ADVANCE REVIEW
WIREs

Theoretical and computational approaches to study crystal nucleation from solution

Aaron R. Finney[1] | **Matteo Salvalaglio**[1]

[1]Department of Chemical Engineering, University College London

Correspondence
Matteo Salvalaglio, Department of Chemical Engineering, University College London, London, United Kingdom
Email: m.salvalaglio@ucl.ac.uk

Funding information
EPSRC, UKRI Frontier Research Guaranteee Grant, EP/X033139/1; EPSRC, Crystallization in the Real World Programme Grant, EP/R018820/1

Nucleation is the initial step towards the formation of crystalline materials from solutions. Various factors, such as environmental conditions, additives, and external forces, can influence its outcomes and rates. Indeed, controlling this rate-determining step towards phase separation can affect the material structure and properties, and it is crucial in a range of scientific fields. In this regard, atomistic simulation methods 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 affects the phase behaviour of nucleating solutions when compared to macroscopic systems. Additionally, a challenge in modelling nucleation from solution is associated with the inadequacy of standard molecular simulations to access the timescales necessary to observe crystal nucleation due to the inherent rareness of these events. In recent decades, simulation methods have emerged to circumvent length- and timescale limitations. However, which simulation method is most suitable for studying crystal nucleation from solution is not always obvious. This review summarises the recent advances in this field, providing an overview of the typical nucleation mechanisms and the suitability of different simulation methods to study them. By doing so, we aim to provide a deeper understanding of the complexities associated with modelling crystal nucleation from solution and identify areas for further research. Our review targets researchers across various scientific fields, including materials science, chemistry, physics and engineering, and will hopefully contribute to developing new strategies for understanding and controlling nucleation.

KEYWORDS
Nucleation, Molecular Simulations, Molecular Dynamics, Enhanced Sampling, Rare Events, Crystallization

Abbreviations: CNT, Classical Nucleation Theory. HON, Homogeneous Nucleation. HEN Heterogeneous Nucleation. PNC, Pre Nucleation Clusters. MD, Molecular Dynamics. API, Active Pharmaceutical Ingredient. US, Umbrella Sampling. WTMetaD, Well-Tempered Metadynamics, TPS, Transition Path Sampling, FFS, Forward Flux Sampling. CV, Collective Variable. MFPT, Mean First-Passage Time. BF, Brute Force (MD simulation). MLD, Modified Liquid Droplet. CμMD, Constant Chemical Potential Molecular Dynamics.

CONTENTS

1 Introduction: why Simulate Crystal Nucleation from Solution? 3

2 Classifying nucleation from solution. 4

3 Dynamic Simulations of Crystallization from solution: Time- and Length-scale Challenges 7
3.1 Timescale limitations . . . 8
3.2 Finite-size dependence of the crystallization driving force . . . 10

4 Modelling approaches to simulate nucleation from solution. 12
4.1 Informing Theory with Molecular Simulations . . . 13
4.1.1 The Seeding Method . . . 13
4.1.2 Prenucleation Species . . . 15
4.2 Molecular Simulations as Computational Experiments . . . 16
4.2.1 Biased enhanced sampling approaches . . . 16
4.2.2 Unbiased enhanced sampling approaches . . . 19
4.2.3 Collective Variables . . . 21

5 Comparing Approaches: NaCl nucleation from aqueous solution 21

6 Perspectives and Conclusions 26

7 graphical abstract 39

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, e.g., construction, pharmaceutical manufacturing, separations, catalysis, and organic electronics. For instance, most active pharmaceutical ingredients (APIs) are formulated as crystalline solids. [1, 2, 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; these include accurate dispersion models in electronic structure calculations[4] and the introduction of quantum effects[5] to calculate lattice energies; the application of machine learning methods to accurately estimate the thermodynamic stability of crystal polymorphs [6]; and the development of novel approaches for the calculation of relative lattice free energies[7, 8, 9, 10, 11, 12]. Establishing the relative stability of crystal polymorphs using computational calculations is extremely valuable in identifying those which are thermodynamically plausible. This has proven particularly useful for the development of APIs, and it is finding increasing applications outside academia [13, 14, 15, 16, 17, 18, 19]. These methods, however, fail to capture the role played by components in a preceding crystal-forming fluid phase in directing phase separation towards a specific outcome. In the case of crystallization from solution, the parent liquid often has a composition radically different to 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 favourable 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, 21, 22, 23, 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, 26, 27] Because the dynamics of 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. The small time scales probed by MD simulations are generally inadequate for studying the complex microscopic steps involved in crystal

nucleation and growth. As we highlight below, however, advanced sampling schemes and theoretical treatments can 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 with studying different aspects of crystallisation 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 modelling 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 (e.g. seeding) with rare events simulations 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). Our review complements the overview from Agarwal et al. [28], which delves into the theoretical background of rate theories applied to nucleation problems, as well as the review by Sosso et al. [29] 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 characterised by a different composition than the parent phase, and that the driving force of the process depends on solute concentration. After defining the remits 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 that enable modelling nucleation from solution with atomistic detail. Finally, we provide a critical comparison of the insight obtained from papers that investigate the nucleation of NaCl(s) from an aqueous solution: a problem that has been tackled with 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.

Primary vs 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.

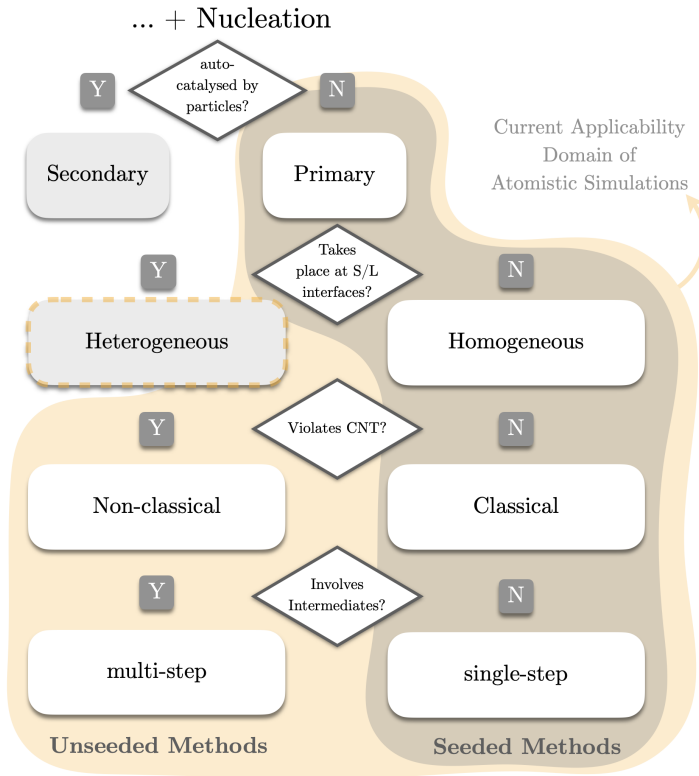

FIGURE 1 To gain insight into nucleation mechanisms and evaluate rates, an appropriate simulation methodology must be chosen to investigate the system at hand. This flow diagram identifies where theory-based simulation methods (based on CNT and its extensions) are most suitable. In particular we highlight that secondary nucleation, typically involves interactions of crystalline particles with the fluid at a scale inaccessible to molecular simulations and, with some notable exceptions [30] 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 are *in principle* applicable examples exist only for crystal nucleation from the melt [31].

Homogeneous vs 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 and impurities may affect the nucleation behaviour. 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 crystallisation 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 i.e. by the recent works on heterogeneous ice nucleation [32, 33, 34, 35, 36],

examples of their application to HEN from solution are currently lacking. As such in Fig. 1 we report HEN as*on the edge* of the applicability domain for molecular simulations of nucleation from solution.

Single-step vs 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 (i.e., the induction time is associated with a single bottleneck towards crystallization). In contrast, multi-step nucleation encompasses a variety of crystallization pathways, which can involve the formation of intermediate precipitate phases, such as amorphous liquids or 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 [37, 38].

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 formed during the first phase transformation in solution. [39, 40] In this case, there are often two bottlenecks to crystallisation; [41, 42, 43] 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. [44, 45, 46] In other words, multi-step nucleation does not necessarily refer to a cascade of single-step nucleation steps.

The process of nucleation in complex solutions may be dependent upon chemical reactions and changes in species stoichiometry occurring locally. Several studies [47, 48, 49] have attributed the multi-step nature of nucleation to these factors. Simulating reactive crystallization events is often beyond the scope of classical molecular simulations. 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.

Classical vs. Non-Classical Nucleation

In section 3.1 we briefly describe CNT, the thermodynamic and kinetic framework for *classical* nucleation. For the purposes of classifying nucleation pathways, we stress that classical nucleation is a 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. [50] 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 which depart from CNT include a surface tension term that is dependent on the cluster size, $\gamma(n)$, which in equation 1 would be included in $\sigma'(n)$.[51, 52, 53] 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. [54] 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. [55]

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 [56, 45]. 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 solutions [45]: 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 *non-classical*. 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 [57, 58, 59, 60, 61].

A paradigmatic example of non-classical nucleation worth mentioning is the *prenucnleation cluster pathway*. The prenucleation cluster (PNC) pathway was first proposed for $CaCO_3$ [62] but has since been attributed to phase separation in a diverse range of systems. [63] 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 $(\text{monomer})x \underset{\text{-monomer}}{\overset{\text{+monomer}}{\rightleftharpoons}} (\text{monomer})_{x+1}$, which is assumed to be equally independent of the value of x. PNCs are highly dynamic, evolving their structure and topology over very short timescales (typically picoseconds).[64] 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. [65] This rate-determining step typically produces a dense liquid phase which 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

Gathering 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, modelling techniques based on MD simulations inherently provide insight into the time evolution of systems with atomistic detail,[66] making them a powerful tool for understanding complex processes such as nucleation.

