A Kinetic-Bias Model for the Dynamic Simulation of Molecular Aggregation. The Liquid, Solute, Solvated-Nanodrop and Solvated-Nanocrystal States of Benzoic Acid

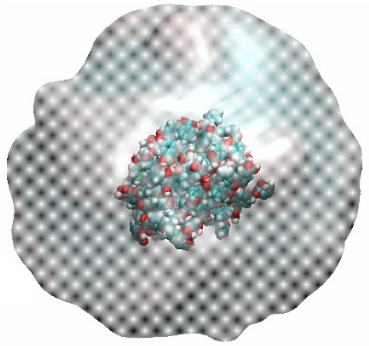
Leonardo Lo Presti,[1,2,*] *Silvia Rizzato*[1,2] *and Angelo Gavezzotti*[3]

1 Department of Chemistry, Università degli Studi di Milano, Milano (MI), Italy

2 Istituto Nazionale di Fisica Nucleare (INFN), Laboratori Nazionali di Frascati, Frascati (Roma) Italy

3 Formerly Professor of Physical Chemistry, Università degli Studi di Milano, Milano (Italy)

Table of content graphics: The KB algorithm enhance cohesion in a nanodroplet of benzoic acid embedded in methanol



ABSTRACT

A kinetically biased Molecular Dynamics (KB-MD) algorithm is developed as an addition to the MiCMoS (Milano Chemistry Molecular Simulation) package. Within a condensed medium, the algorithm sorts out molecular pairs coupled by a strong interaction energy and reduces their kinetic energy by a damping factor, redistributing the resulting excess among other partners within the medium. The aim is to enhance in an iterative manner the incipient intermolecular cohesion, on the way to the formation of recognition aggregates. The algorithm is applied to bulk liquid and crystalline benzoic acid, to homogeneous solutions in methanol, and to liquid or crystalline nanoclusters embedded in methanol solvent. Favourable outcomes are observed in liquid media with formation of large molecular clusters, and in the enhancement of the lifetimes of nanocrystals. Homogeneous solutions are found to require extremely long simulation times to show significant aggregation. Organization into a crystalline structure from liquid precursors is still a faraway simulation goal, but the present approach can be a useful tool, along with the introduction of appropriate collective structural variables, for tackling this long-standing problem at atomic level.

1. INTRODUCTION

How does crystallization from solution begin? Despite recent extraordinary advances in nano-diffractometry,[1] an answer to this apparently simple question is out of reach for current experimental techniques, which are blind to transient supramolecular aggregates that may occur in the no man's land between a homogeneous liquid state and a suspension of crystal nuclei. This is a fundamental problem, because catching the elementary steps of self-assembling is the preliminary condition to understand – and possibly control – the behavior of crystalline matter at the molecular scale.

A predictive theory of crystallization has never been developed, and the general understanding of crystallization phenomena is still fragmentary. So far, thermodynamics and kinetics of nucleation are rationalized within two main frameworks. Classical Nucleation Theory (CNT)[2] suggests an activated single step mechanism, in which well-formed crystalline cores appear abruptly in supersaturated solutions. Despite its conceptual simplicity and wide applicability, CNT largely overestimates the nucleation rates and cannot explain the possible non-monotonicity of kinetic curves at large supersaturations.[3] The Multi-Step Nucleation Theory (MSNT) is a further elaboration, suggesting that nucleation might be preceded by solute-solvent phase separation.[4] The crystal nucleus then develops either in the core of stable liquid-like droplets of solute, or somehow from metastable liquid-like embryos, and is initiated by a density fluctuation.[5,6] Experimental evidences support the validity of MSNT in some systems where CNT fails.[7]

A deeper understanding requires an atomistic description of the very elementary acts that lead to nucleation. Does the solvent favor specific self-recognition modes, or does it provide only the environment that sets the spinodal boundaries? Does the transition from liquid-like embryos to crystalline nuclei start from a point fluctuation, or does it involve a collective structural rearrangement of the whole embryo? Is the polymorphic composition of the nuclei the same as in macroscopic crystals? How do macroscopic variables (temperature, concentration and molecular mobility) influence aggregation phenomena at the molecular scale? So far, no solid clues are available to infer a general picture, leaving space for the unwelcome hypothesis that each system may require its own answers to these very fundamental questions.

In absence of reliable experimental approaches, Molecular Dynamics simulation can shed light on the events that initiate aggregation phenomena. Its application is limited by the timescale of

real nucleation, that could take at least seconds (10^{15} 1-fs steps) or of crystallization, that may require weeks (6 x 10^5 s or 10^{20} 1-fs steps). If a core MD code requires 10^6 flops per step for a system of 10000 atoms, a presently available petaflop computer can handle nucleation in about 280 hours but is still a universe away from crystallization. One could wait for hexaflop computers to come along, but anyway the restriction of simulation progress to the small elite that can have access to supermachines would be improper. When the above figures are scaled to teraflop computers available in normal theoretical chemistry environments, a 1 s "real world" simulation would still take years. Hence the imperative need for biasing and shortcut methods.

Several enhanced sampling techniques[8] are available to speed up the dynamics of processes like crystal nucleation.[3,9] Some methods rely on the definition of collective variables, which are used to drive the system through regions of phase space that would be hardly accessed spontaneously. The potential energy distribution is biased by perturbing the system's Hamiltonian in a way that reduces the height of the barriers to be surmounted by the available kinetic energy, so that the system can explore more quickly and efficiently the potential energy surface. The preparation of appropriate collective variables in terms of simpler chemical or structural concepts is the stumbling point of this approach. Other enhancing methods include the labeling of phase space regions already visited without success, steering the system away from repetitions of such unproductive efforts.[10]

The recently developed MiCMoS platform[11–14] provides several flexible and cheap tools to investigate the evolution of small organic molecules in condensed phases, relying on accurately calibrated intermolecular force fields, having been successfully applied to investigate liquids, crystals, nanodroplets and nanoparticles. It is being upgraded with a novel biasing procedure, as follows. Phase space consists of position (x, y, z) and momenta, or velocities (v_x, v_y, v_z). Rather than acting on positional paths, we explore in this contribution the possibility of biasing molecular velocities, by an energy-dependent redistribution of kinetic energies, *i.e.* slowing down those molecular couplings that show promise for more effective cohesion. Like all biasing algorithms, this KB-MD approach leads molecules through one of the possible paths in which they behave as they are expected to, although it obviously cannot guarantee that it is the path taken in reality. The resulting simulation, like all artificially biased ones, trades some physical rigorousness for chemical information not otherwise obtainable. A major advantage of KB-MD

is that its implementation is straightforward, while application conditions can be tuned for a minimal disturbance of the continuity of the trajectories.

Benzoic acid (BZA) is adopted here as a model entity, well characterized in structure and thermodynamics, and having an easily identifiable intermolecular linkage in its hydrogen-bonding ability. Some Benzoic acid derivatives have been the subject of extensive computational and experimental studies of nucleation.[15,16] Our test systems are pure liquid and crystalline BZA, a benzoic acid solution in methanol, and liquid and crystalline nanoclusters of benzoic acid solvated in methanol. The results prove that the newly devised biasing algorithm provides a viable and promising tool for a faster steering of molecular recognition through phase space, from liquid or disperse systems towards the formation of compact aggregates. A number of relevant and less expected features along the evolution paths also emerge.

