UNIVERSITÀ DEGLI STUDI DI MILANO

PhD Course in Chemistry
38th Cycle

Department of Chemistry

PhD Thesis

Pre-nucleation in molecular liquids: a molecular dynamics perspective

CHEM-02/A

PhD Candidate: Luca Sironi
Matr. R13759
ORCID n. 0000-0003-2839-4243

Supervisor: Prof. Leonardo Lo Presti
Co-supervisor: Prof. Silvia Rizzato
PhD Coordinator: Prof. Daniele Passarella

Academic Year 2024-2025

Table of content

List of figures v

List of tables vii

1 Introduction 1

1.1 General overview 3

1.1.1 Crystals, crystal structure, and polymorphism 3

1.1.2 Nucleation 4

1.1.3 Molecular dynamics 5

1.2 Research scope, objectives and model compounds 7

1.2.1 Scope and objectives 7

1.2.2 Model compounds: benzoic acid and its derivatives 8

2 Methods 10

2.1 In-house software and programs 11

2.1.1 MiCMoS software 11

2.1.2 Confinement algorithm 17

2.1.3 Cluster program 19

Table of content

2.1.4 Stacking program 20

2.1.5 PIXEL 21

2.2 MD simulations 22

2.2.1 MD simulations of confined liquids 24

2.2.2 MD simulations of bulk systems 26

2.3 DFT-D calculations 28

2.3.1 Interaction energy of molecular aggregates 28

2.4 Experimental methods 29

2.4.1 Materials 29

2.4.2 Solubility measurements 29

2.4.3 Nucleation rate determination 30

3 Confinement 31

3.1 Confinement algorithm in short 33

3.2 Confinement of few molecules 33

3.2.1 Formation of molecular aggregates 34

3.2.2 Analysis of cyclic aggregates 38

3.3 Confinement of many molecules 42

3.3.1 Diffusion and viscosity 43

3.3.2 Rotational correlation 45

3.3.3 Structure of the liquid 47

3.3.4 Molecular orientation 51

3.3.5 Barrier stiffness and liquid-liquid transitions 55

3.3.6 Extreme confinement 57

4 Subcritical clusters 61

Table of content

4.1 Cluster program in short 63

4.2 Properties of the liquid 63

4.2.1 Average properties 63

4.2.2 Nanoscale inhomogeneities 66

4.3 Definition and analysis of subcritical clusters 68

4.3.1 Three criteria 68

4.3.2 Analysis of subcritical clusters 70

4.3.3 Size effect 76

5 Aromatic stacking and nucleation rate 78

5.1 Stacking energy of benzoic acid derivatives 79

5.1.1 Database search and stacking energy 79

5.1.2 Dataset analysis and selection of compounds 81

5.2 Nucleation rate determination 83

5.2.1 Solubility of benzoic acid derivatives 84

5.2.2 Nucleation rate of benzoic acid derivatives 87

5.3 Stacking program in short 89

5.4 MD simulations 89

5.4.1 General analysis of intermolecular interactions 90

5.4.2 Analysis of stacking interactions 92

Conclusion 100

Bibliography 105

Appendix A 114

PIXEL calculations 114

Gaussian calculations 115

Table of content

ESP fitting and charge density 115

Energy of molecular aggregates 116

BSSE correction via counterpoise method 117

MiCMoS MD simulations 118

Topology files 118

Bulk liquids parameter files 123

Confined liquids parameter files 124

Gromacs MD simulations 124

Topology file 124

Parameter file 127

Appendix B 129

PIXEL calculations 129

Acknowledgements 134

List of figures

Figure 1.1 9
Figure 3.1 36
Figure 3.2 38
Figure 3.3 40
Figure 3.4 41
Figure 3.5 41
Figure 3.6 42
Figure 3.7 44
Figure 3.8 45
Figure 3.9 47
Figure 3.10 49
Figure 3.11 50
Figure 3.12 51
Figure 3.13 52
Figure 3.14 54
Figure 3.15 56
Figure 3.16 59
Figure 3.17 59

List of figures

Figure 3.18	60
Figure 4.1	64
Figure 4.2	65
Figure 4.3	66
Figure 4.4	67
Figure 4.5	71
Figure 4.6	72
Figure 4.7	73
Figure 4.8	74
Figure 4.9	75
Figure 5.1	80
Figure 5.2	82
Figure 5.3	83
Figure 5.4	85
Figure 5.5	86
Figure 5.6	89
Figure 5.7	91
Figure 5.8	92
Figure 5.9	94
Figure 5.10	97
Figure 5.11	98

List of tables

Table 2.1 25
Table 2.2 26
Table 3.1 46
Table 3.2 58
Table 5.1 83
Table 5.2 84

1 Introduction

This thesis has the aim to investigate partial ordering and aggregation phenomena in molecular liquids, with the goal of gaining deeper insight into the early stages of crystal nucleation. To address this problem, a combination of computational and experimental approaches is employed. Computational work relies primarily on molecular dynamics simulations, while experimental work mainly consisted in crystallization experiments to determine nucleation rate.

The structure of the thesis is outlined as follows. In the *Introduction*, a general overview on crystals, nucleation and molecular dynamics is presented (Section 1.1) along with the scope, objectives and model compounds of the research (Section 1.2). In the *Methods*, first the MiCMoS software and analysis programs are illustrated (Section 2.1); then, the details regarding molecular dynamics simulations (Section 2.2) and DFT-D calculations are given (Section 2.3); finally the experimental methods are illustrated (Section 2.4). The results obtained in this work are divided into three chapters, each of which is preceded by a short introductory section providing the necessary background and context.

First, *Confinement* is exploited as a tool to study molecular aggregation. A newly developed confining algorithm is presented (Section 3.1), then the aggregation of few molecules under confinement is studied (Section 3.2) and the dynamics of

molecular liquids confined in different geometries is shown, including extreme confinement with a single molecular layer (Section 3.3).

Second, *Subcritical clusters* are investigated. A new analysis program is presented (Section 4.1), the average properties of molecular liquids and nanoscale inhomogeneities are studied (Section 4.2), and a new definition of subcritical cluster is illustrated along with their characterization (Section 4.3).

Finally, *Aromatic stacking and nucleation rate* are investigated. This part of the work was carried out during the research period in Prof. Aurora Cruz-Cabeza's group at Durham University and continued at Università degli Studi di Milano. The Cambridge Structural Database is analyzed to retrieve model compounds (Section 5.1) to be used for nucleation rate experiments (Section 5.2), another analysis program is illustrated, this time specifically designed to study aromatic stacking interactions (Section 5.3), and results of molecular dynamics simulations to study aromatic stacking interactions from a computational point of view are presented (Section 5.4).

1.1 General overview

1.1.1 Crystals, crystal structure, and polymorphism

Crystals are remarkably common in everyday life. Only few familiar examples are table salt (sodium chloride), sugars such as sucrose and glucose, lipids, starches, and ice, which is not more than water in a crystalline state. Minerals and most gemstones are crystalline in nature, such as quartz (SiO_2), calcite ($CaCO_3$), and gypsum ($CaSO_4 \cdot 2H_2O$), and also elemental carbon can form crystals, ranging from graphite to diamond. Metals like iron, copper, aluminum, and gold can also arrange in crystalline structures. Crystals are spread in pharmaceuticals[1], food industry[2,3] and materials science, being employed for many advanced technological applications, ranging from semiconductors for electronics and photonics[4,5], piezoelectric materials for actuators[6], metal organic frameworks (MOFs) for gas adsorption[7,8], perovskites for solar energy systems[9], and many others.

But what exactly are crystals? And why are they of such importance? According to the International Union of Crystallography (IUCr) definition, crystals are solids composed of atoms, ions and/or molecules that, on average, form a long-range order arrangement[10]. Hereafter, emphasis is placed on molecular crystals. It is the average long-range order that makes crystals special. Such order can be described by precise mathematical rules that define the crystal structure, which can be experimentally determined through diffraction techniques, using X-rays, neutrons, or electrons, including those available at large-scale facilities such as synchrotrons and free electron lasers (FELs). The crystal structure, together with the mathematical rules regulating it, links molecular and macroscale, giving crystalline materials predictable and tunable properties.

However, a particular molecule can potentially give rise to different polymorphs, crystals having the same molecular composition but different structure and properties[11], such as thermodynamic stability, solubility, bioavailability, and also electronic properties[1]. Predicting which polymorphs are most likely to crystallize is of fundamental importance and lies at the heart of crystal structure prediction (CSP)

methods[12,13]. To date, these methods have largely relied on computationally expensive brute-force approaches, which are not always successful and often generate an excessive number of candidate structures. A deeper understanding of the fundamental principles that govern how ordered crystals nucleate from a disordered phase could significantly enhance CSP, ultimately enabling better control over crystal properties.

1.1.2 Nucleation

The first step of crystallization is crystal nucleation, that is, the formation of a translationally ordered (or partly ordered) cluster of atoms, ions, or molecules at the nanoscale, able to grow and become a stable (or metastable) macroscopic crystal. Nucleation may occur within the bulk liquid or solution, in which case it is referred to as homogeneous nucleation, or at an interface, where it is termed heterogeneous nucleation. It can also be classified as primary, when no pre-existing crystals are present in the nucleating phase, or secondary, when it arises from previously formed crystals.

Nucleation has been described traditionally by classical nucleation theory (CNT), originally developed by Gibbs at the end of the 19th century for condensation of vapor into liquid and later extended to crystals[14,15]. According to CNT, when an aggregate form in the nucleating phase, its free energy can be divided into a stabilizing bulk contribution, due to the formation of a new, stable phase in a metastable one, and a destabilizing surface contribution, explained by the formation of an interface. When the size of the aggregate is small, the surface contribution predominates but as the size becomes larger, the bulk contribution prevails. The size at which there is a 50% chance for the aggregate to grow or dissolve is called critical size and the aggregate is called nucleus. Classical Nucleation Theory (CNT) provides a simplified framework that relies on several strong assumptions. For example, it treats nuclei as miniature versions of the corresponding macroscopic crystals, assuming bulk-like properties even at the molecular scale. This core assumption is sometimes referred to as “capillary approximation”, which formally hypothesizes that macroscopic surface tension of the crystal facets may be seamlessly applied to nanoscale structures like nuclei. It also

assumes that growth proceeds exclusively through monomer attachment, neglecting the possibility of cluster-cluster coalescence, despite both experimental and computational evidence demonstrating its relevance.

For this reason, non-classical nucleation theories have also been proposed to account for the behavior of proteins like lysozyme[16], inorganic compounds like calcium carbonate[17] and sodium chloride[18], small organic molecules like glycine[19] and other systems like metal organic frameworks, covalent organic frameworks, and colloids[20]. In general, non-classical mechanisms include the formation of dense liquid droplets before nucleation can take place, and also the formation of pre-nucleation clusters that are stable, persistent aggregates.

Nucleation often determines the final crystal structure. A deeper understanding of nucleation mechanisms could enable reliable crystal structure prediction, moving beyond brute-force computational approaches. This has immense practical relevance, from optimizing pharmaceuticals to designing advanced functional materials. Despite its importance, nucleation remains notoriously difficult to study. Its inherently stochastic nature makes it challenging to determine when and where a nucleus will first emerge. Experimental techniques, while invaluable, face significant challenges in capturing the transient and nanoscale nature of nucleation events.

1.1.3 Molecular dynamics

Computational approaches offer a complementary route for the understanding of nucleation. The most straightforward computational method able to describe the dynamics of a system is molecular dynamics[21]. Molecular dynamics can be useful to understand nucleation, since it is like an ultra-high-speed camera that records a "movie" of the time evolution of a system, with atomistic resolution[18,22,23]. This is a huge advantage compared to experimental techniques which are unlikely to have such space and time resolution. Ab initio molecular dynamics methods (AIMD), such as Car-Parrinello method, are powerful techniques that rely on density functional theory (DFT)[24]. However, the high accuracy comes with a computational cost that limits the applications of these techniques to short simulations of small systems. For this reason,

when the simulation of larger systems and longer timescales is required, classical molecular dynamics is preferred, since the computational cost is several orders of magnitude lower than AIMD methods.

In classical molecular dynamics, the system size can range from thousands to millions of atoms while the duration is typically of the order of nanoseconds to microseconds. Atoms are usually given point charges to model electrostatic interactions through Coulomb potential while Lennard-Jones potential is commonly used to model dispersion and repulsion interactions. This favors a quick evaluation of intermolecular interactions compared to AIMD methods, allowing to extend the duration of simulations at a reasonable computational cost. As regards intramolecular interactions, bond stretching, bending, and torsion are usually modelled with simple harmonic potentials. The parameters regarding both intermolecular and intramolecular interactions are parametrized against accurate experimental data or high-level quantum mechanical calculations. The analytical expressions describing all the interactions along with the associated parameters which depend on the specific atom type build up the so-called force field, which is the mathematical framework used to describe the potential energy of the system. The workflow of a classical molecular dynamic simulation is simple: after an initialization that assigns a starting velocity to each atom, the interatomic potential is calculated and derived to get the forces acting on each atom, and Newton's equations of motion are integrated numerically to get the new atomic positions. The process is then repeated for millions of steps to collect a trajectory describing the dynamic evolution of the system.

However, despite the powerful capabilities of classical molecular dynamics, it is important to be aware of some disadvantages when it is employed to study non-equilibrium processes like nucleation. First, nucleation is a rare and stochastic event that may occur on much longer timescales than the ones accessible to these simulations. Second, when the system is a solution, depletion effects might arise due to the formation of molecular aggregates that significantly reduce the local supersaturation, which is the driving force of nucleation. Together, these factors make direct

observation of nucleation by standard MD both computationally demanding and often impractical.

A possible way to address the time-scale issue can be to enhance the probability of nucleation by applying a bias. For instance, metadynamics is nowadays a mature technique that adds a biasing potential in the form of Gaussian functions, along some collective variables able to reduce the dimensionality of the system without losing the ability to capture the relevant features of the process of interest[25,26]. In this way, the system can escape free energy minima and explore rare configurations more efficiently, like the formation of a crystal nucleus. To tackle the depletion effects issue, instead, smart methods have been developed to work at constant supersaturation[27–29]. The system is divided into different regions: the inner region is where nucleation takes place; around the inner region there is an intermediate region where supersaturation is artificially kept constant; the outer region acts as a reservoir and is used to feed the intermediate region with new molecules.

Nevertheless, massive unbiased simulations may eventually capture nucleation, but they are feasible only with significant access to high-performance computing (HPC) infrastructure. Some examples are ice nucleation, where several trajectories more than one microsecond long on more than 1,500 atoms systems were run for months[30], and guanine crystallization in water solution that required half microsecond long simulations on 800,000 atoms systems[31].

1.2 Research scope, objectives and model compounds

1.2.1 Scope and objectives

This thesis focuses on the investigation of nucleation phenomena in molecular systems through computer simulations and experiments. The work is limited to unbiased, classical molecular dynamics for describing the dynamical evolution of the system, combined with static quantum-mechanical calculations to provide complementary information. The force fields implemented in the MiCMoS software

are used and assessed for their ability to reproduce the relevant interactions. Experimental studies are also carried out to provide a complementary perspective to computer simulations. All simulations are restricted to single-component molecular crystals and liquids, while experiments are performed on supersaturated solutions, with a particular focus on benzoic acid and its derivatives.

This research has three main objectives: first, to study the effects of confinement on partial ordering of molecular liquids; second, to investigate on the aggregation of molecules in the pre-nucleation phases, aiming to clarify how precursors having subcritical size can influence the nucleation mechanism; third, to explore the role of aromatic stacking interactions, in governing pre-nucleation phenomena in molecular liquids and solutions.

1.2.2 Model compounds: benzoic acid and its derivatives

Choosing appropriate model compounds is crucial to balance realism and computational feasibility. Benzoic acid emerges as an ideal candidate. It is widely regarded as the prototypical aromatic carboxylic acid. Its molecular structure consists of a single carboxyl group, capable of forming both single and double hydrogen bonds, attached to an aromatic ring that can engage in aromatic stacking interactions. Owing to its small number of atoms and limited flexibility, possessing only two rotatable bonds, benzoic acid represents a relatively rigid system. These features, combined with its ability to participate in diverse intermolecular interactions, make it an ideal model compound for investigating fundamental aspects of molecular association, aggregation, and crystal nucleation.

Benzoic acid is monomorphic to date, meaning that, regardless of the crystallization conditions, it always adopts the same crystal form, belonging to $P2_1/c$ space group (Figure 1.1), the most popular among organic compounds. The unit cell contains four molecules ($Z = 4$), with one molecule in the asymmetric unit ($Z' = 1$). It is characterized by two short cell edges a and b (5.5 and 5.1 Å), and the third axis, c, approximately four times longer (21.9 Å), with a β angle of 97°. The carboxylic groups are involved in two hydrogen bonds, forming cyclic hydrogen-bonded dimers related

by inversion symmetry. From a geometric perspective, the crystal structure of benzoic acid can be conveniently described by considering these cyclic dimers as the fundamental building units. Adjacent dimers interact laterally through CH···O interactions along the [110] crystallographic direction, forming 1-D motifs. These motifs further interact via parallel offset aromatic stacking interactions along [010], leading to 2-D assemblies which pack into the 3-D crystal structure through T-shaped aromatic stacking interactions along [001], with an angle of 82.8° between the planes of adjacent dimers.

In this work, benzoic acid along with some of its derivatives are investigated and compared assessing the influence of the additional functional groups.

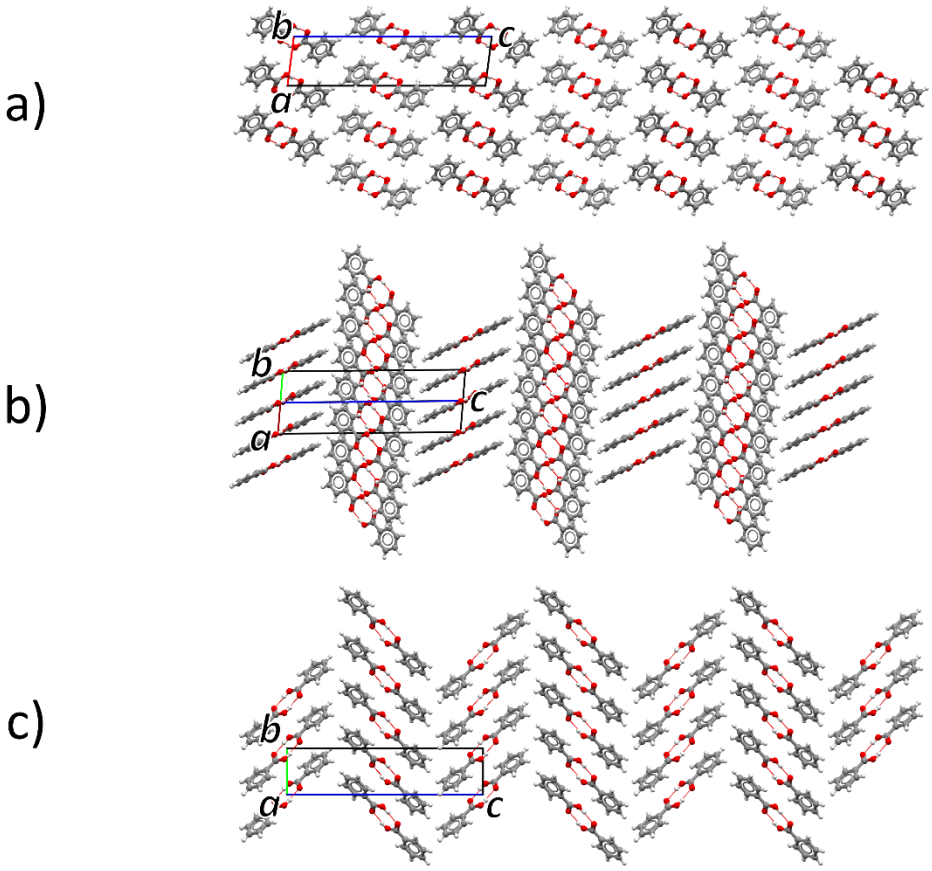

Figure 1.1 Crystal structure of benzoic acid viewed along [010] (a), [110] (b), and [100] (c) crystallographic directions.

2 Methods

In this Section, the computational and experimental methods used for this work are presented. First, an overview of MiCMoS software is given, along with a description of the newly implemented confining algorithm, new analysis programs, and PIXEL method (Section 2.1). Second, the details regarding molecular dynamics simulations of both confined and bulk systems (Section 2.2), and also DFT-D calculations of molecular aggregates (Section 2.3) are provided. Finally, the experimental materials and methods used for solubility measurements and nucleation rate experiments are illustrated (Section 2.4).

2.1 In-house software and programs

2.1.1 MiCMoS software

MiCMoS (Milano Chemistry Molecular Simulation) is a set of programs written in FORTRAN, originally developed by Gavezzotti and Lo Presti, specifically designed for the simulation of condensed matter of small organic molecules[32]. MiCMoS allows to perform static calculations with either force field methods or PIXEL method (Section 2.1.5) and also classical Monte Carlo and molecular dynamics simulations.

MiCMoS allows the direct generation of simulation input files from crystallographic information files (CIFs). It provides dedicated tools for crystalline systems to construct simulation boxes, such as *Boxcry*, which generates periodic crystals, and *Nanocut*, which enables the creation of isolated nanocrystals based on user-defined crystallographic planes. In addition, the package includes specialized analysis programs for crystalline materials. These include *Adp*, which computes the space-time averaged crystal structure and evaluates anisotropic atomic displacement parameters (ADPs), and *Difscat*, which calculates X-ray scattering patterns from the simulated structures.

All calculations are performed in double precision. At present, parallelization is achieved exclusively through OpenMP; MPI parallelization and GPU acceleration are not supported. Consequently, the software is primarily suited for simulations of relatively small systems ($10^4 - 10^5$ atoms) and moderate trajectory lengths (usually 1 – 10 ns).

General workflow

The workflow starts from a CIF file. The *Retcif* program reads the CIF file and converts it into an *.oih* file. This file is then processed by *Retcor*, which generates an *.oeh* file without atomic charges. Subsequently, *Retcha* assigns point charges to the *.oeh* file: charges for the CLP force field (see below) are added automatically, whereas

ESP-fitted charges for the LJC force field must be added manually after performing a quantum calculation.

The *.oeh* file contains the atomic coordinates (either crystallographic, including cell parameters and symmetry operators, or Cartesian for simulations of liquids) together with atom types and point charges. This file serves as the starting point for static calculations performed with the *Crysaa* module for cohesive energy determination and with PIXEL (described in Section 2.1.5).

Starting from the *.oeh* file, it is also possible to generate the topology and simulation box required for Monte Carlo and molecular dynamics simulations. The *Pretop* program is used to generate the topology *.top* file, which includes Cartesian coordinates, atom types and charges, as well as stretching, bending, torsional parameters and intramolecular non-bonded interactions. The simulation box is generated using *Boxcry*, *Boxliq*, or *Boxsol* for crystals, liquids, or solutions, respectively, producing files in *.dat* format for molecular dynamics simulations and *.bxi* format for Monte Carlo simulations. Finally, an instruction file (*.mdi* for molecular dynamics or *.mci* for Monte Carlo simulations) is required to define all steering parameters. The simulations produce a trajectory file (*.dat* for molecular dynamics or *.bxo* for Monte Carlo) and an energy file *.ene*, which are subsequently used for data analysis.

Force fields

MiCMoS includes two all-atom force fields, both calibrated against the experimental enthalpies of sublimation and vaporization of small organic molecules: CLP and LJC[33–35]. CLP (Coulomb-London-Pauli) is characterized by an explicit term that takes into account the polarizability, uses cheap Mulliken population derived point charges, and has the functional form described in Equation 2.1. LJC (Lennard-Jones-Coulomb) is characterized by the use of point charges fitted against the electrostatic potential derived from an MP2/6-31G** wavefunction. It is more suitable for polar or strongly hydrogen-bonded systems and has a functional form described in Equation 2.2.

$$E_{CLP} = E_{Coulomb} - \frac{A4}{R^4} - \frac{A6}{R^6} + \frac{A12}{R^{12}} \quad (2.1)$$

$$E_{LJC} = E_{Coulomb} - \frac{A6}{R^6} + \frac{A12}{R^{12}} \quad (2.2)$$

A4, *A6*, and *A12* are the coefficients of polarization, dispersion, and repulsion terms while *R* is the interatomic distance. Both force fields have an intramolecular potential described by Equation 2.3, 2.4, 2.5.

$$E_{stretching}(R) = \frac{1}{2} k_s (R - R_0)^2 \quad (2.3)$$

$$E_{bending}(\theta) = \frac{1}{2} k_b (\cos\theta - \cos\theta_0)^2 \quad (2.4)$$

$$E_{torsion}(\tau) = k_t [1 + f \cdot \cos(m\tau)] \quad (2.5)$$

R, θ and τ are the bond length, angle and torsion; R_0 and θ_0 are the equilibrium bond length and angle; k_s and k_b are the stretching and bending force constants; k_t is the torsional force constant, corresponding to half of the height of the potential energy barrier; f is a phase factor that can be ± 1; m is the multiplicity regarding the periodicity of the torsional potential. k_s and k_b are automatically determined by MiCMoS and are derived from *ab initio* stretching and bending energy profiles and fitting against 54A7 GROMOS force field parameters[34]. k_t, f, and m must be set manually, according to tabulated values of several classes of compounds for which the torsional potentials have been calculated using MP2/6-31G** wavefunctions. Intramolecular non-bonded interactions can be included and are damped by a factor *factin*, usually set to 0.7.

