Ectors P, Anwar J, Zahn D. Two-step nucleation rather than self-poisoning: an unexpected mechanism of asymmetrical molecular crystal growth. *Cryst Growth Des.* 2015;15(10):5118–23.
83. Lifanov Y, Vorselaars B, Quigley D. Nucleation barrier reconstruction via the seeding method in a lattice model with competing nucleation pathways. *J Chem Phys.* 2016;145(21):211912.
84. Anwar J, Boateng PK. Computer simulation of crystallization from solution. *J Am Chem Soc.* 1998;120:9600–4.
85. Kawska A, Brickmann J, Knier R, Hochrein O, Zahn D. An atomistic simulation scheme for modeling crystal formation from solution. *J Chem Phys.* 2006;124:024513.
86. Agarwal V, Peters B. Solute precipitate nucleation: a review of theory and simulation advances. In: Rice SA, Dinner AR, editors. *Advances in Chemical Physics.* New York: John Wiley & Sons; 2014. p. 97–160.
87. Giberti F, Salvalaglio M, Parrinello M. Metadynamics studies of crystal nucleation. *IUCrJ.* 2015;2(2):256–66.
88. Anwar J, Zahn D. Uncovering molecular processes in crystal nucleation and growth by using molecular simulation. *Angew Chem Int Ed.* 2011;50(9):1996–2013.
89. Zimmermann NE, Vorselaars B, Quigley D, Peters B. Nucleation of NaCl from aqueous solution: critical sizes, ion-attachment kinetics, and rates. *J Am Chem Soc.* 2015;137(41):13352–61.
90. Ectors P, Duchstein P, Zahn D. From oligomers towards a racemic crystal: molecular simulation of dl-norleucine crystal nucleation from solution. *CrstEngComm.* 2015;17:6884–9.
91. Salvalaglio M, Giberti F, Parrinello M. 1,3,5-Tris (4-bromophenyl) benzene prenucleation clusters from metadynamics. *Acta Cryst Sect C.* 2014;70(2):132–6.
92. Salvalaglio M, Perego C, Giberti F, Mazzotti M, Parrinello M. Molecular-dynamics simulations of urea nucleation from aqueous solution. *Proc Natl Acad Sci.* 2015;112(1):E6–E14.
93. Salvalaglio M, Mazzotti M, Parrinello M. Urea homogeneous nucleation mechanism is solvent dependent. *Faraday Discuss.* 2015;179:291–307.
94. Price S, Rimez B, Sun W, Peters B, Christenson H, Hughes C, et al. Nucleation in complex multi-component and multi-phase systems: general discussion. *Faraday Discuss.* 2015;179:503–42.
95. Farkas L. Keimbildungsgeschwindigkeit in übersättigten Dämpfen. *Aus Z Z Phys Chem.* 1927;125U(1):236–42.
96. Volmer M, Weber A. Keimbildung in übersättigten Gebilden. *Z Phys Chem.* 1926;119U(1):277–301.
97. Becker R, Döring W. Kinetische Behandlung der Keimbildung in übersättigten Dämpfen. *Ann Phys.* 1935;416(8):719–52.
98. Sunyaev RA, editor. *10. On the theory of new phase formation.* Cavitation Princeton University Press. 1992;140:120–37.
99. Wedekind J, Reguera D, Strey R. Finite-size effects in simulations of nucleation. *J Chem Phys.* 2006;125(21):214505.
100. Salvalaglio M, Vetter T, Mazzotti M, Parrinello M. Controlling and predicting crystal shapes: the case of urea. *Angew Chem Int Ed.* 2013;52(50):13369–72.
101. Durán-Olivencia MA, Lutsko JF. Mesoscopic nucleation theory for confined systems: a one-parameter model. *Phys Rev E.* 2015;91(2):022402.
102. Perego C, Salvalaglio M, Parrinello M. Molecular dynamics simulations of solutions at constant chemical potential. *J Chem Phys.* 2015;142(14):144113.
103. Ozcan A, Perego C, Salvalaglio M, Parrinello M, Yazaydin O. Concentration gradient driven molecular dynamics: a new method for simulations of membrane permeation and separation. *Chem Sci.* 2017;8(5):3858–65.
104. Karmakar T, Piaggi PM, Parrinello M. Molecular dynamics simulations of crystal nucleation from solution at constant chemical potential. *J Chem Theory Comput.* 2019;15(12):6923–30.