2. METHODS

2.1 Definition of relative energies, forces and velocities in a molecular medium

Let the final frame of a molecular simulation contain M molecules of N atoms each of mass m(i), at position **x**(i), with velocity **v**(i) and acting force **f**(i) (boldface denote vectors). The molecular mass is $W = \sum_i m(i)$. The relationships between distances, energies, velocities and forces between pairs of molecules are analyzed as follows. Let **X**(i) be the center of mass (c.o.m.) position of the i-th molecule, and **V**(i), **F**(i) be the velocity and force vectors at the center of mass:

$$\mathbf{X}(i) = \sum_i m(i)\frac{\mathbf{x}(i)}{W}, \;\; i = 1, N \tag{1}$$

$$\mathbf{V}(i) = \sum_i m(i)\frac{\mathbf{v}(i)}{W}, \;\; i = 1, N \tag{2}$$

$$\mathbf{F}(i) = \sum_i m(i)\frac{\mathbf{f}(i)}{W}, \;\; i = 1, N \tag{3}$$

"Coupled" or paired molecules are defined as when either their cohesive energy, $E(i,j)$, or their distance between c.o.m.'s, $R(i,j)$, is below a given threshold. $N_p(t)$, the number of detected pairs as a function of time, is a prime indicator of ongoing aggregation. The relative velocity and force vectors, **V**(i,j) and **F**(i,j), are obtained by taking the difference between the projections of the respective single vectors along the line joining the two c.o.m.'s. The construction is such that

molecules are coming closer when V(i,j) > 0, and that a force F(i,j) > 0 is repulsive (see Supporting Information, part S1).

2.2 The KB-MD (kinetic-bias Molecular Dynamics) approach

The background of this approach stems from the molecular level proceedings when a pure liquid is on its way to freezing, or when a solute crystallizes out of a solution. Before the crystallization conditions (temperature and saturation) are met, all solute and solvent molecules have an equipartition kinetic energy, and the system is characterized by a solute-solute, solute-solvent and solvent-solvent (or solvation) potential energy, called *E*(uu) and *E*(uv), both per molecule of solute, and *E*(vv), per molecule of solvent: *E*(uu) is small at low concentration in solution, *E*(uv) and *E*(vv) are zero for a monophase. When the crystallization bell "rings", aggregating molecules shed some kinetic energy as they get in potential-energy coupling with their homologues, and the process is self-enhancing until a conglomerate is obtained. In a pure liquid phase, molecules are expected to condense into tighter and tighter clusters; in solution *E*(uu) increases sharply at the expenses of *E*(uv). In either process, the excess kinetic energy is redistributed in a complex manner across the system, or eventually to the surroundings, as “latent heat”. Rather than biasing the potential energy surfaces, the present biasing scheme reduces the kinetic energy of molecular pairs that show a propensity for aggregation. The aim is to prevent stable pairs/clusters from being destroyed by random thermal fluctuations, thus increasing their lifetime and their chances to grow in size.

Being E_{ij} the intermolecular energy of the *i-j* molecular pair and E_{bu}, E_{bl} the user-selected upper and lower limits for biasing, a scaling factor $g \leq 1$ is computed as

$$g = 1 - \frac{\left|E_{ij} - E_{bu}\right|}{\left|E_{ij}\right|}, \qquad E_{bl} < E_{ij} < E_{bu} \tag{4}$$

Outside the [E_{bl}, E_{bu}] interval $g = 1$. The larger the $\left|E_{ij} - E_{bu}\right|$ difference, the smaller is g; when $E_{ij} << E_{bu}$, $g \to 0$, while $g \to 1$ when E_{ij} approaches E_{bu}.

The kinetic energies, T_i and T_j, of molecules involved in an attractive pair are scaled according to:

$$T_i^b = g \cdot T_i \tag{5a}$$

$$T_j^b = g \cdot T_j \tag{5b}$$

For simplicity, we refer just to the i^{th} single molecule hereinafter, the same operations being repeated also for the j^{th} molecule. The kinetic energy change due to the bias is

$$\Delta T_i = T_i^b - T_i = \ g \cdot T_i - T_i = T_i \cdot (g-1) \tag{6}$$

with $\Delta T_i < 0$ as $g < 1$; no change is made in time intervals where the bias does not apply or for molecules outside the selected energy range. ΔT_i is equally distributed across the N_i atoms in the molecule i by the following sequence, so that each atom a of i loses an amount of kinetic energy $\Delta T_{i,a}$:

$$T_i = \sum_a^{N_i} T_{i,a} = \sum_a^{N_i} \frac{1}{2} m_{i,a} v_{i,a}^2 \tag{7}$$

$$\Delta T_{i,a} = \frac{\Delta T_i}{N_i} = T_{i,a}^b - T_{i,a} < 0 \tag{8}$$

$$T_i^b = \sum_a^{N_i} T_{i,a}^b = \sum_a^{N_i} \frac{1}{2} m_{i,a} v_{i,a,b}^2 = g \cdot T_i \tag{9}$$

$$\sum_a^{N_i} \frac{1}{2} m_{i,a} v_{i,a,b}^2 = g \cdot \sum_a^{N_i} \frac{1}{2} m_{i,a} v_{i,a}^2 = \sum_a^{N_i} \frac{1}{2} m_{i,a} g\left(v_{i,a}^2\right) \tag{10}$$

The final result is:

$$v_{i,a,b}^2 = g\left(v_{i,a}^2\right) \tag{11}$$

$$\pm v_{i,a,s}^b = \pm\sqrt{g} \cdot v_{i,a,s} \ , \quad s = x, y, z \tag{12}$$

The kinetic energy reduction ΔT_i must be equally redistributed to all the N_k non-biased solute and solvent molecules to conserve the total kinetic energy of the ensemble. Thus, the k^{th} molecule gains $\Delta T_k = \ -\Delta T_i \ / N_k > 0$, and the rescaling on the k^{th} non-biased molecule goes by a coefficient, B, as:

$$T_k^b = T_k + \Delta T_k = B \cdot T_k \tag{13a}$$

$$B = 1 + \frac{\Delta T_k}{T_k} \tag{13b}$$

$$\pm v_{k,a,s}^{new} = \pm\sqrt{B} \cdot v_{k,a,s} \quad , \quad s = x, y, z \tag{14}$$

This KB-MD correction is applied with a specified step frequency and can be repeatedly switched on/off for predefined time intervals. The same molecule might be biased more than once, if it is involved in multiple interactions that fall in the selected energy range, but in any case, internal checks ensure that the total kinetic energy, T_{tot}, is conserved within numerical accuracy. As the average kinetic energy per degree of freedom, T_{tot} / N_{DOF}, is also conserved, this procedure is consistent with the equipartition theorem and ensures that the kinetic temperature T of the system is not affected:

$$T = \frac{2 \cdot T_{tot}}{k_b \cdot N_{\mathrm{DOF}}} \tag{15}$$

The overall outcome is a drive toward the formation of (meta)stable clusters that might evolve into larger and more complex structures, like nanodroplets and crystal embryos. Clear advantages of our procedure are that its implementation is straightforward, and that no collective variables need be defined, dispensing with symmetry or orientational constraints on the equations of motion. The procedure now features in the latest release of the MD module in MiCMoS.[17]