Thermostats and barostats

MiCMoS provides three thermostats to control the temperature during the simulations: stiff coupling, Berendsen, and CSVR. Stiff coupling simply rescales all velocities by a factor $(T_{set}/T)^{1/2}$, where T_{set} refers to the temperature set on the thermostat while T is the instantaneous temperature of the system. Berendsen thermostat exploits a weak coupling procedure by rescaling all velocities by a factor λ described in Equation 2.6.

$$\lambda(t) = \sqrt{\left[1 + \frac{dt}{\tau} \cdot \left(\frac{T_{set}}{T} - 1\right)\right]} \tag{2.6}$$

τ is the temperature relaxation time and dt is the timestep. The ratio dt/τ is called T_{relax} in MiCMoS and is approximated by an empirical coefficient of 0.5-0.6. Berendsen thermostat tends to suppress kinetic energy fluctuations and does not reproduce a proper canonical ensemble (*NVT*). The Canonical Sampling through Velocity Rescaling (CSVR) thermostat developed by Bussi, Donadio, and Parrinello[36], instead, introduces a stochastic term that enables a correct sampling of kinetic energies that yields a proper canonical ensemble, using a rescaling factor α described in Equation 2.7. K is the actual kinetic energy, K_t represents the target kinetic energy extracted from a canonical ensemble distribution described in Equation 2.8, N_{dof} is the number of degrees of freedom, and $\beta = 1/k_bT$ is the Boltzmann's factor.

$$\alpha = \sqrt{\frac{K_t}{K}} \tag{2.7}$$

$$\bar{P}(K_t)dK_t \propto K_t^{\frac{N_{dof}}{2}-1} e^{-\beta K_t} dK_t \tag{2.8}$$

As regards pressure control, MiCMoS provides a minimal isotropic or anisotropic barostat, and a Parrinello-Rahman (PR) barostat. As for the isotropic minimal barostat, first the kinetic energy E_{kin} and the virial W are computed through Equation 2.9 and Equation 2.10. Then, the pressure p of the system is estimated by the classical virial theorem using Equation 2.11.

$$E_{kin} = \frac{1}{2} \sum_{i=1}^{N_{mol}} m_i v_i^2 \tag{2.9}$$

$$W = -\frac{1}{2} \sum_{i,j=1}^{N_{mol}} \boldsymbol{F}_{ij} \cdot \boldsymbol{R}_{ij} \tag{2.10}$$

$$p = \frac{2}{3V}(E_{kin} - W) \tag{2.11}$$

N_{mol} is the number of molecules; m_i and v_i are the mass and velocity of molecule i; $\boldsymbol{F}_{ij}$ represents the force exerted on molecule i by molecule j and $\boldsymbol{R}_{ij}$ is the vector

connecting the centers of mass of molecules i and j; V is the volume of the system. Finally, a rescaling factor μ defined in Equation 2.12 is applied to box edges. When the anisotropic version of the barostat is employed, the box edges are rescaled according to Equation 2.13, where the subscript k refers to the components x, y, z of the rescaling factor and pressure.

$$\mu = \sqrt[3]{[1 - \mu_0(p_{set} - p)]} \tag{2.12}$$

$$\mu_k = \sqrt[3]{[1 - \mu_0(p_{set} - p_k)]} \tag{2.13}$$

The PR algorithm, instead, considers $\boldsymbol{v}_i$, $\boldsymbol{F}_{ij}$, and $\boldsymbol{R}_{ij}$ as first-rank tensors and calculates kinetic energy $\boldsymbol{E}_{kin}$ and the virial $\boldsymbol{W}$, both second-rank tensors, through tensor products. The resulting pressure $\boldsymbol{P}$ is a second-rank tensor as well, and the corresponding hydrostatic pressure p can be determined simply calculating one-third of the trace of $\boldsymbol{P}$. The equation of motion for the simulation box edges and angles is explicitly solved using Equation 2.14.

$$\boldsymbol{F} = w\frac{d^2\boldsymbol{H}}{dt^2} = [\boldsymbol{P} - p\boldsymbol{I}]\boldsymbol{\sigma} \tag{2.14}$$

$\boldsymbol{F}$ is the force which acts on the simulation box, w is a coupling parameter expressed in units of mass, $\boldsymbol{H}$ represent the matrix of box edges, $\boldsymbol{P}$ and p are the pressure tensor and the scalar hydrostatic pressure, $\boldsymbol{I}$ is the identity matrix, and $\boldsymbol{\sigma}$ is the volume-scaled reciprocal cell matrix.

Integration algorithm

In molecular dynamics simulations, MiCMoS provides two different algorithms for the integration of Newton's equations of motion: leapfrog (Equation 2.15, 2.16) and velocity Verlet (Equation 2.17, 2.18).

$$\boldsymbol{v}\left(t + \frac{1}{2}\Delta t\right) = \boldsymbol{v}\left(t - \frac{1}{2}\Delta t\right) + \frac{\Delta t}{m}\boldsymbol{F}(t) \tag{2.15}$$

$$\boldsymbol{r}(t + \Delta t) = \boldsymbol{r}(t) + \Delta t \cdot \boldsymbol{v}\left(t + \frac{1}{2}\Delta t\right) \tag{2.16}$$

$$\boldsymbol{v}\left(t + \frac{1}{2}\Delta t\right) = \boldsymbol{v}(t) + \frac{1}{2}\frac{\Delta t}{m}\boldsymbol{F}(t) \tag{2.17}$$

$$\boldsymbol{r}(t+\Delta t) = \boldsymbol{r}(t) + \Delta t \cdot \boldsymbol{v}\left(t + \frac{1}{2}\Delta t\right) \tag{2.18}$$

t and m are time and mass, while $\boldsymbol{v}$, $\boldsymbol{F}$, and $\boldsymbol{r}$ are velocity, force, and position.

Analysis programs

MiCMoS includes several analysis programs, but only the main ones are presented here. A full overview is provided in MiCMoS user's manual[37]. The *Analys* module is used to compute the center-of-mass and any atom-atom radial distribution functions (RDFs) for each frame of a molecular dynamics trajectory and the *Distrib* module is used to perform a time average. The RDF $g(r)$ quantifies the local structural ordering in a system by measuring the probability of finding a molecule (or atom) at a distance r from a reference molecule (or atom), compared to a random distribution, and is described by Equation 2.19.

$$g(r_i) = \left(\frac{n_i}{V_i}\right) \cdot \left(\frac{n}{V}\right)^{-1} = \frac{n(r_i)}{4\pi r_i^2 dr} \cdot \left(\frac{n}{V}\right)^{-1} \tag{2.19}$$

Here, r_i denotes the distance of the i^{th} bin; n_i is the number of distances found in the spherical shell of thickness dr and volume V_i; the quantity n/V represent the average number density of distances and serves as a normalization factor. The *Correl* module is used to determine the translational mean squared deviation of molecules (Equation 2.20) and also the rotational correlation function $C(t)$ (Equation 2.21).

$$msd = \frac{\sum_{k=1}^{N_{mol}} |\boldsymbol{r}_k(t) - \boldsymbol{r}_k(0)|^2}{N_{mol}} = \langle |\boldsymbol{r}_k(t) - \boldsymbol{r}_k(0)|^2 \rangle \tag{2.20}$$

$$C(t) = \frac{\sum_{k=1}^{N_{mol}} \boldsymbol{u}_k(t) \cdot \boldsymbol{u}_k(0)}{N_{mol}} = \langle \boldsymbol{u}_k(t) \cdot \boldsymbol{u}_k(0) \rangle \tag{2.21}$$

$\boldsymbol{r}_k(t)$ and $\boldsymbol{r}_k(0)$ are the positions while $\boldsymbol{u}_k(t)$ and $\boldsymbol{u}_k(0)$ are the orientation vectors of molecule k at time t and time 0; N_{mol} is the total number of molecules. The translational *msd* is related to the diffusion coefficient $D(t)$ by Equation 2.22. In addition, the dynamic viscosity η of the system is related to $D(t)$ by Stokes-Einstein equation (Equation 2.23).

$$D(t) = \frac{1}{6t} \langle |\boldsymbol{r}_k(t) - \boldsymbol{r}_k(0)|^2 \rangle \tag{2.22}$$

$$D = \frac{k_B T}{6\pi\eta r} \tag{2.23}$$

k_B is Boltzmann's constant, T is the absolute temperature of the system, and r is the hydrodynamic radius of the molecule. The *Denflu* module is able to compute the local density fluctuations of the system, spotting inhomogeneities in liquids, possibly related to early aggregation phenomena. The simulation box is divided into a grid of volume elements and, in each of them, the number of molecules is computed, along with the local density ρ_i. The mean squared density deviation $\langle \Delta^2 \rho_i \rangle$ for each volume element i can be calculated by Equation 2.24, where brackets indicate a time average. When a space average in considered, instead, the quantity $\langle \Delta^2 \rho \rangle / \langle \rho_{box}^2 \rangle$ is amenable to experimental verification through light diffuse scattering experiments.

$$\langle \Delta^2 \rho_i \rangle = \langle |\rho_i - \langle \rho_i \rangle|^2 \rangle \tag{2.24}$$

The new modules specifically developed for this work, now fully implemented in the latest MiCMoS release, are described in dedicated sections (Section 2.1.2, Section 2.1.3, Section 2.1.4).

2.1.2 Confinement algorithm

An original outcome of this work is the introduction, in the release 2.2 of MiCMoS, of a new capability for simulating systems in confined spaces[38]. This required the development of a dedicated module, *Confbox*, for generating the initial simulation box, along with the addition of new subroutines to the main code. The source code is freely available when downloading the MiCMoS package[37].

The *inano* option was added to the input *.mdi* file to allow for the simulation of systems confined into nanolayers (*inano* = 1), square nanotubes (*inano* = 2), and cubic nanocavities (*inano* = 3). Confinement is obtained thanks to van der Waals barriers made of rectangular grids of pseudoatoms with Lennard-Jones parameters of any atom type, tunable radius, no mass and no charge. These pseudoatoms do not interact with one another, thus force constants are employed to avoid unphysical deformations of the barriers caused by the pressure of the system.

Confbox module

The *Confbox* module takes as input a pre-equilibrated, liquid simulation box and produces as output a modified, confined box and a parameter *.par* file, containing all the information needed to perform a confined molecular dynamics simulation. This include the type of confinement (*inano* option, that must match the one reported in the input *.mdi* file), the confining planes to be employed (XY, XZ, or YZ), the starting distance between opposite barriers (*thickness*) and an offset distance (*offset*), the radius and atom type of pseudoatoms, a flag (*iattr*) determining whether to use or not the attractive part of the Lennard-Jones potential to simulate full van der Waals or repulsive only barriers, the scaling factors (*dampk*) to be applied to force constants to tune the stiffness of barriers. In addition, the parameter *.par* file contains the information to control the target equilibrium distance between barriers (*zacsize*) along with the target equilibrium volume of the simulation box and its packing efficiency. A parameter *nmolzacu* is used to calculate the equilibrium volume of the simulation box (V_{box}), assuming a packing efficiency C_{pack} of 66%, which is the theoretical limit for close packing of random spheres[39]. Thus, defining V_{mol} the molecular volume, $V_{box} = nmolzacu \cdot V_{mol}/C_{pack}$. If *nmolzacu* is greater than the number of molecules in the system, a greater volume and thus a lower packing efficiency are obtained. The equilibrium length l of a cubic nanocavity is simply $l = \sqrt[3]{V_{box}}$. For a nanolayer, the equilibrium distance between the barriers is equal to *zacsize* while the other two edges have equilibrium length $l = \sqrt[2]{V_{box}/zacsize}$. For square nanotubes, instead, the two equilibrium distances between barriers are equal to *zacsize* while the third one is $l = V_{box}/zacsize^2$.

Workflow of confined simulations

Pseudoatoms are arranged in rectangular grids and are placed externally with respect to the simulation box and tangentially to its faces. The number of pseudoatoms in each direction is constant during the simulation and is calculated rounding down the ratio between the equilibrium size of the simulation box and the pseudoatom diameter. All the molecules having atoms in contact with the starting box faces, placed at an

initial distance *thickness* plus twice the *offset*, are deleted during the *Confbox* procedure. The *offset* parameter can be tuned to avoid the deletion of too many molecules, placing the barriers further away at the beginning of the simulation. When the simulation begins, the pressure control makes the simulation box expand or contract in specific directions. Since pseudoatoms are non-interacting, a fictitious counterpression is added, proportional to the difference between the actual box size and the equilibrium box size, divided by the equilibrium area of the face orthogonal to that direction. In this way, barriers can stretch or compress around their equilibrium size, avoiding unphysical deformations. A force constant of a Csp^2-Csp^2 aromatic bond ($3400\ kJ\ mol^{-1}\ Å^{-2}$) is employed and can be tuned by scaling parameters (*dampk*). If the initial *offset* is too large, an excessively high counterpressure may be generated and the box can overshrink, causing steric clashes at the beginning of the simulation. For this reason, a maximum change of 5% was imposed on the expansion or contraction of box edges, to prevent such a collapse. Every time the barostat is called, the pressure is calculated along with the counterpressure, then the box edges are updated accordingly, and the pseudoatom positions are recalculated based on the length of box edges.

2.1.3 Cluster program

Cluster input

Another outcome of this work, added to release 2.3 of MiCMoS, is the new analysis program, *Cluster*, to detect the formation of molecular aggregates during a molecular dynamics trajectory. Also in this case, the source code is freely available downloading the MiCMoS package[37]. *Cluster* analyzes molecular dynamics trajectories and also requires a topology file as input. *Cluster* offers two options to determine whether two molecules are bonded, belonging to the same aggregate, or not. The first option is geometric and is thought for D-H···A hydrogen bonds: the donors D, hydrogens H, and acceptors A atoms need to be specified, along with a threshold H···A distance, based on the van der Waals distance scaled by a damping factor, and a DHA angle. The second option is energetic and only requires an energy threshold for intermolecular interactions. Molecular pairs characterized by an interaction energy

more negative than the energy threshold are considered bonded, regardless of the type of interaction. The choice of the energy threshold is crucial since a too high and negative value would result in a system composed of monomers only, while a too small and negative value would lead to a single, large aggregate including all the molecules of the system. To help the choice of the energy threshold, a distribution of interaction energies during the simulation is also printed.

Cluster outputs

Regardless of whether intermolecular interactions are defined using geometric or energetic criteria, a list of aggregates with associated descriptors is generated for each frame. For each aggregate, the output includes the list of molecules that constitute it, the cohesive energy per molecule within the aggregate, the interaction energy with the surrounding molecules normalized by the number of molecules in the aggregate, and the cohesive energy per molecule of all molecules outside the aggregate. These energies can be used to calculate the excess energy defined in Section 4.3.1. In addition, the eigenvalues of the inertia tensor are given to have an idea on the aggregate shape (spherical, oblate, or prolate), along with a label that classifies aggregates in linear, cyclic, or mixed (i.e. containing branches or a mixture of linear and cyclic regions), and the global asymmetry index G proposed by Gavezzotti and Lo Presti[40]. These results can be easily converted into a Pandas dataframe for subsequent analysis with Python libraries, if needed. Moreover, two additional output files are printed: one containing the distribution of interaction energies and the other one with the distribution of lifetimes of intermolecular interactions.

2.1.4 Stacking program

Stacking analysis program was specifically developed for this work. The current version is designed for benzoic acid derivatives, but a more general version will be made available in the next MiCMoS release.

Stacking input

Stacking analyzes a molecular dynamics trajectory looking for all aromatic stacking interactions. Molecular topology is also required, ring atoms and carbonyl

carbon atoms must be indicated, along with two orientation vectors: one that is orthogonal to the ring and the other one lying on the ring plane, passing through the carbonyl carbon atom. A third orientation vector is automatically determined by the vector product of the previous two. Aromatic stacking interactions are defined according to geometric criteria, with user defined thresholds. In this work, relatively loose thresholds were initially adopted in order to collect raw, unfiltered data, which were then further refined through post-processing. The first condition is based on the distance between ring centroids (<8.0 Å). The second condition is on the angle between normal vectors to the rings (<30°). The third condition regards the longest distance between ring centroid and ring plane (<5.5 Å). The fourth and final condition is based on the degree of overlap between the two interacting molecules (≥1 carbon atom). The overlap is estimated by projecting the ring atoms and the carbonyl carbon of one molecule onto the ring plane of the other and counting how many atoms lie within the van der Waals distance. The procedure is then repeated by projecting the second molecule onto the first one, and the smaller of the two counts is taken as a measure of the molecular overlap.

Stacking output

For each frame, all molecular pairs involved in aromatic stacking interactions are reported, together with their interaction energy and its coulombic and dispersion contributions. In addition, several geometric descriptors are provided, including the ones used to define stacking interactions, the angles between the other two orientation vectors, the minimum distance between the carbonyl carbon atom and the ring centroid. In addition, a flag indicates stacked pairs that are also engaged in hydrogen bonding.

2.1.5 PIXEL

PIXEL is a smart algorithm developed by Gavezzotti and implemented in MiCMoS, for the calculation of cohesive energies and intermolecular interaction energies between molecules in a crystal structure[41]. The key advantage of PIXEL is its balance between high accuracy and low computational cost.

PIXEL workflow

The idea behind the algorithm is simple: the space around the molecule of interest is divided into k volume elements (3D pixels, or voxels) and, in each of them, the charge density ρ_k is derived from an MP2/6-31G** wavefunction; second, for each volume element V_k of the grid, the charge q_k is calculated by $q_k = \rho_k \cdot V_k$; third, the pixels are contracted into super-pixels to reduce the computational cost and are assigned to specific atoms; finally the intermolecular interaction energy is determined by pixel-pixel calculations that account for Coulomb energy and also polarization, dispersion, and repulsion energies that depend on empirical parameters. These parameters have been optimized based on agreement with experimental heats of sublimations and ab initio interaction energies. More details can be found in MiCMoS user's manual[37].

PIXEL calculations

For PIXEL calculations, *Pixelc* module of MiCMoS was used. From a structural *.oeh* file, obtained by applying the *Retcif* and *Retcor* procedures on a *.cif* file, a *.gjf* input file for Gaussian[42], and a *.inp* input file for PIXEL were generated by the *Pixmt2* routine. Hydrogen atom positions were renormalized to neutron distances to get more accurate bond lengths. For the subsequent PIXEL calculations with *Pixelc*, the Gaussian density *.den* file, the PIXEL *.inp* file and a standard parameter *.par* file with contraction level of 4 were used. For a complete reproducibility, input files are reported in *Appendix A*. The *pix.pri* output file provides a list of the most favorable intermolecular interaction energies, decomposed into coulombic, polarization, dispersion, and repulsion components, between symmetry-related molecules. This information is reported along with the corresponding symmetry operators, translation vectors, and center-of-mass distances.

2.2 MD simulations

For this work, LJC force field was always adopted for both Monte Carlo and molecular dynamics simulations, since benzoic acid and its derivatives are systems

characterized by strong hydrogen bonds. Thus, the electrostatic potential (ESP) derived point charges were used accordingly. The topology of benzoic acid was obtained from the experimental crystal structure in the Cambridge Structural Database (refcode BENZAC02[43], space group $P2_1/c$, T = 300 K) using the *Retcif* and *Retcor* modules (see MiCMoS user's Manual[37]). The same procedure was applied on the benzoic acid derivatives 4-bromobenzoic acid (refcode BRBZAP01[44], space group $P2_1/n$, T = 100 K) and 4-(methoxycarbonyl)benzoic acid (refcode QIYRUO[45], space group $P2_1/c$, T = 296 K). Due to the intrinsic low accuracy of X-ray diffraction techniques, to get the correct position of hydrogen atoms, bonds involving hydrogens were reset with standard procedures to get accurate, neutron diffraction bond lengths. ESP charges were derived from an MP2/6-31G** wavefunction obtained with Gaussian[42], using an input like the one deposited in *Appendix A*.

All the simulations were performed at constant number of molecules, pressure, and temperature in the *NpT* ensemble, with the exception of simulations confined into nanocavities where an *NVT*-like ensemble was adopted (see Section 2.2.1), and also the simulations of few molecules confined in a nanocavity and under extreme confinement in a nanolayer where an *NVT* ensemble was simulated. Unless otherwise specified, a leap-frog integrator was employed, with a timestep of 1 fs and a cutoff of 16 Å for both electrostatic and van der Waals interactions.

As benzoic acid is a neutral molecule, long-range electrostatics was not evaluated using algorithms like Ewald summation or Particle Mesh Ewald (PME). A cutoff radius of 16 Å was enough to capture 99.7% of the cohesive energy of the system, with full convergence reached at 20 Å, while ensuring a reasonable computational cost. Interactions between molecules separated by distances larger than 16 Å, all weaker than $-0.5\ kJ\ mol^{-1}$, were therefore neglected.

Coordinates and energies were saved every 0.5 ps. The thermostats employed were either Berendsen or CSVR[36], having a T_{relax} of 0.6 and a frequency of every 100 steps, while the barostats employed were either the isotropic or anisotropic minimal barostat, set at 1 atm and having a coefficient μ_0 of $0.4 \cdot 10^{-9} Pa^{-1}$ and a frequency of every 50 steps. The center of mass of the system was reset every 100 steps to suppress

unphysical translational drifts, with the exception of the simulations of few molecules (Section 3.2) where the center-of-mass reset was not performed. Intramolecular non-bonded interactions were considered with a dampening factor *factin* of 0.7 between atoms capable of forming a 6-member ring.

2.2.1 MD simulations of confined liquids

All the simulations of confined liquids were performed on benzoic acid[38]. The *inano* option in the input *.mdi* file was set to 1 for simulations into nanotubes, 2 for nanolayers and 3 for nanocavities. The additional parameter *.par* file was prepared and modified for each simulation. Some examples of *.mdi* and *.par* files are provided in *Appendix A*. Confinement was applied by the use of barriers parallel to the XY plane for nanolayers, XY, XZ planes for nanotubes and XY, XZ, and YZ planes for nanocavities. Pseudoatoms with van der Waals radius of 1.77 Å and Lennard-Jones parameters of carbon atom type (number 12 in MiCMoS) were used to form hydrophobic barriers.

Confinement of few molecules

A cubic simulation box of size 20 Å with a number of randomly oriented molecules ranging from 2 to 8 was evolved for 2 ns at 350 K using a Berendsen thermostat with frequency every 2000 steps. Due to the limited size of the system, no cutoff was used and interactions between all molecules were considered. Coordinates and energies were saved every 0.1 ps to allow for a better characterization of transient intermediate states. Full van der Waals barriers were employed (*iattr* = 1).

Confinement of many molecules

A simulation box of 432 randomly oriented molecules was prepared with the *Boxliq* module and hard contacts were disposed of by 1×10^6 Monte Carlo steps at 293 K. The resulting simulation box, approximately cubic with dimensions of ~47 Å, was employed as starting point for all the simulations. An *offset* of 6 Å was applied to confined simulations to avoid the removal of any molecule. The confined simulations were 0.5 ns long, for a total of 1001 frames. The simulations were performed in the *NpT* ensemble except for nanocavity simulations, where the volume was restrained

(*NVT*-like ensemble) by applying force constants to the barriers. A Berendsen thermostat was set to 350 K, and an anisotropic minimal barostat was set to 1 atm. The effect of the attractive part of the Lennard-Jones potential was investigated by varying the *iattr* parameter in the *.par* file to simulate both repulsive only barriers (*iattr* = 0) and full van der Waals barriers (*iattr* = 1). In addition, different adimensional scaling factors s for the barrier force constants were tested: $s = 0.025$ for soft barriers and $s = 0.050$ for stiff barriers. Table 2.1 reports the details of the simulations performed.

Table 2.1 List of simulations indicating confinement geometry, barrier type that is either full van der Waals (vdW) or repulsive-only (rep), scaling factor applied to force constants (s), and number of molecules in the simulation box (N_{mol}).

Geometry	Barrier type	s	N_{mol}
Nanolayer	vdW	0.025	432
Nanotube			
Nanocavity			
Nanolayer	Rep		
Nanotube			
Nanocavity			
Nanolayer	vdW		864
Nanotube			
Nanocavity			
Nanolayer	Rep		
Nanotube			
Nanocavity			
Nanolayer	vdW	0.050	432
Nanotube			
Nanocavity			
Nanolayer	Rep		
Nanotube			
Nanocavity			

Extreme confinement

Simulations in extremely confined spaces were performed using the nanolayer geometry (*inano* = 1) with full van der Waals barriers (*iattr* = 1), using a nominal distance between barriers (*zacsize*) of 3.6 Å. The starting simulation box was prepared from the final frame of the unconfined, bulk liquid simulation, that was replicated three times along x and y directions. Then, using *Confbox*, a thin layer of molecules was

obtained, bearing either 108 or 216 molecules. The *nmolzacu* parameter was fixed at 285 for all the systems, resulting in two different close packing efficiencies. This also lowered the viscosity, thereby favoring molecular aggregation among benzoic acid molecules. The systems underwent 100 ps *NpT* equilibration at 350 K using a Berendsen thermostat, an anisotropic minimal barostat and damping parameters *dampk(XY)* = 0.10, *dampk(XZ)* = 0.01, *dampk(YZ)* = 0.01. The purpose of this equilibration was to squeeze the simulation box to reach the desired thickness. Finally, production runs of 5 ns were performed in the *NVT* ensemble, to avoid oscillations in the nanolayer thickness, employing a CSVR thermostat[36]. Coordinates and energies were saved every 5 ps. To explore the effect of temperature, production runs were performed both at 300 K and 350 K. Table 2.2 reports the parameters characterizing the simulations.

Table 2.2 List of simulations under extreme confinement reporting the nominal thickness of the simulation box in the z direction, real thickness obtained after the equilibration step, number of molecules (N_{mol}), and temperature.

Nominal thickness / Å	Real thickness / Å	N_{mol}	T / K
3.6	3.71	108	300
			350
	3.80	216	300
			350

2.2.2 MD simulations of bulk systems

Bulk crystal

A periodic bulk simulation of crystalline benzoic acid was performed for comparison with the corresponding liquid systems. The starting simulation box was prepared using *Boxcry* module of MiCMoS[37], replicating the experimental (BENZAC02[43]) crystallographic cell 6×9×2 times along *a*, *b*, and *c*, obtaining a box of ~33×46×44 Å with an angle β of 97° and a total of 432 molecules. The system was evolved for 500 ps at 350 K using a Berendsen thermostat and an anisotropic minimal barostat.

Bulk liquid for confinement (MiCMoS)

The starting simulation box with 432 benzoic acid molecules and size of ~47 Å obtained after 10^6 MC steps and described in Section 2.2.1 was further equilibrated with a 0.5 ns long *NpT* trajectory at 350 K with a Berendsen thermostat and an anisotropic minimal barostat set at 1 atm, using the same parameters as the confined simulations for consistency, but with *inano*=0 to simulate an unconfined bulk liquid.

Bulk liquid for confinement (Gromacs)

A molecular dynamics simulation of bulk liquid was also performed with Gromacs 2018.4 for comparison[46]. The pre-equilibrated box obtained from the respective MiCMoS simulation was further equilibrated for 100 ps with *NVT* and then *NpT* trajectories, and a production run of 1 ns was performed. A Berendsen barostat and an anisotropic velocity rescaling thermostat were employed at 1 bar at 350 K. A timestep of 2 fs was used, with a leap-frog integrator. The LINCS algorithm was employed to constrain covalent bonds to their equilibrium lengths, a cutoff of 10 Å was used for both coulombic and van der Waals interactions and Particle Mesh Ewald (PME) was exploited to treat long-range electrostatics. As regards molecular topology and atomic point charges, the Automated Topology Builder (ATB) webserver was employed[47], choosing GROMOS 54A7 force field.

Bulk liquids for subcritical clusters (MiCMoS)

The starting simulation box with 432 benzoic acid molecules and size of ~47 Å obtained after 10^6 MC steps and described in Section 2.2.1 was further equilibrated with a 1 ns long *NpT* trajectory at 400 K with a Berendsen thermostat and an anisotropic minimal barostat. Then, 2 ns long *NpT* trajectories were performed at 350 K, 400 K, and 430 K with a CSVR thermostat[36]. An additional trajectory at 350 K comprising 1024 molecules was also run to assess size consistency.