105. Liu C, Wood GP, Santiso EE. Modelling nucleation from solution with the string method in the osmotic ensemble. *Mol Phys.* 2018;116(21–22):2998–3007.
106. Liu C, Cao F, Kulkarni SA, Wood GPF, Santiso EE. Understanding polymorph selection of Sulfamerazine in solution. *Cryst Growth Des.* 2019;19(12):6925–34.
107. Reguera D, Bowles RK, Djikaev Y, Reiss H. Phase transitions in systems small enough to be clusters. *J Chem Phys.* 2003;118(1):340–53.
108. Li L, Paloni M, Finney AR, Barducci A, Salvalaglio M. Nucleation of biomolecular condensates from finite-sized simulations. *J Phys Chem Lett.* 2023;14(7):1748–55.
109. Lutsko JF, Durán-Olivencia MA. Classical nucleation theory from a dynamical approach to nucleation. *J Chem Phys.* 2013;138(24):244908.
110. Karmakar T, Finney AR, Salvalaglio M, Yazaydin AO, Prego C. Non-equilibrium modeling of concentration-driven processes with constant chemical potential molecular dynamics simulations. *Acc Chem Res.* 2023;56(10):1156–67.
111. Han D, Karmakar T, Bjelobrk Z, Gong J, Parrinello M. Solvent-mediated morphology selection of the active pharmaceutical ingredient isoniazid: experimental and simulation studies. *Chem Eng Sci.* 2019;204:320–8.
112. Bjelobrk Z, Piaggi PM, Weber T, Karmakar T, Mazzotti M, Parrinello M. Naphthalene crystal shape prediction from molecular dynamics simulations. *CrstEngComm.* 2019;21(21):3280–8.
113. Praprotnik M, Site LD, Kremer K. Multiscale simulation of soft matter: from scale bridging to adaptive resolution. *Ann Rev Phys Chem.* 2008;59(1):545–71.
114. Espinosa JR, Vega C, Valeriani C, Sanz E. Seeding approach to crystal nucleation. *J Chem Phys.* 2016;144(3):034501.
115. Lamas C, Espinosa J, Conde M, Ramírez J, de Hijes PM, Noya EG, et al. Homogeneous nucleation of NaCl in supersaturated solutions. *Phys Chem Chem Phys.* 2021;23(47):26843–52.
116. Zimmermann NE, Vorselaars B, Espinosa JR, Quigley D, Smith WR, Sanz E, et al. NaCl nucleation from brine in seeded simulations: sources of uncertainty in rate estimates. *J Chem Phys.* 2018;148(22):222838.
117. Pedevilla P, Fitzner M, Sosso GC, Michaelides A. Heterogeneous seeded molecular dynamics as a tool to probe the ice nucleating ability of crystalline surfaces. *J Chem Phys.* 2018;149(7):072327.
118. Espinosa J, Young J, Jiang H, Gupta D, Vega C, Sanz E, et al. On the calculation of solubilities via direct coexistence simulations: investigation of NaCl aqueous solutions and Lennard-Jones binary mixtures. *J Chem Phys.* 2016;145(15):154111.
119. Knott BC, Molinero V, Doherty MF, Peters B. Homogeneous nucleation of methane hydrates: unrealistic under realistic conditions. *J Am Chem Soc.* 2012;134(48):19544–7.
120. Mandal T, Larson RG. Nucleation of urea from aqueous solution: structure, critical size, and rate. *J Chem Phys.* 2017;146(13):134501.
121. Hwang H, Cho YC, Lee S, Lee YH, Kim S, Kim Y, et al. Hydration breaking and chemical ordering in a levitated NaCl solution droplet beyond the metastable zone width limit: evidence for the early stage of two-step nucleation. *Chem Sci.* 2021;12(1):179–87.
122. Garcia NA, Malini RI, Freeman CL, Demichelis R, Raiteri P, Sommerdijk NAJM, et al. Simulation of calcium phosphate prenucleation clusters in aqueous solution: association beyond ion pairing. *Cryst Growth Des.* 2019;19(11):6422–30.
123. Habraken WJEM, Tao J, Brylka LJ, Friedrich H, Bertinetti L, Schenk AS, et al. Ion-association complexes unite classical and non-classical theories for the biomimetic nucleation of calcium phosphate. *Nat Commun.* 2013;4(1):1507.