In MD, the forces that determine atomic motion are modelled using classical mechanics [67]. Here, chemical bonds are typically approximated as classical 'springs' whose displacement from an equilibrium bond distance is modelled 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 is not

well-suited to simulate chemical reactions and dynamical changes to chemical speciation, though reactive force fields [68] and the advent of machine learning methods to force field design [69, 70] make this a possibility. Particularly in the case of ionic systems, polarisability can be simulated using a relatively cheap treatment of charge displacement from atomic nuclei, e.g., 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. [67, 71] 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 crystallisation, 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 modelled 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 modelling 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 [72, 73]. 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 [44, 74, 75, 76, 77] and towards the in-depth investigation of nucleation in monocomponent molecular systems, such as pure water [78]. In recent years, an evolution towards systems with increased complexity and practical relevance has begun, enabling e.g., the study of solute precipitate nucleation. These tend to focus on model systems [79, 28, 80], such as two-component Potts-lattice models, or inorganic systems such as NaCl(aq) [81, 82, 83, 84, 85, 31, 86]. Attempts to simulate organic solids nucleating from solution have been successful in highlighting the structural features of early-stage crystallization precursors[87, 88] and to extract qualitative information on nucleation mechanisms.[89, 90] Significant progress was also made in simulating crystal growth from solution by combining MD simulations, Kinetic Monte Carlo, and enhanced sampling [91, 92, 93, 94].

Origin of the timescale separation in nucleation: Nucleation Free Energy Barriers and Classical Nucleation Theory.

Classical Nucleation Theory (CNT) provides a quasi-mechanistic description of the nucleation process, connecting crystallization equilibrium thermodynamics and nucleation kinetics.[50] First applied by Gibbs 150 years ago to describe the formation of liquid droplets from saturated vapours, CNT was developed over the following century and is now often applied 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 favourable, 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, 95].

The size of the nucleus was originally described using a linear descriptor, i.e., 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:

$$F(n) = -n\Delta\mu_{\ell\to xtal} + \sigma' n^{2/3} \tag{1}$$

where $\Delta\mu_{\ell\to xtal} = \mu_{\ell} - \mu_{xtal}$ 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 favourable, and the system is *supersaturated*. The surface term instead scales as $n^{2/3}$, where σ' is the product of a shape factor and the surface tension, σ.

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^* can be computed as:

$$F(n^*) = \frac{4\sigma'^3}{27\Delta\mu^2_{\ell\to xtal}} = \frac{1}{2}\Delta\mu_{\ell\to xtal} n^* \tag{2}$$

where the *critical* size of the crystal nucleus is

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

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, therefore, that a single energy barrier separates the parent liquid from a bulk crystal in equilibrium with a solution. It also assumes that the interfacial energy for the smallest clusters of monomers representing the new phase can be described by scaling the surface energy of the bulk phase at equilibrium. This means that the surface tension and the shape factor equating σ' are unchanging as a function of n. This is the so-called capillary approximation which likely fails for the smallest crystals whose shape and faceting are expected to deviate from that of the bulk crystal. [38] Extensions of the theory can be made to account for varying surface tension.[51, 52, 53]

The crystal embryo growth is assumed to occur by the transfer of *monomers* from the solution to the crystal. [96] This is justified by considering the probability of finding clusters of size n: $p(n) \propto \exp(-\Delta F(n)/k_{\mathrm{B}}T)$, 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: [97, 98]

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

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). [99, 50] Due to the exponential term in Eq. 4, small changes to $\Delta F(n^*)$, as determined by $\Delta\mu_{\ell\to xtal}$ and σ', can result in orders of magnitude changes to crystal nucleation rates.

Following units for the rate as $\mathrm{m}^{-3}\ \mathrm{s}^{-1}$, 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 $J = 10^{20}\ \mathrm{m}^{-3}\ \mathrm{s}^{-1}$. 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 one hour. 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, $\Delta\mu_{\ell\to xtal}$. [100, 101, 28, 89, 72] 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.[89, 72, 102] 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 $\Delta\mu_{\ell\to xtal}$ or the development of specialised methods to mimic open boundary conditions,[103, 104, 105, 106, 107] unless sufficiently large systems can be simulated to minimise the finite-size effects.

Nucleation free energy in small systems

A rationale for the effect of confinement on the phase behaviour of metastable liquids is provided by developing a model for crystal nucleation in confined volumes analogous to the modified liquid droplet (MLD) mode developed by Reguera et al. [108, 100] to describe depletion effects on the thermodynamics of nucleation of liquid droplets [109, 89] Consider a two-component supersaturated solution, ℓ. The chemical potential of solute in solution, i, can be

written as,

$$\mu_\ell = \mu_0 + k_B T \ln a^i \approx \mu_0 + k_B T \ln x^i \tag{5}$$

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 behaviour. When a crystal forms, the chemical potential of this phase equals the chemical potential of the solution at equilibrium, such that,

$$\mu_{xtal} = \mu_0 + k_B T \ln a^{i,*} \approx \mu_0 + k_B T \ln x^{i,*} \tag{6}$$

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 crystallisation is not diffusion limited). In a simulation, however, where the total number of monomers is fixed (N), the depletion changes the driving force for crystallisation and must be accounted for in an additional term to Eq. 1:

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

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. On the other hand, one could take advantage of the limitations imposed by finite size by performing simulations initiated with different N and V and then analysing the steady-state properties of the system obtained from simulations to evaluate $\Delta F(n)$. For example, Li et al.[110] 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 Eq. 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 [89, 90].

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 crystallisation step, making precise estimates of σ difficult. Based on the local density, alternative formulations of CNT have also been constructed, [111] and the finite size effects can be included here to account for the changing thermodynamic driving force. [102]

Simulating condensed matter systems with pseudo open boundaries

A limited number of simulation strategies have been developed to avoid the effects of solute depletion that shift the thermodynamic driving force for steady-state crystallisation in simulations. The constant chemical potential MD method (CμMD)[112, 113, 114] 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, e.g., urea,[103] isoniazid [115] and naphthalene, [116] which demonstrate crystal facet and solvent dependent growth rates. In addition, crystal nucleation can be studied using a spherical variant of CμMD that was applied to simulate NaCl crystallisation. [113]

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. [117]. This was successfully applied to understand the nucleation and polymorph selection of sulfamerazine. [107] Unlike CμMD, the reaction coordinate for crystallisation 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. [118] 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 modelled 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 modelled with high resolution. There are no known examples of applying such approaches to study crystallisation. Both the coarse-grained model and the method by which the molecule representations are transformed must be performed carefully to ensure the properties of the system are maintained.

4 | MODELLING APPROACHES TO SIMULATE NUCLEATION FROM SOLUTION.

As highlighted in Sections 2 and 3, modelling 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 utilised 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 the variety of such methods in relation to their reliance on CNT and its underpinning assumptions.

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, prioritising 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

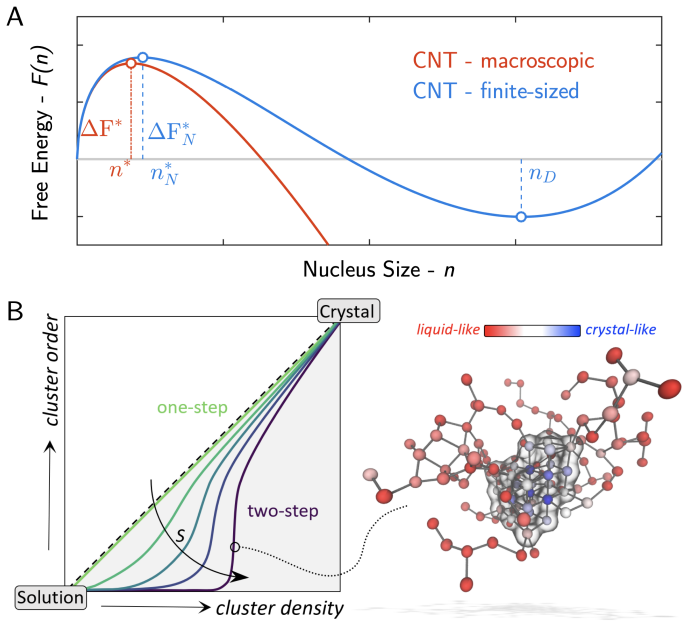

FIGURE 2 A) The finite size effects of closed simulations manifest in a free energy minimum for the stable thermodynamic phase resulting from nucleation. This is different to the macroscopic case due to a bounded partition function and a depletion of crystal building blocks that transfer from the parent phase to the nucleating phase. B) Crystal nucleation pathways from solution to crystal represented on a two-dimensional reaction coordinate characterising 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 which 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, is increased far into the metastable zone for phase separation.

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 facilitates an assessment of the suitability of these approaches to investigate crystal nucleation *in silico* more generally.

4.1 | Informing Theory with Molecular Simulations

4.1.1 | The Seeding Method

In the seeding method, introduced and popularised by the Espinosa, Vega, Valeriani, Sanz [119, 120], Quigley[80], and Peters [86, 121] 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. [121], 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).