Bulk liquids for subcritical clusters (Gromacs)

Molecular dynamics simulations of bulk liquids were also performed with Gromacs 2018.4 for comparison[46]. Input parameters were chosen as close as possible to the ones used for MiCMoS simulations. A Berendsen barostat and an anisotropic

velocity rescaling thermostat were employed to simulate an *NpT* ensemble at 1 bar and 350, 400, or 430 K, using as a starting point the pre-equilibrated box at 430 K containing 432 molecules obtained from the MiCMoS simulation. A timestep of 2 fs was used for a total duration of 3 ns, using a leap-frog integrator. The LINCS algorithm was employed to constrain covalent bonds to their equilibrium lengths, a cutoff of 16 Å was used for both coulombic and van der Waals interactions and Particle Mesh Ewald (PME) was exploited to treat long-range electrostatics. Molecular topology and atomic point charges were obtained from the Automated Topology Builder (ATB) webserver[47], choosing GROMOS 54A7 force field.

Bulk liquids for stacking interactions

Benzoic acid (BA, $T_{mp} = 395\ K$), 4-bromobenzoic acid (4-Br BA, $T_{mp} = 526\ K$), and 4-(methoxycarbonyl)benzoic acid (4-COOMe BA, $T_{mp} = 493\ K$) were simulated at four different temperatures around the melting point T_{mp} to study aromatic stacking interactions: $T_{mp} + 30\ K$, T_{mp}, $T_{mp} - 10\ K$, and $T_{mp} - 30\ K$. The following procedure was used to prepare the simulation boxes and equilibrate the system. A simulation box containing 432 molecules was generated with the *Boxliq* procedure. 10^8 Monte Carlo steps at $T_{mp} + 30\ K$ were performed to dispose of hard contacts and a 0.5 ns long *NpT* molecular dynamics simulation at $T_{mp} + 30\ K$ with a CSVR thermostat[36] was performed to thermalize the system above the melting point. Then, the system was thermalized at T_{mp}, $T_{mp} - 10\ K$, and $T_{mp} - 30\ K$ performing 0.5 ns long simulations. Finally, 2 ns long simulations were performed at each of the four temperatures, for each compound.

2.3 DFT-D calculations

2.3.1 Interaction energy of molecular aggregates

Geometry optimization of molecular aggregates extracted from molecular dynamics trajectories was performed with Gaussian 16[42], using B3LYP level of theory and 6-311G** basis set, along with Grimme's D3 empirical dispersion correction[48].

Frequency calculations were performed to confirm that the optimized structures correspond to true minima on the potential energy surface. Counterpoise calculations were then conducted to correct the aggregate energies for basis set superposition error (BSSE)[49,50]. Finally, the interaction energy of aggregates of *N* molecules was calculated by subtracting *N* times the energy of a monomer after geometry optimization. The resulting energy was normalized per molecule to facilitate a comparison among the aggregates. Example of input files are reported in *Appendix A*.

2.4 Experimental methods

2.4.1 Materials

4-Fluorobenzoic acid (4-F BA, CAS 456-22-4), 4-bromobenzoic acid (4-Br BA, CAS 586-76-5) and 4-(1H-1,2,3-triazol-1-yl)benzoic acid (4-HTrz BA, CAS 216959-87-4) were purchased from Fluorochem while 4-(methoxycarbonyl)benzoic acid (4-COOMe BA, CAS 1679-64-7) from ThermoFisher. 4-F BA, 4-Br BA, and 4-COOMe had a purity level of 99% while 4-HTrz was 97% pure. All the materials were used without further purification.

2.4.2 Solubility measurements

A Crystalline instrument (Technobis) was employed to determine solubilities by gravimetric method, after temperature recalibration. For each compound, 6 mL of saturated solutions in isopropyl alcohol (IPA) were stirred at 20.0 °C for at least 48 hours using a stirring rate of 400 rpm. The excess solid was allowed to settle for 1 hour and 4 mL of solution were first withdrawn using a syringe, and then poured in a clean vial, after the application of a syringe filter (13 mm, 0.45 μm membrane PTFE, VWR International). Afterwards, 3 aliquots of 1 mL of clear, saturated solution were collected in pre-weighted vials using a micropipette and the amount of solution ($m_{solution}$) was determined using an analytical balance. The solvent was evaporated in the oven at 43 °C under vacuum (0.05 MPa) and the amount of solid was determined by weight (m_{solid}). The amount of solvent was obtained by difference ($m_{solvent} =$

$m_{solution} - m_{solid}$) and the average solubility (S^*) was determined as the molar ratio $S^* = n_{solid}/n_{solvent}$.

2.4.3 Nucleation rate determination

The nucleation rate experiments were performed using a Crystal16 multi-reactor crystallizer (Technobis), after temperature recalibration. 29 mL of IPA were mixed with the proper amount of solute to reach the desired supersaturation. For each compound, five different supersaturations were studied. The solution was stirred for 1 hour at 40 °C to guarantee a complete dissolution. Then, a micropipette was used to withdraw aliquots of 1.5 mL and pour the solution into 16 weighted vials with a stirrer bar and a screw cap (9 mm polypropylene natural rubber, ThermoFisher). To get exactly 1.5 mL at 20 °C, thermal expansion of the 40 °C solution was taken into account. To prevent solvent evaporation, Parafilm® was applied around the caps. For the nucleation rate experiments, the procedure described elsewhere was used[51]. All the experiments were performed at 20.0 °C with a stirring rate of 900 rpm. First, the temperature was raised up to 40 °C using a heating rate of 5 °C per minute and held for 1 hour to ensure a complete dissolution. Then, the vials were crash-cooled decreasing the temperature to 20 °C with a cooling rate of -5 °C per minute and the temperature was held at 20 °C for 8 hours. Laser recalibration in each vial was performed on the clear solution, before the cooling stage. The temperature cycle was repeated 5 times to get 80 replicas of the same experiment, for each supersaturation. In total, 400 experiments were performed for each compound. During the experiments, the transmissivity in each vial was monitored at time intervals of 1 s to detect nucleation, marked by a sharp decrease in transmissivity.

3 Confinement

Confinement in nanospaces has been widely employed to accelerate chemical reactions, enhance the selectivity, stabilize reactive species and unstable aggregates, and even change electronic properties[52]. Another important application of confinement is crystallization. Confinement has been broadly applied to control crystallization, affecting crystal morphology, nucleation rate, and also the polymorphic form[53–57]. Micelles, droplets, vesicles, nanopores, nanochannels, and interstices between beads of block copolymers have been exploited as confining scaffolds[58–62]. The dynamics of a confined liquid before nucleation is altered with respect to bulk systems. Surface effects become increasingly important as the confined space becomes smaller, since the number of surface molecules increases, becoming comparable to that of bulk molecules. Surfaces can foster local ordering, and even heterogeneous nucleation. In addition, the spatial restrain given by confinement can lead to the crystallization of specific polymorphs and the stabilization of metastable phases. The size, shape, and chemistry of confining vessels can be tuned to obtain tailored materials.

Experimentally, the structure and dynamics of molecular liquids were investigated through X-ray diffraction[63], diffuse X-ray scattering[64], neutron scattering[65], and NMR[66]. Confinement was also exploited to study nucleation of NaCl with time-resolved TEM, using carbon nanotubes as confining vessels[67], and also

nucleation of glycine with micro-Raman, where confinement was achieved by the use of optical tweezers[68,69]. However, nucleation involves unstable intermediates and metastable structures that are difficult to detect with experimental techniques. Thus, there are many examples in the Literature where computational techniques like molecular dynamics are exploited to study systems under confinement[70,71].

In this Section, confinement is applied to undercooled liquids of benzoic acid to study molecular ordering and aggregation. To this purpose, a confinement algorithm described in detail in the *Methods* (Section 2.1.2) was developed and is summarized here for convenience (Section 3.1). The results of molecular dynamics simulations in confined spaces are presented and compared with the unconfined bulk liquid. First, the behavior of few molecules under confinement is investigated (Section 3.2); then, the number of molecules is increased to simulate liquids confined in different geometries, including extreme confinement, where molecular layers having a thickness comparable to molecular dimensions are simulated (Section 3.3).

3.1 Confinement algorithm in short

The confinement algorithm developed and implemented in MiCMoS is described in detail in the *Methods* (Section 2.1.2). Here is reported the essential information to understand the underlying logic. Nanoconfinement is obtained by exploiting van der Waals barriers, composed of a rectangular grid of massless, neutral pseudoatoms. The Lennard Jones parameters of these pseudoatoms can be chosen among all the atom types implemented in MiCMoS; however, for this work, carbon atom type of LJC force field is employed, to simulate a hydrophobic environment. The repulsive term of Lennard Jones is always taken into account, while the attractive term can be switched on or off, producing a full Van der Waals or a repulsive-only barrier. Harmonic potentials are employed to avoid unphysical deformations of barriers, that can stretch and compress according to the pressure of the system. In addition, the value of harmonic force constants can be tuned to simulate confined environments having different stiffness.

3.2 Confinement of few molecules

What happens when two benzoic acid molecules are confined into a box of nanometric size? Most likely, they will not remain monomers for a long time, as they can interact with one another forming a dimeric structure made up by strong hydrogen bonds. But what happens when the number of molecules is increased up to three? Different types of aggregates might form: either dimer + monomer or a single trimer. The complexity increases when four molecules are considered: tetramer, trimer + monomer, dimer + dimer. In this Section, preliminary simulations are performed to explore the possible interaction modes of up to 8 benzoic acid molecules confined in cubic boxes with full van der Waals barriers. These simulations are intended as exploratory tests, carried out mainly to observe qualitative trends and gain intuition about the system's behavior, rather than to draw universal or quantitatively accurate conclusions, which would require extensive statistical sampling.

3.2.1 Formation of molecular aggregates

Benzoic acid has a carboxylic group capable of forming hydrogen bonds with surrounding molecules. These hydrogen bonds can be either single or double, and in both cases are particularly strong. Therefore, benzoic acid is likely to form aggregates. Confinement in a small cubic box (edge size 20 Å) can be exploited to induce aggregation in rather short timeframes. Below, the molecular aggregates obtained during the dynamics of some (2-8) benzoic acid molecules is reported.

Two to six molecules

Randomly oriented benzoic acid molecules tend to organize in aggregates to maximize favorable interactions that lead to a high and negative cohesive energy. When two molecules are constrained in a confined space, they form a dimeric aggregate held together by a cyclic hydrogen bond (Figure 3.1a). This interaction mode is the one also found in the crystal structure of benzoic acid[43] and is common in many carboxylic acids, both aromatic and aliphatic. Indeed, the cyclic COOH dimer is considered to be a supramolecular synthon and is widely exploited in the field of crystal engineering[72]. The single hydrogen bond dimer is less stable and is only found as intermediate state (Figure 3.1b).

When three molecules are inserted in the cubic nanocavity, they eventually form a cyclic trimer (Figure 3.1c). This interaction mode is very rare in the crystalline state, occurring only in a single crystal structure in the entire Cambridge Structural Database (CSD), FIZFEB[73], which crystallizes in the uncommon space group $R\bar{3}$ (<1%)[74], but might be rather common in liquid benzoic acid (see Section 4.3.2). A possible explanation could be the difficulty to pack trimers efficiently. However, intermediate structures are also found. One example is an aggregate formed by a double hydrogen-bonded dimer interacting with another molecule through a single hydrogen bond (Figure 3.1d). In this case, all hydrogen bond donors are employed but one acceptor is not used. Another example is the linear trimer, which has one donor and one acceptor unused (Figure 3.1e).

In the simulation with four molecules, a cyclic tetramer is formed eventually (Figure 3.1f). In this supramolecular aggregate, all hydrogen bond donors and acceptors are satisfied. In addition, aromatic rings interact in pairs through aromatic stacking interactions, constituting a distorted tetrahedron geometry having approximately a S_4 point symmetry. Obviously, this geometry is not maintained during the simulation since aromatic stacking interactions are intermittent, and the stacking partners frequently exchange over time. Intermediate, short-lived states are also detected like a linear tetramer (Figure 3.1g) and a cyclic hydrogen-bonded trimer interacting with a monomer through single hydrogen bond (Figure 3.1h). In both cases, not all hydrogen bonds are satisfied.

When five molecules are considered, the number of molecules is odd but a cyclic pentamer forms anyway (Figure 3.1i). This structure is clearly frustrated, since only two pairs of molecules at a time can make aromatic stacking interactions, leaving the other one spectating. Like the case of the tetramer, these interactions exchange partner frequently. In this case, multiple intermediate structures can be detected such as cyclic trimer plus cyclic dimer (Figure 3.1j), cyclic trimer with a linear dimer interacting through a single hydrogen bond (Figure 3.1k), cyclic tetramer interacting with a monomer through hydrogen bond (Figure 3.1l) and finally linear pentamer (Figure 3.1m). In this latter case, "linear" refers to the network of hydrogen-bonded molecules. However, the geometric configuration of this aggregate is somehow folded to maximize cohesion and to avoid clashes with the barriers of the cubic box.

Adding one more molecule in the cubic box, a cyclic hexamer forms (Figure 3.1n). The number of molecules is even in this case, however there are two distinct groups of three molecules lying on opposite sides of the ring constituted by hydrogen bonds. Thus, intermolecular interactions are maximized when two molecules interact through parallel aromatic stacking while the unpaired molecule interacts through T-shape aromatic interactions. It goes without saying that all the interactions are frequently exchanged. During the trajectory, the hexamer splits into a cyclic tetramer and a cyclic dimer, and the system remains for ~750 ps in this configuration, which is

less stable than the cyclic hexamer (Figure 3.1o). Other intermediate structures are a linear hexamer (Figure 3.1p) and a cyclic pentamer plus one monomer (Figure 3.1q).

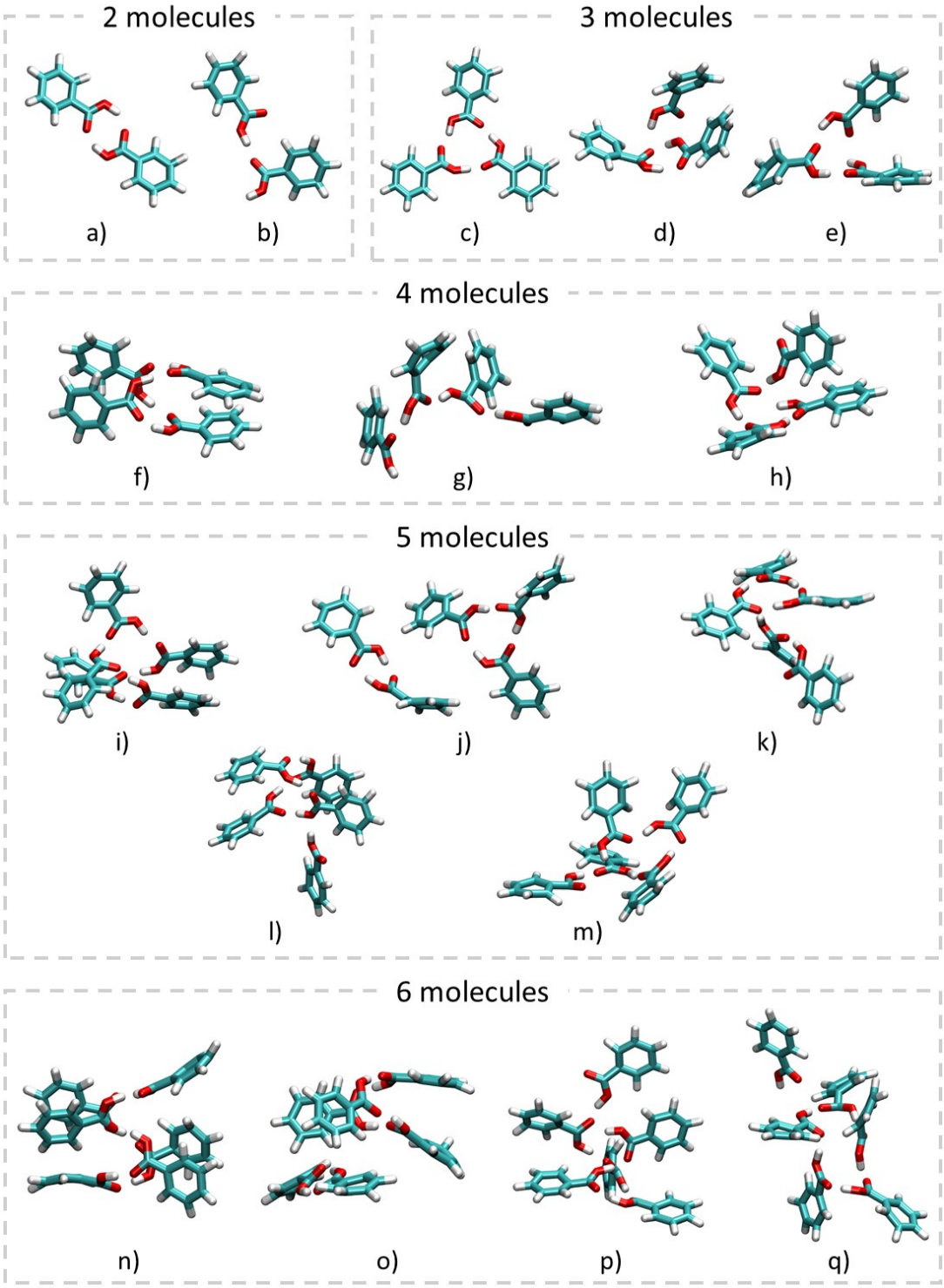

Figure 3.1 Collection of molecular aggregates in the simulations containing 2-6 molecules.

More than six molecules

When more than six molecules are considered, cyclic structures do not form. This could be explained by the small size of the cubic box, the higher entropy penalty to form a single aggregate compared to different smaller aggregates, and the short simulation time. Nevertheless, these simulations further confirm the stability of the small aggregates described before and show how these pack together. In other words, it seems to be an upper limit (or an upper range) of sizes for cyclic aggregates, at least for these confined simulations.

Seven molecules arrange mainly in either one cyclic trimer and two cyclic dimers (Figure 3.2a) or one cyclic tetramer and one cyclic trimer (Figure 3.2b) for almost the entire simulation. In both cases, the different aggregates tend to interact with one another through aromatic stacking interactions. Another intermediate structure that is also found is three cyclic dimers plus one monomer interacting through a single hydrogen bond (Figure 3.2c).

When eight molecules are confined together, several aggregates form, including four cyclic dimers tending to stack on one another, resembling the crystal structure of benzoic acid (Figure 3.2d), two cyclic trimers plus one cyclic dimer (Figure 3.2e), one linear tetramer plus two cyclic dimers (Figure 3.2f), one linear hexamer plus one cyclic dimer (Figure 3.2g), one linear pentamer interacting through a single hydrogen bond with a cyclic trimer (Figure 3.2h), and a linear octamer (Figure 3.2i). In general, the most favorable interaction between these small aggregates occurs through parallel aromatic stacking. Even in the case of the linear octamer where a single aggregate is present, the catemeric structure is folded to allow non-consecutive molecules to interact through aromatic stacking.

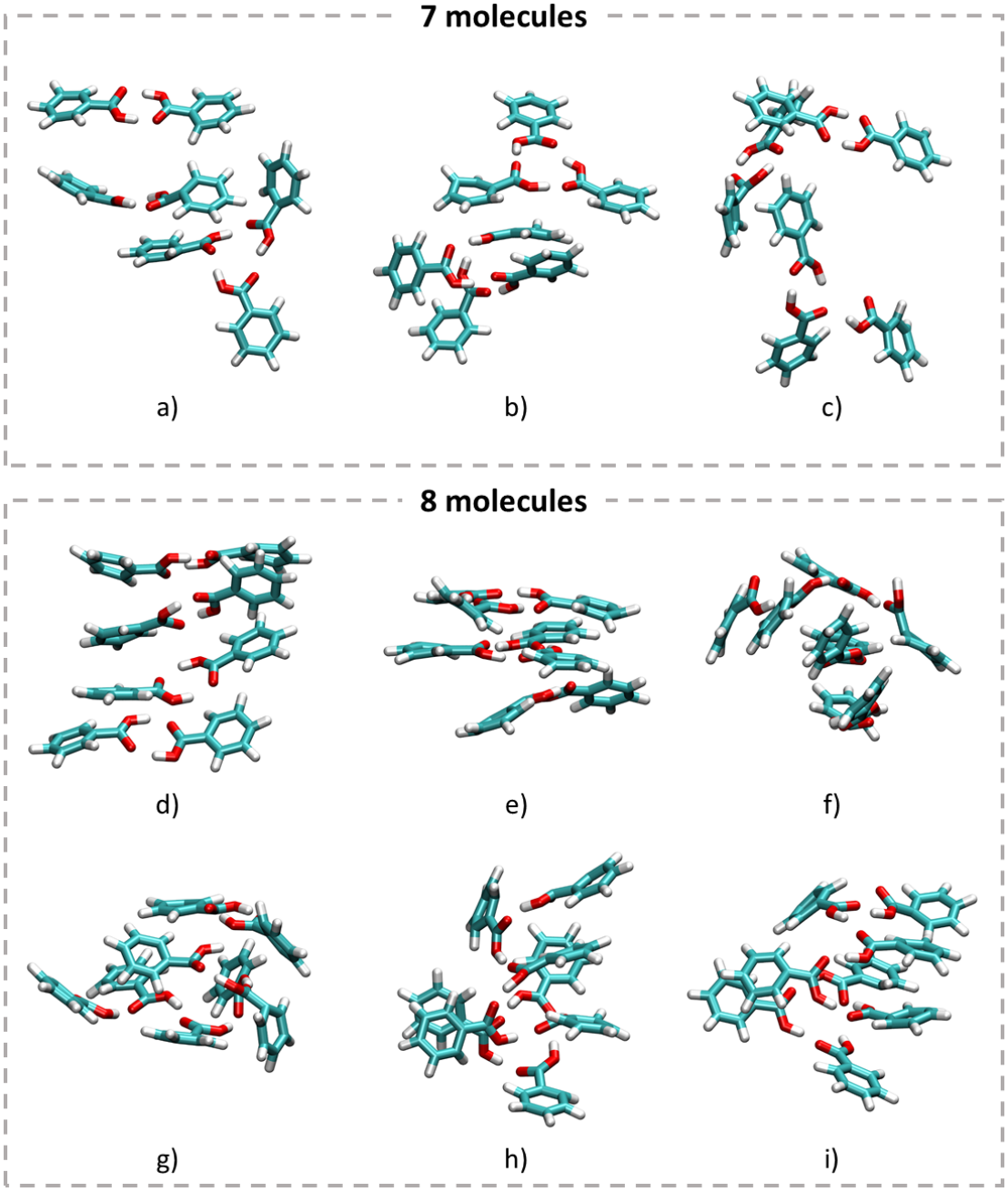

Figure 3.2 Collection of molecular aggregates in the simulations containing 7-8 molecules.

3.2.2 Analysis of cyclic aggregates

Cyclic aggregates are by far the most persistent ones, since they have all their hydrogen bond donors and acceptors involved in hydrogen bonds. In this Section, these

aggregates are analyzed more in detail, studying intermolecular interactions and cohesive energy.

Radial distribution function

Radial distribution function (RDF) $g(r)$, defined by Equation 2.19, is a useful tool to study the local structure of periodic systems. Here it is employed to study specific molecular aggregates containing only few molecules. Therefore, intensities are normalized with respect to the highest peak for each aggregate, and results are discussed only qualitatively (Figure 3.3a,b). Though, the peaks position remains unaltered, since it does not depend on normalization factors.

The center-of-mass (c.o.m.) RDFs of the cyclic dimer and trimer are characterized by a single peak with maximum at 6.95 and 7.25 Å respectively. However, while the peak of the dimer is sharp, the peak related to the trimer is broader, presenting a shoulder at shorter distances. This is due to distorted geometries allowed by the less rigid structure of the trimer, having each pair of molecules interacting through single hydrogen bonds rather than double. In both aggregates, aromatic stacking interactions are geometrically forbidden. The tetramer has a $g(r)$ that is almost identical to the one of the trimer, apart from an additional broad peak with maximum at 4.35 Å, indicating the presence of aromatic stacking interactions, allowed by greater flexibility. The RDF of the pentamer has a broader peak at 7.25 Å with a shoulder at lower distances, due to the presence of an even less rigid structure. The broad peak corresponding to aromatic stacking interactions is shifted at longer distances with respect to the tetramer, with maximum at 4.45 Å, indicating the presence of non-optimal interactions due to the frustrated structure with two pairs of molecules stacked and a molecule unmatched. This latter gives an additional weak, albeit characteristic, signal at 8.95 Å. Finally, the RDF of the hexamer has a peak at 7.15 Å like the pentamer. The feature relative to aromatic stacking interactions is a broad, double peak with maxima at 4.05 Å, indicating parallel stacking, and 5.05 Å, signaling the presence of T-shape stacking. Each molecule, on average, interacts with two molecules through hydrogen bonds and two molecules through parallel and T-shape

stacking. The remaining molecule is farther apart and does not interact strongly, being responsible of the shoulder at ~8-9 Å.

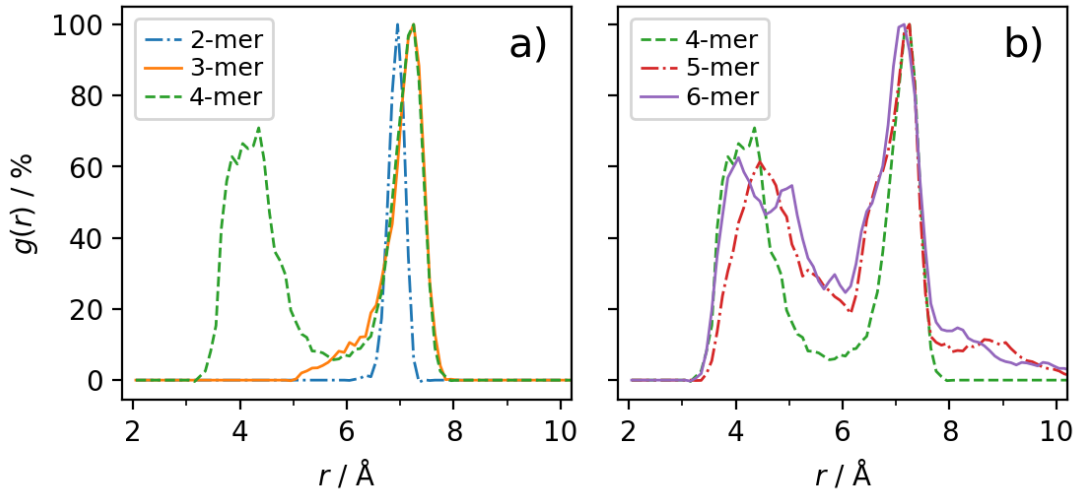

Figure 3.3 Center-of-mass radial distribution functions of cyclic molecular aggregates composed of 2-4 molecules (a), and 4-6 molecules (b). $g(r)$ values are normalized to 100% and are averaged over the 500-1000 ps range (5000 frames), where each cyclic aggregate is present. The data are split into two subplots to avoid overcrowding. The tetramer (4-mer) curve is included in both panels to facilitate comparison.