2.3 Description of the computational systems

The choice of benzoic acid (pure or solute) and methanol (solvent) was dictated by the relatively predictable dimerization mode of the solute, by the modest size of the solvent, that nevertheless allows some hydrogen bonding solute-solvent competition, and by the excellent performance of the atom-atom LJC (Lennard-Jones-Coulomb) potential scheme[12] on such common organic compounds. File manipulations to prepare our model systems were carried out within the framework of the MiCMoS v2.0 suite of programs, and the description of the necessary modules (cited here in *Italic*) is available in Supporting Information (part S2) and, in more detail, in the MiCMoS manual.[17]

The benzoic acid topology was built from the experimental crystal structure at 300 K ($P2_1/c$; Cambridge Structural Database refcode BENZAC02);[18] a model for methanol was prepared with standard geometry **(details of the geometries and force field parameters for both molecules are in SI, part S3).** For all benzoic acid and methanol systems, in MC the only variables were

the C–COO and C–OH bonds treated as rotatable, the rest of the molecule being rigid; MD runs under the full inter- and intramolecular MiCMoS force field (*Pretop*). Periodic boundary conditions apply in MC and MD, except for isolated clusters or nanodroplets. In MD, a leap-frog integrator at time steps of 1 fs applies, temperature is controlled by a weak coupling thermostat (relaxation time τ = 0.6 ps) and pressure (when applicable for periodic systems) by an isotropic Berendsen barostat (compressibility coefficient $\mu_0 = 0.4$ atm^{-1}, reference pressure 1 atm).

Intermolecular energy summations run on entire neutral molecular units to ensure convergence of Coulombic terms. The cutoff of 15-16Å then applies between molecular centers of mass. Cutoffs have been calibrated together with force field parameters to be as short as possible, compatibly with plateau performance. They are in a way a measure of the force field range, so that using shorter values would be improper and we trust that using higher values would not change the results.

Benzoic acid monophases. The pure benzoic acid liquid was modeled by a box of 432 molecules **of same topology as above** (*Boxliq*), equilibrated by MC at 350 K (melting point = 395 K) and then subject to MD runs of up to 200 ps.

Homogeneous solutions. Liquid boxes of 432 benzoic acids and 2000 methanols were generated (*Boxliq*) and equilibrated by 1 or 2 million Monte Carlo steps at ambient conditions. Densities and cohesive energies agree with available experiment for these species that were among the calibration standards of the LJC potential scheme. The solute and solvent boxes were merged (*Solution*) to produce a cubic box containing a homogeneous 0.5-molar solution. Dynamics were run for ~ 190 ps.

Liquid nanodroplets of benzoic acid in liquid methanol. An isolated spherical liquid nanodroplet of solute (116 molecules) was produced (*Excbox*) from the MC-equilibrated box of liquid benzoic acid. The droplets were embedded in MC-equilibrated methanol (*Nanosolv*, *Excbox*) so that the final system had an approximate spherical shape, with a final number of 1420 solvent molecules. The MD trajectories were 150 ps long.

Crystalline nanoparticles of benzoic acid in liquid methanol. A benzoic acid crystallite with 150 molecules was cut (*Nanocut*) out of the experimental periodic structure. The BFDH morphology[19] predicted by Mercury[20] was chosen to select the (*hkl*) boundaries. Then, the crystallite nanoparticle was embedded in equilibrated liquid methanol (*Nanosolv*) cutting out

molecules at corners of the solvent box (*Excbox*) to produce roughly spherical nanodroplets. The final model consisted of 1246 methanol + 150 benzoic acid, without periodic boundary conditions.

Unless otherwise specified, the bias was applied every 10 MD moves and never switched off. For systems without periodic boundaries, the tethering algorithm (tethering distance: 40 Å, scaling factor: 0.98) was adopted to deal with evaporation, while default parameters were selected to suppress spurious cluster rotational momentum. Detail of these last two procedures can be found in the MiCMoS documentation.[13,17]

2.4 Reproducibility

Molecular geometries, force fields, and MD input files are deposited in the Supporting Information part S3. Starting and final simulation boxes in dat format can be obtained from the authors upon request. The MiCMoS package can be downloaded at no charge upon registration from https://sites.unimi.it/xtal_chem_group/, under the specified citation and use conditions. This ensures complete reproducibility of all the calculations here described.

3. RESULTS

3.1 Pure liquid benzoic acid

Prior to the discussion of long MD simulations and of the effects of kinetic bias, a survey was conducted on an equilibrated frame of liquid benzoic acid, exploring the relationships between the single-molecule and pairing quantities described in section 2.1. The limiting values imposed in the analysis of molecular pairs were $E(i,j) < -15$ kJ mol^{-1} or $R(i,j) < 4$ Å.

Figure 1 shows the plot of energy *vs*. distance, with a clear separation between ***a) hydrogen bonded pairs, whose energy goes from -30 to -80 kJ mol^{-1} depending on formation of single or double-cyclic O-H...O bonds over the carboxyl function, respectively, and b) pairs related by dispersion whose energy is -20 to -10 kJ mol^{-1} and*** whose distance can be as short as the ring stacking distance of about 3.5 Å.

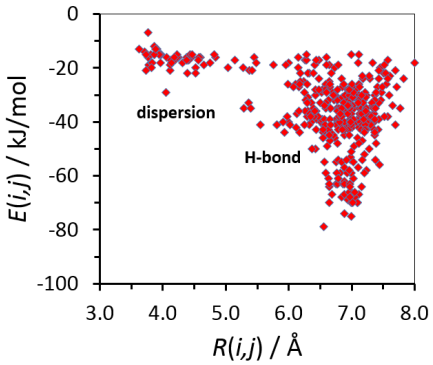


Figure 1. Molecule-molecule energies as a function of molecule-molecule distance in the benzoic acid liquid droplet at 350 K.

Figure 2a shows the distribution of absolute velocities of the two partners in each pair. The order of magnitude is 100-500 $m \cdot s^{-1}$, the correct range of about 1/10 of that of gas molecules. The distribution is uniform without reduction for more cohesive pairs. This is the feature on which the bias is going to act, cutting down the kinetic energy of molecules which are engaged in more cohesive interactions.

Figure 2b shows the plot for the relative velocity and force against cohesive energy. The forces are of the order of 30 $kJ \cdot mol^{-1} \cdot A^{-1}$ or 6 x 10^{-10} N = 0.6 nN (nanonewton). These are of the expected order of magnitude, as estimated by accurate AFM measurements of intermolecular forces on biological samples.[21,22] The only bias appearing in this plot is that a high cohesion energy brakes the relative velocity of separation ($\mathbf{v}(i,j) < 0$), but does not restrict the velocity of closer approach ($\mathbf{v}(i,j) > 0$), in agreement with dynamics and common sense. The plot of relative velocity against c.o.m. distance (Supporting Information, part S4) shows instead a total dispersion.