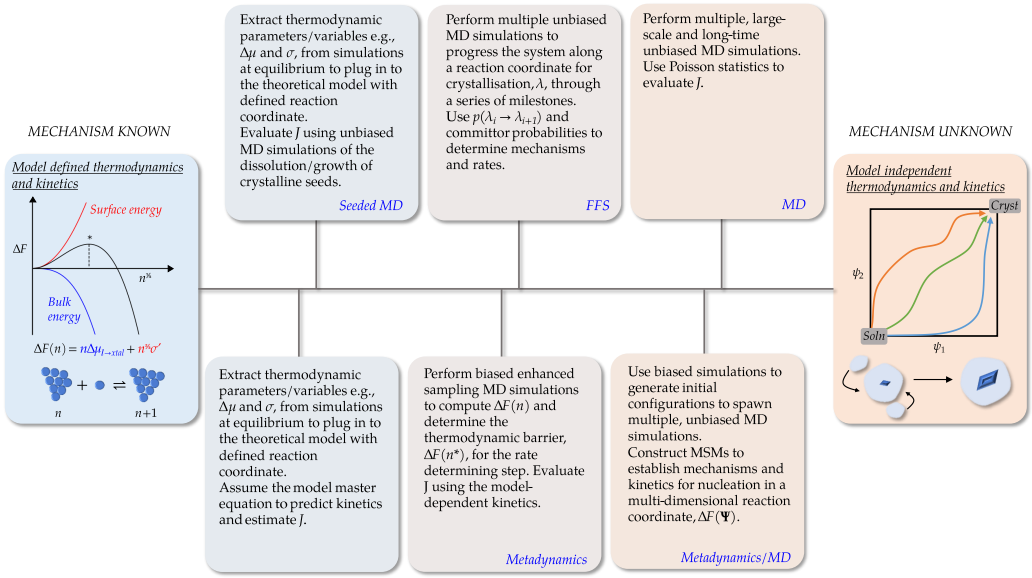

FIGURE 3 The range of molecular simulation techniques that can be applied to study nucleation from solution goes from completely theory based (left side) to exploration driven computational experiments (right side). The insight available and its associated computational cost vary significantly across 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 the simulation method of choice.

The key expression in the seeding method is the CNT nucleation rate, i.e. Eq. 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 Eq. 4 [119, 121, 120].

To solve the rate equations, the parameters that need to be estimated at a given T are the solute density ρ, the thermodynamic driving force $\Delta\mu_{\ell\to xtal}$, 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 $\Delta\mu_{\ell\to xtal}$, which has to be independently obtained at the composition of interest.

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 initialised 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 simulation campaigns for 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 over-damped Langevin dynamics; Zimmerman et al. [109, 121] have shown that the attachment frequency is limited by the desolvation process. Under this condition $f^+ = f^+_{des}$:

$$f^+_{des} = 4\pi k_s \sigma_s \rho \tag{8}$$

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.[122]:

$$f^+_{AF} = \frac{\langle (n(t) - n(0))^2 \rangle}{2t} \tag{9}$$

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[120], 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. [80].

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 Eq. 4, while enhanced sampling methods are used to independently compute ΔF^*[105].

Notable applications of the seeding method to the investigation of nucleation from solutions include, i.e., the study of the nucleation of methane hydrates [123], and the study of urea nucleation from aqueous solution [124].

4.1.2 | Prenucleation Species

Many simulations 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 behaviour 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. [64, 46, 125]

For example, Demichelis et al.[126] 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.[127] Simulations have also shed light on structural building units that act as precursors in the nucleation of metal-organic frameworks [128, 129, 130] and the complexes that assemble to form inorganic functional materials.

It was hypothesised 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. [131, 132, 133, 134, 135] In simple 2D models of flexible chiral molecules, Carpenter and Grünwald [136] recently demonstrated how bulk crystal structures are related to the organisation 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 phase separate following the PNC pathway and other nonclassical crystallisation routes. In particular, many studies have considered the structure and stability of $CaCO_3$ assemblies that emerge in solution and the effects of additives and solution environment on their properties. [64, 137, 138, 139, 140, 141, 142, 143] These studies have demonstrated that entropy

drives the formation of PNCs [139] and questioned the realtionship between PNCs and microscopic precursors to dense liquids [140, 143].

Predicting the phase behaviour in systems that follow PNC separation requires evaluation of the equilibrium constant for monomer association, as discussed in section 2. [65] Enhanced sampling simulations, particularly Umbrella Sampling and metadynamics, [142, 144, 138] have been successfully applied to evaluate these constants which corroborate experimental measurements. [139] 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. [64, 141, 143] 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 classical nucleation theory. 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 clear solution where a nucleus is absent. As such, in unseeded simulations, the mechanism of nucleation 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 behaviour 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 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 the free energy barrier information 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 is only practical in simple cases [72].

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-dependent 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

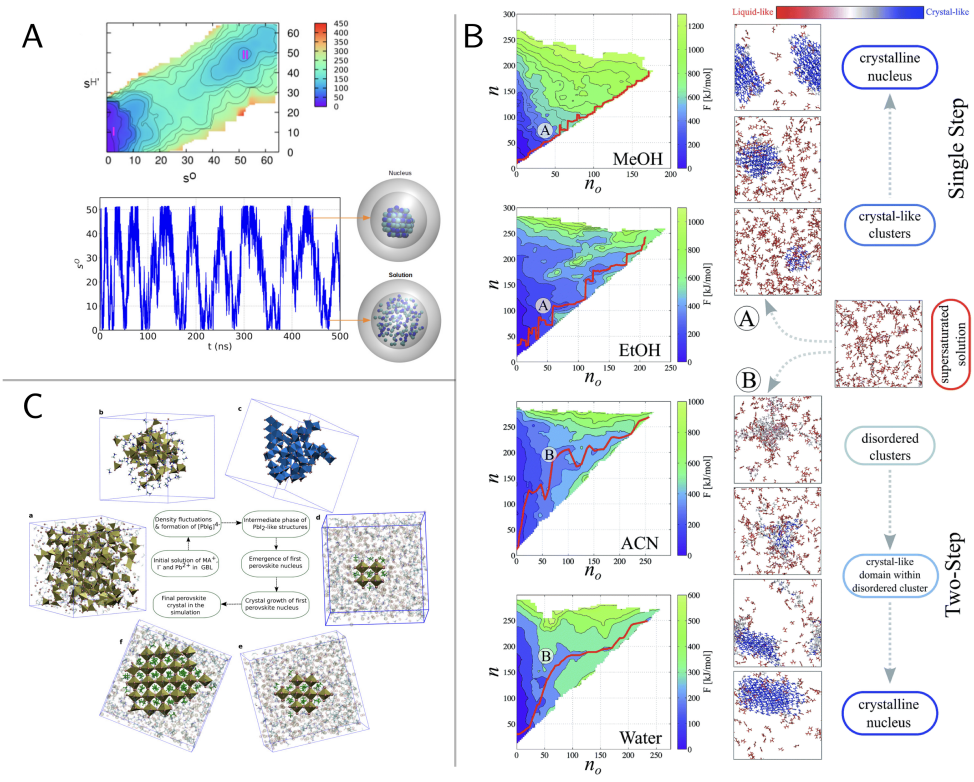

FIGURE 4 MetaD simulations of crystal nucleation from solutions. A) Combining CμMD [105, 114] and WTMetaD Karmakar et al compute the free energy surface associated to NaCl nucleation from solution limiting the finite-size effects due to confinement. Here the CV used enhanced the sampling of crystal structures following with a rocksalt structure, and the method was focussed on the calculation of the free energy barrier to nucleation. Adapted from Ref. [105]. 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. Adapted from Ref. [90]. C) Metadynamics enables the discovery of a two-step nucleation mechanism for the synthesis of Methylammonium Lead Iodide Perovskites from solution. Adapted from Ref [145].

do not report on sampling methods based on constrained dynamics such as the string method in CV space, used by Santiso et al. to model crystallization both from the melt [146] and from solution [117], and we refer the interested reader to the original publications for an overview of this method, related to US and MetaD. Biased enahnced sampling methods depend crucially on the choice of low-dimensional descriptors of the system configuration - i.e. the collective variables (CVs) - that in biased sampling are used to define the bias potential [147, 43**?**]. 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. [84] and Neha et al. [148]. 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 software[149]

which interfaces with most MD engines and offers a wide range of CVs, including the ones applicable 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 their favourite 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[150, 151, 152]. 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 using Zwanzig perturbation theory[153], and a global free energy profile is then obtained by employing either the weighted histogram analysis (WHAM) [154, 151] the multiple Bennett Acceptance Ratio (mBAR) [155], or Umbrella Integration (UI) [156]. 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, i.e. ΔF^*, without resorting to a theoretical formulation of the barrier dependence on thermodynamic parameters such as the solubility and surface tension [157]. 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 expressions [122] 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 [158, 109].

Metadynamics

Metadynamics (MetaD) is a molecular simulation technique to study the thermodynamics and mechanism of rare events [159, 160]. 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. This helps uncover poorly-explored regions of configuration space that might not be easily accessible using traditional molecular dynamics simulations. 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.[150, 159, 161, 162] In MetaD such a potential is adaptively constructed as a sum of Gaussian kernels defined in a low-dimensional set of collective variables (CVs), usually indicated as $\mathbf{s}$. CVs are formulated as continuous and differentiable functions of the microscopic coordinates of a system.[160, 163, 84, 164] In recent years, several adaptations of MetaD have been proposed, the most relevant being Well-Tempered metadynamics (WTmetaD) introduced by Barducci et al. [160] In WTmetaD, the external repulsive potential is iteratively updated as:

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

where $n1$ 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 homogeneous 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 behaviour. 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: $V(\mathbf{s}, t \to \infty) = -\frac{\Delta T}{T+\Delta T} F(\mathbf{s}) + c(t)$, where $F(\mathbf{s})$ is the free energy in the set collective variables $\mathbf{s}$, and $c(t)$ is a time-dependent constant [165].