Cohesive energy

The molecular recognition process underlying the formation of aggregates is fast, taking no longer than 50 ps. This is evident from the analysis of cohesive energy of the systems, normalized per molecule (Figure 3.4a-e). Indeed, there is an energy change from around 0 (almost non-interacting molecules) to negative values down to $-60\ kJ\ mol^{-1}$ per molecule (strongly hydrogen-bonded aggregates). Significant changes from the cyclic aggregate geometries are reflected in a less negative cohesive energy, like the transient dissociation of the tetramer at 1490 ps, of the pentamer at 1500 and 1900 ps, and the transition from hexamer to tetramer plus dimer at 1280 ps. Since the structure of the aggregates during the simulation is dynamic, their cohesive energies were estimated by averaging 5000 frames between 500 and 1000 ps, ensuring that the aggregates persisted throughout this time interval.

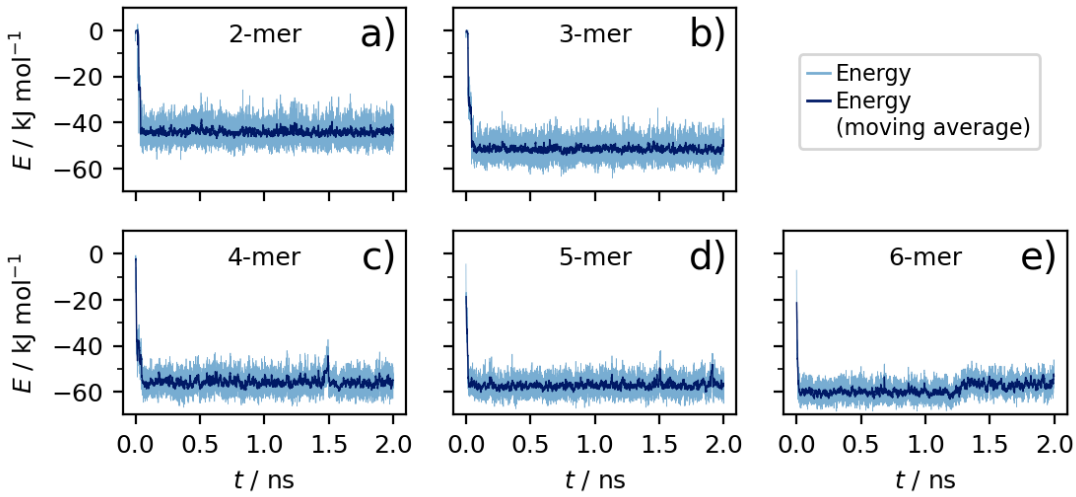

Figure 3.4 Time evolution of cohesive energy of the systems containing 2-6 molecules. Energy values are normalized per molecule. The curves are smoothed by applying a moving average computed over a 3 ps (30 frames) time window.

DFT-D calculations

Cyclic aggregates from 2 to 6 molecules were extracted from the molecular dynamics trajectories to perform DFT calculations using B3LYP functional, 6-311G** basis set and Grimme's D3 empirical dispersion correction[48] (Figure 3.5).

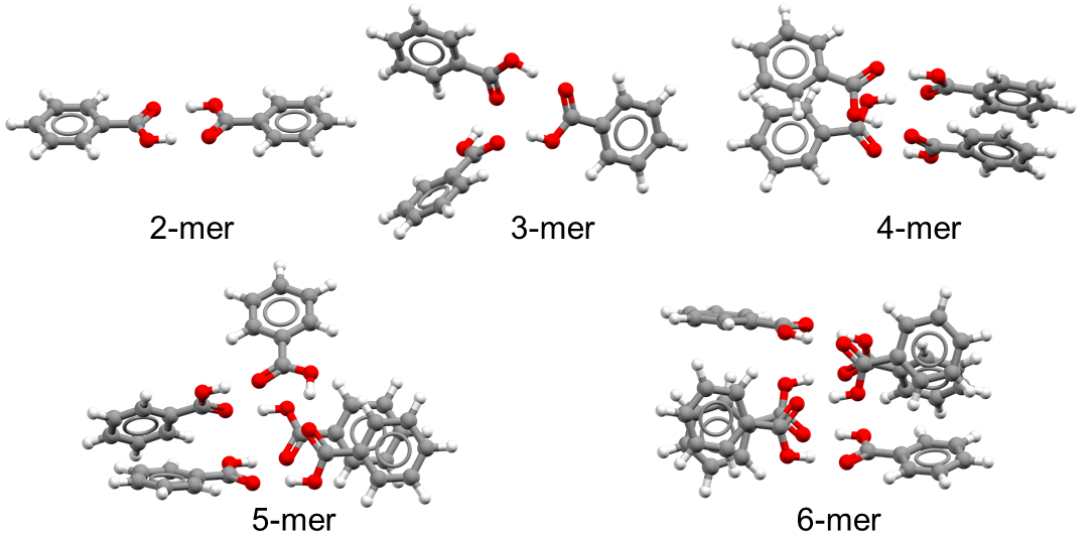

Figure 3.5 DFT-D optimized geometries of cyclic molecular aggregates containing 2-6 molecules.

The geometry was optimized, and the energy was corrected for the basis set superposition error (BSSE) using the counterpoise method. The electronic energy per molecule of cyclic aggregates was compared to the cohesive energy derived from the

LJC force field employed in the molecular dynamics simulations (Figure 3.6). It is important to consider that the two energies are not directly comparable. On the one hand, electronic energy is calculated on a static, optimized structure at 0 K using an expensive, quantum mechanical method. On the other hand, the cohesive energy derived from molecular dynamics simulations considers temperature and dynamical effects but relies on cheaper force field methods. However, both methods give in general the same qualitative trend, with more negative energies associated with larger aggregates, with the only exception of the electronic energy of the cyclic trimer which is less negative than the one of the cyclic dimer by less than 2 $kJ\ mol^{-1}$. For the other aggregates, the difference between DFT-D and LJC energies is, on average, around 2 standard deviations, demonstrating also a good quantitative agreement.

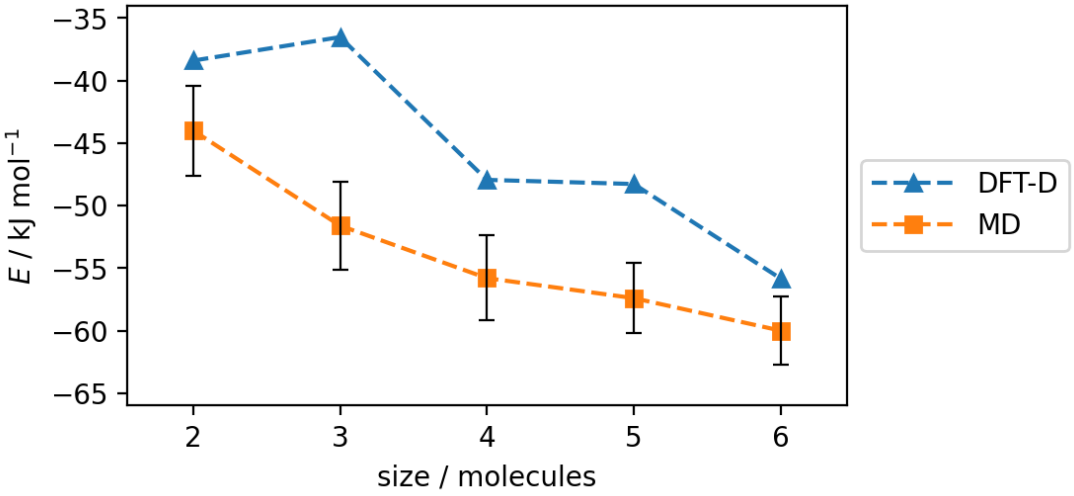

Figure 3.6 Comparison between DFT-D and MD energy of cyclic molecular aggregates containing 2-6 molecules. DFT-D energy values refer to optimized structures and are corrected for BSSE, while MD ones come from averages over 500 ps (5000 frames). Standard deviations are also reported. In both cases, energy values are normalized per molecule. Solid lines are shown only as a guide to the eye.

3.3 Confinement of many molecules

In the previous Section, the behavior of few molecules confined into a cubic box has been shown. In this Section, the number of molecules is increased up to 432, constituting a liquid, and confined spaces having different periodicity are utilized: nanolayers (2D), nanotubes (1D), and nanocavities (0D). Both repulsive-only and full

van der Waals barriers are employed. The results are compared with bulk liquid, which is not confined and has periodic boundary conditions in the three dimensions. Finally, simulations in extremely confined nanolayers are also performed on either 216 or 108 molecules to investigate new aggregates (Section 3.3.6).

3.3.1 Diffusion and viscosity

Diffusion in a liquid is governed by intermolecular interactions and is directly related to dynamic viscosity, a key macroscopic property that can be measured experimentally. From a computational point of view, the self-diffusion coefficient D can be determined from the molecular mean squared displacement (Equation 2.22) and can be used to estimate the dynamic viscosity η using the Stokes-Einstein equation (Equation 2.23). As a reference, a typical value of D for hydrocarbons is of the order of $10^{-8}\ m^2 s^{-1}$ while one of the bulk liquid of benzoic acid at 350 K is 2 orders of magnitude lower, that is reasonable due to the strong hydrogen bonds that characterize the system and the 45 K of undercooling[34]. Therefore, η obtained from molecular dynamics simulations with MiCMoS is high, reaching a value of $0.00262(1)\ Pa\ s$, that is in good agreement with Gromacs estimate of $0.00290(6)\ Pa\ s$. Both values are also very close to the experimental one of $0.00315\ Pa\ s$ at 343.57 K, deposited in public repositories[75], and $0.00196\ Pa\ s$ at 350 K, extrapolated from a linear regression on data ranging from 400 to 450 K[76]. Also the density of bulk liquid (MiCMoS: $1.109(8)\ g\ cm^{-3}$; Gromacs: $1.156(1)\ g\ cm^{-3}$) is in agreement with the experimental value of $1.150\ g\ cm^{-3}$, indicating the reliability of the force field.

When the system is simulated in confined spaces, D drops and η increases (Figure 3.7a,b). The greatest effect can be seen in the nanocavity simulations where self-diffusion is the lowest and dynamic viscosity is the highest. This indicates that the greater the degree of confinement, the stronger the impact on the dynamics of the system. The degree of confinement can be described by the number of non-periodic directions N_{NP}, which corresponds to the number of pairs of parallel barriers introduced in the system: bulk liquid (0), nanolayer (1), nanotube (2), nanocavity (3). When full van der Waals barriers are considered, the decrease of D and the increase of η are less

evident. A possible explanation is that the degree of ordering is higher in confined liquids with full van der Waals barriers, leading to stronger molecular correlations that allow for more concerted movements, resulting in a higher self-diffusion coefficient. To gain more information on this matter, the rotational correlation and the structure of the liquid are investigated in Section 3.3.2 and Section 3.3.3.

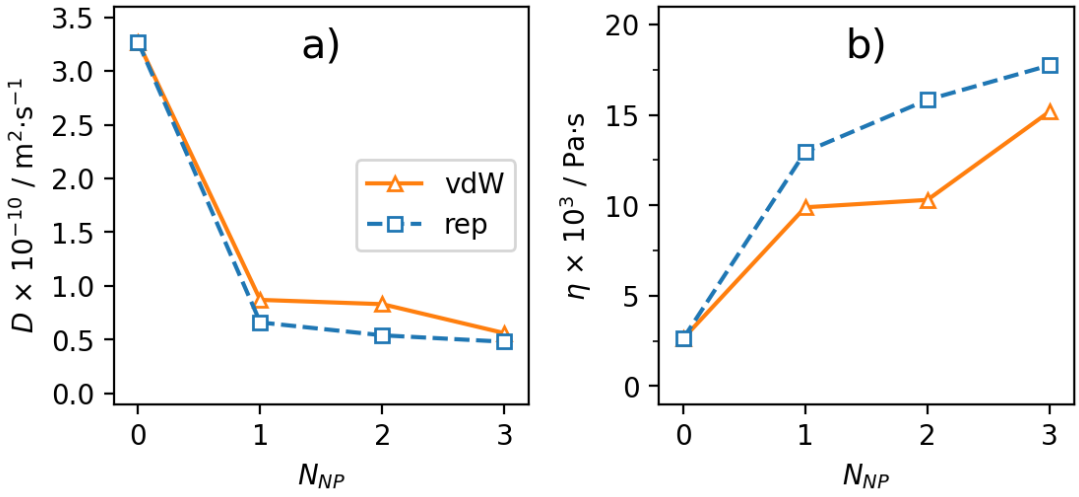

Figure 3.7 Self-diffusion coefficient (a) and dynamic viscosity (b) as a function of the number of non-periodic directions N_{NP} (0 = bulk liquid, 1 = nanolayer, 2 = nanotube, 3 = nanocavity). Results obtained with full van der Waals (vdW) and repulsive-only (rep) barriers are compared. Solid lines are shown only as a guide to the eye.

Size effect

Another key parameter that influences the dynamics of the liquid is the size of the confined space. The packing efficiency was maintained at 66%, the volume and the number of molecules in the simulation box were doubled and the distance between parallel barriers, initially set at ~42 Å, was increased by a factor 1.5 for nanolayers and nanotubes and by a factor of $\sqrt[3]{2}$ (~1.26) for nanocavities. Densities remain almost unchanged while self-diffusion coefficients increase, and dynamic viscosities decrease. This behavior is expected since the higher distance between barriers makes the system more similar to the bulk liquid[38].

3.3.2 Rotational correlation

Beyond self-diffusion and dynamic viscosity, rotational correlation is another fundamental parameter for characterizing the dynamics of a system. The rotational correlation function $C(t)$ is defined by Equation 2.21. Since the choice of reference vectors can significantly affect $C(t)$, two vectors were used for comparative analysis (Figure 3.8): $\boldsymbol{v}_1$ that is parallel to the aromatic ring and along the C1-C4 direction, and $\boldsymbol{v}_2$ that is perpendicular to the aromatic ring. From this point on, quantities referring to $\boldsymbol{v}_1$ will be denoted with the subscript ∥, and those associated with $\boldsymbol{v}_2$ with the subscript ⊥.

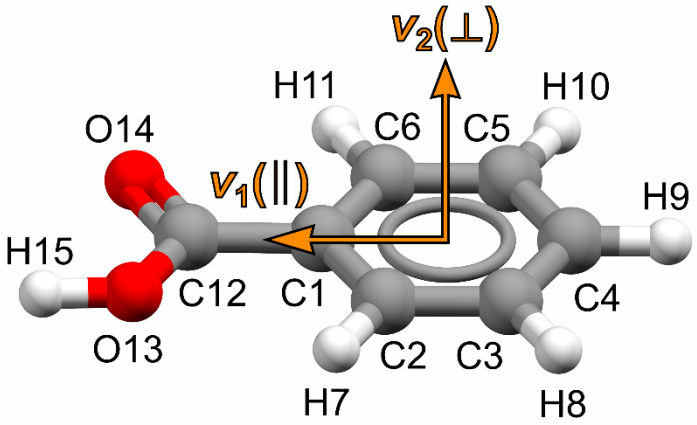

Figure 3.8 Scheme of benzoic acid molecule with atomic numbering. Vectors v_1 (along the C1-C4 direction, parallel to the aromatic ring) and v_2 (orthogonal to the aromatic ring) are also shown.

The rotational correlation function can be modeled with Kohlrausch-Williams-Watt function or stretched exponential[77,78], described in Equation 3.1.

$$C(t) = C(0)e^{-(t/\tau_c)^{\beta}} \tag{3.1}$$

τ_c is the rotational correlation time, representing the time required by a molecule to rotate by one radian. β is the Kohlrausch exponent, an empirical parameter ranging from 0 to 1 that quantifies the deviation from purely exponential behavior, corresponding to a value of 1. A value of β significantly lower than 1, instead, is related to a slow molecular reorientation due to correlated or collective motions[79]. This behavior is already observed in the bulk liquid of benzoic acid, where strong, directional hydrogen bonds occur between the carboxylic functional groups.

Accordingly, translational diffusion is less hindered than rotational diffusion, as is commonly observed in undercooled liquids[80].

As for confined liquids, the relaxation is significantly slower, particularly when $\boldsymbol{v}_1$ is considered. $C(t)$ decays very slowly, remaining between 70 and 80% even after 500 ps, compared to 40% of bulk liquid. When $\boldsymbol{v}_2$ is considered, instead, $C(t)$ reaches 40-50% in 500 ps for confined liquids and falls below 20% in bulk liquid[38]. This behavior can be explained by the presence of a rotatable C-C bond which does not change the orientation of $\boldsymbol{v}_1$, but has a massive impact on $\boldsymbol{v}_2$, provoking a quicker decay of $C(t)$. To get τ_c and β values for bulk and confined liquids, $C(t)$ curves were fitted with stretched exponentials (Equation 3.1) using a non-linear Levenberg-Marquardt least-squares algorithm implemented in Fityk 1.3.1[81].

Table 3.1 List of rotational correlation times (τ_c) and Kohlrausch exponents (β) referred to vector v_1 ($\parallel$) and v_2 ($\perp$) for bulk liquid and confined systems with either full van der Waals (vdW) or repulsive-only (rep) barriers.

Barrier type	System	Vector	τ_c / ps	β
None	Bulk liquid	$v_1(\parallel)$	398(2)	0.829(1)
		$v_2(\perp)$	122(4)	0.56(1)
vdW	Nanolayer	$v_1(\parallel)$	1200(20)	0.86(1)
		$v_2(\perp)$	390(4)	0.64(1)
	Nanotube	$v_1(\parallel)$	6300(400)	0.435(1)
		$v_2(\perp)$	304(6)	0.54(1)
	Nanocavity	$v_1(\parallel)$	6300(500)	0.435(2)
		$v_2(\perp)$	568(7)	0.54(1)
Rep	Nanolayer	$v_1(\parallel)$	4900(300)	0.59(2)
		$v_2(\perp)$	508(5)	0.69(2)
	Nanotube	$v_1(\parallel)$	6000(300)	0.48(1)
		$v_2(\perp)$	724(9)	0.57(1)
	Nanocavity	$v_1(\parallel)$	2600(100)	0.66(2)
		$v_2(\perp)$	653(7)	0.56(1)

In both bulk liquid and confined simulations, $\tau_{c\parallel} > \tau_{c\perp}$ and the lowest values are associated with the bulk liquid, as expected (Table 3.1). In addition, $\tau_{c\parallel}$ values have an inverse correlation with $\beta_{\parallel}$, meaning that higher rotational correlation times correspond to a higher deviation from an ideal exponential decay of $C(t)$. This is due to the presence of strong, directional hydrogen bonds between carboxylic groups along

the $\boldsymbol{v}_1$ direction. The rotation of a hydrogen-bonded dimer is disfavored, as is the reorientation of a monomer following hydrogen bond breakage. In contrast, weaker interactions are present along the $\boldsymbol{v}_2$ direction, thus rotational correlation times are much lower and no obvious correlation is found between $\tau_{c\perp}$ and $\beta_\perp$.

3.3.3 Structure of the liquid

Cohesive energy

The cohesive energy is closely related to the liquid structure and is obtained by summing all coulombic and dispersion intermolecular interactions among benzoic acid molecules, taking care to avoid double counting of pairwise interactions and neglecting the interactions with the barriers. The corresponding value in the bulk liquid is $-85.3\ kJ\ mol^{-1}$ per molecule. When barriers are introduced in the system, the cohesive energy of the liquid becomes less negative, and the more the number of non-periodic directions N_{NP}, the less negative the cohesive energy (Figure 3.9).

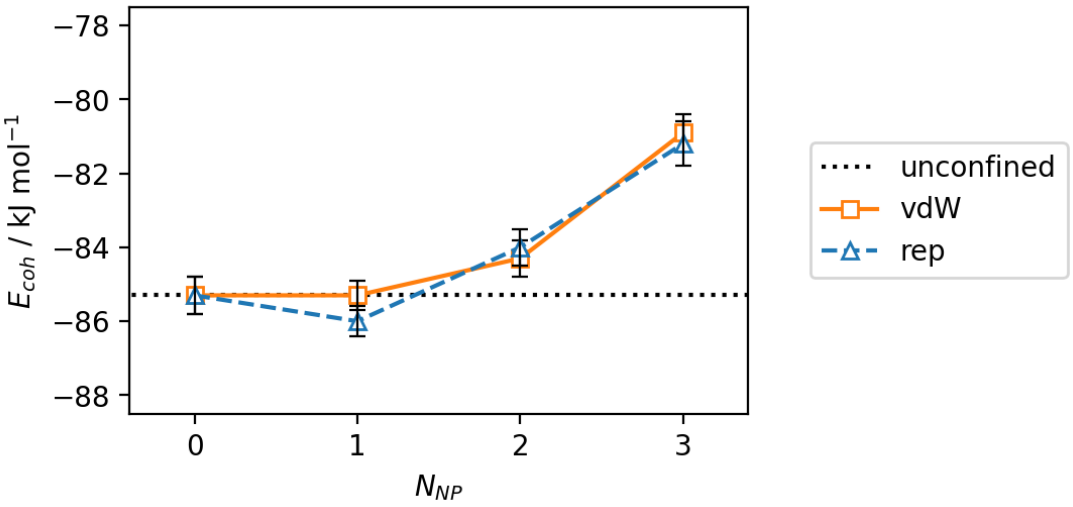

Figure 3.9 Cohesive energy E_{coh} per molecule as a function of the number of non-periodic directions N_{NP} (0 = bulk liquid, 1 = nanolayer, 2 = nanotube, 3 = nanocavity). Results obtained with full van der Waals (vdW) and repulsive-only (rep) barriers are compared. The unconfined bulk liquid is used as a reference. Solid lines are shown only as a guide to the eye.

This is due to the under-coordination of molecules adjacent to the barriers. Indeed, the greatest effect is reached in the nanocavity simulations, where N_{NP} is maximum and the cohesion energy is $4\ kJ\ mol^{-1}$ per molecule less negative,

regardless of the potential used to model the barriers. In nanolayer simulations, instead, where N_{NP} is only 1, the effect is less evident and the cohesive energy is statistically identical to the bulk liquid.

Radial distribution functions

Another way to characterize the average structure of the liquid is the radial distribution function (RDF), indicated by $g(r)$ and defined in Equation 2.19. Center-of-mass (c.o.m.) and OH···O RDFs of confined liquids are compared to the ones of the bulk liquid and thermalized crystal of benzoic acid (Figure 3.10). RDFs were averaged over the last 100 ps of each trajectory (200 frames). To enable a consistent comparison of the RDFs, periodic boundary conditions were applied, if not already present, prior to the calculation of the corresponding $g(r)$ functions. As a result, all curves share the same normalization and are fully comparable. Any observed differences can thus be attributed solely to the distinct behavior of benzoic acid in confined versus unconfined environments. All liquids exhibit a similar hydrogen bond structure (Figure 3.10b), characterized by a pronounced peak at ~1.5 Å, which arises from close OH···O contacts between carboxylic groups. This peak is slightly shifted to lower distances compared to the corresponding feature in the crystal. The broad, unstructured band centered at ~3.5 Å corresponds to molecules that are within the same aggregate, but not directly linked through hydrogen bonds. The c.o.m. RDFs of the confined systems are qualitatively similar to that of the bulk liquid (Figure 3.10a). All $g(r)$ functions show the same broad peak, consisting of a sharp maximum at ~7 Å and a wide shoulder extending down to ~3.5 Å. Comparison with c.o.m. RDF of the crystal indicates that the peak at ~7 Å corresponds to hydrogen-bonded benzoic acid cyclic dimers. In contrast, the shoulder at lower distances likely reflects random close packing of molecules, which, in the crystal, corresponds to short-range, translation-dependent, aromatic stacking interactions. Notably, the ~8 Å peak observed in the crystal, associated with long-range translational correlations, is entirely absent in all liquids.

The subtle differences between confined liquids and bulk liquid can be caught calculating the $\Delta g(r)$, using the bulk liquid as a reference (Figure 3.10c). The positive

$\Delta g(r)$ at ~4 Å indicates a higher number of molecules that interact via aromatic stacking when confinement is active. In particular, the effect increases with the degree of confinement, regardless of the type of barriers used. However, it is more pronounced when purely repulsive barriers are employed. Instead, the negative $\Delta g(r)$ at ~7 Å signals the presence of a fewer number of hydrogen-bonded dimers. A deeper analysis of aromatic stacking interactions is provided in Section 5.

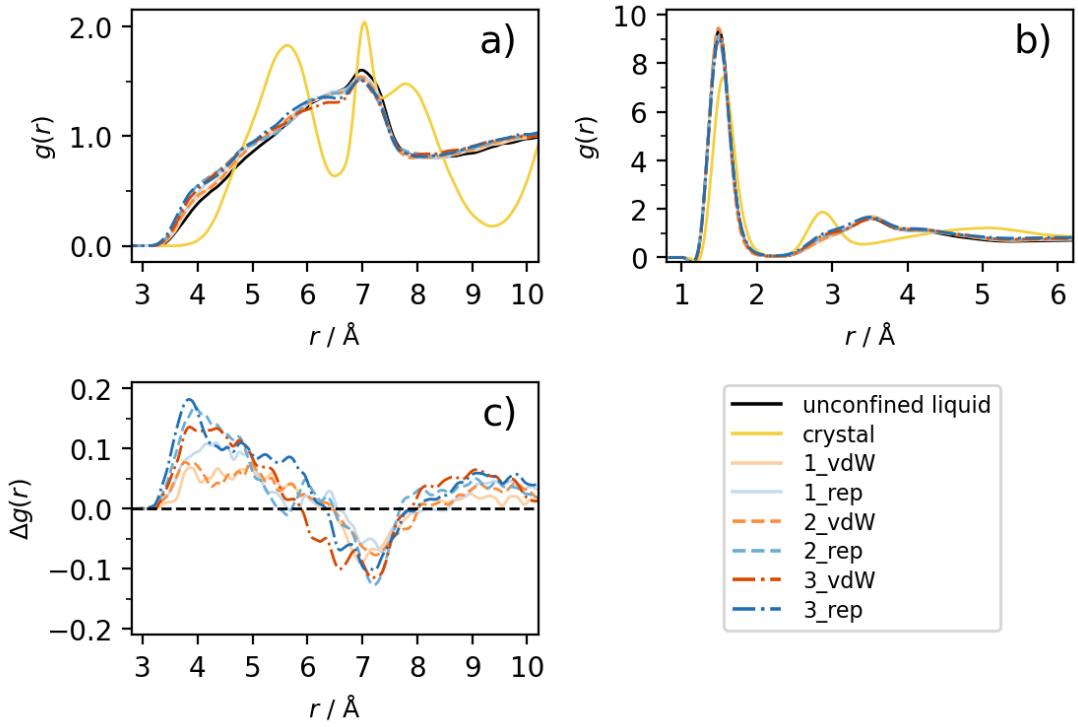

Figure 3.10 Center-of-mass RDF (a) and OH···O RDF (b) of nanolayers (1), nanotubes (2), and nanocavities (3) with either full van der Waals (vdW) or repulsive-only (rep) barriers. The unconfined bulk liquid and crystal are used as a reference. The center-of-mass RDF difference between confined liquids and unconfined bulk liquid is also shown (c).