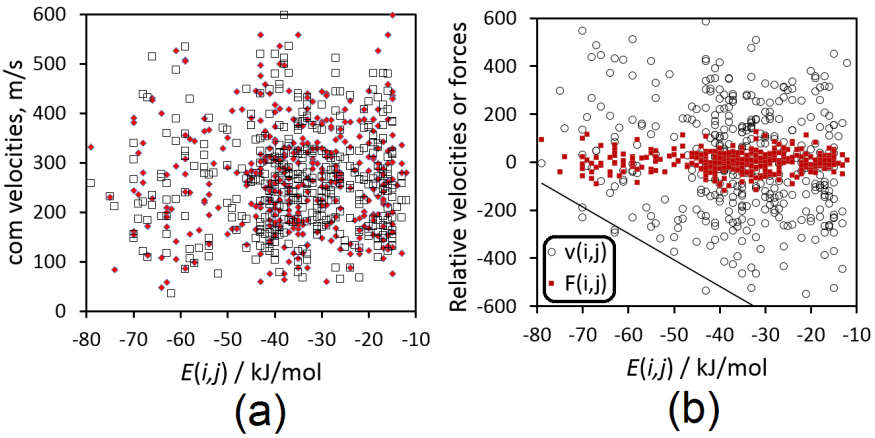


Figure 2. (a) Center-of-mass absolute velocities ($m \cdot s^{-1}$) of the two partners (a dot and a square) in a molecular pair as a function of cohesion energy. (b) Relative velocity ($m \cdot s^{-1}$) and force (10^{-10} nN) between pairs of molecules. Molecules approaching when $\mathbf{V}(i,j) > 0$, and a force $\mathbf{F}(i,j) > 0$ repulsive (see Scheme 1, Supporting Information part S1). The zone below the bar is strong stabilization braking pair separation.

3.2 Kinetic bias on pure liquid

The biased simulation of benzoic acid pure liquid was carried out with bias applied in the $-20 \leq E(i,j) \leq -15$ kJ mol^{-1} range (see Eq. 4) every 250 MD moves. The collective variables that help describing the aggregation modes are hydrogen-bond "clusters", ensembles of any shape in which molecules are continuously linked by COOH···O contacts shorter than the sum of atomic radii (2.68 Å), and hydrogen-bond "cycles", where molecules are connected as above but are organized in a cyclic fashion. ***Cycles can be substructures embedded in larger clusters, or can stand alone, in which case they are also counted as clusters.*** These features are recognized in a new module (*aggreginv*) developed *ad hoc*.

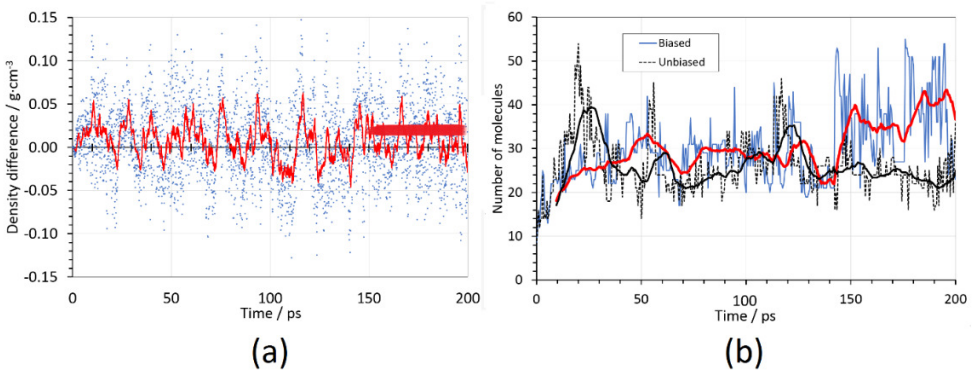


Figure 3. MD simulations of the BZA liquid **at T = 350** K. (a) Step-by-step difference of bulk density $D = d$(biased)–d(unbiased). The red line is the rolling average over a period of 20 fs, and the heavy red streak denotes the period of net difference increase. (b) Time evolution of the maximum number of molecules in HB clusters. The black (unbiased) and red (biased) curves are interpolations provided as guides to the eye.

The biased run produces a small but significant increase in density (Figure 3a, about +0.014(2) $g \cdot cm^{-3}$). This bulk difference mirrors a tendency to aggregation in the inner structure of the biased liquid (Figure 3b): both runs show wide oscillations in the maximum number of molecules in a cluster, but the biased simulation eventually shows a net tendency to formation of larger clusters. Figure 4 shows a wider visualization of the evolution of cluster sizes in simulation time: the biased run shows a number of large clusters ($N > 40$) towards the end of the simulation, while continuous columns denote cluster families able to survive for longer times. On average, the kinetic bias has a stabilizing effect on clusters of intermediate ($20 < N < 40$) size.

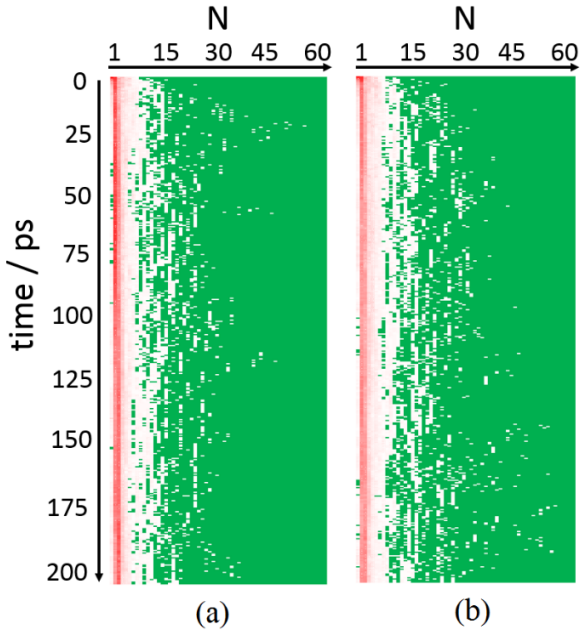


Figure 4. Time evolution of the distribution of HB clusters in liquid benzoic acid at 350 K (200 ps at time steps of 0.001 ps). (a): unbiased run. (b): biased. *N* is the total number of molecules contained in the cluster (cluster dimension). Each (*N*, *t*) point is colored according with the occurring frequency of clusters of dimension *N*. Red: high frequency (> 5 up to ~60, the hotter the higher); white: low frequency (from 1 to 5). Green pixels mark 0 frequency (no clusters of that dimension at time *t*)

More insight can be gained by analyzing the stability trends of aggregates of different size (examples are given in Figure 5, more detail for all cluster sizes in Supporting Information, part S5). The number of monomers drops very quickly to only ~ 1-2 % of the total in both simulations, presumably under the action of pure dynamics, but the introduction of the kinetic bias leads to a significant increase in the number of trimers. Table 1 shows a compendium in approximate numbers of the essential results on cluster size evolution.

The number of dimers temporarily increases in the unbiased simulation, but eventually both unbiased and biased trajectories come to the same end, namely a large decrease of dimers. The unbiased simulation includes a larger proportion of trimers, and also a minor increase in medium size, 5-10 member clusters. Large clusters ($N > 10$) are a minority in both simulations, but the biased simulation has more very large aggregates (up to $N = 55$, see Figure 6) than the unbiased simulation, and these become increasingly frequent at larger times. On the contrary, very large aggregates form very quickly ($t < 50$ ps) in the unbiased simulation (Figure 3b), but they are rapidly destroyed and become less and less frequent as the simulation proceeds. These very large aggregates may be considered as proto-droplets within a high density liquid phase, a form of density fluctuation that might precede nucleation.