In the context of the study of nucleation processes *from solution*, metadynamics has been applied to explore the configurational landscapes associated with the nucleation of small organic molecules from liquid solutions [89, 90, 88, 166, 167], the nucleation of salts from aqueous solution [168, 105, 46], the nucleation of methane clathrates [169], and the assembly of perovskites [145]. 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[170, 171] and Forward Flux Sampling [172, 158] that has recently been used to simulate NaCl nucleation from aqueous solution at moderate supersaturations[173] (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 i.e., by slightly modifying the momenta, thus generating a new trajectory via a backward and forward propagation in time. If the new trajectory, proceeding through the set of configurations x in n steps, connects reactants and products, it is accepted with a path weight that depends on the equilibirum phase space probability density of the initial point in the trajectory $\rho(x_0)$ as:

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

where $p(x_i \to x_{i+1})$ 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 Refs. [174, 158, 175] 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 as[171]:

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

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 probabilty 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 [176, 175]), have been 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 $\lambda_{i+1} > \lambda_{i+1}$ 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 Eq: 12. However, it differs in the computational approach adopted for the calculation of the term $P(\lambda_{i+1}|\lambda_i)$, a key difference that provides the attibute *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}.[76] Different implementations of the FFS algorithm differ in the specifics of the algorithm implemented to generate configurations at interfaces and thus in the details associated with the calculation of the $P(\lambda_{i+1}|\lambda_i)$, and additional information can be found in the original publications and in a number of reviews covering the specifics 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[177]. Furthermore, Hall et al. [172] 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 intersections 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 done with all MD packages and using script to automate the spawning and analysis of trajectories. Of course, the CV must be implemented in order to use the method. In this regard, SSAGES [178] 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 ...) that are inherently equivalent and that come together to form a nucleus of a new phase characterised 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 characterising 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 functions of the atomic coordinates that are used to define the bias potential added to the Hamiltonian in biased enhanced sampling or that are used to mark the progress between reactants and products in unbiased enhanced sampling and MD. In this context, 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 reactants, products and, ideally, configurations belonging to the transition state ensembles.[179] Moreover, to be used in biased enhanced sampling simulations, CVs should be continuous and differentiable functions of the atomic coordinates. As demonstrated by Peters et al. for single-step, classical nucleation mechanisms, CVs should approximate the size of the largest crystalline nucleus in solution, which is often an excellent approximation of the reaction coordinates for these types of systems[83, 179]. 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[43, 46, 89, 90], which also lend themselves to a theoretical description of two-step nucleation[45]. 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[43, 148, 180, 181, 182], that allows for the application of biased enhanced sampling by driving the polymorph-specific crystal nucleation.[167]

The definition of effective CVs for describing and enhancing 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 atomic or sing-particle-based crystals using bond-orientational OPs such as the Steinhardt order parameters[183, 184, 148, 84] it remains a challenge for molecular crystals. Approaches based on the calculation of generalised pair distribution functions[185] or on the calculation of properties of the distributions of order parameters [186, 187, 188] demonstrate promise but still require significant improvements in terms of computational efficiency, generalisability to molecular systems with a significant degree of conformational complexity [189] and with hundreds of putative polymorphs[190, 191]. 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 mineralising solutions and

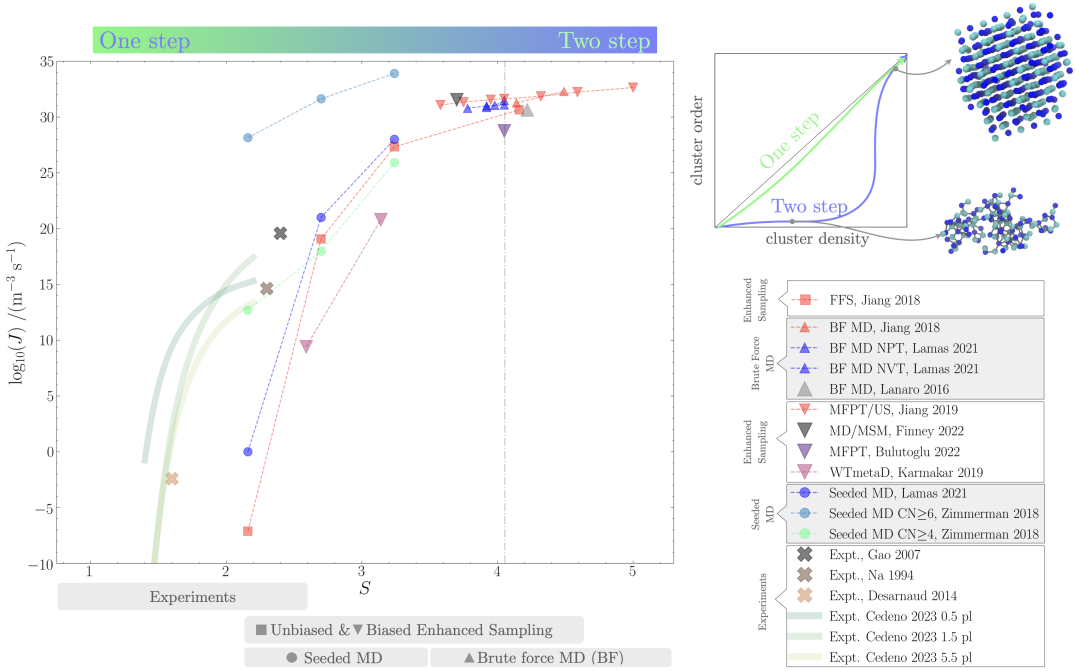

FIGURE 5 Rates for NaCl crystal nucleation from aqueous solution taken from the literature. Here, BF refers to brute force MD simulation of in which crystal nuclei emerge spontaneously. The limit of solution stability is indicated by the vertical dashed line at $S = 4.05$. The rate for Bulutoglu was evaluated using their literature rate value of 4×10^3 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 Lamas et al. [120]. All simulations are performed at 298 K, except in the case of Karmakar et al. [113], where simulations were performed at 350 K. The data reported in this plot is available at `://github.com/mme-ucl/NaCl_water_Nucleation_Rates.git`

represents one of the earliest case studies for crystallisation [85, 192], 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 compositions. 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 Fig. 5. The forcefield of reference in the studies described in this section, unless otherwise stated, is the NaCl Joung-Cheatham (JC) force field coupled with the SPC/E water model.

Importantly, for this system, 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 [193, 194, 195, 196, 197, 198, 199, 200, 122]. 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) [86, 121, 120], and to consistently attribute a supersaturation level to simulations performed with theory-agnostic methods such as FFS, and MSM. [201, 157, 173, 46]. 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. [121], this estimate is based on the values independently

estimated by Moucka et al. (3.64±0.2 m), Mester et al. (3.71±0.04 m), Benavides et al. (3.71±0.20 m), Kolafa et al. (3.6±0.4 m) and Espinosa et al. (3.7±0.4 m), and is the product of a process that has seen the refinement of simulation approaches and correction of earlier inaccurate estimates.

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 ($S = b/b^{sat}$) ranging between 2.1 and 4.5 [173]. 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[173]. 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.

Nucleation Kinetics within the remits of CNT

The quality of the dataset provided by Jiang et al. [173] 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 4.1.1). Two papers [121, 120] have provided independent attempts at comparing nucleation kinetics of seeding with the FFS dataset. Zimmerman et al. [86, 121] exploit that the ion attachment frequency f^+ is dominated by ion desolvation and adopt the theoretical expression of the nucleation rate prefactor reported in Eq. 8). Moreover, in contrast with other seeding studies, [122] the authors adopt the Girshick-Chiu correction [203] $\Gamma = \exp\left(F(1)/k_BT\right)$ (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 consensus value of $\Delta\mu_{\ell\rightarrow xtal}$ [121], 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. [120] who adopted the Auer and Frenkel expression of the attachment frequency f^+ (see Eq. 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 Fig. 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. [121] and Lamas et al. [120]. In particular, Zimmerman et al. show that nucleation rates are extremely sensitive to the number of crystal-like neighbours necessary to consider an ion part of the crystal nucleus. In this case, their conservative choice of having at least five crystalline neighbours leads to underestimating the critical nucleus size, thus overestimating the nucleation rate. 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 resorting to nucleation rates in good agreement with FFS results.

Overcoming solution depletion

Karmakar et al. [113], compute the nucleation barrier at two distinct supersaturation conditions from unseeded simulations (see section 4.2) by coupling WTmetaD[160] with CμMD[112]. 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 behaviour of nucleating systems in small volumes (see section 3.2)[89, 31]. Depletion artefacts 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 $n^* << V(c - c^{sat})$[86], 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 CμMD to compute the exponential term of the nucleation rate expression (Eq. 4). The prefactor is then estimated with the approach adopted by Espinosa et al. [122]. The values of the nucleation rates obtained by Karmakar et al. following this route are reported in Fig. 5. It should be noted that while the forcefield used in this study is the same adopted in the seeding papers discussed above, the results are not directly comparable as calculations were performed at 350 K, rather than at room temperature.

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 phase[90], and in particular on its degree of supersaturation. Studies based on seeding by construction cannot lead to the discovery of pathways departing from CNT-compliant ones.In contrast, unseeded simulations can reveal departures from CNT (see section2).

Using unseeded simulations, Panatgiotopoulos and coworkers [157] discovered that also in the case of NaCl nucleation from aqueous solution, the mechanism of nucleation 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 Fig. 5 as a vertical dashed line). The dense *amourphous 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 pathway[157] (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. [120] 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 Fig. 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 non-classical, multi-step pathway has been further corroborated by the work of Bulutoglu et al.[205], where the nucleation mechanism of NaCl is analysed by performing free energy calculations and by developing a theoretical approach based on the composite-cluster model [206] 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 favoured compared to the aqueous solution, thus further validating the existence of a two-step

nucleation pathway for NaCl at high supersaturation.

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 interface [207] at supersaturation levels significantly lower than the limit of solution stability discovered by Jiang et al. [157], and by recent experimental and simulation observations suggesting significant ion-ion correlations occurring in supersaturated NaCl; [125] 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[46].

By initialising hundreds of brute force MD simulations from uniformly distributed configurations in the n, $n(q_6)$ space, we then constructed a Markov State model able to yield model-free estimates of the nucleation rate of the committor surface in order parameter space, thus providing a quantitative description of the emergent nucleation mechanism. 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 favourable. Interestingly, the analysis of the committor probability surface in the n, $n(q_6)$ space suggests that at these conditions, one may need 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 towards a macroscopic crystal or dissolving.

As well as characterising 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 favourable 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.

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 [168]. Performing WTMetaD simulations of NaCl nucleation from an aqueous solution, where the NaCl ions were modelled with the GROMOS forcefield,[208] Giberti et al. discovered that small NaCl clusters might preferentially adopt structures that differ from that of bulk rock salt.[168] By employing a CV designed to enhance local density fluctuations without favouring a specific crystal structure[168, 43], 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 favourable 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 solubility[173] 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 as emergent collective processes from a direct sampling obtained with enhanced MD techniques.