Hydrogen bonds

To deepen the understanding of the liquid structure, the dynamics of hydrogen bonds was explored, providing a more complete picture of molecular interactions under confinement. In all benzoic acid liquids, most hydrogen bonds fall within the range of 1.55 to 1.75 Å. This is significantly shorter than the sum of the van der Waals radii (2.68 Å), confirming the tendency of the system to form strong hydrogen bonds. The number of hydrogen bonds and their persistence were analyzed, using a threshold

distance of 90% of the sum of van der Waals radii of donor and acceptor. This threshold allows to exclude long, weak contacts that are also intermittent, with a lifetime of 0.5-1.0 ps. The average number of hydrogen bonds per frame is ~433 in all the simulations. Considering that the total number of molecules is 432, this indicates that, on average, each molecule is involved in a hydrogen bond, confirming the high hydrogen bonding propensity of the carboxylic functions. To study the time persistence, all the hydrogen bonds present the last frame of the trajectory, chosen as a reference, were determined. The choice of reference frame has no significant impact on the results, as the same trend is consistently observed regardless of the selected frame. Each time the OH···O distance raises above the defined threshold, the hydrogen bond is assumed to be broken, and the time persistence is registered (Figure 3.11). Most hydrogen bonds are destroyed in 2 ps or less, but also persistent interactions are present, that last more than 200 ps. Compared to the bulk liquid, the confined systems exhibit a higher number of short-lived hydrogen bonds (≤10 ps) and a lower number of long-lived ones (>50 ps). This suggests that the presence of barriers enhances the frequency of hydrogen bond breaking and reformation events. No clear dependence on either the confinement geometry or the type of barriers is observed.

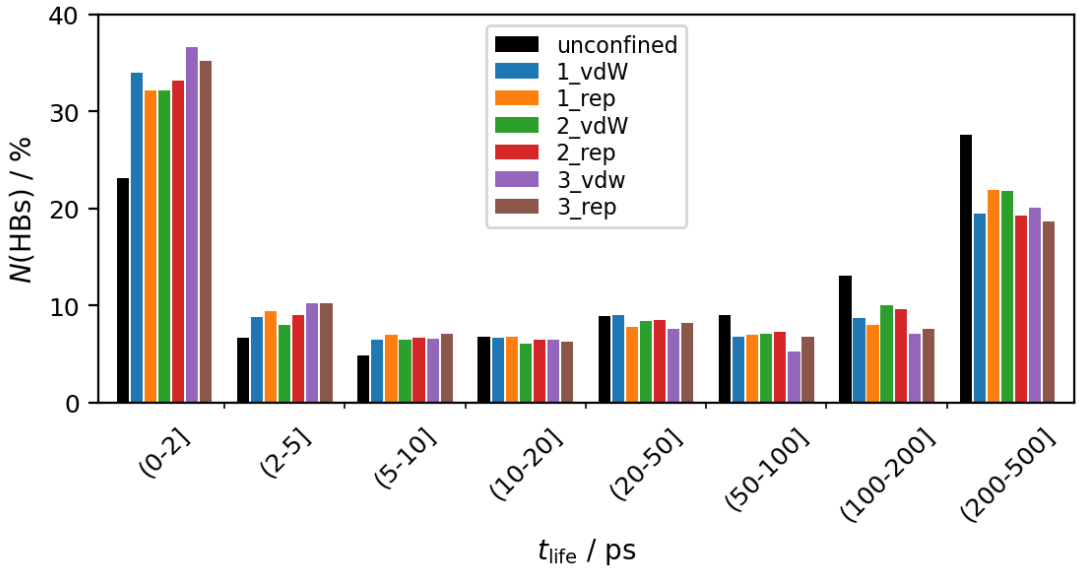

Figure 3.11 Percentage of hydrogen bonds as a function of their lifetime in the simulations confined into nanolayers (1), nanotubes (2), and nanocavities (3) with either full van der Waals (vdW) or repulsive-only (rep) barriers. The unconfined bulk liquid is also shown as a reference.

Number density

A common feature of confined liquids is the inhomogeneous spatial distribution of density. The average local number density $\langle \rho_{loc} \rangle$ was calculated by counting the number of molecular centers of mass within 0.1 Å-thick slices oriented parallel to the confining barriers of nanocavity simulations (Figure 3.12). This quantity was averaged over the three Cartesian directions and computed over the final 300 ps of the simulations (600 frames). At ~3 Å there is the main peak indicating that there is a first layer of molecules adsorbed to the barriers. At larger distances, $\langle \rho_{loc} \rangle$ exhibits a damped oscillatory behavior that is also found in confined liquids of van der Waals spheres[82]. When repulsive-only barriers are employed, the peaks are shifted at slightly higher distances to the lack of the attractive part of the potential, but the behavior remains mostly unchanged.

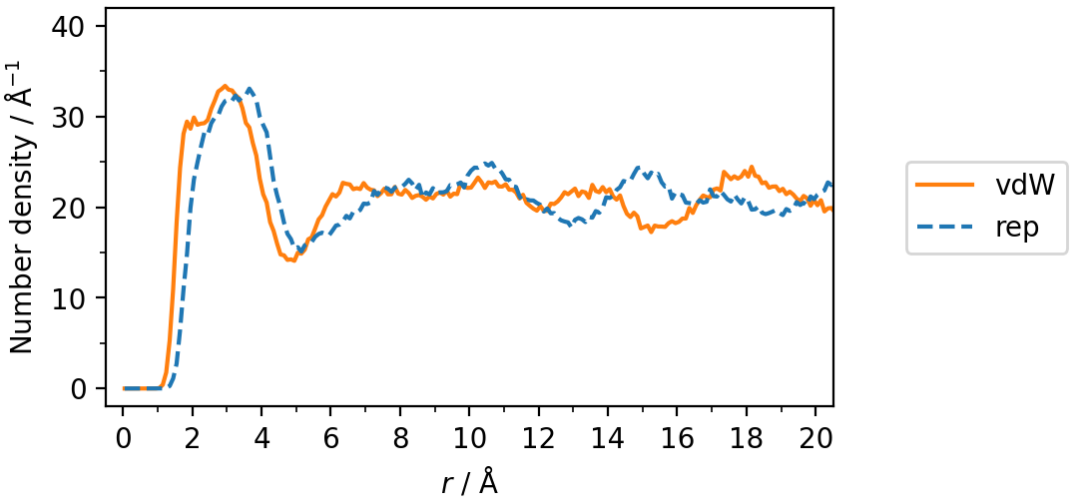

Figure 3.12 Average linear number density in the nanocavity simulations with full van der Waals (vdW) and repulsive-only (rep) barriers. Results were obtained from the last 300 ps (600 frames) and averaging over the three directions orthogonal to the barriers.

3.3.4 Molecular orientation

Molecular orientation is key to the understanding of the behavior of molecules when barriers are introduced. The angle between two randomly oriented vectors in three-dimensional space follows a symmetric distribution between 0° and 180°, with a maximum probability at 90°. Benzoic acid molecules are far from being randomly oriented, due to the strong, directional hydrogen bonds and the anisotropic molecular

shape which favors stacking interactions. For this reason, deviations from random distributions are highlighted to understand the molecular orientation behavior with respect to both surrounding molecules and barriers. For the analysis, vectors $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ (see Figure 3.8) are employed. The results of the last 300 ps (600 frames) of the nanocavity simulation with repulsive-only barriers are presented, since it is the one with the highest degree of confinement. Analogue results are obtained when full van der Waals barriers are employed[38].

Molecule-molecule orientation

The orientation between molecules was analyzed for each of them within a radius of 10 Å, thereby considering only the first coordination shells. Two regions were distinguished according to the distance from the barriers, to highlight the effect of confinement: the "barrier" region and the "bulk" region. The distance threshold was set to 6.4 Å, that corresponds to the distance at which the interactions with barriers are negligible ($0.001\ kJ\ mol^{-1}$).

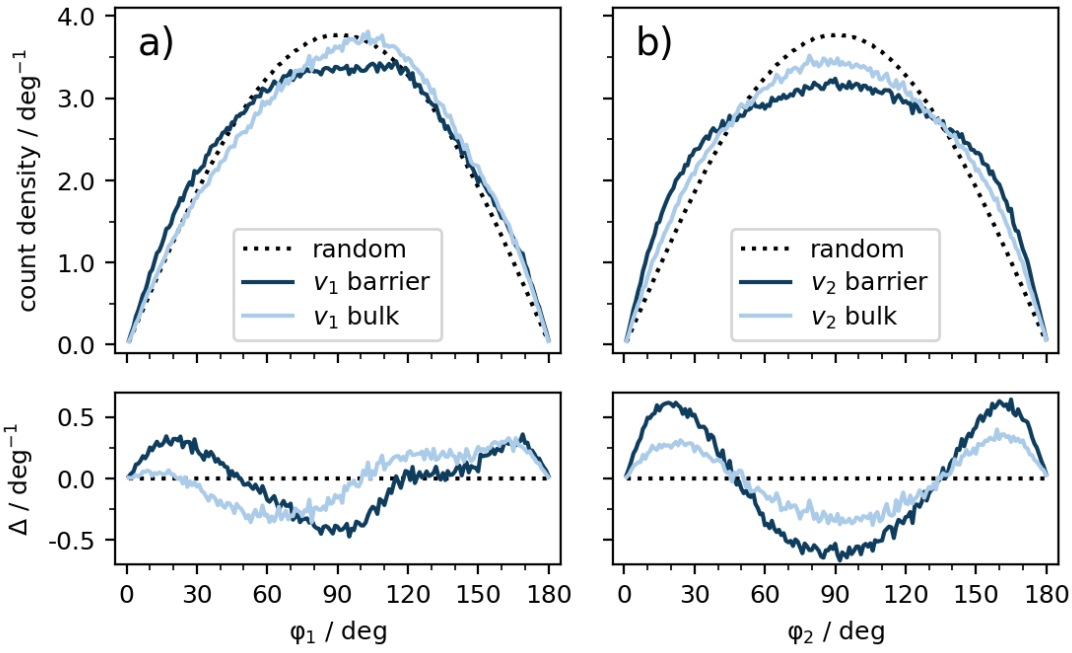

Figure 3.13 Molecular orientation based on the angle between vectors $\boldsymbol{v}_1$ (a) or $\boldsymbol{v}_2$ (b) in the nanocavity simulation with repulsive-only barriers. Results in the bulk and barrier regions are compared. Values are expressed as number of molecules per degree. Random distributions are indicated through dotted lines, and differences between actual and random distributions are reported in the subplots below.

The angle φ_1 between $\boldsymbol{v}_1$ vectors is preferentially around 180°, which corresponds to both cyclic dimers and consecutive molecules in hydrogen-bonded chains, or catemers, 120°, suggesting the presence of cyclic trimers, and 0°, mostly due to non-consecutive molecules in catemers (Figure 3.13a). When the angle φ_2 between $\boldsymbol{v}_2$ vectors is considered, instead, the distribution is symmetric and the most populated orientations are around 0° and 180°, which correspond to parallel and antiparallel stacking interactions (Figure 3.13b). Notably, in the barrier region, the deviation from a random orientation is more pronounced, indicating a partial ordering given by an increased occurrence of stacked molecular arrangements, likely promoted by the presence of barriers.

Molecule-barrier orientation

To investigate the molecular orientation relative to the barriers, the distribution of angles between each molecule's orientation vector and the vector normal to a reference barrier is computed as a function of the distance from the barriers (Figure 3.14). Such distance is calculated between the molecular center of mass and the plane passing through the centers of pseudoatoms forming the barriers. A random distribution is subtracted to focus on the relevant differences and results are averaged over all six barriers of the cubic nanocavity. ϑ_1 and ϑ_2 are defined as the angles between vector $\boldsymbol{v}_1$ or vector $\boldsymbol{v}_2$ respectively, and the normal to the barrier, that points towards the center of the nanocavity. Even though the barrier is modeled by a repulsive-only potential, there is a first layer of molecules that are adsorbed onto the barrier. This is evident from the two peaks at ~4 Å shifted towards 0° and 180° in ϑ_2 that signal the presence of benzoic acid molecules that are parallel or antiparallel to the normal, and thus flat with respect to the surface (Figure 3.14b). The other two peaks at the same ϑ_2 angles and at ~8 Å from the barrier indicate the presence of a second layer of molecules, interacting with the first one via stacking interactions. This is consistent with the greater propensity of confined liquids to form stacking interactions with respect to bulk liquid, observed through the radial distribution function (Figure 3.10), and also the increased local density at short distances (Figure 3.12). These four peaks at 0° and 180° in ϑ_2

correspond to two peaks in the ϑ_1 distribution, both at 90°, one at ~4 Å and the other one at ~8 Å. This behavior naturally arises from the orthogonality of $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ vectors.

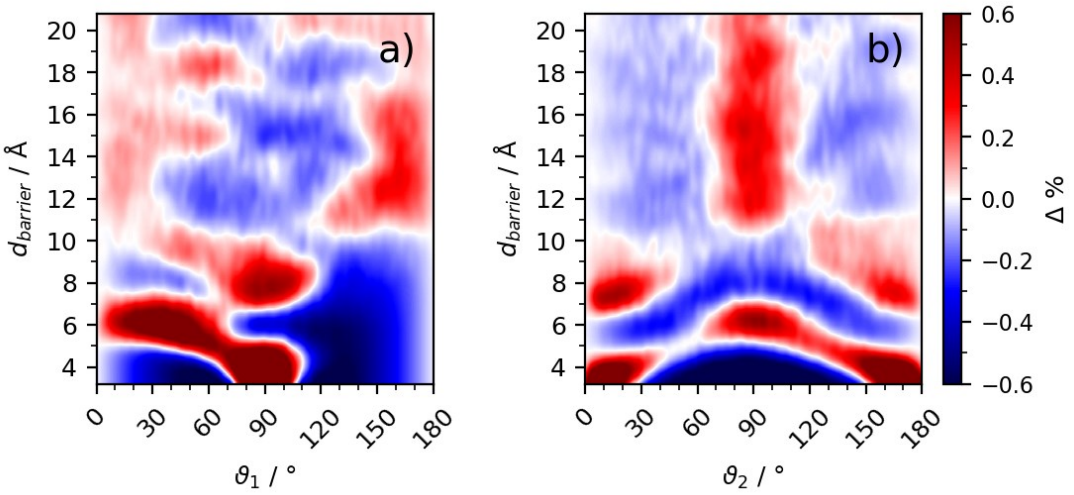

Figure 3.14 Molecular orientation based on the angle between the vectors orthogonal to the barriers and vectors $\boldsymbol{v}_1$ (a) or $\boldsymbol{v}_2$ (b) and the distance from the barriers in the nanocavity simulation with repulsive-only barriers. The intensity is expressed as the difference between the actual distribution and a completely random distribution.

The peak at ~6 Å at 90° in ϑ_2, corresponding to the one shifted towards 0° in ϑ_1, is a result of the geometry of confinement. Indeed, due to the finite size of the nanocavity barriers, a subset of molecules lies near the edges where two barriers intersect; thus, molecules that are oriented parallel to one barrier are therefore orthogonal to the other one, resulting in a 90° angle in ϑ_2. Not surprisingly, these molecules prefer to adopt an angle ϑ_1 closer to 0° than 180°, to point the carboxylic groups away from the hydrophobic barriers, favoring the formation of hydrogen bonds with other benzoic acid molecules. The broad peak at distances greater than 11 Å at 90° in ϑ_2 is another feature due to the confinement geometry, and refers to all the molecules that are flat with respect to adjacent barriers, an thus orthogonal to the reference one. This time, there is not a corresponding peak in the ϑ_1 distribution, since these molecules are too far away from the reference barrier to have a favored orientation of their $\boldsymbol{v}_1$ vectors. Finally, there is a weaker peak at ~13 Å shifted towards 180° in ϑ_1 that is due to benzoic acid molecules that are hydrogen bonded to the ones at ~6 Å, thus adopting an opposite orientation. The distance difference of ~7 Å

corresponds to most of the hydrogen-bonded molecules, as shown in the radial distribution function analysis (Figure 3.10).

3.3.5 Barrier stiffness and liquid-liquid transitions

The effect of the barrier stiffness on the properties of the system is explored by using a scaling factor s with double the magnitude (0.050) of previous simulations (see Section 2.2.1 for more details). The nanocavity geometry is adopted, using both repulsive-only and full van der Waals barriers, since it is the one with the highest degree of confinement. An oscillatory behavior is observed between two liquid states characterized by different densities. To ensure reproducibility, three replicas of each simulation were performed starting from identical initial configurations[38]. Minor variations caused by numerical approximations and multithreading are sufficient to drive the system along slightly different evolutionary paths. Nonetheless, the differences among the replicas are small, with density variations of $\sim 0.001\ g\ cm^{-3}$ and a cohesive energy difference of $0.3\ kJ\ mol^{-1}$ per molecule, both well within the standard deviations of $0.03\ g\ cm^{-3}$ and $2\ kJ\ mol^{-1}$. Importantly, the transition between the two liquid states is consistently observed in all replicas. However, since the microscopic behavior near a transition is governed by nanoscale fluctuations, the timing and direction of the transition can vary: in some cases, a high-to-low density transition occurs first, while in others the reverse takes place. The trajectories with the most evident transitions that last for enough time to extract meaningful information are here commented.

Analysis of liquid-liquid transition

The nanocavity simulation with full van der Waals barriers shows a low-to-high density transition with an onset at 134 ps while the one with repulsive-only barriers shows a high-to-low density transition at 255 ps (Figure 3.15). The density difference between the two states is $\sim 0.04\ g\ cm^{-3}$ and the transition occurs almost instantaneously, in less than 10 ps (Figure 3.15d). When considering the cohesive energy E_{coh}, it is clear that the high density (HD) state is metastable, being $\sim 4\ kJ\ mol^{-1}$ per molecule less stable than the low density (LD) state and showing

larger fluctuations (Figure 3.15a). Cohesive energy can be decomposed into electrostatic (E_{Coul}) and dispersion/repulsion (E_{vdW}) terms, such that $E_{coh} = E_{Coul} + E_{vdW}$ (Figure 3.15b,c).

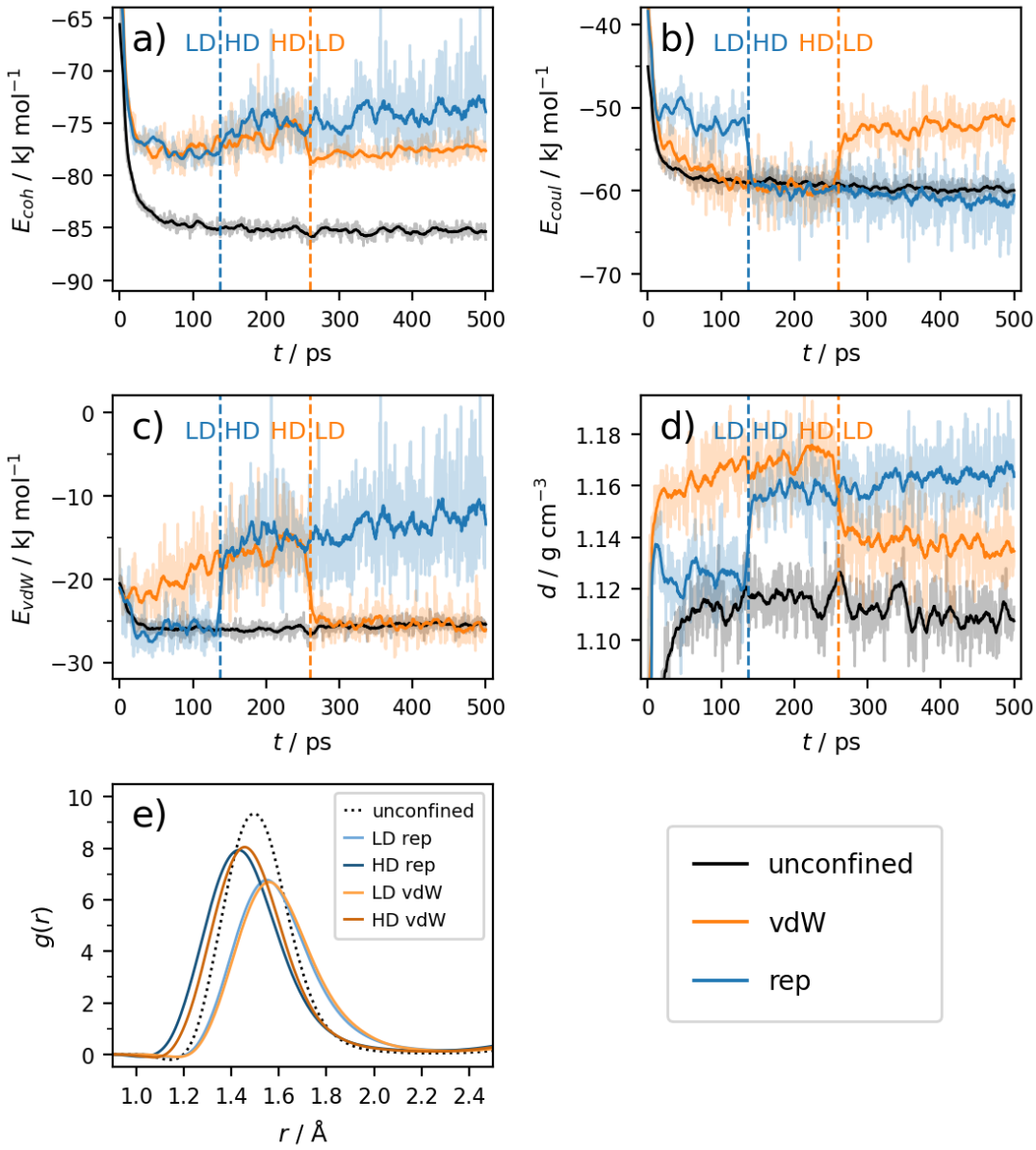

Figure 3.15 Cohesive energy (a), coulombic energy (b), dispersion energy (c), and density (d) of the systems confined into nanocavities with full van der Waals (vdW) and repulsive-only (rep) barriers with stiff force constants (scaling factor s = 0.050), compared to the unconfined bulk liquid, used as a reference. Moving averages are calculated over 10 ps time windows (20 frames). The OH···O RDFs before and after the transitions are also shown, indicating the high and low density states as HD and LD (e).

The lower stability of the HD state results from a more favorable E_{Coul} which is outweighed by a less favorable E_{vdW}. The overall destabilization is therefore dominated by the dispersion/repulsion contribution. The large fluctuations observed in E_{coh} are primarily driven by fluctuations in E_{vdW}. Since the electrostatic interactions of benzoic acid are largely governed by hydrogen bonding, the HD state likely features strong hydrogen bonds which keep the molecules in closer contact. This enhanced proximity increases the density but also the repulsive part of E_{vdW}, leading to a destabilization of E_{vdW} and, consequently, E_{coh}.

Radial distribution function (RDF) $g(r)$, defined in Equation 2.19, is a useful tool for revealing structural differences between the two liquid states. RDFs of OH···O contacts in the nanocavity simulation with repulsive-only barriers were computed over 100 ps (200 frames) for both the LD and HD states, and compared to that of the bulk liquid, used as a reference (Figure 3.15e). Almost identical results are obtained when considering the two liquid states obtained in the nanocavity simulation with full van der Waals barriers. The $g(r)$ of the HD liquid exhibits a peak at approximately 1.5 Å, which is more intense and also shifted by ~0.2 Å towards shorter distances compared to that of the LD liquid. This suggests a greater strength of hydrogen bonds, that is in line with the more negative E_{Coul}.

3.3.6 Extreme confinement

In this Section, confinement is pushed to the limit by reducing the distance between barriers of nanolayers down to a nominal thickness of 3.6 Å. This nanolayer thickness was deliberately selected to modulate the rotational freedom of the aromatic ring. The out-of-plane reorientation is sterically hindered, and the ring can only reorient within its molecular plane. In this configuration, linear aggregates, previously observed mainly as transition states of more stable cyclic aggregates in Section 3.2, become favored. The purpose of these simulations was therefore to explore the potential formation of novel structures or interaction modes of benzoic acid when the space is restricted to one molecular layer, with consequent restriction of degrees of freedom. To contrast the low molecular self-diffusion generated by the extremely confined system

and favor molecular aggregation, lower packing efficiencies than the maximum theoretical limit for random packing (66%) are employed[39]: one that is 50% (216 molecules) and the other one that is 25% (108 molecules). The details regarding these simulations are reported in Section 2.2.1.

Molecular diffusion and rotational correlation

Systems with a packing efficiency of 25% display substantially greater translational diffusion, exhibiting self-diffusion coefficients roughly five times higher than those of the systems with 50% packing efficiency. Higher temperatures further amplify diffusion, yielding almost a twofold increase (Table 3.2). The coefficients are not directly comparable with those of thicker nanolayers since diffusion under extreme confinement is restricted to a single plane and the mean square displacement in the direction orthogonal to the nanolayer gives a negligible contribution.

The rotational correlation function $C(t)$ among $\boldsymbol{v}_1$ vectors has the usual decay, accelerated by the lower packing efficiency of the system (Figure 3.16a). A lower packing efficiency corresponds to a higher loss of rotational correlation, as expected. When $\boldsymbol{v}_2$ is taken as reference vector, $C(t)$ remains between 90 and 94%, indicating a strong rotational correlation due to the hindered reorientation (Figure 3.16b). Temperature, in this case, has not a strong influence and its effect is overcome by the different packing efficiency.

Table 3.2 Number of molecules (N_{mol}), packing coefficient (C_{pack}), temperature (T), and self-diffusion coefficients (D) calculated over the last 4 ns of each simulation under extreme confinement.