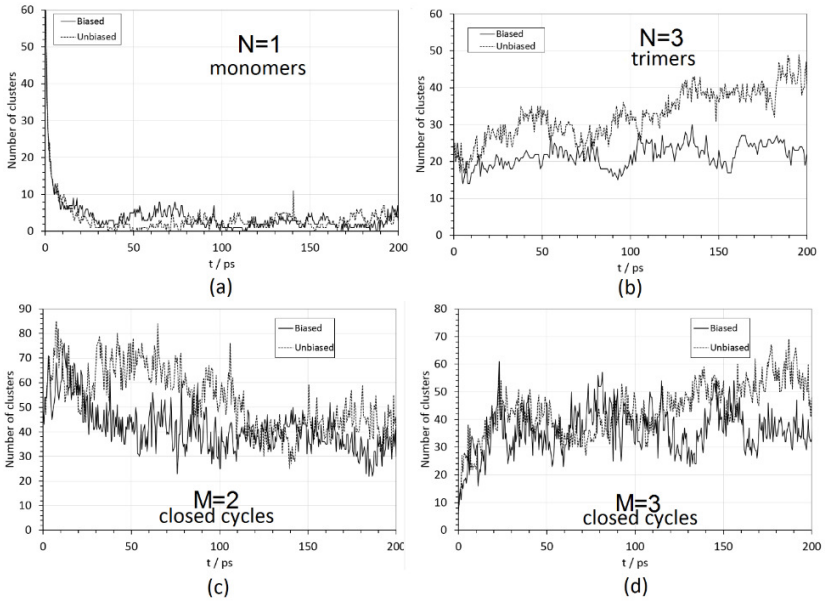


Figure 5. (a-b) Evolution of the number of monomers and trimer clusters in the biased and unbiased simulation of liquid benzoic acid. (c-d) The total number of cyclic dimers and trimers containing *M* molecules in liquid BZA at 350 K as a function of time. Comparison between unbiased and biased MD runs.

Table 1. Evolution of *n*(*N*), the number of clusters of size *N*, during the biased and unbiased simulations. Zero time (start) is the MC equilibrated liquid.

N range	*n*(*N*) start	*n*(*N*) end, unbiased	*n*(*N*) end, biased
1	68	3	4
2	58	24	24
3	20	41	30
4	18	14	11
5-10	20	18	18
>10	0	5	8

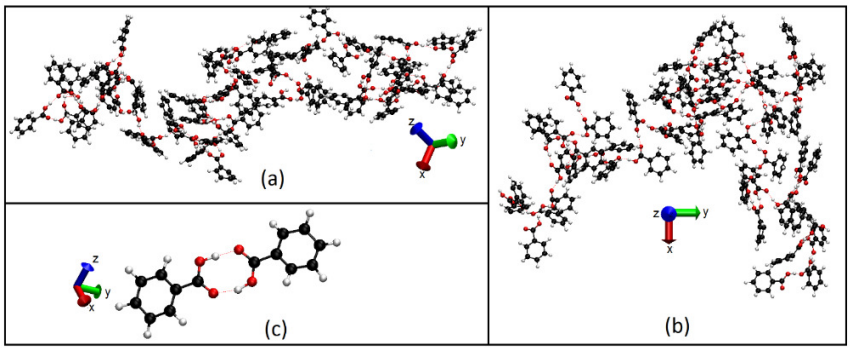


Figure 6. Benzoic acid hydrogen-bonded clusters found in the first 200 ps of the MD trajectories. (a) 350 K biased liquid, t = 176 ps, 55 molecule-large cluster with an approximate volume of 10.9 nm^3 (b) 350 K liquid, unbiased run, t = 20 ps, 54 molecule-large cluster with approximate volume of 10.4 nm^3. (c) Cyclic dimer in the benzoic acid crystal, 298 K. Red dashed lines highlight geometry-defined hydrogen bonded contacts.

Inspection of the number and frequency of cyclic clusters may provide further hints on the molecular recognition process. The basic packing unit in crystalline benzoic acid is a cyclic dimer of inversion-related molecules, as expected in carboxylic acids.[23,24] In liquid benzoic acid, cyclic arrangements of –COOH groups are set up either as dimers or as part of larger aggregates (an example is given in Supplementary Information, part S6). A set of M molecules is said to form a cycle if the M^{th} one connects the first one through a H···O hydrogen bond; the *total* number of cyclic clusters, isolated or merged, was monitored. This number (Figure 5c,d) is about the same in both liquids at equilibrium, while the number of cyclic trimers follows the same trend discussed in general for N=3 aggregates. Thus, the biased simulation tends to develop, on average, less clusters with larger dimensions.

The former simulations suggest that the cyclic dimer is neither the only nor the more frequent self-recognition mode at equilibrium in liquid benzoic acid. These results issue a warning in applying chemical common sense when trying to predict the lowest hierarchy of supramolecular nucleation and growth units just on consideration of the most tightly bound motifs observed in crystal phases. Rather than well characterized dimer precursors, liquid benzoic acid aggregates are dynamic entities that continuously switch among complex branched motifs; the largest ones grow by annexing clusters of lower dimensions in their bulk structure and are all but an ensemble of dimers (Figure 6). Such high-density droplets may well be the entities postulated in the two-

stage nucleation model, by which the next kinetic step should be a collective reorganization, a structure fluctuation that leads to a nucleus resembling the crystal structure. In this model, dimers can be hardly considered as basic supramolecular building units of the crystal, as the main kinetic step is not dimerization, but rather coalescence of denser HB clusters in the bulk liquid matrix.

Table 2 collects the cohesive energies per molecule of the largest clusters found in biased and unbiased simulations, along with those from MD of the bulk liquid and bulk crystal. ***Energies per molecule are obtained by dividing the total configurational energy of a given molecular ensemble by the number of molecules in the ensemble. As a first approximation (neglecting the minor contribution of kinetic energy differences and assuming zero interaction in the vapor phase), stabilizing energies per molecule in a bulk phase simulation should compare with vaporization or sublimation enthalpies, while those in a small cluster are just a measure of cohesive strength.***

The enthalpies of vaporization and sublimation of benzoic acid **(professionally averaged values in webbook.nist.gov/chemistry) are 7**9 and 90 kJ mol^{-1}, respectively and they compare favorably with calculated bulk cohesive energies of liquid and crystal. The first relevant feature in the data of Table 2 is that the Coulombic contribution is constant irrespective of size and structure, presumably because this energy is mostly if not exclusively localized on the hydrogen bonds. Interestingly, the isolated in-crystal cyclic dimer comes up with a repulsive dispersive balance, likely due to the proximity of the two carboxyl groups, held together by a very strong pair of hydrogen bonds. The attractive dispersive term in clusters and bulk structures, on the contrary, is due to collective and non-local interactions among whole molecular charge densities, and its variation on going from liquid clusters to bulk liquid and crystal is spectacular. On the one hand, this result suggests that the clusters found in our simulations are still a long way from actual condensation into a compact nucleus; on the other hand, it appears that reorganization into a micelle or even a crystal nucleation precursor is driven by non-directional forces having nothing to do with hydrogen bonding. This is in keeping with the crystal structure of benzoic acid, that does not show HB networking but consists of a juxtaposition of self-contained HB dimers.

Table 2. MD-derived cohesive energies per molecule (E_{coh}), and their decomposition into Lennard–Jones (E_{LJ}) and Coulomb (E_c) terms, of BZA clusters and bulk phases. All values in kJ·mol^{-1}.

Cluster	E_{coh}	E_{LJ}	E_c
55mol, biased liquid[a]	–60.6	–5.3	–55.3
54mol, unbiased liquid[a]	–61.3	–2.8	–58.5
2mol, cyclic, crystal[b]	–47.3	–7.8	–55.1
Bulk liquid, biased[c]	–77(3)	–22(5)	–55(5)
Bulk liquid, unbiased[c]	–76(3)	–23(5)	–53(5)
Bulk crystal[c]	–88(1)	–38.9(9)	–49(1)

[a] Largest clusters found in the first 200 ps of each simulation (Figure 3).