6 | PERSPECTIVES AND CONCLUSIONS

Molecule and particle base 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 categorised 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 e.g., 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 crystallisation 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 [209], there appears to be reasonable consensus in the rates from unbiased and biased simulation strategies. 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. [121] 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 mislabelling analysis of Lamas et al. [120]. Taking the latter seeding results along with the forward flux sampling (FFS) rates from Jiang et al. [173], 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 matching 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. [125, 46, 205, 157]

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 have been 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 realised beyond small molecules. Nevertheless, increasing computational power and the progress made in machine learning techniques may make the routine simulation of crystal nucleation realisable 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 crystallisation pathways that are evident in different systems. Indeed, simulations are utilised 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 some systems might never be achieved at the atomic level. For instance, some systems, such as proteins, are just too big; others have such low solubility that simulating crystallisation directly from monomers in solution at experimental concentrations requires system' sizes that are simply beyond our reach. 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. In turn, this will enable accounting for crucial out-of-equilibrium effects in computationally-assisted material design.

ACKNOWLEDGEMENTS

We acknowledge funding from the *Crystallization in the Real World* EPSRC Programme Grant (Grant EP/R018820/1). MS acknowledges support from the EPSRC via the UKRI Frontier Research Guaranteee Grant (EP/X033139/1).

REFERENCES

[1] Price SL. Control and prediction of the organic solid state: a challenge to theory and experiment. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 2018;474(2217):20180351.

[2] Price SL. Predicting crystal structures of organic compounds. Chemical Society Reviews 2014;43(7):2098–2111.

[3] Day GM. Current approaches to predicting molecular organic crystal structures. Crystallography Reviews 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. Science Advances 2019;5(1). https://advances.sciencemag.org/content/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. Physical review letters 2016;117(11):115702.

[6] Kapil V, Engel EA. A complete description of thermodynamic stabilities of molecular crystals. arXiv preprint arXiv:210213598 2021;.

[7] Li L, Totton T, Frenkel D. Computational methodology for solubility prediction: Application to the sparingly soluble solutes. The Journal of Chemical Physics 2017;146(21):214110.

[8] Cheng B, Ceriotti M. Computing the absolute Gibbs free energy in atomistic simulations: Applications to defects in solids. Physical Review B 2018;97(5):054102.

[9] Li L, Totton T, Frenkel D. Computational methodology for solubility prediction: Application to sparingly soluble organic/inorganic materials. The Journal of chemical physics 2018;149(5):054102.

[10] Kapil V, Engel E, Rossi M, Ceriotti M. Assessment of Approximate Methods for Anharmonic Free Energies. Journal of Chemical Theory and Computation 2019;15(11):5845–5857.

[11] Abraham NS, Shirts MR. Adding anisotropy to the standard quasi-harmonic approximation still fails in several ways to capture organic crystal thermodynamics. Crystal Growth & Design 2019;19(12):6911–6924.

[12] Abraham NS, Shirts MR. Statistical mechanical approximations to more efficiently determine polymorph free energy differences for small organic molecules. Journal of Chemical Theory and Computation 2020;16(10):6503–6512.

[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 Crystallographica Section B: Structural Science, Crystal Engineering and Materials 2016;72(4):439–459.

[14] Zhang P, Wood GP, Ma J, Yang M, Liu Y, Sun G, et al. Harnessing Cloud Architecture for Crystal Structure Prediction Calculations. Crystal Growth & Design 2018;18(11):6891–6900.

[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. Communications Chemistry 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. Crystal Growth & Design 2020;20(8):5211–5224.

[17] Kendrick J, Leusen FJ, Neumann MA, van de Streek J. Progress in crystal structure prediction. Chemistry–A European Journal 2011;17(38):10736–10744.

[18] Cruz-Cabeza AJ, Reutzel-Edens SM, Bernstein J. Facts and fictions about polymorphism. Chemical Society Reviews 2015;44(23):8619–8635.

[19] Nyman J, Reutzel-Edens SM. Crystal structure prediction is changing from basic science to applied technology. Faraday Discussions 2018;211:459–476.

[20] Srinivasan K. Crystal growth of α and γ glycine polymorphs and their polymorphic phase transformations. J Cryst Growth 2008;311(1):156 – 162.

[21] Price SL. Why don't we find more polymorphs? Acta Crystallogr Sec B 2013 Aug;69(4):313–328. https://doi.org/10.1107/S2052519213018861.

[22] Gadewar SB, Doherty MF. A dynamic model for evolution of crystal shape. J Cryst Growth 2004;267:239–250.

[23] Gadewar SB, Hofmann HM, Doherty MF. Evolution of crystal shape. Cryst Growth Des 2004;4:109–112.

[24] Zhang Y, Sizemore JP, Doherty MF. Shape evolution of 3-dimensional faceted crystals. AIChE J 2006;52:1906–1915.

[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;6:1788–1796.

[26] Davey RJ, Schroeder SLM, Ter Horst JH. Nucleation of organic crystals - A molecular perspective. Angewandte Chemie - International Edition 2013;52(8):2167–2179.

[27] Sun CC, Sun W, Price S, Hughes C, Ter Horst J, Veesler S, et al. Solvent and additive interactions as determinants in the nucleation pathway: general discussion. Faraday Discussions 2015;179:383–420.

[28] Agarwal V, Peters B. Nucleation near the eutectic point in a Potts-lattice gas model. J Chem Phys 2014;140(8):084111.

[29] 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. Chemical Reviews 2016 1;116(12):7078–7116. https://pubs.acs.org/doi/10.1021/acs.chemrev.5b00744.

[30] Anwar J, Khan S, Lindfors L. Secondary crystal nucleation: nuclei breeding factory uncovered. Angewandte Chemie International Edition 2015;54(49):14681–14684.

[31] 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. Chemical reviews 2016;116(12):7078–7116.

[32] Sosso GC, Tribello GA, Zen A, Pedevilla P, Michaelides A. Ice formation on kaolinite: Insights from molecular dynamics simulations. The Journal of chemical physics 2016;145(21):211927.

[33] Sosso GC, Sudera P, Backes AT, Whale TF, Fröhlich-Nowoisky J, Bonn M, et al. The role of structural order in heterogeneous ice nucleation. Chemical Science 2022;13(17):5014–5026.

[34] Factorovich MH, Naullage PM, Molinero V. Can clathrates heterogeneously nucleate ice? The Journal of Chemical Physics 2019;151(11):114707.

[35] Glatz B, Sarupria S. Heterogeneous ice nucleation: Interplay of surface properties and their impact on water orientations. Langmuir 2018;34(3):1190–1198.

[36] Lata NN, Zhou J, Hamilton P, Larsen M, Sarupria S, Cantrell W. Multivalent surface cations enhance heterogeneous freezing of water on muscovite mica. The Journal of Physical Chemistry Letters 2020;11(20):8682–8689.

[37] Mullin JW. Crystallization. 4th ed. Butterworth-Heinemann; 2001.

[38] De Yoreo JJ, Vekilov PG. Principles of Crystal Nucleation and Growth. Reviews in Mineralogy and Geochemistry 2003 jan;54(1):57–93. https://pubs.geoscienceworld.org/rimg/article/54/1/57-93/87490.

[39] Erdemir D, Lee AY, Myerson AS. Nucleation of Crystals from Solution: Classical and Two-Step Models. Accounts of Chemical Research 2009 may;42(5):621–629. https://pubs.acs.org/doi/10.1021/ar800217x.

[40] Vekilov PG. The two-step mechanism of nucleation of crystals in solution. Nanoscale 2010;2(11):2346. http://xlink.rsc.org/?DOI=c0nr00628a.

[41] Schöpe HJ, Bryant G, van Megen W. Two-Step Crystallization Kinetics in Colloidal Hard-Sphere Systems. Physical Review Letters 2006 may;96(17):175701. https://link.aps.org/doi/10.1103/PhysRevLett.96.175701.

[42] Vatamanu J, Kusalik PG. Observation of two-step nucleation in methane hydrates. Physical Chemistry Chemical Physics 2010;12(45):15065. http://xlink.rsc.org/?DOI=c0cp00551g.

[43] Finney AR, Salvalaglio M. A variational approach to assess reaction coordinates for two-step crystallization. The Journal of Chemical Physics 2023 mar;158(9):094503. https://aip.scitation.org/doi/10.1063/5.0139842.

[44] ten Wolde PR, Frenkel D. Enhancement of protein crystal nucleation by critical density fluctuations. Science 1997;277:1975–1978.

[45] Kashchiev D. Classical nucleation theory approach to two-step nucleation of crystals. Journal of Crystal Growth 2020 jan;530:125300. https://linkinghub.elsevier.com/retrieve/pii/S0022024819305159.

[46] Finney AR, Salvalaglio M. Multiple pathways in NaCl homogeneous crystal nucleation. Faraday Discussions 2022;235:56–80. http://xlink.rsc.org/?DOI=D1FD00089F.

[47] 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. Research Square 2022;https://doi.org/10.21203/rs.3.rs-1743346/v1.

[48] 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. Materials Advances 2021;2(20):6532–6568. http://xlink.rsc.org/?DOI=D1MA00222H.

[49] 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. Chemistry of Materials 2018 may;30(9):2895–2904. https://pubs.acs.org/doi/10.1021/acs.chemmater.7b04245.

[50] Kashchiev D. Nucleation. Elsevier; 2000.

[51] Dillmann A, Meier GEA. A refined droplet approach to the problem of homogeneous nucleation from the vapor phase;94(5):3872–3884. http://aip.scitation.org/doi/10.1063/1.460663.

[52] Ford IJ, Laaksonen A, Kulmala M. Modification of the Dillmann–Meier theory of homogeneous nucleation;99(1):764–765. http://aip.scitation.org/doi/10.1063/1.465756.