N_{mol}	C_{pack}	T / K	$D \times 10^{-10}$ / m^2 s^{-1}
108	25%	300	0.0693(6)
		350	0.1214(7)
216	50%	300	0.01365(7)
		350	0.02575(6)

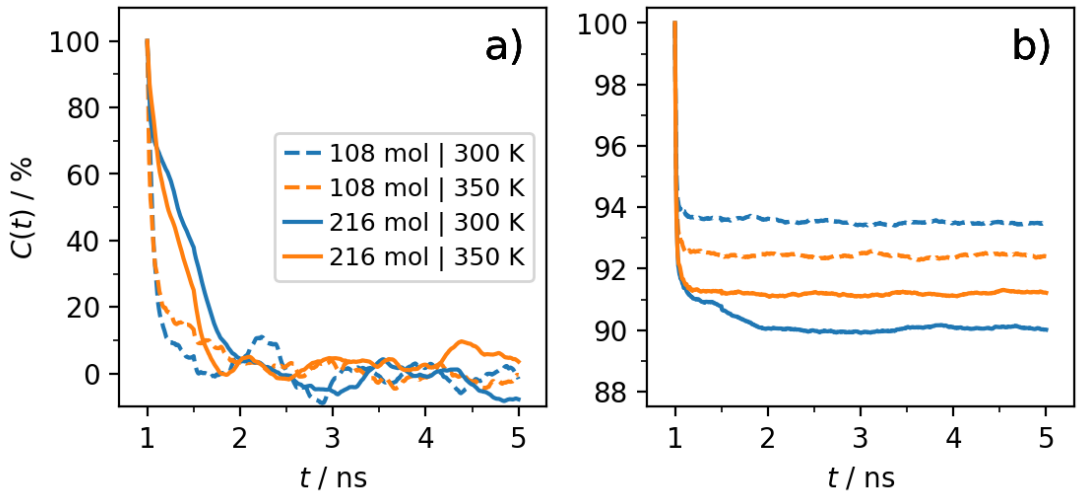

Figure 3.16 Rotational correlation functions computed for vectors $\boldsymbol{\nu}_1$ (a) and $\boldsymbol{\nu}_2$ (b). Moving averages are calculated in the last 4 ns of the simulations, using time windows of 500 ps (100 frames). Comparisons among simulations with lower (25%, 108 mol) and higher (50%, 216 mol) packing efficiency at 300 and 350 K are shown.

Aggregates

Benzoic acid molecules, initially arranged almost randomly and with void spaces evenly distributed, form hydrogen-bonded aggregates that tend to interact with one another to maximize dispersion interactions, leading to high concentrated regions. The result of spontaneous aggregation is particularly evident when comparing the initial and final frame of simulations with lower packing efficiency (Figure 3.17a,b).

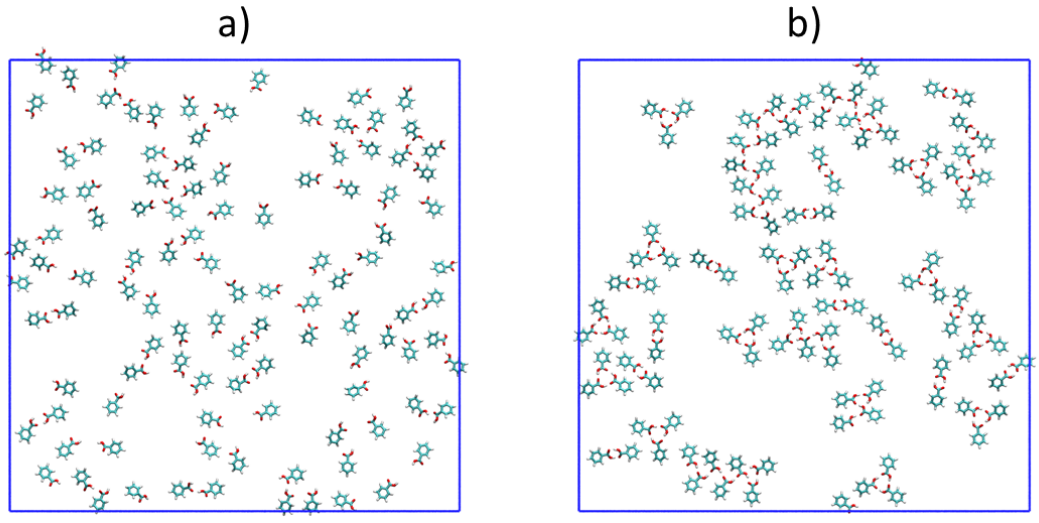

Figure 3.17 Initial (a) and final (b) frame of the simulation with lower packing efficiency at 300 K.

Extreme confinement induces the formation of flat aggregates. For this reason, the large cyclic aggregates having 3D structures studied in Section 3.2, characterized by aromatic stacking interactions, must adopt a different structure or do not form anymore. In the simulation with higher packing efficiency at 350 K, the largest cyclic aggregate is found, consisting of five molecules (Figure 3.18a). In the same simulation, the largest aggregate is also found, which is composed of 26 molecules and is prevalently catemeric (Figure 3.18b). At one end, it has an oxygen atom that is not involved in any hydrogen bond while on the other end it is constituted by a cyclic trimer that allows to saturate hydrogen bond donors and acceptors. Such linear catemers, although containing some defects that make the chain bend, resemble the structural motif identified in a recent crystal structure prediction study on benzoic acid[13] (Figure 3.18c). This suggests that extreme confinement might induce the formation of metastable polymorphs of compounds like benzoic acid, of which no other polymorphs are known to date.

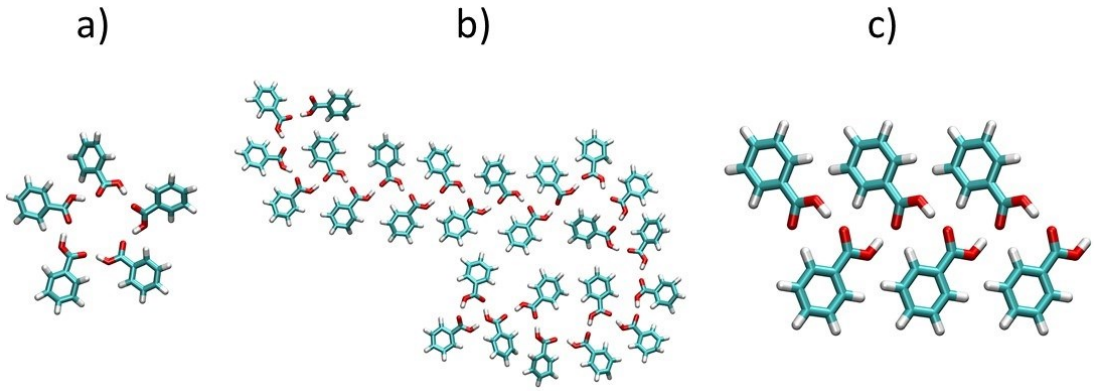

Figure 3.18 Largest cyclic aggregate (a) and largest aggregate (b) both detected in the simulation with higher packing efficiency at 350 K. The catemeric motif (c) found in a crystal structure prediction study[13] is also shown for comparison.

4 Subcritical clusters

In Section 3, confinement was exploited to favor aggregation, and the formation of different types of aggregates was detected. Now, the behavior of bulk liquids is investigated through molecular dynamics simulations, focusing on the formation of stable molecular aggregates prior to the nucleation event. These aggregates, having a smaller size than the critical one, are commonly named subcritical clusters or pre-nucleation clusters[83].

According to classical nucleation theory (CNT), a crystal nucleus forms from a homogeneous liquid through the sequential attachment of monomers. However, real liquids and solutions are not perfectly homogeneous; they exhibit local inhomogeneities in density or concentration that can be attributed to the formation of supramolecular aggregates[84]. Numerous systems provide evidence of deviations from CNT, following non-classical nucleation pathways[85,86], including proteins like lysozyme[16], inorganic compounds like calcium carbonate and calcium phosphate[87,88], zeolites[89], and even simple amino acids[90]. Unlike the mechanism proposed by CNT, disordered, metastable aggregates may form and coalesce, giving rise to denser or more concentrated amorphous regions, from which an ordered nucleus can eventually emerge after structural rearrangement.

There are also many examples of systems where subcritical clusters were found through molecular dynamics simulations like organic molecules[91], methane hydrate[92], and calcium carbonate[93].

In this Section, subcritical clusters are studied in liquid benzoic acid. The newly developed *Cluster* program described in detail in the *Methods* (Section 2.1.3) is briefly summarized (Section 4.1). Then, the liquid structure and dynamics of benzoic acid is analyzed with particular attention to nanoscale inhomogeneities (Section 4.2); finally, a tentative definition of subcritical clusters is proposed, and a detailed analysis is provided (Section 4.3).

4.1 Cluster program in short

The *Cluster* program developed and implemented in MiCMoS is described in detail in Section 2.1.3. Here is reported the essential information to understand the underlying logic. This program is employed to detect molecular aggregates and determine their properties.

First of all, to define a molecular aggregate, it is important to establish when two molecules are bonded. There are two options: the geometric criterion, which is based on hydrogen bonds, and the energetic criterion, which is based on interaction energy. In this work, the energy criterion is employed, thus, two molecules having an interaction energy more negative than a user defined threshold, are considered bonded.

For each frame of a molecular dynamics trajectory, the program prints the list of aggregates with their properties, such as cohesive energy of the aggregate and the interaction energy with the surroundings. In addition, the distribution of interaction energies and lifetimes of intermolecular interactions are also given as global descriptors of the system.

4.2 Properties of the liquid

The crystal structure of benzoic acid is characterized by the presence of a single type of hydrogen-bonded aggregate, consisting of cyclic dimers[43]. The LJC force field estimates the strength of this specific interaction roughly $-90\ kJ\ mol^{-1}$ per molecule, that is much higher than any other type of interaction. Hydrogen bonds are expected to have an important role in the liquid as well, however, several interaction modes could coexist.

4.2.1 Average properties

The average properties of the system were extracted from the last 0.5 ns of each simulation. Comparisons with experimental data, when available, were also carried out.

Viscosity and vaporization enthalpy

To assess the quality of the force field, the dynamic viscosity of the system, calculated through Equation 2.22, 2.23, and the vaporization enthalpy were compared to experimental data[75,76,94] and Gromacs simulations (Figure 4.1a,b). MiCMoS is close to both Gromacs results and experimental data.

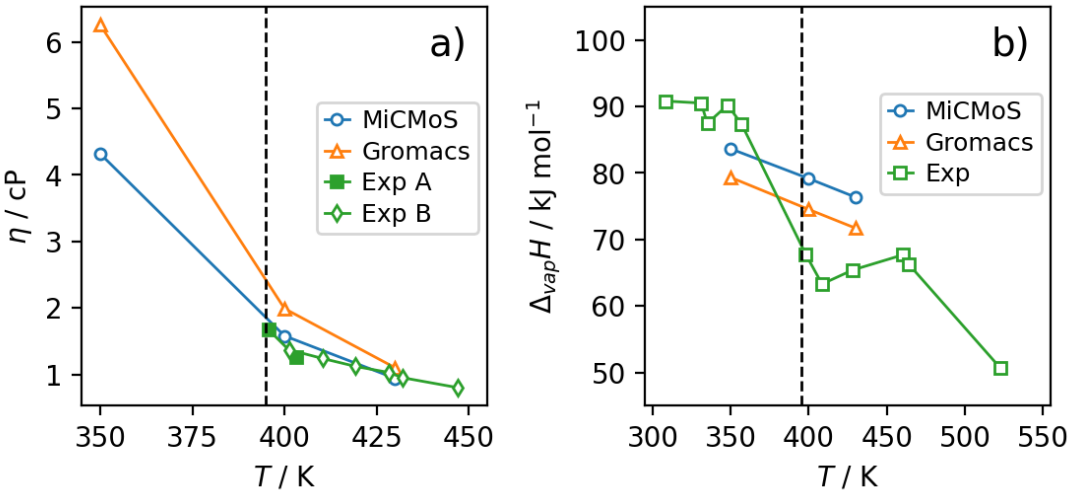

Figure 4.1 Comparison of MiCMoS, Gromacs, and experimental data (Exp A[94], Exp B[76], Exp[75]) for dynamics viscosity (a) and vaporization enthalpy (b). Solid lines are shown only as a guide to the eye.

Molecular orientation

The molecular orientation of benzoic acid molecules in the liquid phase can give much information on the average structural properties of the liquid. The angle distribution between vectors $\boldsymbol{v}_1$ (see Figure 3.8), is calculated as a function of the molecular center-of-mass distance d_{com} (Figure 4.2a-c). The intensity is a difference with respect to a random distribution (see Section 3.3.4). Red regions correspond to the angles characterizing the most populated interaction modes. Regions A_1 ($d_{com} \leq 6$ Å; $0° \leq \vartheta \leq 50°$) and A_2 ($d_{com} \leq 6$ Å; $140° \leq \vartheta \leq 180°$) refer to stacking interactions; regions B (6 Å $\leq d_{com} \leq 8$ Å; $90° \leq \vartheta \leq 150°$) and C ($d_{com} \cong 7$ Å; $160° \leq \vartheta \leq 180°$) correspond to cyclic trimers and dimers respectively. Similar results are obtained at different temperatures. This geometric classification does not allow a clear distinction between the undercooled liquid and the liquid above the

melting point. Further analysis is required to uncover the nanoscale inhomogeneities of the liquid (Section 4.2.2).

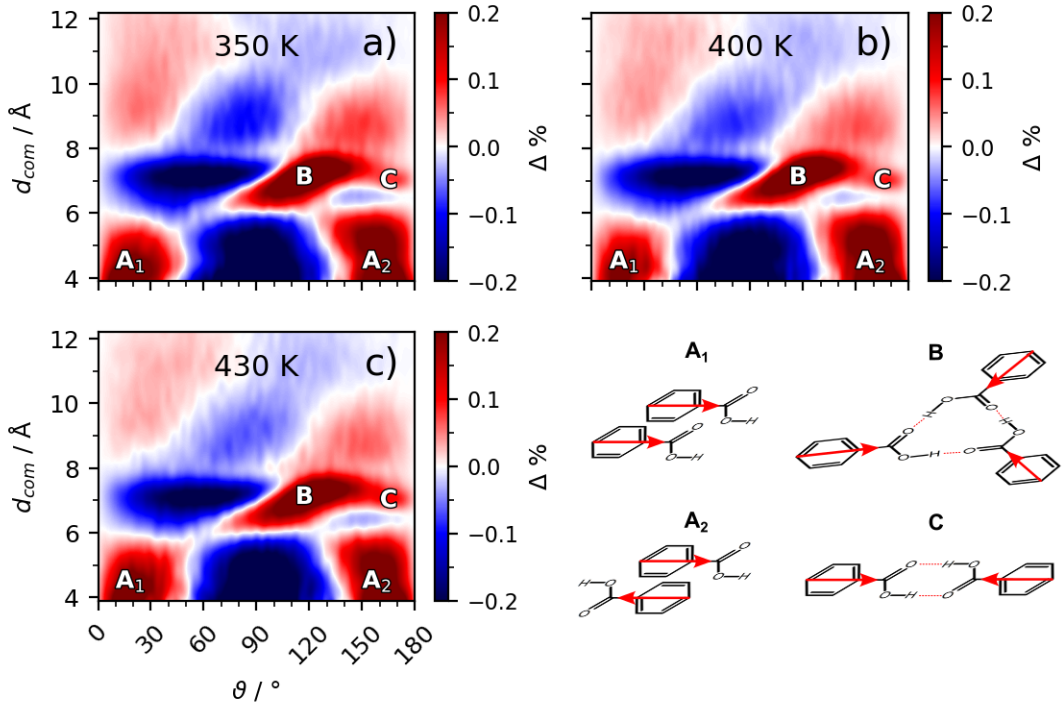

Figure 4.2 Molecular orientation expressed as angle between vectors $\boldsymbol{v}_1$ (ϑ) and center-of-mass distance (d_{com}) for the simulations at 350 (a), 400 (b), and 430 K (c). The intensity is expressed as the difference between the actual distribution and a completely random distribution. The main interaction modes are also shown schematically.

Strength and persistence of intermolecular interactions

Interaction strength is evaluated by examining the distribution of interaction energies E_{ij} over the final 0.5 ns of each trajectory (Figure 4.3). Beyond the broad contribution of weak, long-range interactions ($E_{ij} > -20\ kJ\ mol^{-1}$), the distribution reveals two distinct peaks. The lower-intensity peak at more negative energies, at $-90\ kJ\ mol^{-1}$, is unequivocally attributed to double hydrogen bonds in cyclic dimers. The other one, at $-50\ kJ\ mol^{-1}$, is mainly associated with interaction modes that involve a single hydrogen bond. At higher temperatures, the average kinetic energy increases, causing both peaks to shift toward less negative interaction energies. This effect is more pronounced for the peak at $-50\ kJ\ mol^{-1}$, which also decreases in intensity. An interaction energy of $-25\ kJ\ mol^{-1}$, located near the minimum between

the single hydrogen bond peak and the weak interaction region, is used as a threshold to identify tightly bound molecular pairs.

The energy threshold of $E_{ij} = -25\ kJ\ mol^{-1}$ can be used to get an estimate of the timescale Δt of thermal fluctuations, which can be defined as the average time required for a molecule to climb a potential well[95]. Most hydrogen bonds persist for only a few picoseconds, and the distribution of their lifetimes t_{live} exhibits an approximately exponential decay with time. On average, Δt is 42 ps, 30 ps, 22 ps at 350 K, 400 K and 430 K respectively, showing a clear reduction at higher temperatures. This trend reflects the enhanced thermal fluctuations at elevated temperatures, which lead to a higher frequency of strong interaction breakage.

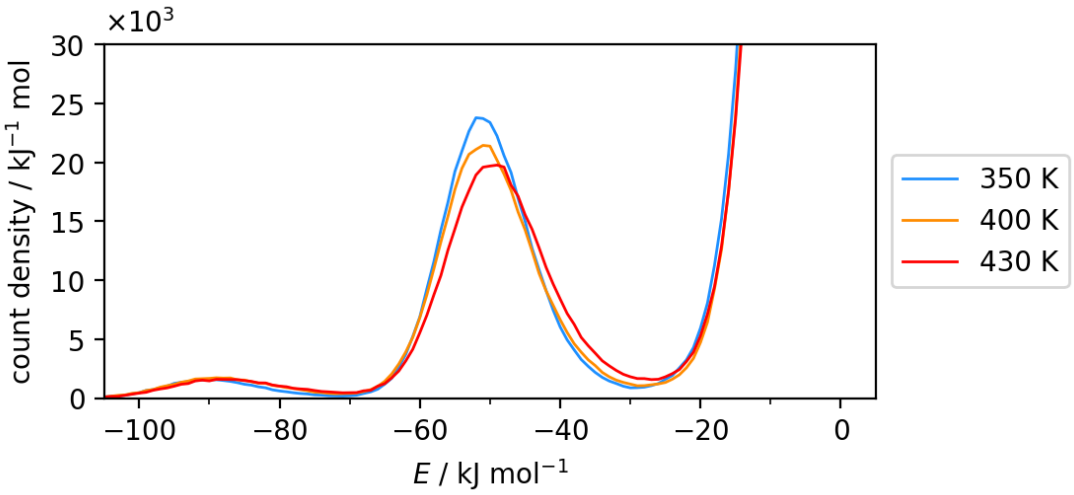

Figure 4.3 Density distribution of interaction energies in the simulations at 350, 400, and 430 K.

4.2.2 Nanoscale inhomogeneities

Liquids are not perfectly homogeneous at the nanoscale; instead, they exhibit some inhomogeneities. The emergence of (meta)stable, persistent molecular aggregates produces locally denser regions, accompanied by areas of reduced density. A straightforward approach to identify nanoscale inhomogeneities in the liquid consists in examining the spatial distribution of density throughout the simulation box.

Density inhomogeneities

The simulation box was divided into 10^3 volume elements dV_i, and in each element, the local space-time density was computed. This was done by identifying all molecules whose center of mass lies within each volume element and making an average over the last 0.5 ns of the simulation. The histograms of time-averaged deviations between local density ρ_i and total density of the simulation box ρ_{box} were fitted by Gaussian functions (Figure 4.4).

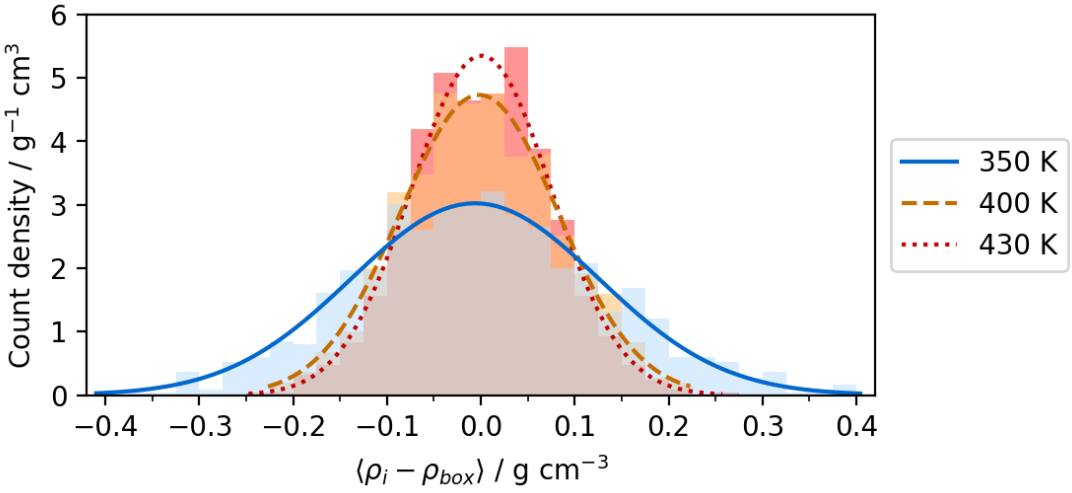

Figure 4.4 Density distribution of local density deviations from average density in the simulations at 350, 400, and 430 K. Histograms are fitted with Gaussian functions.

At 400 K and 430 K, above the melting point, the density fluctuations are similar, with maximum deviations of roughly $\pm 0.2\ g\ cm^{-3}$ and a standard deviation σ of 0.08 and 0.07 $g\ cm^{-3}$ respectively. In contrast, the undercooled liquid at 350 K exhibits maximum variations of $\pm 0.4\ g\ cm^{-3}$, double with respect to higher temperature liquids, and a σ of $0.13\ g\ cm^{-3}$. Density fluctuations can also be quantified by the ratio between the mean squared density deviations over all the volume elements, $\langle \Delta^2 \rho_i \rangle$, and the square of the density of the system, ρ_{box}^2 as defined in Equation 2.24. The quantity $\langle \Delta^2 \rho \rangle / \rho_{box}^2$ could potentially be determined experimentally, being related to diffuse light scattering[96], and its estimates are 1.4230(9), 1.4788(8), and 1.4815(7) at 350, 400, and 430 K respectively.

A possible explanation for the presence of high-density regions within the simulation box is the formation of relatively large, persistent molecular aggregates, facilitated by the higher viscosity and reduced kinetic energy characteristic of the undercooled liquid.

4.3 Definition and analysis of subcritical clusters

Nanoscale inhomogeneities are naturally present in liquids well above their melting point, manifesting as regions where the local density deviates from the thermodynamic average. The interest here is devoted to molecular aggregates of benzoic acid molecules that affect the overall properties of the liquid, the so-called subcritical clusters. In this context, a subcritical cluster is a molecular aggregate smaller than the critical size required to overcome the energy barrier and transition into a stable crystalline phase. These clusters are metastable and tend to dissolve rather than grow. However, a precise and universally accepted definition of a subcritical cluster is still lacking, particularly in liquids consisting of a single molecular species, where purely geometric, proximity-based criteria are insufficient. In Section 4.3.1, a tentative definition is proposed based on three criteria. In Section 4.3.2, molecular aggregates are analyzed to find subcritical clusters.

4.3.1 Three criteria

The new definition of subcritical clusters is based on three criteria that account for molecule-molecule interaction energy, time persistence of the molecular aggregate, and excess energy of the aggregate, defined below.

First criterion: interaction energy

First, subcritical clusters must be molecular aggregates, that is, groups of molecules held together by strong interactions. In this context, two molecules are considered bonded if their interaction energy E_{ij} is more negative than a defined energy cutoff E_{cut}. This straightforward approach was employed about three decades ago to study liquid argon, modeled as a system of van der Waals spheres[97]. Based on the

rationale discussed in Section 4.2.1, a threshold $E_{cut} = -25\ kJ\ mol^{-1}$ is adopted to identify tightly bound molecular pairs. Thus, an aggregate A can be defined as any group of N_A molecules in which each molecule i has at least one interaction energy $E_{ij} \leq E_{cut}$ with another molecule j belonging to the same aggregate. The existence of an interaction energy $E_{ij} \leq E_{cut}$ implies that molecules i and j belong aggregate A as summarized in Equation 4.1.

$$\exists E_{ij} \leq E_{cut} \Rightarrow i, j \in A \tag{4.1}$$

Second criterion: time persistence

Every time the interaction energy E_{ij} becomes less negative than E_{cut}, the corresponding interaction is considered broken. If a molecule no longer has any interaction below E_{cut} with other molecules of the aggregate, it is no longer considered part of it. This condition implies that the aggregate is disrupted. The lifetime of an aggregate A can be determined by counting the number of consecutive frames in which all the molecules belonging to A remain part of it. The lifetime t_{live} is then compared with the lifetime of thermal fluctuations Δt defined in Section 4.2.1. Subcritical clusters must be aggregates having a lifetime t_{live} that is longer than that of thermal fluctuations Δt, as stated in Equation 4.2.

$$t_{live} > \Delta t \tag{4.2}$$

Third criterion: excess energy

The total number of molecules of the system is denoted as N. For a given aggregate A, the number of molecules it contains is defined as N_A, while the number of molecules not belonging to A is N_O, where the subscript O refers to the "outer" region relative to the aggregate. This allows the system to be partitioned into two disjoint subgroups, A and O, such that $N = N_A + N_O$, with $N_O \gg N_A$. The cohesive energy of the aggregate, E_A, is defined as the sum of all the pairwise interaction energies E_{ij} between molecules within A. Similarly, the cohesive energy of the outer region, E_O, is the sum of all E_{ij} between molecules within O. The total cohesive energy of the system, E_{coh}, is then the sum of E_A, E_O, and an additional cross-term E_{AO}, which

accounts for interactions between molecules in A and those in O. Typically, all these cohesive energy terms are negative, and $E_{coh} \approx E_O$, especially when N_A is small. This implies that E_A is less negative than E_O. When normalized by the number of molecules in each region, the molecular cohesive energies denoted as $E_{O,m} = E_O/N_O$ and $E_{A,m} = E_A/N_A$, also reveal that $E_{A,m}$ is still less negative than $E_{O,m}$. This is because the outer region O closely resembles the bulk liquid, while the aggregate A contains many undercoordinated molecules. The cross-term $E_{AO,m} = E_{AO}/N_A$ is normalized by the number of molecules of the aggregate and can thus be compared with $E_{A,m}$. To this end, a new quantity is here introduced: the excess energy ΔE_{exc}, defined in Equation 4.3 as the difference between the cohesive energy of the aggregate, $E_{A,m}$, and the interaction energy between the aggregate and its surroundings, $E_{AO,m}$, normalized per molecule within the aggregate. A negative ΔE_{exc} indicates that on average, molecules belonging to aggregate A experience a greater cohesion with other molecules within A than with those in the outer region O. The internal structure of molecular aggregates continuously changes due to thermal fluctuations and so does ΔE_{exc}. To consider these dynamical effects, the final criterion to define a subcritical cluster considers a time average over the lifetime t_{live} of the aggregate. A subcritical cluster is then characterized by a negative, time-averaged ΔE_{exc}, as described in Equation 4.4.

$$\Delta E_{exc} = E_{A,m} - E_{AO,m} \tag{4.3}$$

$$\langle \Delta E_{exc} \rangle < 0 \tag{4.4}$$

4.3.2 Analysis of subcritical clusters

Size distribution

The size distribution of molecular aggregates at each temperature shows a high occurrence of dimers, which are the basic building units found in the crystal structure of benzoic acid[43] (Figure 4.5a). Interestingly, an even greater number of trimers is observed, while for larger sizes, the frequency decreases almost exponentially. Similar trends are observed when examining the distribution of molecules across aggregates (Figure 4.5b). The predominance of molecules in trimers over dimers becomes even

more pronounced, and the presence of molecules belonging to larger aggregates, such as those of size 14 or 17 observed at 350 K, is more evident.