[b] Typical dimer from the last frame of the crystal simulation.

[c] Average (esd) over the 100-200 ps interval.

3.3 Homogeneous solution of 43 benzoic acid + 1520 CH_3OH (0.5M)

The study of a homogeneous solution in which solutes are distributed at random and are widely separated is a challenging enterprise, because the mutual forces are weak and presumably very long timescales are implied for coalescence of solutes. We present here only an **assay** with different biasing protocols. Table 3 shows that aggregation is minimal within the afforded timescales, with very small numbers of solute HB formed; the large *E*(uv), or solvation energy per solute molecule, indicates a strong, multiple hydrogen bond grip of the solvent on the solutes. The biased simulation does however produce a larger solute-solute interaction, although this is but a fraction of the carboxylic acid HB energy. A longer simulation (500 ps) produces higher *E*(uu) and also a stronger solute-solvent interaction.

Table 3. Last frames of the simulation of a 0.5 molar solution of benzoic acid in **CH_3OH**. The number of HB's is the number of =O···H–O contacts below 90% of the atomic radii. Energies in kJ mol^{-1}.

	ps	*E*(uu)	*E*(vv)	*E* (uv)	solute disp.	solute Coul	solute HB's
Biased, $E_b < -10$	100	-13.4	-38.9	-121.1	-3.7	-9.7	7
Unbiased	500	-5.7	-36.0	-111.1	-0.9	-4.9	3
Biased $-20<E_b<-10$	500	-18.5	-39.6	-135.7	-8.6	-9.8	6

Due to dispersion in the solvent matrix, monomers are the most common form in solution. A small number of dimers and trimers is present, while cyclic structures are just sparse. In general, the structural differences between biased and unbiased liquid phases are minimal. Thus, there seems to be no indication that the most stable interaction mode found in the crystal (cyclic HB dimer) forms by recognition in liquid methanol, also because the solvent can favorably interact with the solute, preventing the formation of stable cycles. On the other hand, phase separation could provide a more favorable environment for benzoic acid self-recognition. To check this hypothesis, that is, the possible next step along the nucleation path, a simulation of the dynamics of liquid nanodroplets of benzoic acid in methanol was undertaken.

3.4 Liquid droplet of 116 benzoic acid molecules in 1420 CH_3OH solvents

The main purpose of this simulation was to test the system rearrangement under the kinetic bias drive once phase separation had occurred in the solution at nanoscopic level. A relevant feature of the KB algorithm is that the user can select the energy range in which the bias is applied (Eq. 4). The benzoic acid crystal structure consists of slanted columns of tightly **bonded** HB dimers (Table 2), whose individual cohesive energy lies in the –45 to –50 kJ mol^{-1} range. As expected, this most stable interaction mode appears already in the unbiased liquid (Figure 5c). Biased runs where the E_{bu} limit was set at –80 kJ mol^{-1}, about twice the cohesive energy per molecule, had no **significant** impact on the dynamical evolution of the system, probably because the electrostatic drive toward cyclic HB's is strong enough by itself to allow the formation of stable cyclic dimers. The second highest-ranking interaction in crystalline benzoic acid is the stacking of aromatic rings, occurring between inversion–related molecules at low centre of mass distance (4.98 Å) with a cohesive contribution of –20 kJ mol^{-1}, largely by dispersion. This specific interaction was selectively enhanced by setting the bias boundaries at –20 and –10, thus

aiming the bias at the secondary packing factor in the crystal. A higher bias frequency (10 MD steps, 10 fs) than in the simulations of liquid was also employed to speed up the process.

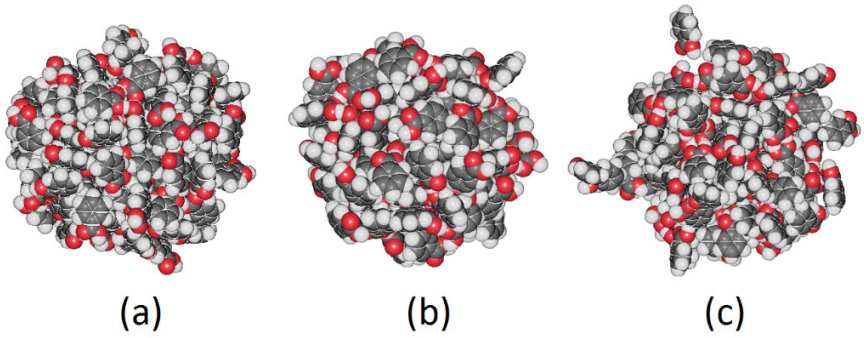


Figure 7. First and last frames of liquid droplet trajectories for 116 benzoic acid + 1420 **CH_3OH** (solute only). (a): First frame, equal for both the simulations. (b): last frame, biased. (c): Last frame, unbiased.

Figure 7 shows the first and last frames of the MD simulation for a liquid droplet of benzoic acid in methanol. The last frame of the biased simulation is more compact also on visual inspection, while the last frame of the unbiased simulation shows also some incipient transition of solute molecules into the solvent. The use of rotational correlation functions and of radial distribution functions is problematic for such a small system of only 116 molecules. Anyway, biased and unbiased simulations show a sharp peak for the OH···O= hydrogen bond at around 1.8 Å, but the number of short O···H contacts in the unbiased simulation (79) is larger than that in the biased simulation (61). Table 4 shows that the cluster after bias is more compact and overall, more stable; it is also more oriented to space-filling interactions, as confirmed by the higher dispersion energy and smaller Coulombic term. The kinetic bias seems to be acting in a bulk-volume, non-directional manner, as expected when setting the boundary conditions of the bias.

Table 4. Last frame of the simulation of a liquid droplet of 116 solutes in 1420 solvents. **Solute-solute (uu), solvent-solvent (vv) and solute-solvent (uv) total, Lennard-Jones and Coulombic** energies per molecule ($kJ \cdot mol^{-1}$).

	*E(*uu)	*E*(vv)	*E*(uv)	solute LJ	solute Coul.	n. of H-Bonds

biased	-63.7	-46.3	-58.7	-25.0	-38.8	61
unbiased	-59.6	-40.0	-53.8	-15.6	-42.8	79

Figure 8 shows that the kinetic bias performs as expected, producing a more compact cluster with a large increase of short centre of mass contacts, and leading to a minor but significant improvement of the cohesive energies. In a further confirmation, the distribution of pair energies against distance (Figure 9) reveals a number of short contacts in the 4.5-6.5 Å range, with smaller pairing energies than the hydrogen bonding ones. The application of the kinetic bias has a small effect on the distribution of *absolute* centre of mass velocities (Supporting Information, part S7), presumably because it acts as an instantaneous perturbation at every specified number of moves and the system can re-equilibrate the velocities when the bias is not active. A difference appears in the distribution of *relative* velocities *vs.* centre of mass distances (Supporting Information, part S7), confirming that under the bias molecules are packed more closely, the intermediate $R(i,j)$ range being more populated than in the unbiased simulation.