[53] Prestipino S, Laio A, Tosatti E. Systematic Improvement of Classical Nucleation Theory. Physical Review Letters 2012 May;108(22):225701. https://link.aps.org/doi/10.1103/PhysRevLett.108.225701.

[54] Ten Wolde PR, Frenkel D. Computer simulation study of gas–liquid nucleation in a Lennard-Jones system. The Journal of Chemical Physics 1998 Dec;109(22):9901–9918. http://aip.scitation.org/doi/10.1063/1.477658.

[55] Lutsko JF, Durán-Olivencia MA. A two-parameter extension of classical nucleation theory. Journal of Physics: Condensed Matter 2015 Jun;27(23):235101. https://iopscience.iop.org/article/10.1088/0953-8984/27/23/235101.

[56] Iwamatsu M. Free-energy landscape of nucleation with an intermediate metastable phase studied using capillarity approximation. The Journal of Chemical Physics 2011 Apr;134(16):164508. http://aip.scitation.org/doi/10.1063/1.3583641.

[57] Sear RP. The non-classical nucleation of crystals: microscopic mechanisms and applications to molecular crystals, ice and calcium carbonate. International Materials Reviews 2012 Nov;57(6):328–356. http://www.tandfonline.com/doi/full/10.1179/1743280411Y.0000000015.

[58] Lee J, Yang J, Kwon SG, Hyeon T. Nonclassical nucleation and growth of inorganic nanoparticles. Nature Reviews Materials 2016 Jun;1(8):16034. https://www.nature.com/articles/natrevmats201634.

[59] Karthika S, Radhakrishnan TK, Kalaichelvi P. A Review of Classical and Nonclassical Nucleation Theories. Crystal Growth & Design 2016 Nov;16(11):6663–6681. https://pubs.acs.org/doi/10.1021/acs.cgd.6b00794.

[60] De Yoreo J. A Perspective on Multistep Pathways of Nucleation. In: Zhang X, editor. ACS Symposium Series, vol. 1358 Washington, DC: American Chemical Society; 2020.p. 1–17. https://pubs.acs.org/doi/abs/10.1021/bk-2020-1358.ch001.

[61] Jun YS, Zhu Y, Wang Y, Ghim D, Wu X, Kim D, et al. Classical and Nonclassical Nucleation and Growth Mechanisms for Nanoparticle Formation. Annual Review of Physical Chemistry 2022 Apr;73(1):453–477. https://www.annualreviews.org/doi/10.1146/annurev-physchem-082720-100947.

[62] Gebauer D, Volkel A, Colfen H. Stable Prenucleation Calcium Carbonate Clusters. Science 2008 Dec;322(5909):1819–1822. https://www.science.org/doi/10.1126/science.1164271.

[63] Gebauer D, Kellermeier M, Gale JD, Bergström L, Cölfen H. Pre-nucleation clusters as solute precursors in crystallisation. Chem Soc Rev 2014;43(7):2348–2371. http://xlink.rsc.org/?DOI=C3CS60451A.

[64] Demichelis R, Raiteri P, Gale JD, Quigley D, Gebauer D. Stable prenucleation mineral clusters are liquid-like ionic polymers. Nature Communications 2011 Dec;2(1):590. https://www.nature.com/articles/ncomms1604.

[65] Avaro JT, Wolf SLP, Hauser K, Gebauer D. Stable Prenucleation Calcium Carbonate Clusters Define Liquid–Liquid Phase Separation. Angewandte Chemie International Edition 2020 Apr;59(15):6155–6159. https://onlinelibrary.wiley.com/doi/10.1002/anie.201915350.

[66] Palmer JC, Debenedetti PG. Recent advances in molecular simulation: A chemical engineering perspective. AIChE Journal 2015;61(2):370–383. http://dx.doi.org/10.1002/aic.14706.

[67] Frenkel D, Smit B. Understanding Molecular Simulation. Elsevier; 2002. https://linkinghub.elsevier.com/retrieve/pii/B9780122673511X50007.

[68] 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 Computational Materials 2016 mar;2(1):15011. https://www.nature.com/articles/npjcompumats201511.

[69] Meuwly M. Reactive molecular dynamics: From small molecules to proteins. WIREs Computational Molecular Science 2019 jan;9(1). https://onlinelibrary.wiley.com/doi/10.1002/wcms.1386.

[70] Meuwly M. Machine Learning for Chemical Reactions. Chemical Reviews 2021 aug;121(16):10218–10239. https://pubs.acs.org/doi/10.1021/acs.chemrev.1c00033.

[71] Tuckerman ME. Statistical Mechanics: Theory and Molecular Simulation. 2nd ed. Oxford Graduate Texts;.

[72] 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, in press 2016;.

[73] Diemand J, Angélil R, Tanaka KK, Tanaka H. Large scale molecular dynamics simulations of homogeneous nucleation. The Journal of chemical physics 2013;139(7):074309.

[74] Auer S, Frenkel D. Prediction of absolute crystal-nucleation rate in hard-sphere colloids. Nature 2001;409:1020–1023.

[75] Valeriani C, Sanz E, Frenkel D. Rate of homogeneous crystal nucleation in molten NaCl. J Chem Phys 2005;122:194501.

[76] 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.

[77] 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.

[78] Haji-Akbari A, Debenedetti PG. Direct calculation of ice homogeneous nucleation rate for a molecular model of water. ProcNat Acad Sci 2015;112(34):10582–10588.

[79] 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):5118–5123.

[80] Lifanov Y, Vorselaars B, Quigley D. Nucleation barrier reconstruction via the seeding method in a lattice model with competing nucleation pathways. The Journal of Chemical Physics 2016;145(21):211912.

[81] Anwar J, Boateng PK. Computer simulation of crystallization from solution. J Am Chem Soc 1998;120:9600–9604.

[82] Kawska A, Brickmann J, Kniep R, Hochrein O, Zahn D. An atomistic simulation scheme for modeling crystal formation from solution. J Chem Phys 2006;124:024513.

[83] Agarwal V, Peters B. In: Solute Precipitate Nucleation: A Review of Theory and Simulation Advances John Wiley & Sons, Inc.; 2014. p. 97–160.

[84] Giberti F, Salvalaglio M, Parrinello M. Metadynamics studies of crystal nucleation. IUCrJ 2015 Mar;2(2):256–266. https://doi.org/10.1107/S2052252514027626.

[85] Anwar J, Zahn D. Uncovering Molecular Processes in Crystal Nucleation and Growth by Using Molecular Simulation. Angewandte Chemie International Edition 2011;50(9):1996–2013. http://dx.doi.org/10.1002/anie.201000463.

[86] Zimmermann NE, Vorselaars B, Quigley D, Peters B. Nucleation of NaCl from aqueous solution: Critical sizes, ion-attachment kinetics, and rates. Journal of the American Chemical Society 2015;137(41):13352–13361.

[87] Ectors P, Duchstein P, Zahn D. From oligomers towards a racemic crystal: molecular simulation of dl-norleucine crystal nucleation from solution. CrystEngComm 2015;17:6884–6889. http://dx.doi.org/10.1039/C4CE02078B.

[88] Salvalaglio M, Giberti F, Parrinello M. 1, 3, 5-Tris (4-bromophenyl) benzene prenucleation clusters from metadynamics. Acta Cryst Sect C 2014;70(2):132–136.

[89] Salvalaglio M, Perego C, Giberti F, Mazzotti M, Parrinello M. Molecular-dynamics simulations of urea nucleation from aqueous solution. Proc Nat Acad Sci 2015;112(1):E6–E14.

[90] Salvalaglio M, Mazzotti M, Parrinello M. Urea homogeneous nucleation mechanism is solvent dependent. Faraday Discuss 2015;179:291–307.

[91] Piana S, Reyhani M, Gale JD. Simulating micrometre-scale crystal growth from solution. Nature 2005;438:70–73.

[92] Piana S, Gale JD. Three-dimensional kinetic Monte Carlo simulation of crystal growth from solution. J Cryst Growth 2006;294:46–52.

[93] 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;127:1975–1982.

[94] Piana S, Jones F, Gale JD. Aspartic acid as a crystal growth catalyst. CrystEngComm 2007;9:1187–1191.

[95] 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 Discussions 2015;179:503–542.

[96] Farkas L. Keimbildungsgeschwindigkeit in übersättigten Dämpfen;125U(1):236–242. https://www.degruyter.com/document/doi/10.1515/zpch-1927-12513/html.

[97] Volmer M, Weber A. Keimbildung in ubersattigten Gebilden;119U(1):277–301. https://www.degruyter.com/document/doi/10.1515/zpch-1926-11927/html.

[98] Becker R, Döring W. Kinetische Behandlung der Keimbildung in übersättigten Dämpfen;416(8):719–752. https://onlinelibrary.wiley.com/doi/10.1002/andp.19354160806.

[99] In: Sunyaev RA, editor. 10. On the Theory of New Phase Formation. Cavitation Princeton University Press;. p. 120–137. https://www.degruyter.com/document/doi/10.1515/9781400862979.120/html.

[100] Wedekind J, Reguera D, Strey R. Finite-size effects in simulations of nucleation. J Chem Phys 2006;125(21):–.

[101] Salvalaglio M, Vetter T, Mazzotti M, Parrinello M. Controlling and predicting crystal shapes: The case of urea. Angew Chem Int Ed 2013;52(50):13369–13372.

[102] Durán-Olivencia MA, Lutsko JF. Mesoscopic nucleation theory for confined systems: A one-parameter model. Physical Review E 2015 Feb;91(2):022402. https://link.aps.org/doi/10.1103/PhysRevE.91.022402.

[103] Perego C, Salvalaglio M, Parrinello M. Molecular dynamics simulations of solutions at constant chemical potential. J Chem Phys 2015;142(14):144113.

[104] 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):3858–3865.