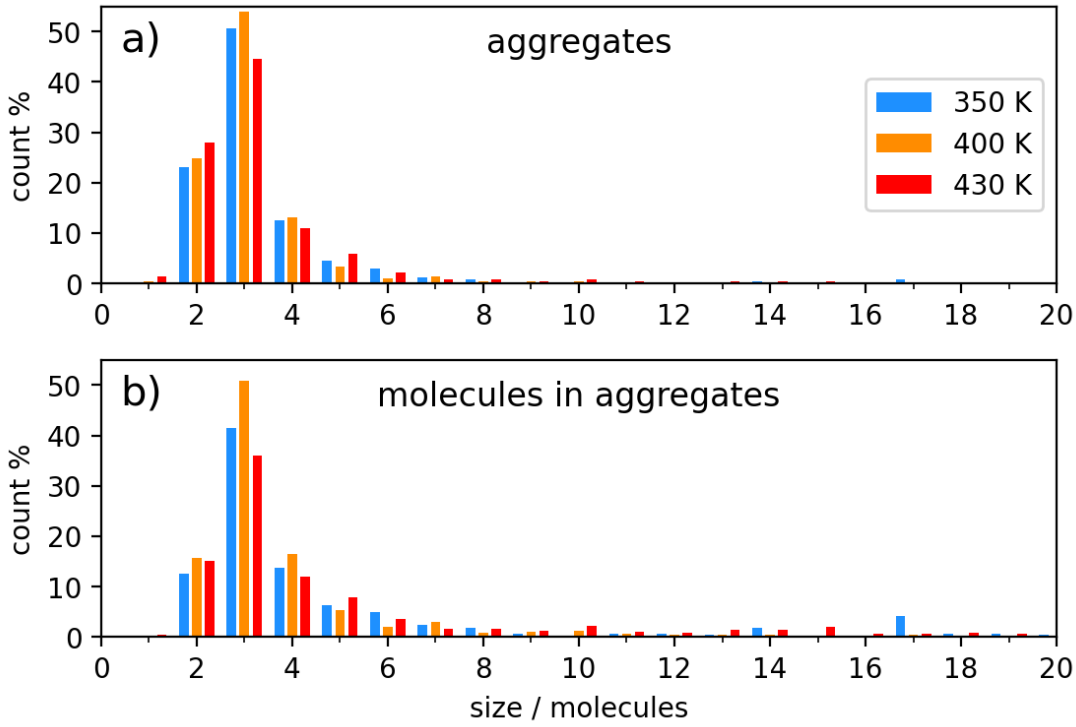

Figure 4.5 Distributions of aggregates (a) and molecules in aggregates (b) as a function of their size, in the simulations at 350, 400, and 430 K. Counts are expressed as percentages and aggregates spanning multiple frames are counted multiple times. Only aggregates with size up to 20 molecules are shown.

Time persistence

Molecular aggregates are dynamic entities that can form, dissociate, and reform over time. To analyze their time persistence, it is convenient to consider only unique, independent aggregates, ensuring each is counted a single time regardless of its dynamic behavior. The presence of large aggregates does not necessarily imply the formation of subcritical clusters. Indeed, most of them have lifetimes t_{live} of a few picoseconds, well below the lifetime of thermal fluctuations Δt (Figure 4.6), not fulfilling the criterion imposed in Equation 4.2. A few exceptions are particularly evident in the undercooled liquid, where aggregates of 14 and 17 molecules have t_{live} greater than 100 ps. Smaller aggregates like dimers and trimers are characterized by longer lifetimes. At higher temperatures, molecular aggregates generally have shorter

t_{live} due to an increased thermal motion. Therefore, large and persistent aggregates are unlikely.

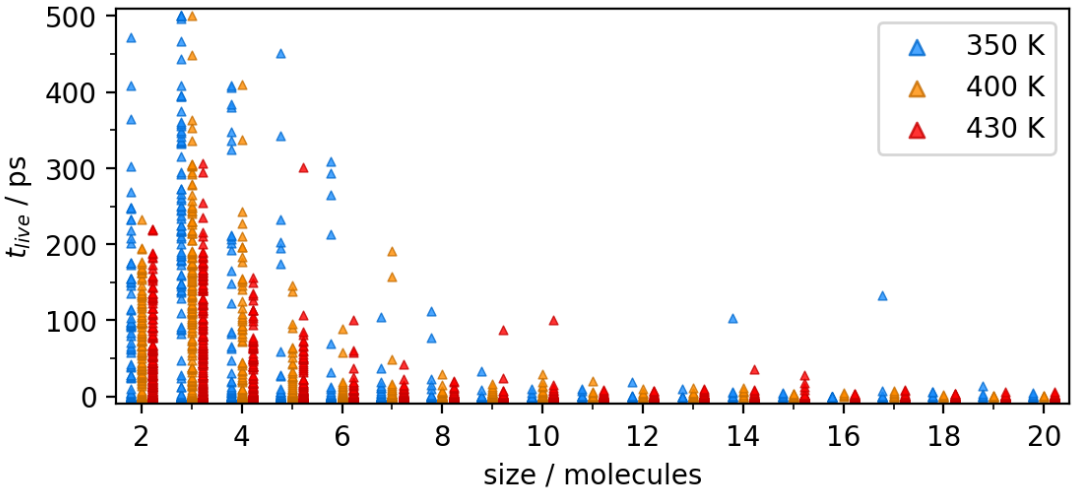

Figure 4.6 Maximum lifetimes of unique molecular aggregates as a function of their size, in the simulations at 350, 400, and 430 K. Only aggregates with size up to 20 molecules are shown.

Excess energy

To identify subcritical clusters, the last criterion based on the time-averaged excess energy $\langle \Delta E_{exc} \rangle$, described in Equation 4.3, 4.4, must be also taken into account. The distribution of t_{live} with respect to $\langle \Delta E_{exc} \rangle$ was analyzed (Figure 4.7a-c). All markers in the plots are unique molecular aggregates, satisfying the first criterion (Equation 4.1). Persistent aggregates that satisfy the second criterion (Equation 4.2) lie on the right side while aggregates having a negative averaged excess energy, satisfying the third criterion (Equation 4.4), lie on the bottom part of the plots. Subcritical clusters satisfy the three criteria and correspond to markers in the bottom right quadrants.

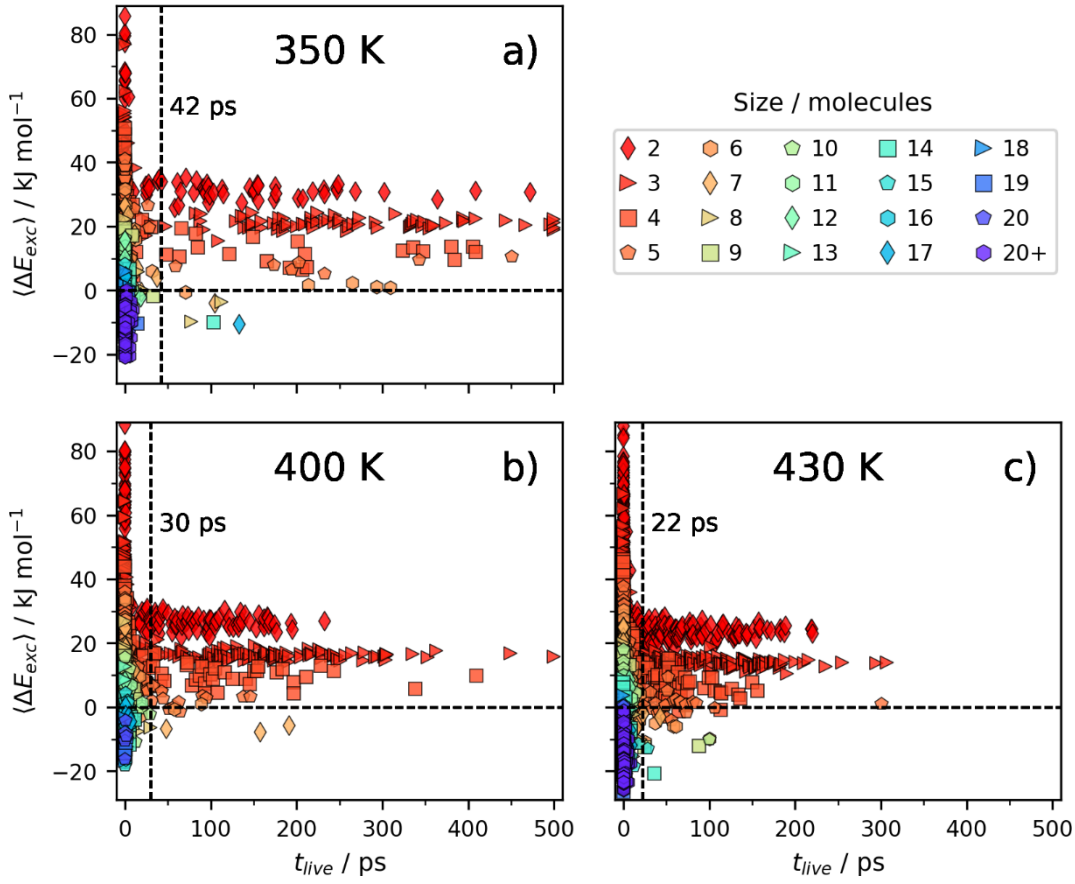

Figure 4.7 Average excess energy normalized per molecule as a function of lifetime and size of unique molecular aggregates at 350 (a), 400 (b), and 430 K (c). Vertical dashed lines indicate the lifetime of thermal fluctuations.

Critical size

The critical size of a molecular aggregate $N_{A,c}$ defining a subcritical cluster is the smallest N_A that satisfies the three established criteria. Interestingly, $N_{A,c}$ falls within the range of 4 to 6, regardless of temperature (Figure 4.8a-c). Most of the aggregates having a critical size are destroyed very quickly and only a few of them have significant internal cohesion greater than the surrounding liquid, and long time persistence. More likely, these metastable and persistent aggregates have a size that is much larger than $N_{A,c}$. These transient aggregates are subcritical clusters and likely play an important role in determining the properties of the liquid.

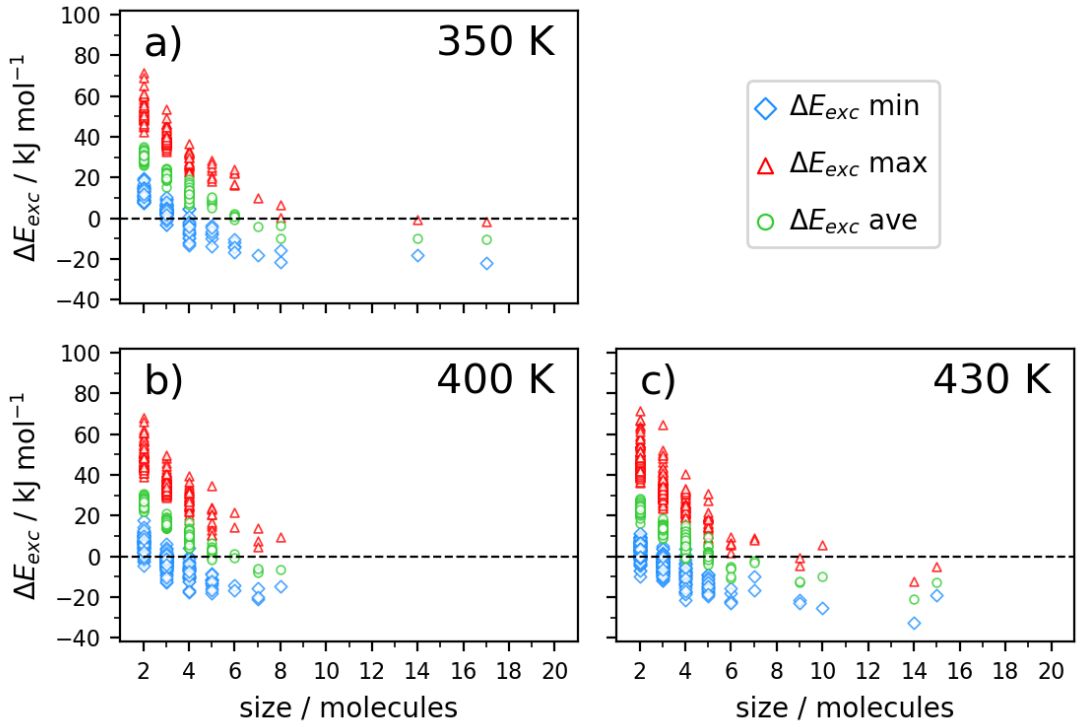

Figure 4.8 Minimum, maximum, and average excess energy of persistent aggregates as a function of their size, in the simulations at 350 (a), 400 (b), and 430 K (c). Energy values are normalized per molecule. The critical size is the smallest size for which an aggregate has a negative average excess energy (ΔE_{exc} ave).

Structural properties

The inner structure of the largest ($N_A = 17$) and most persistent ($t_{live} > 130\ ps$) subcritical cluster in the undercooled liquid was analyzed in terms of radial distribution function (RDF) (Figure 4.9a,b), described in Equation 2.19. The RDFs $g(r)$ of the subcritical cluster were rescaled by a multiplicative factor to allow a comparison with the liquid and thermalized crystal. The OH···O RDF is almost identical to the one of the bulk liquid, with a main peak at ~1.6 Å, indicating the presence of hydrogen bonds (Figure 4.9a). This confirms the ability of the first criterion (Equation 4.1) to capture hydrogen bonds, despite relying solely on an energetic threshold without any geometric descriptors. The center-of-mass radial distribution function (RDF) of the subcritical cluster shows a broad shoulder in the 4-6 Å range (Figure 4.9b). This feature is indicative of stacking interactions between non-consecutive molecules along the hydrogen-bonded chains. A prominent peak appears

at ~7.1 Å, which is also observed in the RDFs of the bulk liquid and the crystalline phase. This peak is associated with two structural motifs: cyclic hydrogen-bonded dimers and single hydrogen bond interactions, where the two molecules are oriented with vectors $\boldsymbol{v}_1$ at ~180°. In addition, a distinct peak at ~6.8 Å is observed in the RDF of the subcritical cluster, which is not visible in the RDF of the bulk crystal and is largely masked in the bulk liquid due to structural averaging. This newly resolved peak arises from a specific spatial arrangement in which consecutive hydrogen-bonded molecules adopt a twisted configuration that brings their centers of mass slightly closer. This arrangement resembles a helical structure, reminiscent of a protein-like motif, which helps the system maximize cohesive energy within the cluster. Although this motif exists in the bulk liquid as well, it becomes more pronounced when isolating the subcritical cluster, where such conformations are statistically enriched.

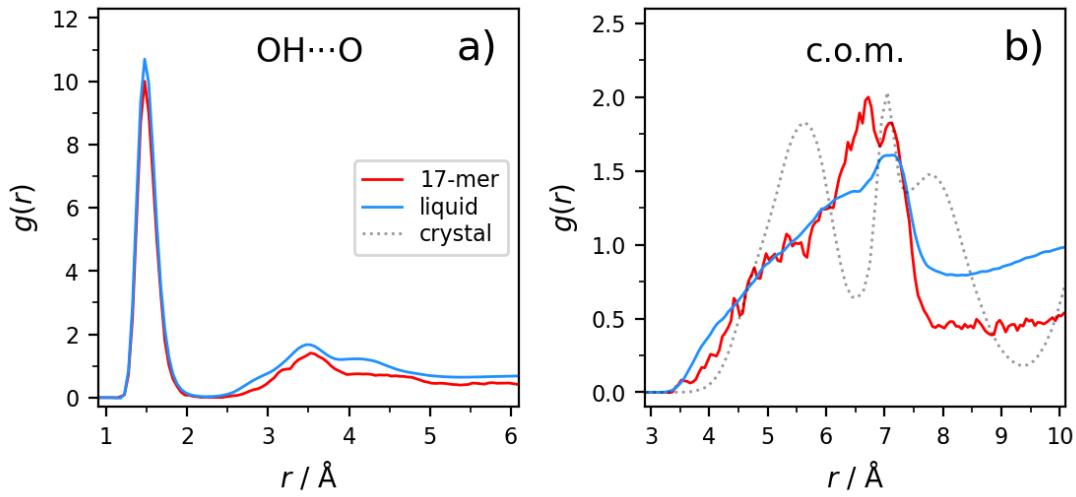

Figure 4.9 OH···O RDF (a) and center-of-mass RDF (b) of the largest subcritical cluster composed of 17 molecules (17-mer) compared to the ones of bulk liquid and crystal.

Dynamical properties

While hydrogen bonding represents the main structural motif responsible for the cohesion of the subcritical cluster, several additional, highly dynamic interactions are established among benzoic acid molecules. Parallel aromatic stacking interactions within the largest subcritical cluster of 17 molecules were therefore analyzed more in detail exploiting the *Stacking* program described in Section 2.1.4. Over 268 frames, a

total of 352 aromatic stacking interactions were found, corresponding to an average of 1.3 interactions per frame. Only 25 unique interactions are observed and are rather weak, short-lived, and intermittent. Indeed, the average interaction energy is $-15\ kJ\ mol^{-1}$, and the maximum time persistence does not exceed 3.5 ps. Moreover, individual interactions break and reform multiple times, with an average of 7.2 reforming events and a maximum of 28 over the 133.5 ps lifetime of the subcritical cluster. In contrast, hydrogen-bonded molecular pairs are significantly stronger ($-48\ kJ\ mol^{-1}$) and persist throughout the entire lifetime of the subcritical cluster.

Overall, the tightly bound hydrogen-bonded chain provides structural stability to the subcritical cluster, while the weaker aromatic stacking interactions enable limited structural rearrangements, allowing a restrained exploration of the conformational space. This is confirmed by the ratios between the eigenvalues of the moment of inertia tensor, whose small standard deviations indicate that the overall shape of the cluster is preserved: $I_3/I_1 = 1.80(14)$ and $I_2/I_1 = 1.32(16)$. The limited structural rearrangements are reflected in small changes in cohesive energy, which is on average $-61(2)\ kJ\ mol^{-1}$.

4.3.3 Size effect

To assess the dependence of the results on system size, a new simulation of the undercooled liquid was carried out at 350 K, increasing the number of molecules from 432 to 1024 and expanding the box dimensions from 48×42×40 Å to 54×53×64 Å[98]. Despite this significant increase in size, key physical properties such as density and viscosity ($1.120(5)\ g\ cm^{-3}$ and $4.30\ cP$) remain statistically unchanged compared to the smaller system ($1.112(8)\ g\ cm^{-3}$ and $4.32\ cP$). Similarly, the molecular orientation as a function of the center-of-mass distance retains the same qualitative features, and both the distribution of pairwise interaction energies and the lifetimes of thermal fluctuations exhibit consistent trends.

Local density inhomogeneities also follow the same Gaussian distribution in both systems, with a standard deviation of $0.12\ g\ cm^{-3}$ in the larger simulation, closely matching the previous value of $0.13\ g\ cm^{-3}$. The size distribution of molecular

aggregates is likewise preserved, with only minor differences: in the larger system, a slightly higher number of trimers is observed, offset by a corresponding decrease in dimers.

The only meaningful differences emerge when focusing on individual aggregates. Naturally, specific aggregates with long lifetimes and large sizes are not reproducible between simulations due to the stochastic nature of molecular dynamics. In addition, the larger system, due to the higher number of molecules, samples a greater number of unique aggregates within the same simulation time. This improved statistical coverage enhances the observation of local inhomogeneities. For instance, in the large simulation, the largest aggregate has 127 molecules ($t_{live} = 0.5\ ps$), while the largest subcritical clusters are constituted by 45, 18, and 11 molecules. In the smaller system, the highest number of molecules in an aggregate is 53 ($t_{live} < 0.5\ ps$) and the largest subcritical clusters have 17, 14, and 8 molecules.

Nonetheless, the general conclusions about subcritical clusters remain robust. The characteristics and behavior of these transient molecular structures are reproduced consistently across different system sizes, confirming the reliability of the findings. Indeed, the critical size $N_{A,c}$ remains unvaried at 6 molecules and the same observations regarding the structural properties remain valid.

5 Aromatic stacking and nucleation rate

Aromatic stacking plays a pivotal structural role in systems like DNA (between the aromatic bases), proteins (between most aromatic side chains of amino acids), protein-ligand complexes and crystal structures of organic molecules like many pharmaceuticals. Aromatic stacking was also suggested to play a key role in protein folding and crystal nucleation of organic molecules[99,100].

The importance of aromatic stacking in nucleation was demonstrated using benzoic acid derivatives[100], tuning the strength of aromatic stacking interactions through the addition of substituents to the aromatic ring. The substituents not only affect the electronic properties of the aromatic ring, but also add new interactions between one another[101,102].

The strength of aromatic stacking interactions is often much lower than that of hydrogens bonds, however stacking interactions have the advantage of building periodic aggregates along one direction. This is not always the case for hydrogen bonds. Aromatic carboxylic acids are more prone to form cyclic dimers rather than catemers, due to the presence of bulky aromatic rings[103]. Cyclic dimers of carboxylic acids are double hydrogen-bonded structures usually related by inversion. This

symmetry operation, in absence of other substituents capable of forming strong interactions, leads to the formation of zero-periodical entities that are unable to propagate infinitely. Aromatic stacking interactions, although weaker than hydrogen bonds, are still capable of binding dimers together and contributing to build up the crystal structure.

In this Section, first a database search is performed to find structures of benzoic acid derivatives and the PIXEL method[41] is used to rank them according to their stacking energies (Section 5.1). After a first selection of four compounds, solubility measurements are performed and two compounds are selected for nucleation rate experiments (Section 5.2). Then, the stacking program described in detail in the *Methods* (Section 2.1.4) is here summarized (Section 5.3). Finally, molecular dynamics simulations are performed on benzoic acid and the two derivatives to study aromatic stacking interactions from a dynamic point of view and define an aromatic stacking propensity (Section 5.4).

5.1 Stacking energy of benzoic acid derivatives

Carboxylic acids usually crystallize in structures containing either infinite chains of single hydrogen-bonded molecules or double hydrogen-bonded dimers[103]. In this work, emphasis is put on benzoic acid derivatives with structures containing dimers only, which pack through parallel aromatic stacking interactions. First, a database search is performed to collect crystal structures and the PIXEL method is employed to estimate the strength of aromatic stacking interactions (Section 5.1.1). Afterwards, the dataset of crystal structures is analyzed, and few structures are selected as potential candidates for nucleation rate experiments (Section 5.2.2).

5.1.1 Database search and stacking energy

Database search of dimeric benzoic acids

The Cambridge Structural Database[104] (CSD) was mined to retrieve the crystal structures of benzoic acid derivatives as potential candidates for nucleation rate

experiments. 408 meta- and para-substituted benzoic acids were found in the "best R" subset of the CSD (version 5.45, November 2023), with number of residues equal to 1, $R < 5\%$, no disorder, and no metals options. Ortho-substituted benzoic acids were not taken into account to avoid interferences with hydrogen bonds between carboxylic groups. Then, the motif search functionality in Mercury[105] revealed the presence of 266 dimeric structures, of which only 227 having $Z' \leq 1$ were selected. After excluding all the structures containing different atoms from H, C, N, O, S, P, F, Cl, and Br, 202 structures were left.

Determination of stacking energies with PIXEL

For these 202 structures, the charge density was calculated from an MP2/6-31G** wavefunction using Gaussian[42], for subsequent PIXEL calculations[41] (see Section 2.1.5). A few structures had to be discarded because of issues either in Gaussian or PIXEL calculations. A total of 191 valid structures were retained for analysis (Figure 5.1) and the full list is reported in *Appendix B*.

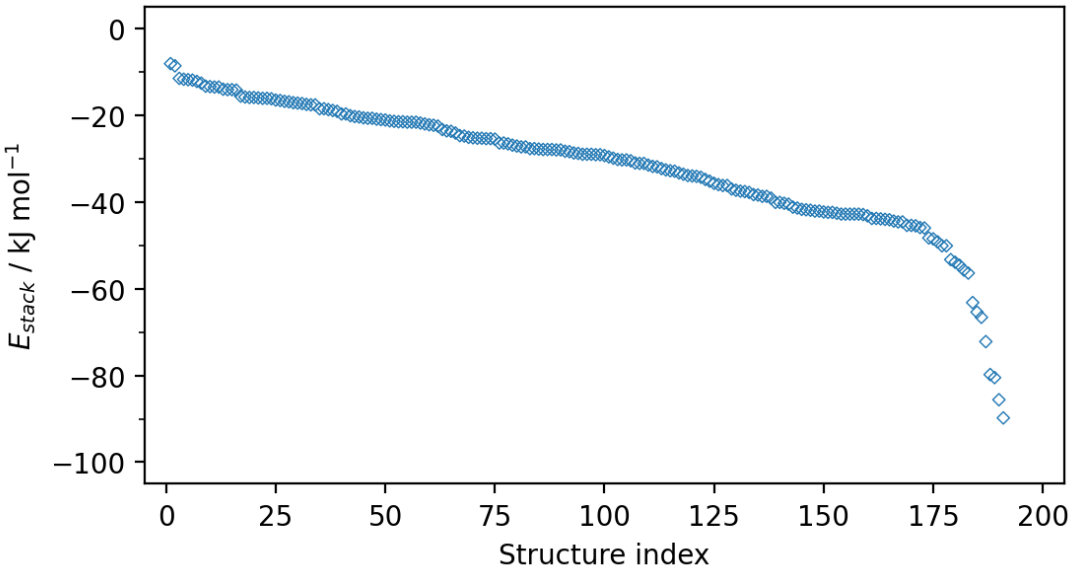

Figure 5.1 Stacking energy of 191 benzoic acid derivatives ranked from least to most negative.