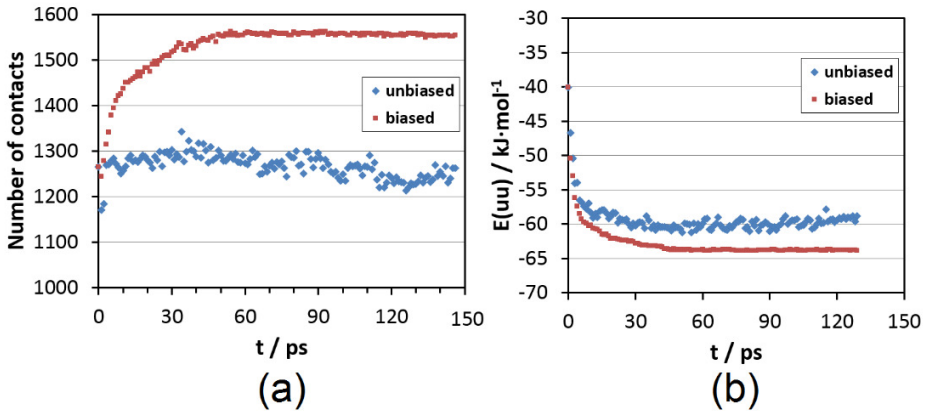


Figure 8. Comparison of the evolution of the number of com-com distances < 12 Å (a) and of solute-solute energies (b) during the biased and unbiased simulation of a liquid droplet of 116 solutes in 1420 solvents.

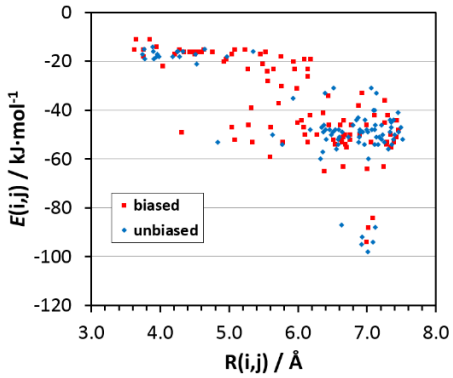


Figure 9. Plot of pair energy *vs*. com distance. The biased simulation reduces the number of strong hydrogen bonds in favor of closer dispersive interaction in the 4.5-6.5 Å range.

The figure in Supporting Information, part S8, shows the evolution of the number of the crucial cyclic dimers in the biased and unbiased simulation of the embedded liquid nanodroplet. Both trajectories conclude at about 10 such cycles after some oscillation. The proportion of 20 engaged molecules to a total of 116 is not so different from the proportion of 80 molecules (Figure 5c, 40 dimers) to a total of 432 in the bulk liquid.

3.5 Simulations of a solvated crystal slab

There is a striking difference between the unbiased and biased runs for a crystal slab of 100 molecules embedded in 1420 methanol molecules. In the unbiased simulation the number of short contacts decreases sharply in time (Figure 10), and the crystal shape is quickly lost. This lack of stability of crystalline clusters, either isolated or solvated, has been repeatedly verified in our test calculations,[14] in a distinct if reverse support of the two-step nucleation mechanism that invokes formation of liquid like micelles as a first step in nucleation. On the contrary, the application of the KB algorithm imparts a continuing stability to the crystalline structure, with a constant number of short contacts and a preservation of crystal shape (Figure 11). Taken at face value, this result is a tautology, since removal of kinetic energy is precisely what is needed to preserve a structure from melting; however, its physical message is that a molecular aggregate evolving towards a crystalline structure must find a way of shedding thermal energy to its surroundings, something that the biasing algorithm does by simulation, in a further confirmation of its correct performance.

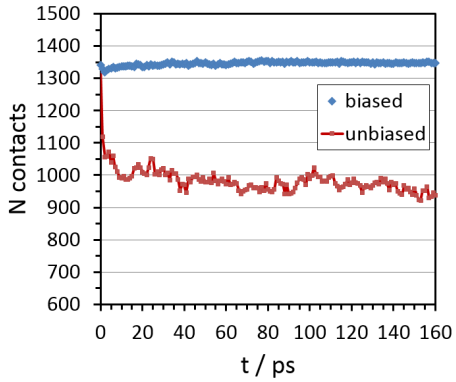


Figure 10. Number of contacts (c.o.m. distance < 12 Å) in the biased and unbiased runs for a 100-molecule crystal slab solvated in a methanol cage.

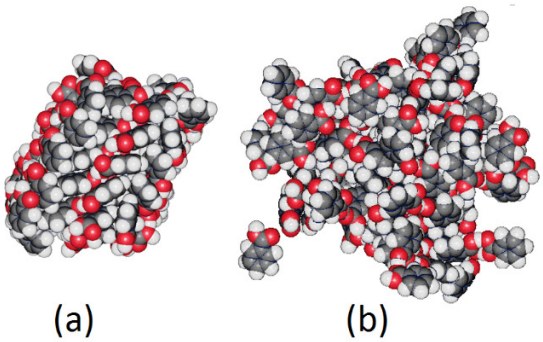


Figure 11. Last frame (solute only) of the biased (a) and unbiased (b) runs for a crystal slab of benzoic acid solvated in methanol.

4. SUMMARY AND OUTLOOK

Tracing a continuous trajectory from a liquid state to a crystalline configuration, either in bulk or dissolved phases, has been the aim of many recent computational efforts, in parallel with attempts of an experimental observation of structural or thermodynamic events at nanoscopic level. Dynamic simulation is confronted with a time and size barrier, limited as it is in length of the simulation - order of milliseconds -and number of evolving particles, where even a thousand may be short by orders of magnitude. Biasing algorithms must come into play, although all they

can do is surmount these barriers by some mathematical constraint. Biases affecting position variables to trace a path in phase space by coupling to some collective structural variable are by definition blind guesses, the argument being circular: knowing the path one would not need a bias. The approach taken in this paper is to explore the modes and the results of a bias affecting velocities rather than position. The intermolecular part of the configurational energy is partitioned over molecular pairs, something that can be done in an atom-atom potential formulation at zero computational cost; the kinetic energy of molecular pairs in selected potential energy layers is then reduced by a disposable fudge factor, in a computational encouragement to enhanced cohesion. Computational care is of course taken to re-distribute the excess kinetic energy through the whole system in a physically acceptable manner. The choice of the potential energy layer's limits and of the fudge factor is what gives the method its flexibility.

The scope of the present application of the kinetic bias is a first test of its performance over a sample of crystallization candidates or intermediates, bulk phases or droplet and nanocrystals embedded in solvent. Benzoic acid is a suitable candidate because its total cohesive energy can be easily separated into dispersive and hydrogen-bonding layers. Under the biasing restrictions, pure liquid benzoic acid condenses by increasing its density and forming hydrogen-bonded clusters of irregular shape and multiform size, without a clear preference for cyclic dimers. A solvated liquid droplet freezes to an amorphous semi-solid, into a compact form held together mainly by the dispersive potentials. A microcrystal slab preserves its symmetric stability under biased dynamic conditions, while an unbiased simulation quickly runs into a liquid drop. This last point strongly suggests (one could say demonstrates) that isolated clusters cannot be crystalline, contrary to hypotheses embedded into classical nucleation theory. The simulation of a homogeneous solution of benzoic acid in methanol is obviously too short to allow the

observation of true condensation, but some preliminary clues on the early aggregation modes into dimers can be obtained by biasing.

A final appreciation of results on a rather wide landscape of molecular scale events should take into account the time and size limits, and above all, the choice of the key parameters in the kinetic bias: on the one hand, the layer limits allow a precise intervention of different chemical phenomena, namely, in benzoic acid, the hydrogen bonding and the dispersive factors generated by aromatic rings; on the other hand, the fudge factor is responsible for the "speed" of the bias, and should represent a careful compromise between effectiveness (that requires a high value) and relaxation (that needs a moderately tuned value to avoid undesirable freezing). Further progress may come from coupling the kinetic bias with collective variables, whose formulation is quite a task with respect to space group symmetry, some of which have been used to bias Monte Carlo into a direct evolution from liquid to crystal for the simplest space group.[25] The present results are overall encouraging and further extensive tests are being planned to find the best boundary conditions to the formation of symmetrized clusters, true solid-state precursors, that is still a faraway goal.