[105] Karmakar T, Piaggi PM, Parrinello M. Molecular dynamics simulations of crystal nucleation from solution at constant chemical potential. Journal of chemical theory and computation 2019;15(12):6923–6930.

[106] Liu C, Wood GP, Santiso EE. Modelling nucleation from solution with the string method in the osmotic ensemble. Molecular Physics 2018;116(21-22):2998–3007.

[107] Liu C, Cao F, Kulkarni SA, Wood GPF, Santiso EE. Understanding Polymorph Selection of Sulfamerazine in Solution. Crystal Growth & Design 2019 Dec;19(12):6925–6934. https://pubs.acs.org/doi/10.1021/acs.cgd.9b00576.

[108] Reguera D, Bowles RK, Djikaev Y, Reiss H. Phase transitions in systems small enough to be clusters;118(1):340–353. http://aip.scitation.org/doi/10.1063/1.1524192.

[109] Agarwal V, Peters B. Solute precipitate nucleation: A review of theory and simulation advances. Advances in Chemical Physics: Volume 155 2014;p. 97–160.

[110] Li L, Paloni M, Finney AR, Barducci A, Salvalaglio M. Nucleation of Biomolecular Condensates from Finite-Sized Simulations;p. 1748–1755. https://pubs.acs.org/doi/10.1021/acs.jpclett.2c03512.

[111] Lutsko JF, Durán-Olivencia MA. Classical nucleation theory from a dynamical approach to nucleation. The Journal of Chemical Physics 2013 Jun;138(24):244908. http://aip.scitation.org/doi/10.1063/1.4811490.

[112] Perego C, Salvalaglio M, Parrinello M. Molecular dynamics simulations of solutions at constant chemical potential. The Journal of Chemical Physics 2015 Apr;142(14):144113. http://aip.scitation.org/doi/10.1063/1.4917200.

[113] Karmakar T, Piaggi PM, Parrinello M. Molecular Dynamics Simulations of Crystal Nucleation from Solution at Constant Chemical Potential. Journal of Chemical Theory and Computation 2019 Dec;15(12):6923–6930. https://pubs.acs.org/doi/10.1021/acs.jctc.9b00795.

[114] ST K, A F, M S, AO Y, C P. Non-Equilibrium Modelling of Concentration-Driven Processes with Constant Chemical Potential Molecular Dynamics Simulations. Accounts of Chemical Research 2023;.

[115] Han D, Karmakar T, Bjelobrk Z, Gong J, Parrinello M. Solvent-mediated morphology selection of the active pharmaceutical ingredient isoniazid: Experimental and simulation studies. Chemical Engineering Science 2019 Aug;204:320–328. https://linkinghub.elsevier.com/retrieve/pii/S0009250918307358.

[116] Bjelobrk Z, Piaggi PM, Weber T, Karmakar T, Mazzotti M, Parrinello M. Naphthalene crystal shape prediction from molecular dynamics simulations. CrystEngComm 2019;21(21):3280–3288. http://xlink.rsc.org/?DOI=C9CE00380K.

[117] Liu C, Wood GPF, Santiso EE. Modelling nucleation from solution with the string method in the osmotic ensemble. Molecular Physics 2018 Nov;116(21-22):2998–3007. https://www.tandfonline.com/doi/full/10.1080/00268976.2018.1482016.

[118] Praprotnik M, Site LD, Kremer K. Multiscale Simulation of Soft Matter: From Scale Bridging to Adaptive Resolution;59(1):545–571. https://www.annualreviews.org/doi/10.1146/annurev.physchem.59.032607.093707.

[119] Espinosa JR, Vega C, Valeriani C, Sanz E. Seeding approach to crystal nucleation. The Journal of chemical physics 2016;144(3):034501.

[120] Lamas C, Espinosa J, Conde M, Ramírez J, de Hijes PM, Noya EG, et al. Homogeneous nucleation of NaCl in supersaturated solutions. Physical Chemistry Chemical Physics 2021;23(47):26843–26852.

[121] 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. The Journal of chemical physics 2018;148(22):222838.

[122] 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. The Journal of chemical physics 2016;145(15):154111.

[123] Knott BC, Molinero V, Doherty MF, Peters B. Homogeneous nucleation of methane hydrates: Unrealistic under realistic conditions. Journal of the American Chemical Society 2012;134(48):19544–19547.

[124] Mandal T, Larson RG. Nucleation of urea from aqueous solution: Structure, critical size, and rate. The Journal of Chemical Physics 2017;146(13):134501.

[125] 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. Chemical Science 2021;12(1):179–187. http://xlink.rsc.org/?DOI=D0SC04817H.

[126] 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. Crystal Growth & Design 2019 Nov;19(11):6422–6430. https://pubs.acs.org/doi/10.1021/acs.cgd.9b00889.

[127] 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. Nature Communications 2013 Feb;4(1):1507. https://www.nature.com/articles/ncomms2490.

[128] 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. Journal of the American Chemical Society 2019;141(14):6073–6081.

[129] Kollias L, Rousseau R, Glezakou VA, Salvalaglio M. Understanding metal–organic framework nucleation from a solution with evolving graphs. Journal of the American Chemical Society 2022;144(25):11099–11109.

[130] Balestra SR, Semino R. Computer simulation of the early stages of self-assembly and thermal decomposition of ZIF-8. The Journal of Chemical Physics 2022;157(18):184502.

[131] 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. The Journal of Physical Chemistry B 2006 Feb;110(7):3323–3329. https://pubs.acs.org/doi/10.1021/jp055982e.

[132] 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. The Journal of Physical Chemistry B 2008 Jul;112(26):7794–7802. https://pubs.acs.org/doi/10.1021/jp7106582.

[133] Chen J, Trout BL. A Computational Study of the Mechanism of the Selective Crystallization of - and -Glycine from Water and MethanolWater Mixture. The Journal of Physical Chemistry B 2010 Nov;114(43):13764–13772. https://pubs.acs.org/doi/10.1021/jp1039496.

[134] 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. Crystal Growth & Design 2021 Sep;21(9):4823–4836. https://pubs.acs.org/doi/10.1021/acs.cgd.0c01605.

[135] Bobrovs R, Drunka L, Auzins AA, Jaudzems K, Salvalaglio M. Polymorph-Selective Role of Hydrogen Bonding and – Stacking in p -Aminobenzoic Acid Solutions. Crystal Growth & Design 2021 Jan;21(1):436–448. https://pubs.acs.org/doi/10.1021/acs.cgd.0c01257.

[136] Carpenter JE, Grünwald M. Pre-Nucleation Clusters Predict Crystal Structures in Models of Chiral Molecules. Journal of the American Chemical Society 2021 Dec;143(51):21580–21593. https://pubs.acs.org/doi/10.1021/jacs.1c09321.

[137] 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. Physical Chemistry Chemical Physics 2021;23(48):27253–27265. http://xlink.rsc.org/?DOI=D1CP04226B.

[138] 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. The Journal of Physical Chemistry B 2020 Apr;124(17):3568–3582. https://pubs.acs.org/doi/10.1021/acs.jpcb.0c01582.

[139] Kellermeier M, Raiteri P, Berg JK, Kempter A, Gale JD, Gebauer D. Entropy Drives Calcium Carbonate Ion Association. ChemPhysChem 2016 Nov;17(21):3535–3541. https://onlinelibrary.wiley.com/doi/10.1002/cphc.201600653.

[140] Smeets PJM, Finney AR, Habraken WJEM, Nudelman F, Friedrich H, Laven J, et al. A classical view on nonclassical nucleation. Proceedings of the National Academy of Sciences 2017 Sep;114(38). https://pnas.org/doi/full/10.1073/pnas.1700342114.

[141] Finney AR, Rodger PM. Probing the structure and stability of calcium carbonate pre-nucleation clusters. Faraday Discussions 2012;159:47. http://xlink.rsc.org/?DOI=c2fd20054f.

[142] Finney AR, Innocenti Malini R, Freeman CL, Harding JH. Amino Acid and Oligopeptide Effects on Calcium Carbonate Solutions. Crystal Growth & Design 2020 May;20(5):3077–3092. https://pubs.acs.org/doi/10.1021/acs.cgd.9b01693.

[143] 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 Aug;341(6148):885–889. https://www.science.org/doi/10.1126/science.1230915.

[144] Raiteri P, Demichelis R, Gale JD. Thermodynamically Consistent Force Field for Molecular Dynamics Simulations of Alkaline-Earth Carbonates and Their Aqueous Speciation. The Journal of Physical Chemistry C 2015 Oct;119(43):24447–24458. https://pubs.acs.org/doi/10.1021/acs.jpcc.5b07532.

[145] Ahlawat P, Dar MI, Piaggi P, Gratzel M, Parrinello M, Rothlisberger U. Atomistic mechanism of the nucleation of methylammonium lead iodide perovskite from solution. Chemistry of Materials 2019;32(1):529–536.

[146] Santiso EE, Trout BL. A general method for molecular modeling of nucleation from the melt. The Journal of Chemical Physics 2015;143(17):174109.

[147] Tribello GA, Giberti F, Sosso GC, Salvalaglio M, Parrinello M. Analyzing and driving cluster formation in atomistic simulations. Journal of chemical theory and computation 2017;13(3):1317–1327.

[148] Neha, Tiwari V, Mondal S, Kumari N, Karmakar T. Collective Variables for Crystallization Simulations from Early Developments to Recent Advances. ACS omega;.

[149] Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. PLUMED 2: New feathers for an old bird. Computer Physics Communications 2014;185(2):604–613.

[150] Torrie GM, Valleau JP. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. Journal of Computational Physics 1977;23(2):187–199.

[151] Roux B. The calculation of the potential of mean force using computer simulations. Computer physics communications 1995;91(1-3):275–282.

[152] Kästner J. Umbrella sampling. Wiley Interdisciplinary Reviews: Computational Molecular Science 2011;1(6):932–942.

[153] Zwanzig R. From classical dynamics to continuous time random walks. J Stat Phys 1983;30(2):255–262.