124. Kollias L, Cantu DC, Tubbs MA, Rousseau R, Glezakou VA, Salvalaglio M. Molecular level understanding of the free energy landscape in early stages of metal–organic framework nucleation. *J Am Chem Soc.* 2019;141(14):6073–81.
125. Kollias L, Rousseau R, Glezakou VA, Salvalaglio M. Understanding metal–organic framework nucleation from a solution with evolving graphs. *J Am Chem Soc.* 2022;144(25):11099–109.
126. Balestra SR, Semino R. Computer simulation of the early stages of self-assembly and thermal decomposition of ZIF-8. *J Chem Phys.* 2022;157(18):184502.
127. Hamad S, Moon C, Catlow CRA, Hulme AT, Price SL. Kinetic insights into the role of the solvent in the polymorphism of 5-fluorouracil from molecular dynamics simulations. *J Phys Chem B.* 2006;110(7):3323–9.
128. Chen J, Trout BL. Computational study of solvent effects on the molecular self-assembly of tetrolic acid in solution and implications for the polymorph formed from crystallization. *J Phys Chem B.* 2008;112(26):7794–802.
129. Chen J, Trout BL. A computational study of the mechanism of the selective crystallization of α - and β -glycine from water and methanol–water mixture. *J Phys Chem B.* 2010;114(43):13764–72.
130. Berziņš A, Semjonova A, Actiņš A, Salvalaglio M. Speciation of substituted benzoic acids in solution: evaluation of spectroscopic and computational methods for the identification of associates and their role in crystallization. *Cryst Growth Des.* 2021;21(9):4823–36.
131. Bobrovs R, Drunka L, Auzins AA, Jaudzems K, Salvalaglio M. Polymorph-selective role of hydrogen bonding and π - π stacking in *p*-aminobenzoic acid solutions. *Cryst Growth Des.* 2021;21(1):436–48.
132. Carpenter JE, Grünwald M. Pre-nucleation clusters predict crystal structures in models of chiral molecules. *J Am Chem Soc.* 2021;143(51):21580–93.
133. Schuitemaker A, Aufort J, Koziara KB, Demichelis R, Raiteri P, Gale JD. Simulating the binding of key organic functional groups to aqueous calcium carbonate species. *Phys Chem Chem Phys.* 2021;23(48):27253–65.
134. Raiteri P, Schuitemaker A, Gale JD. Ion pairing and multiple ion binding in calcium carbonate solutions based on a polarizable AMOEBA force field and ab initio molecular dynamics. *J Phys Chem B.* 2020;124(17):3568–82.
135. Kellermeier M, Raiteri P, Berg JK, Kempter A, Gale JD, Gebauer D. Entropy drives calcium carbonate ion association. *ChemPhysChem.* 2016;17(21):3535–41.

136. Smeets PJM, Finney AR, Habraken WJEM, Nudelman F, Friedrich H, Laven J, et al. A classical view on nonclassical nucleation. *Proc Natl Acad Sci*. 2017;114(38):E7882–90.
137. Finney AR, Rodger PM. Probing the structure and stability of calcium carbonate pre-nucleation clusters. *Faraday Discuss*. 2012;159:47.
138. Finney AR, Innocenti Malini R, Freeman CL, Harding JH. Amino acid and oligopeptide effects on calcium carbonate solutions. *Cryst Growth Des*. 2020;20(5):3077–92.
139. Wallace AF, Hedges LO, Fernandez-Martinez A, Raiteri P, Gale JD, Waychunas GA, et al. Microscopic evidence for liquid–liquid separation in supersaturated CaCO_3 solutions. *Science*. 2013;341(6148):885–9.
140. Raiteri P, Demichelis R, Gale JD. Thermodynamically consistent force field for molecular dynamics simulations of alkaline-earth carbonates and their aqueous speciation. *J Phys Chem C*. 2015;119(43):24447–58.
141. Santiso EE, Trout BL. A general method for molecular modeling of nucleation from the melt. *J Chem Phys*. 2015;143(17):174109.
142. Tribello GA, Giberti F, Sosso GC, Salvalaglio M, Parrinello M. Analyzing and driving cluster formation in atomistic simulations. *J Chem Theory Comput*. 2017;13(3):1317–27.
143. Neha, Tiwari V, Mondal S, Kumari N, Karmakar T. Collective variables for crystallization simulations from early developments to recent advances. *ACS Omega*. 2023;8(1):127–46.
144. Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. PLUMED 2: new feathers for an old bird. *Comput Phys Commun*. 2014;185(2):604–13.
145. Torrie GM, Valleau JP. Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J Comput Phys*. 1977;23(2):187–99.
146. Roux B. The calculation of the potential of mean force using computer simulations. *Comput Phys Commun*. 1995;91(1–3):275–82.
147. Kästner J. Umbrella sampling. *WIREs Comput Mol Sci*. 2011;1(6):932–42.
148. Zwanzig RW. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J Chem Phys*. 1954;22(8):1420–6.
149. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem*. 1992;13(8):1011–21.
150. Shirts MR, Chodera JD. Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys*. 2008;129(12):124105.
151. Kästner J, Thiel W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: “umbrella integration”. *J Chem Phys*. 2005;123(14):144104.
152. Jiang H, Debenedetti PG, Panagiotopoulos AZ. Nucleation in aqueous NaCl solutions shifts from 1-step to 2-step mechanism on crossing the spinodal. *J Chem Phys*. 2019;150(12):124502.
153. Allen RJ, Valeriani C, Ten Wolde PR. Forward flux sampling for rare event simulations. *J Phys Condens Matter*. 2009;21(46):463102.
154. Laio A, Parrinello M. Escaping free-energy minima. *Proc Natl Acad Sci*. 2002;99(20):12562–6.
155. Barducci A, Bussi G, Parrinello M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys Rev Lett*. 2008;100(2):020603.
156. Bonomi M, Parrinello M. Enhanced sampling in the well-tempered ensemble. *Phys Rev Lett*. 2010;104(19):190601.
157. Bussi G, Gervasio FL, Laio A, Parrinello M. Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics. *J Am Chem Soc*. 2006;128(41):13435–41.
158. Barducci A, Bonomi M, Parrinello M. Metadynamics. *WIREs Comput Mol Sci*. 2011;1(5):826–43.
159. Valsson O, Tiwary P, Parrinello M. Enhancing important fluctuations: rare events and metadynamics from a conceptual viewpoint. *Annu Rev Phys Chem*. 2016;67:159–84.
160. Tiwary P, Parrinello M. A time-independent free energy estimator for metadynamics. *J Phys Chem B*. 2014;119(3):736–42.
161. Giberti F, Salvalaglio M, Mazzotti M, Parrinello M. 1,3,5-Tris(4-bromophenyl)-benzene nucleation: from dimers to needle-like clusters. *Cryst Growth Des*. 2017;17(8):4137–43.
162. Zou Z, Beyerle ER, Tsai ST, Tiwary P. Driving and characterizing nucleation of urea and glycine polymorphs in water. *Proc Natl Acad Sci*. 2023;120(7):e2216099120.
163. Giberti F, Tribello GA, Parrinello M. Transient polymorphism in NaCl. *J Chem Theory Comput*. 2013;9(6):2526–30.
164. Lauricella M, Meloni S, English NJ, Peters B, Ciccotti G. Methane clathrate hydrate nucleation mechanism by advanced molecular simulations. *J Phys Chem C*. 2014;118(40):22847–57.
165. Ahlwat P, Dar MI, Piaggi P, Gratzel M, Parrinello M, Rothlisberger U. Atomistic mechanism of the nucleation of methylammonium lead iodide perovskite from solution. *Chem Mater*. 2019;32(1):529–36.
166. Arjun BTA, Bolhuis PG. Unbiased atomistic insight in the competing nucleation mechanisms of methane hydrates. *Proc Natl Acad Sci*. 2019;116(39):19305–10.
167. Arjun A, Bolhuis PG. Rate prediction for homogeneous nucleation of methane hydrate at moderate supersaturation using transition interface sampling. *J Phys Chem B*. 2020;124(37):8099–109.
168. Hall SW, Díaz Leines G, Sarupria S, Rogal J. Practical guide to replica exchange transition interface sampling and forward flux sampling. *J Chem Phys*. 2022;156(20):200901.
169. Jiang H, Haji-Akbari A, Debenedetti PG, Panagiotopoulos AZ. Forward flux sampling calculation of homogeneous nucleation rates from aqueous NaCl solutions. *J Chem Phys*. 2018;148(4):044505.
170. Bolhuis PG, Chandler D, Dellago C, Geissler PL. Transition path sampling: throwing ropes over rough mountain passes, in the dark. *Annu Rev Phys Chem*. 2002;53(1):291–318.