SUPPORTING INFORMATION AVAILABLE: Calculation of relative velocities and forces; Description of auxiliary modules in MiCMoS; Geometries, force field files, run control run prameters; Distribution of relative velocities vs. centre of mass distance; Definition of cyclic dimers; Distribution of c.o.m. velocities and relative velocities; Evolution of cyclic dimers in BZA nanodroplets.

CORRESPONDING AUTHOR: leonardo.lopresti@unimi.it; Current Address: Department of Chemistry, Università degli Studi di Milano, Via Golgi 19, 20133 Milano

REFERENCES

(1) Broadhurst, E. T.; Xu, H.; Parsons, S.; Nudelman, F. Revealing the Early Stages of Carbamazepine Crystallization by CryoTEM and 3D Electron Diffraction. *IUCrJ* **2021**, *8* (6), 860–866.

(2) Volmer, M. Weber, A. Z. Nucleus Formation in Supersaturated Systems. *Zeitschrift für Phys. Chemie* **1926**, *119*, 277–301.

(3) Sosso, G. C.; Chen, J.; Cox, S. J.; Fitzner, M.; Pedevilla, P.; Zen, A.; Michaelides, A. Crystal Nucleation in Liquids: Open Questions and Future Challenges in Molecular Dynamics Simulations. *Chem. Rev.* **2016**, *116* (12), 7078–7116.

(4) Erdemir, D.; Lee, A. Y.; Myerson, A. S. Nucleation of Crystals from Solution: Classical and Two-Step Models. *Acc. Chem. Res.* **2009**, *42* (5), 621–629.

(5) Wolde, P. R. ten; Frenkel, D. Enhancement of Protein Crystal Nucleation by Critical Density Fluctuations. *Science (80-.).* **1997**, *277* (5334), 1975–1978.

(6) Vekilov, P. G. Dense Liquid Precursor for the Nucleation of Ordered Solid Phases from Solution. *Cryst. Growth Des.* **2004**, *4* (4), 671–685.

(7) Davey, R. J.; Schroeder, S. L. M.; ter Horst, J. H. Nucleation of Organic Crystals-A Molecular Perspective. *Angew. Chemie Int. Ed.* **2013**, *52* (8), 2166–2179.

(8) Yang, Y. I.; Shao, Q.; Zhang, J.; Yang, L.; Gao, Y. Q. Enhanced Sampling in Molecular Dynamics. *J. Chem. Phys.* **2019**, *151* (7), 070902.

(9) Kirchner, B.; di Dio, P. J.; Hutter, J. Real-World Predictions from Ab Initio Molecular Dynamics Simulations; Springer, Berlin, Heidelberg, 2011; pp 109–153.

(10) Giberti, F.; Salvalaglio, M.; Parrinello, M. Metadynamics Studies of Crystal Nucleation. *IUCrJ* **2015**, *2* (2), 256–266.

(11) Rizzato, S.; Gavezzotti, A.; Lo Presti, L. Molecular Dynamics Simulation of Molecular Crystals under Anisotropic Compression: Bulk and Directional Effects in Anthracene and

Paracetamol. *Cryst. Growth Des.* **2020**, *20* (11), 7421–7428.

(12) Gavezzotti, A.; Lo Presti, L.; Rizzato, S. Mining the Cambridge Database for Theoretical Chemistry. Mi-LJC: A New Set of Lennard-Jones–Coulomb Atom–Atom Potentials for the Computer Simulation of Organic Condensed Matter. *CrystEngComm* **2020**, *22* (43), 7350–7360.

(13) Gavezzotti, A.; Lo Presti, L. Molecular Dynamics Simulation of Organic Crystals: Introducing the CLP-Dyncry Environment. *J. Appl. Crystallogr.* **2019**, *52* (6), 1253–1263.

(14) Gavezzotti, A.; Lo Presti, L. Dynamic Simulation of Liquid Molecular Nanoclusters: Structure, Stability and Quantification of Internal (Pseudo)Symmetries. *New J. Chem.* **2019**, *43* (5).

(15) Du, W.; Cruz-Cabeza, A. J.; Woutersen, S.; Davey, R. J.; Yin, Q. Can the Study of Self-Assembly in Solution Lead to a Good Model for the Nucleation Pathway? The Case of Tolfenamic Acid. *Chem. Sci.* **2015**, *6* (6), 3515–3524.

(16) Sullivan, R. A.; Davey, R. J.; Sadiq, G.; Dent, G.; Back, K. R.; Ter Horst, J. H.; Toroz, D.; Hammond, R. B. Revealing the Roles of Desolvation and Molecular Self-Assembly in Crystal Nucleation from Solution: Benzoic and p -Aminobenzoic Acids. *Cryst. Growth Des.* **2014**, *14* (5), 2689–2696.

(17) Lo Presti, L.; Gavezzotti, A. MiCMOs (MIlano Chemistry MOlecular Simulation) 2.0. Università degli Studi di Milano: Milano 2021, p https://sites.unimi.it/xtal_chem_group/.

(18) Feld, R.; Lehmann, M. S.; Muir, E. W.; Speakman, J. C. The Crystal Structure of Benzoic Acid: A Redetermination with X-Rays at Room Temperature; a Summary of Neutron-Diffraction Work at Temperatures down to 5 K. *Zeitschrift fur Krist. - New Cryst. Struct.* **1981**, *157* (3–4), 215–231.

(19) Donnay, J. D. H.; Harker, D. A New Law of Crystal Morphology Extending the Law of Bravais. *Am. Miner.* **1937**, *22*, 446–467.

(20) Macrae, C. F.; Sovago, I.; Cottrell, S. J.; Galek, P. T. A.; McCabe, P.; Pidcock, E.;

Platings, M.; Shields, G. P.; Stevens, J. S.; Towler, M.; et al. Mercury 4.0 : From Visualization to Analysis, Design and Prediction. *J. Appl. Crystallogr.* **2020**, *53* (1), 226–235.

(21) Florin, E. L.; Moy, V. T.; Gaub, H. E. Adhesion Forces between Individual Ligand-Receptor Pairs. *Science* **1994**, *264* (5157), 415–417.

(22) Moy, V. T.; Florin, E. L.; Gaub, H. E. Intermolecular Forces and Energies between Ligands and Receptors. *Science* **1994**, *266* (5183), 257–259.

(23) Gavezzotti, A.; Colombo, V.; Lo Presti, L. Facts and Factors in the Formation and Stability of Binary Crystals. *Cryst. Growth Des.* **2016**, *16* (10), 6095–6104.

(24) Gavezzotti, A.; Lo Presti, L. Building Blocks of Crystal Engineering: A Large-Database Study of the Intermolecular Approach between C–H Donor Groups and O, N, Cl, or F Acceptors in Organic Crystals. *Cryst. Growth Des.* **2016**, *16* (5), 2952–2962.

(25) Gavezzotti, A. Molecular Level Insights on the Liquid–Solid Transition of Large Organics by Biased Monte Carlo Simulations. *Cryst. Growth Des.* **2013**, *13* (8), 3801–3815.