[154] Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. Journal of computational chemistry 1992;13(8):1011–1021.

[155] Shirts MR, Chodera JD. Statistically optimal analysis of samples from multiple equilibrium states. The Journal of chemical physics 2008;129(12):124105.

[156] Kästner J, Thiel W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method:"Umbrella integration". The Journal of chemical physics 2005;123(14):144104.

[157] Jiang H, Debenedetti PG, Panagiotopoulos AZ. Nucleation in aqueous NaCl solutions shifts from 1-step to 2-step mechanism on crossing the spinodal. The Journal of chemical physics 2019;150(12):124502.

[158] Allen RJ, Valeriani C, Ten Wolde PR. Forward flux sampling for rare event simulations. Journal of physics: Condensed matter 2009;21(46):463102.

[159] Laio A, Parrinello M. Escaping free-energy minima. Proc Nat Acad Sci 2002;99(20):12562–12566.

[160] Barducci A, Bussi G, Parrinello M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. Phys Rev Lett 2008;100(2):020603.

[161] Bonomi M, Parrinello M. Enhanced sampling in the well-tempered ensemble. Physical review letters 2010;104(19):190601.

[162] Bussi G, Gervasio FL, Laio A, Parrinello M. Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics. Journal of the American Chemical Society 2006;128(41):13435–13441.

[163] Barducci A, Bonomi M, Parrinello M. Metadynamics. WIREs: Comput Mol Sci 2011;1(5):826–843.

[164] Valsson O, Tiwary P, Parrinello M. Enhancing important fluctuations: Rare events and metadynamics from a conceptual viewpoint. Annual review of physical chemistry 2016;67:159–184.

[165] Tiwary P, Parrinello M. A time-independent free energy estimator for metadynamics. J Phys Chem B 2014;119(3):736–742.

[166] Giberti F, Salvalaglio M, Mazzotti M, Parrinello M. 1, 3, 5-tris (4-bromophenyl)-benzene Nucleation: From Dimers to Needle-like Clusters. Crystal Growth & Design 2017;17(8):4137–4143.

[167] Zou Z, Beyerle ER, Tsai ST, Tiwary P. Driving and characterizing nucleation of urea and glycine polymorphs in water. Proceedings of the National Academy of Sciences 2023;120(7):e2216099120.

[168] Giberti F, Tribello GA, Parrinello M. Transient polymorphism in NaCl. Journal of chemical theory and computation 2013;9(6):2526–2530.

[169] Lauricella M, Meloni S, English NJ, Peters B, Ciccotti G. Methane clathrate hydrate nucleation mechanism by advanced molecular simulations. The Journal of Physical Chemistry C 2014;118(40):22847–22857.

[170] Arjun, Berendsen TA, Bolhuis PG. Unbiased atomistic insight in the competing nucleation mechanisms of methane hydrates. Proceedings of the National Academy of Sciences 2019;116(39):19305–19310.

[171] Arjun A, Bolhuis PG. Rate Prediction for Homogeneous Nucleation of Methane Hydrate at Moderate Supersaturation Using Transition Interface Sampling. The Journal of Physical Chemistry B 2020;124(37):8099–8109.

[172] Hall SW, Díaz Leines G, Sarupria S, Rogal J. Practical guide to replica exchange transition interface sampling and forward flux sampling. The Journal of Chemical Physics 2022;156(20):200901.

[173] Jiang H, Haji-Akbari A, Debenedetti PG, Panagiotopoulos AZ. Forward flux sampling calculation of homogeneous nucleation rates from aqueous NaCl solutions. The Journal of chemical physics 2018;148(4):044505.

[174] Bolhuis PG, Chandler D, Dellago C, Geissler PL. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. Annual review of physical chemistry 2002;53(1):291–318.

[175] Van Erp TS, Bolhuis PG. Elaborating transition interface sampling methods. Journal of computational Physics 2005;205(1):157–181.

[176] Van Erp TS, Moroni D, Bolhuis PG. A novel path sampling method for the calculation of rate constants. The Journal of chemical physics 2003;118(17):7762–7774.

[177] Haji-Akbari A. Forward-flux sampling with jumpy order parameters. The Journal of chemical physics 2018;149(7):072303.

[178] Sidky H, Colón YJ, Helfferich J, Sikora BJ, Bezik C, Chu W, et al. SSAGES: Software Suite for Advanced General Ensemble Simulations. The Journal of Chemical Physics 2018 Jan;148(4):044104. https://pubs.aip.org/aip/jcp/article/75408.

[179] Peters B. Reaction Coordinates and Mechanistic Hypothesis Tests. Ann Rev Phys Chem 2016;67:669–690.

[180] Bonati L, Piccini G, Parrinello M. Deep learning the slow modes for rare events sampling. Proceedings of the National Academy of Sciences 2021 Nov;118(44):e2113533118. https://pnas.org/doi/full/10.1073/pnas.2113533118.

[181] Elishav O, Podgaetsky R, Meikler O, Hirshberg B. Collective Variables for Conformational Polymorphism in Molecular Crystals. The Journal of Physical Chemistry Letters 2023 Feb;14(4):971–976. https://pubs.acs.org/doi/10.1021/acs.jpclett.2c03491.

[182] Rogal J. Reaction coordinates in complex systems-a perspective. The European Physical Journal B 2021;94:1–9.

[183] Steinhardt PJ, Nelson DR, Ronchetti M. Bond-orientational order in liquids and glasses. Physical Review B 1983;28(2):784.

[184] Lechner W, Dellago C. Accurate determination of crystal structures based on averaged local bond order parameters. The Journal of chemical physics 2008;129(11):114707.

[185] Santiso EE, Trout BL. A General Set of Order Parameters for Molecular Crystals. J Chem Phys 2011;134(6):064109.

[186] 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):959–972.

[187] Gimondi I, Salvalaglio M. CO2 Packing Polymorphism Under Confinement in Cylindrical Nanopores. Mol Syst Des Eng 2018;3(1):243–252.

[188] Piaggi PM, Parrinello M. Predicting polymorphism in molecular crystals using orientational entropy. Proceedings of the National Academy of Sciences 2018;115(41):10251–10256.

[189] Marinova V, Dodd L, Lee SJ, Wood GP, Marziano I, Salvalaglio M. Identifying conformational isomers of organic molecules in solution via unsupervised clustering. Journal of Chemical Information and Modeling 2021;61(5):2263–2273.

[190] F NF, Price LS, Nyman J, Price SL, Salvalaglio M. Systematic finite-temperature reduction of crystal energy landscapes. Crystal Growth & Design 2020;20(10):6847–6862.

[191] Francia NF, Price LS, Salvalaglio M. Reducing crystal structure overprediction of ibuprofen with large scale molecular dynamics simulations. CrystEngComm 2021;23(33):5575–5584.

[192] Zahn D. Atomistic mechanism of NaCl nucleation from an aqueous solution. Physical review letters 2004;92(4):040801.

[193] Aragones J, Sanz E, Vega C. Solubility of NaCl in water by molecular simulation revisited. The Journal of Chemical Physics 2012;136(24):244508.

[194] Benavides A, Portillo M, Abascal J, Vega C. Estimating the solubility of 1: 1 electrolyte aqueous solutions: the chemical potential difference rule. Molecular Physics 2017;115(9-12):1301–1308.

[195] Nezbeda I, Moučka F, Smith WR. Recent progress in molecular simulation of aqueous electrolytes: Force fields, chemical potentials and solubility. Molecular Physics 2016;114(11):1665–1690.

[196] Moučka F, Nezbeda I, Smith WR. Molecular simulation of aqueous electrolytes: Water chemical potential results and Gibbs-Duhem equation consistency tests. The Journal of Chemical Physics 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. Journal of Chemical Theory and Computation 2013;9(11):5076–5085.

[198] Moucka F, Nezbeda I, Smith WR. Chemical potentials, activity coefficients, and solubility in aqueous NaCl solutions: Prediction by polarizable force fields. Journal of Chemical Theory and Computation 2015;11(4):1756–1764.

[199] Moucka 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. The Journal of Physical Chemistry B 2011;115(24):7849–7861.

[200] Mester Z, Panagiotopoulos AZ. Mean ionic activity coefficients in aqueous NaCl solutions from molecular dynamics simulations. The Journal of Chemical Physics 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? General discussion. Faraday Discussions 2022;235:219–272.

[202] Benavides AL, Aragones JL, Vega C. Consensus on the solubility of NaCl in water from computer simulations using the chemical potential route. The Journal of Chemical Physics 2016 Mar;144(12):124504. https://pubs.aip.org/aip/jcp/article/910454.

[203] Girshick SL, Chiu CP. Kinetic nucleation theory: A new expression for the rate of homogeneous nucleation from an ideal supersaturated vapor. The journal of chemical physics 1990;93(2):1273–1277.

[204] Chkonia G, Wölk J, Strey R, Wedekind J, Reguera D. Evaluating nucleation rates in direct simulations. The Journal of chemical physics 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] Iwamatsu M. Free-energy landscape of nucleation with an intermediate metastable phase studied using capillarity approximation. The Journal of chemical physics 2011;134(16):164508.

[207] Finney AR, McPherson IJ, Unwin PR, Salvalaglio M. Electrochemistry, ion adsorption and dynamics in the double layer: a study of NaCl (aq) on graphite. Chemical science 2021;12(33):11166–11180.

[208] 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;25(13):1656–1676. https://onlinelibrary.wiley.com/doi/10.1002/jcc.20090.

[209] Blow KE, Quigley D, Sosso GC. The seven deadly sins: When computing crystal nucleation rates, the devil is in the details. The Journal of Chemical Physics 2021;155(4):040901.

Aaron R. Finney is a Senior Research Fellow in the Department of Chemical Engineering, University College London and member of the Thomas Young Centre. 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. His current research focuses on interfaces and crystallisation.

Matteo Salvalaglio is an Associate Professor of Chemical 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, 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.

GRAPHICAL ABSTRACT