171. Van Erp TS, Bolhuis PG. Elaborating transition interface sampling methods. *J Comput Phys*. 2005;205(1):157–81.

172. Van Erp TS, Moroni D, Bolhuis PG. A novel path sampling method for the calculation of rate constants. *J Chem Phys.* 2003;118(17):7762–74.
173. Haji-Akbari A. Forward-flux sampling with jumpy order parameters. *J Chem Phys.* 2018;149(7):072303.
174. Sidky H, Colón YJ, Helfferich J, Sikora BJ, Bezik C, Chu W, et al. SSAGES: software suite for advanced general ensemble simulations. *J Chem Phys.* 2018;148(4):044104.
175. Peters B. Reaction coordinates and mechanistic hypothesis tests. *Ann Rev Phys Chem.* 2016;67:669–90.
176. Bonati L, Piccini G, Parrinello M. Deep learning the slow modes for rare events sampling. *Proc Natl Acad Sci.* 2021;118(44):e2113533118.
177. Elishav O, Podgaetsky R, Meikler O, Hirshberg B. Collective variables for conformational polymorphism in molecular crystals. *J Phys Chem Lett.* 2023;14(4):971–6.
178. Rogal J. Reaction coordinates in complex systems—a perspective. *Eur Phys J B.* 2021;94:1–9.
179. Steinhardt PJ, Nelson DR, Ronchetti M. Bond-orientational order in liquids and glasses. *Phys Rev B.* 1983;28(2):784–805.
180. Lechner W, Dellago C. Accurate determination of crystal structures based on averaged local bond order parameters. *J Chem Phys.* 2008;129(11):114707.
181. Santiso EE, Trout BL. A general set of order parameters for molecular crystals. *J Chem Phys.* 2011;134(6):064109.
182. Gobbo G, Bellucci MA, Tribello GA, Ciccotti G, Trout BL. Nucleation of molecular crystals driven by relative information entropy. *J Chem Theory Comput.* 2018;14(2):959–72.
183. Gimondi I, Salvalaglio M. CO₂ packing polymorphism under confinement in cylindrical nanopores. *Mol Syst Des Eng.* 2018;3(1):243–52.
184. Piaggi PM, Parrinello M. Predicting polymorphism in molecular crystals using orientational entropy. *Proc Natl Acad Sci.* 2018;115(41):10251–6.
185. Marinova V, Dodd L, Lee SJ, Wood GP, Marziano I, Salvalaglio M. Identifying conformational isomers of organic molecules in solution via unsupervised clustering. *J Chem Inf Model.* 2021;61(5):2263–73.
186. Francia NF, Price LS, Nyman J, Price SL, Salvalaglio M. Systematic finite-temperature reduction of crystal energy landscapes. *Cryst Growth Des.* 2020;20(10):6847–62.
187. Francia NF, Price LS, Salvalaglio M. Reducing crystal structure overprediction of ibuprofen with large scale molecular dynamics simulations. *CrstEngComm.* 2021;23(33):5575–84.
188. Zahn D. Atomistic mechanism of NaCl nucleation from an aqueous solution. *Phys Rev Lett.* 2004;92(4):040801.
189. Joung IS, Cheatham TE. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J Phys Chem B.* 2008;112(30):9020–41.
190. Berendsen HJC, Grigera JR, Straatsma TP. The missing term in effective pair potentials. *J Phys Chem.* 1987;91(24):6269–71.
191. Smith WR, Nezbeda I, Kolafa J, Moučka F. Recent progress in the molecular simulation of thermodynamic properties of aqueous electrolyte solutions. *Fluid Phase Equilibria.* 2018;466:19–30.
192. Nezbeda I, Moučka F, Smith WR. Recent progress in molecular simulation of aqueous electrolytes: force fields, chemical potentials and solubility. *Mol Phys.* 2016;114(11):1665–90.
193. Dočkal J, Lísal M, Moučka F. Molecular force field development for aqueous electrolytes: 2. Polarizable models incorporating crystal-line chemical potential and their accurate simulations of halite, hydrohalite, aqueous solutions of NaCl, and solubility. *J Chem Theory Comput.* 2020;16(6):3677–88.
194. Aragonés J, Sanz E, Vega C. Solubility of NaCl in water by molecular simulation revisited. *J Chem Phys.* 2012;136(24):244508.
195. Benavides A, Portillo M, Abascal J, Vega C. Estimating the solubility of 1:1 electrolyte aqueous solutions: the chemical potential difference rule. *Mol Phys.* 2017;115(9–12):1301–8.
196. Moučka F, Nezbeda I, Smith WR. Molecular simulation of aqueous electrolytes: water chemical potential results and Gibbs–Duhem equation consistency tests. *J Chem Phys.* 2013;139(12):124505.
197. Moučka F, Nezbeda I, Smith WR. Molecular force field development for aqueous electrolytes: 1. Incorporating appropriate experimental data and the inadequacy of simple electrolyte force fields based on Lennard–Jones and point charge interactions with Lorentz–Berthelot rules. *J Chem Theory Comput.* 2013;9(11):5076–85.
198. Moučka F, Nezbeda I, Smith WR. Chemical potentials, activity coefficients, and solubility in aqueous NaCl solutions: prediction by polarizable force fields. *J Chem Theory Comput.* 2015;11(4):1756–64.
199. Moučka F, Lísal M, Škvor J, Jirsák J, Nezbeda I, Smith WR. Molecular simulation of aqueous electrolyte solubility. 2. Osmotic ensemble Monte Carlo methodology for free energy and solubility calculations and application to NaCl. *J Phys Chem B.* 2011;115(24):7849–61.
200. Mester Z, Panagiotopoulos AZ. Mean ionic activity coefficients in aqueous NaCl solutions from molecular dynamics simulations. *J Chem Phys.* 2015;142(4):044507.
201. Anderson MW, Bennett M, Cedeno R, Cölfen H, Cox SJ, Cruz-Cabeza AJ, et al. Understanding crystal nucleation mechanisms: where do we stand? *Gen Discuss Faraday Discuss.* 2022;235:219–72.
202. Benavides AL, Aragonés JL, Vega C. Consensus on the solubility of NaCl in water from computer simulations using the chemical potential route. *J Chem Phys.* 2016;144(12):124504.
203. Girshick SL, Chiu CP. Kinetic nucleation theory: a new expression for the rate of homogeneous nucleation from an ideal supersaturated vapor. *J Chem Phys.* 1990;93(2):1273–7.
204. Chkonia G, Wölk J, Strey R, Wedekind J, Reguera D. Evaluating nucleation rates in direct simulations. *J Chem Phys.* 2009;130(6):064505.

205. Bulutoglu PS, Wang S, Boukerche M, Nere NK, Corti DS, Ramkrishna D. An investigation of the kinetics and thermodynamics of NaCl nucleation through composite clusters. *PNAS Nexus*. 2022;1(2):pgac033.
206. Finney AR, McPherson IJ, Unwin PR, Salvalaglio M. Electrochemistry, ion adsorption and dynamics in the double layer: a study of NaCl (aq) on graphite. *Chem Sci*. 2021;12(33):11166–80.
207. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. *J Comput Chem*. 2004;25(13):1656–76.
208. Blow KE, Quigley D, Sosso GC. The seven deadly sins: when computing crystal nucleation rates, the devil is in the details. *J Chem Phys*. 2021;155(4):040901.

AUTHOR BIOGRAPHIES

Aaron R. Finney is a Lecturer in Materials Science at The University of Liverpool. He completed his PhD in the Centre for Scientific Computing and Department of Chemistry, University of Warwick. Following this, he was a Doctoral Prize Fellow and Research Fellow in the Materials Science Department, University of Sheffield, and a Senior Research Fellow in the Department of Chemical Engineering, University College London. His current research focuses on interfaces and crystallization.

Matteo Salvalaglio is a Professor of Molecular Systems Engineering at University College London, member of the Thomas Young Centre and the Sargent Centre for Process Systems Engineering. He received his PhD from Politecnico di Milano (Italy). Before joining UCL in 2015, MS was a postdoctoral researcher at ETH Zurich (CH) in a joint appointment between Chemistry and Mechanical Engineering. MS current research focuses on using computational methods to understand nucleation, polymorphism, and crystal growth.

How to cite this article: Finney AR, Salvalaglio M. Molecular simulation approaches to study crystal nucleation from solutions: Theoretical considerations and computational challenges. *WIREs Comput Mol Sci*. 2023. e1697. <https://doi.org/10.1002/wcms.1697>