In each structure, all aromatic stacking interactions were identified based on the following geometric criteria: (i) the distance between the centroids of the aromatic rings, $d_{centroids}$, must be less than 7.5 Å; (ii) the perpendicular distance from the

centroid of one ring to the plane of another ring, d_{plane}, must be less than 4.0 Å; and (iii) the acute angle ϑ between the normal vectors to the aromatic ring planes must be less than 10°, ensuring parallel or near-parallel orientation. For each identified stacked pair, the center-of-mass distance d_{com} was also recorded. Subsequently, all pairwise intermolecular interaction energies were computed using the PIXEL method, regardless of the interaction type. This analysis provided interaction energies along with d_{com} values for each molecular pair. To isolate stacking interactions from the PIXEL output (*pix.pri* file), molecular pairs were matched based on their d_{com} values. For each matched pair, the total interaction energy and its individual components, coulombic, polarization, dispersion, and repulsion, were extracted, along with symmetry operators and translation vectors associated with the interaction. Finally, the stacking energy E_{stack} associated with each compound was estimated as the strongest stacking interaction.

5.1.2 Dataset analysis and selection of compounds

Dataset overview

The dataset contains 191 structures of a large variety of molecules, having 15 to 69 atoms (Figure 5.2a), 1 to 4 aromatic rings (Figure 5.2b), and 1 to 4 carboxylic groups (Figure 5.2c). The structures belong to eight different space groups (Figure 5.2d): $P2_1/c$ and equivalent settings (51.8%), $P\bar{1}$ (29.8%), $C2/c$ and equivalent settings (13.1%), $Pbca$ (2.1%), $P2_1$ (1.0%), $P2_1/m$ (1.0%), $P2/c$ and equivalent settings (0.5%), and $Pbcn$ (0.5%) (*Appendix B*). In particular, $P2_1/c$, $P\bar{1}$, $C2/c$ are more populated than the general statistics performed on the entire CSD (33.9%, 25.2%, and 8.2% respectively)[74]. The main symmetry operators that relate stacked molecules are translation (54%) and inversion (42%) (Figure 5.2e). The stacking energies obtained with PIXEL span one order of magnitude, ranging from $-8.1\ kJ\ mol^{-1}$ to $-89.8\ kJ\ mol^{-1}$ (Figure 5.2f).

This huge difference is mainly due to the presence of multiple aromatic rings and specific interactions like hydrogen bonds in addition to aromatic stacking. To make an example, the strongest aromatic stacking interaction between benzoic acid

derivatives having a single aromatic ring is found in QQQBSS02[106] ($-48.5\ kJ\ mol^{-1}$) which also has a sulfonamide group with bulky n-propyl substituents that make a strong dispersion contribution.

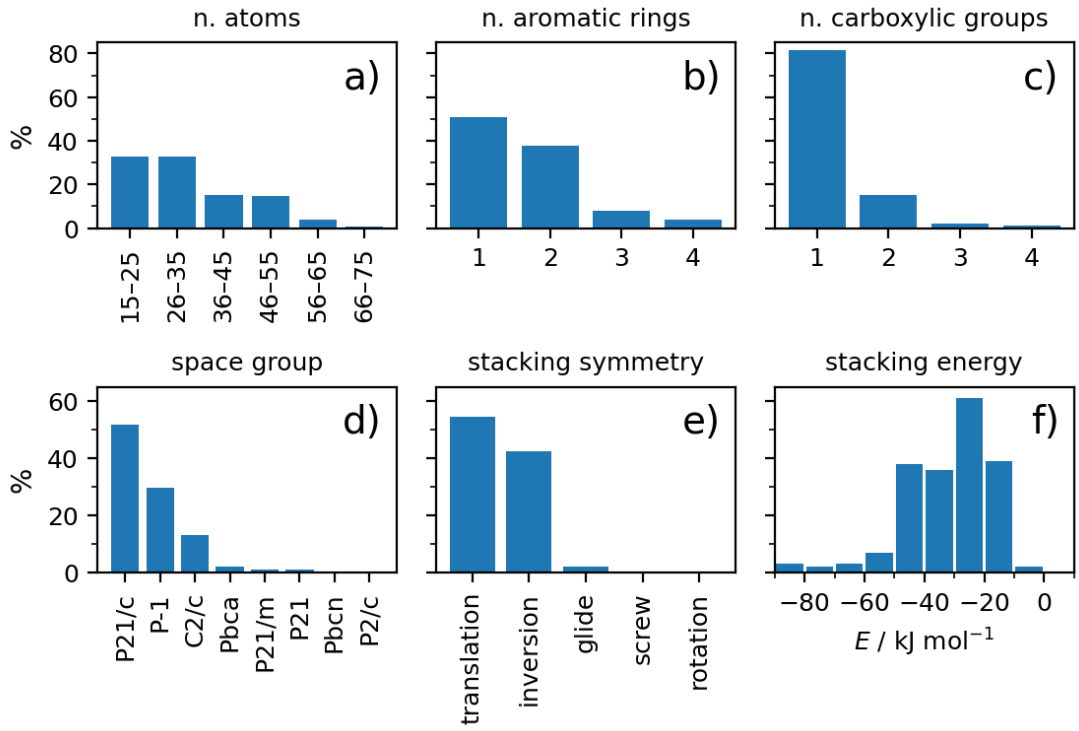

Figure 5.2 Dataset overview of benzoic acid derivatives. Distributions of number of atoms (a), aromatic rings (b), and carboxylic groups (c), along with space groups (d), symmetry operations relating stacked molecules (e), and stacking energy (f). All values are expressed as percentages.

Selection of benzoic acid derivatives

To reduce complexity and facilitate comparison among structurally similar compounds, candidate systems for nucleation rate experiments were selected from the most populated groups to ensure representativeness: space group 14 only ($P2_1/c$ and equivalent settings), maximum 25 atoms, 1 aromatic ring only, 1 carboxylic group only, stacked molecules related by translation only. Among the 17 compounds left (*Appendix B*), four of them were chosen with different stacking energies: 4-fluorobenzoic acid, 4-bromobenzoic acid, 4-(methoxycarbonyl)benzoic acid, and 4-(1H-1,2,3-triazol-1-yl)benzoic acid. Stacking energies, abbreviations and CSD refcodes of the four derivatives are reported in Table 5.1, along with benzoic acid as a reference. Aromatic stacking interactions between dimers are also shown (Figure 5.3).

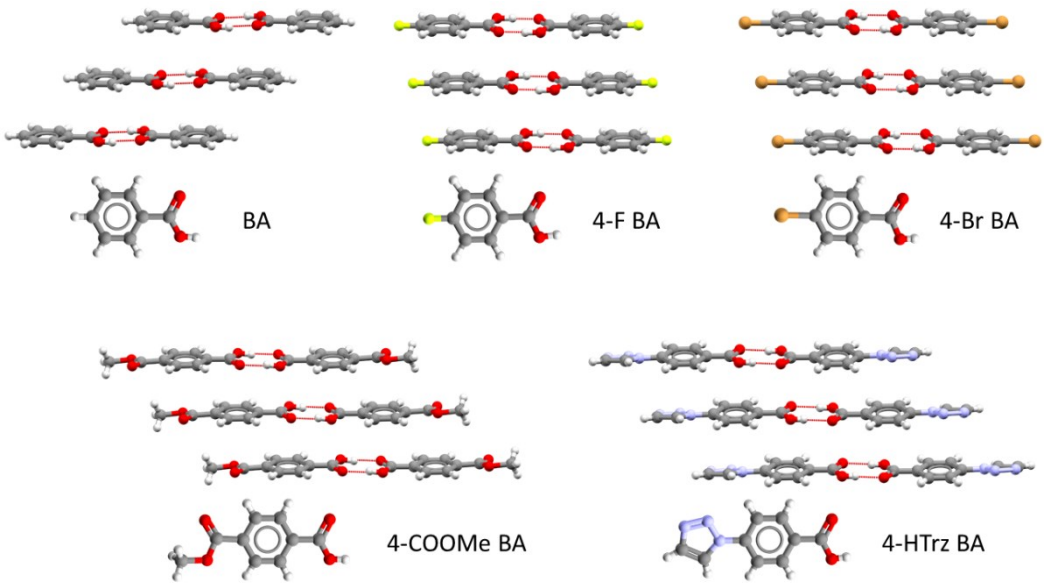

Figure 5.3 Stacked dimers in the crystal structures of benzoic acid (BA), 4-fluorobenzoic acid (4-F BA), 4-bromobenzoic acid (4-Br BA), 4-(methoxycarbonyl)benzoic acid (4-COOMe BA), and 4-(1H-1,2,3-triazol-1-yl)benzoic acid (4-HTrz BA).

Table 5.1 List of the four potential candidates for experiments with abbreviations, CSD refcodes, and stacking energies calculated with PIXEL (E_{stack}). Benzoic acid is also reported as a reference.

Compound name	Abbreviation	Refcode	E_{stack} / kJ mol^{-1}
Benzoic acid	BA	BENZAC12	-11.8
4-fluorobenzoic acid	4-F BA	PFBZAD15	-11.7
4-bromobenzoic acid	4-Br BA	BRBZAP01	-15.9
4-(methoxycarbonyl)benzoic acid	4-COOMe BA	QIYRUO	-18.5
4-(1H-1,2,3-triazol-1-yl)benzoic acid	4-HTrz BA	GEFVEV	-31.1

5.2 Nucleation rate determination

Nucleation rate is strongly dependent on supersaturation, which is why nucleation experiments are typically conducted at multiple supersaturation levels. Since relative supersaturation is defined as the ratio between concentration and solubility, accurate knowledge of solubility is essential. In Section 5.2.1, the solubility of benzoic acid derivatives is assessed, followed by the determination of nucleation rates from induction times in Section 5.2.2.

5.2.1 Solubility of benzoic acid derivatives

The solubility of 4-F BA, 4-Br BA, 4-COOMe BA, and 4-HTrz BA was initially screened in various solvents to obtain a rough estimate. Based on this preliminary evaluation, the most suitable solvent for nucleation rate experiments was selected and used for accurate solubility determination.

Preliminary solvent screening

A practical approach to roughly assess the kinetic solubility of a compound involves incrementally adding solvent to ~10 mg of solid until full dissolution occurs, thereby identifying the minimum volume required. Three solvents with markedly different physicochemical properties were selected, following previous studies[100,107]: isopropyl alcohol (IPA) (polar, protic), acetonitrile (polar, aprotic), and toluene (non-polar, aprotic). The selection was also guided by their relatively high boiling points, all above 80 °C, which help minimize solvent evaporation during nucleation rate experiments. The results of the preliminary solvent screening conducted at room temperature (~19 °C), are collected in Table 5.2. 4-F BA is the only compound that is soluble in all the three solvents, with a solubility in IPA > acetonitrile > toluene. 4-Br BA and 4-COOMe BA are soluble only IPA, while 4-HTrz BA is insoluble in all the three solvents.

Table 5.2 Outcomes of the preliminary solubility screening reporting solvent, benzoic acid (BA) derivative, and semi-quantitative estimate of kinetic solubility.

Solvent	BA derivative	Solubility
IPA	4-F	◆ ◆ ◆ ◆ ◆
	4-Br	◆ ◆ ◆
	4-COOMe	◆ ◆
	4-HTrz	Insoluble
Acetonitrile	4-F	◆ ◆ ◆ ◆
	4-Br	Insoluble
	4-COOMe	Insoluble
	4-HTrz	Insoluble
Toluene	4-F	◆ ◆ ◆
	4-Br	Insoluble
	4-COOMe	Insoluble
	4-HTrz	Insoluble

Attempt to use DMF as solvent

Several solvents were tested to dissolve 4-HTrz BA, with the aim of identifying a common solvent suitable for all four selected compounds. Acetone, acetonitrile, chloroform, dichloromethane, diethyl ether, ethanol, ethyl acetate, hexane, isopropanol, methanol, tetrahydrofuran, water, and 1,4-dioxane revealed ineffective. Dimethylformamide (DMF) was the only solvent capable of fully dissolving 4-HTrz BA, also exhibiting a remarkably high solubility. Since DMF is a solvent well known for its ability to form solvates, its compatibility with the four benzoic acid derivatives was assessed through liquid-assisted grinding (LAG) experiments, followed by powder X-ray diffraction. As for the LAG experiments, 125 mg of substance were milled with 100 µL of DMF for 40 minutes, using a Retsch MM400 Mixer Mill equipped with stainless steel milling jars, at a frequency of 30 Hz. The resulting slurries, deposited onto sample holders and covered with Kapton® tape, were analyzed through powder X-ray diffraction. The diffraction patterns were compared with the ones calculated by Mercury (Figure 5.4a-d).

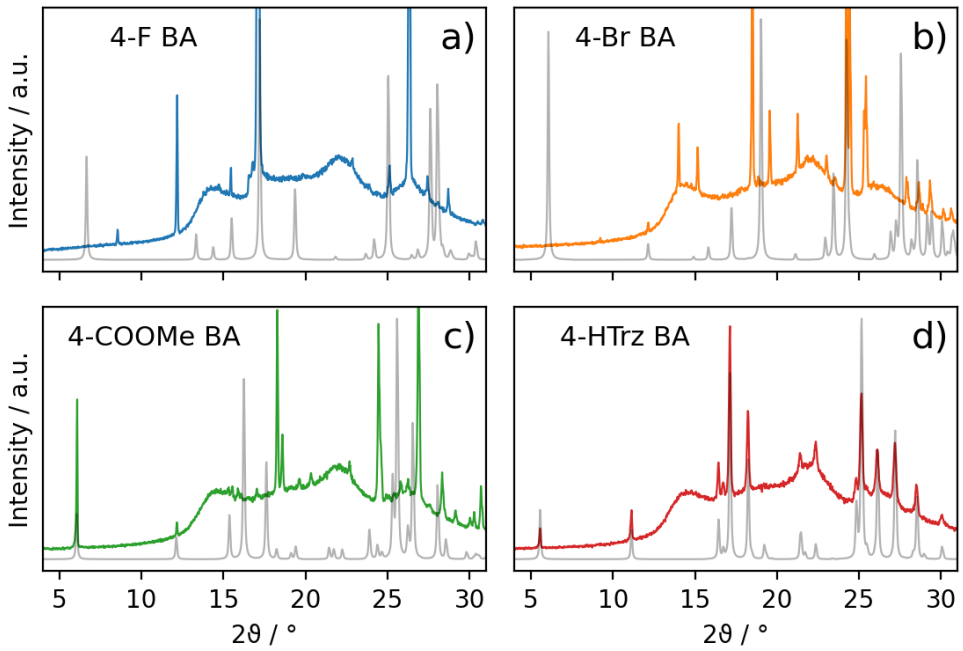

Figure 5.4 Raw experimental PXRD patterns of 4-F BA (a), 4-Br BA (b), 4-COOMe BA (c), and 4-HTrz BA (c) after LAG experiments with DMF (colored lines), compared to calculated patterns using Mercury software (semi-transparent lines). The amorphous peaks contributing to the background are due to the Kapton® tape.

4-HTrz BA shows no evidence of solvation with DMF (Figure 5.4d), while 4-F BA, 4-Br BA, and 4-COOMe BA have different patterns from the calculated ones, suggesting the formation of solvates, confirmed by single-crystal X-ray diffraction. The formation of solvates would completely alter the crystal structure and thus also stacking energy, and nucleation rate. For this reason, the idea of using DMF as a common solvent was abandoned. Nevertheless, the solubility of 4-HTrz BA was measured in IPA, in the hope that its thermodynamic solubility would exceed the kinetic one.

Solubility measurements in IPA

The thermodynamic solubility of 4-F BA, 4-Br BA, 4-COOMe BA, and 4-HTrz BA in IPA at 20 °C was accurately determined gravimetrically, using a Crystalline instrument and following the procedure described in detail in Section 2.4.2. The use of a rather mild vacuum was needed for 4-F BA due to the risk of sublimation. The thermodynamic solubility follows the same trend as the kinetic solubility: 4-F > 4-Br > 4-COOMe > 4-HTrz BA (Figure 5.5).

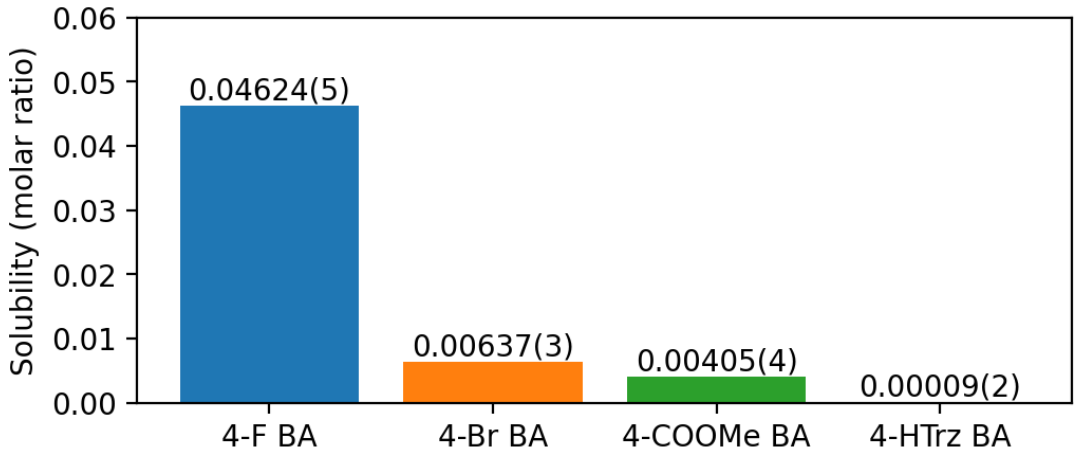

Figure 5.5 Solubility of 4-F BA, 4-Br BA, 4-COOMe BA, and 4-HTrz BA expressed as molar ratio. Solubility values, along with their standard deviations, are reported on top of the bars.

This trend is opposite to that observed for the strength of stacking. The solubility of 4-HTrz BA in IPA expressed as molar ratio is 0.00009(2) that is 45 times lower than 4-COOMe BA, which is the second least soluble compound. Thus, 4-HTrz was ultimately excluded from the study for its poor solubility in IPA that would not

guarantee a correct detection of the nucleation event in subsequent nucleation rate experiments. 4-F BA, instead, was excluded due to time reasons.

5.2.2 Nucleation rate of benzoic acid derivatives

Nucleation is a stochastic phenomenon, meaning that it is inherently unpredictable in both space and time. Consequently, a statistical approach is required, and multiple repetitions of the same experiment must be carried out. Moreover, nucleation is an activated process that involves overcoming an energy barrier. This results in a delay before nucleation occurs, known as the induction time. By collecting induction times from a series of experiments conducted under identical conditions, it is possible to estimate the nucleation rate. Crystal16 multi-reactor crystallizer was employed for the nucleation rate experiments, following the procedure described in detail in Section 2.4.3. Among the selected benzoic acid derivatives, due time reasons and the poor solubility of 4-HTrz BA, only 4-Br BA and 4-COOMe BA were investigated.

Induction times

Induction times are not directly accessible with nucleation rate experiments. What is possible to measure is the detection time t_D, that is the time required by the instrument to realize that nucleation has happened. This time include also the growth time t_g, that is the time required for the crystal nuclei to grow to a sufficient size to induce a decrease in transmissivity that is then measured by the instrument. The t_g can be estimated as the shortest detection time in each set of 80 crystallization experiments. This simple choice is also reliable, as demonstrated in a previous work[51]. In general, higher supersaturation leads to shorter induction time. The five supersaturation levels for each compound were selected to ensure that induction times, and thus also detection times, were neither too short nor too long. Indeed, if the detection time is too short, the accuracy of the results would be compromised, while too long detection times would make it impractical to achieve enough nucleation events within a reasonable experimental timeframe[51]. Supersaturations levels for 4-Br BA were selected in the range 1.27–1.34 while for 4-COOMe BA in the range 1.18–1.26.

Nucleation rate

Nucleation rate J is related to the cumulative probability of nucleation as a function of detection time, $P(t_D)$, by Equation 5.1, where V represents the volume of solution used and t_g corresponds to the growth time. Thus, it is possible to fit the experimental data with Equation 5.1 to obtain a value of J. The fit is performed for each supersaturation, using the data points of the 80 repetitions (Figure 5.6a,b).

$$P(t_D) = 1 - \exp\left(-JV\left(t_D - t_g\right)\right) \tag{5.1}$$

The optimal supersaturation range for 4-COOMe BA, based on the detection time, is 1.18–1.26 (Figure 5.6b). This range lies entirely below that of 4-Br BA, which spans 1.27–1.34 (Figure 5.6a). Moreover, the cumulative probability curves of 4-COOMe BA are generally steeper, suggesting that it nucleates more easily than 4-Br BA, requiring both lower supersaturation and shorter induction times. Consistently, the nucleation rate of 4-COOMe BA is overall higher than that of 4-Br BA, despite the lower superaturation.

The nucleation rate J of 4-Br BA and 4-COOMe BA were also compared to the one of BA, as reported by Cruz-Cabeza and coworkers[100] (Figure 5.6c). Since J of the three compounds was determined in different supersaturation ranges, a direct comparison is not meaningful. Instead, a nucleation propensity can be defined as the supersaturation needed to reach a given value of J, as previously proposed in the Literature[100]. For this work, a reference nucleation rate $J = 100\ m^{-3}s^{-1}$ was chosen, as it lies within the experimentally accessible range for all the three compounds. The corresponding supersaturation is defined as S_{100}. A lower S_{100} denotes a higher nucleation propensity, since less supersaturation is required to reach the same nucleation rate. Values of S_{100} for the three compounds were obtained by interpolating the experimental data using a simple exponential function $f(x) = ae^{bx}$. The resulting values for BA, 4-Br BA, and 4-COOMe BA are 1.88, 1.31, and 1.21, indicating that nucleation propensity follows the trend BA < 4-Br BA < 4-COOMe BA.

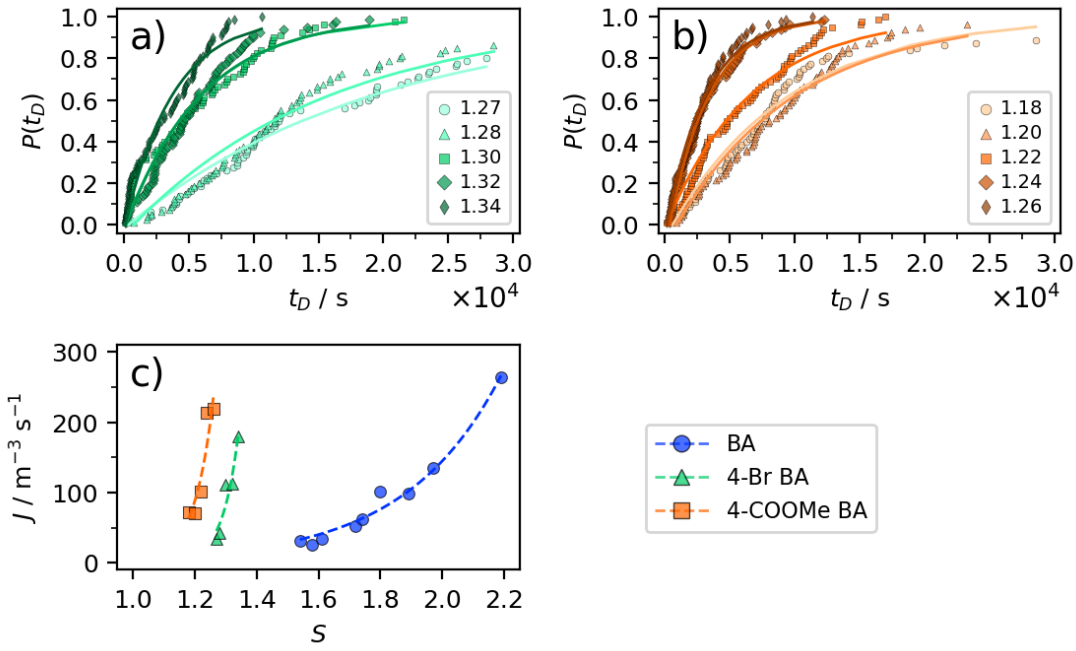

Figure 5.6 Cumulative probability of nucleation of 4-Br BA (a), and 4-COOMe BA (b) as a function of detection time at 5 different supersaturations. Data points are fitted using Equation 5.1. The nucleation rates of 4-Br BA and 4-COOMe BA are also shown (c) and compared with BA data taken from the Literature[100]. Nucleation rate data points are fitted by simple exponential functions.

5.3 Stacking program in short

The *Stacking* program was specifically designed for this work and is described in detail in Section 2.1.4. Here is reported the essential information to understand the underlying logic. *Stacking* detects aromatic stacking interactions between benzoic acid derivatives exploiting four geometric criteria, which are based on the distance between ring centroids, the angle between normal vectors to the aromatic rings, the distance between the ring centroid of one molecule and the ring plane of the other, and the overlap between two interacting molecules. The resulting data can be further analyzed using Pandas library in Python.

5.4 MD simulations

Molecular dynamics simulations were carried out on benzoic acid (BA), taken as a reference, as well as on 4-Br BA and 4-COOMe BA, employed in the nucleation

rate experiments described in Section 5.2.2. The purpose of these simulations was not to reproduce the nucleation rate measurements. Such a task would require prohibitively long simulations on much larger systems to capture nucleation events in solution, while also accounting for depletion effects to maintain constant supersaturation (see *Introduction*, Section 1.1.3). Instead, the present simulations aim to investigate intermolecular interactions, focusing on aromatic stacking, to quantify stacking propensity and assess whether it correlates with the observed nucleation propensity based on nucleation rate experiments. Accordingly, the liquid phase of the three compounds was simulated at four different temperatures around their experimental melting points, ensuring adequate sampling of aromatic stacking interactions. A general analysis of intermolecular interactions (Section 5.4.1) is followed by a specific analysis of stacking interactions (Section 5.4.2), where the aromatic stacking propensity is defined. The details regarding the simulations are reported in Section 2.2.2.

5.4.1 General analysis of intermolecular interactions

Distribution of interaction energies

The distribution of interaction energies in BA, 4-Br BA, and 4-COOMe BA is compared to highlight differences in the interaction modes (Figure 5.7). BA has two distinct peaks for single and double hydrogen bonds and a feature corresponding to weaker interactions. Although BA is the compound with the weakest stacking interactions, it presents the strongest hydrogen bonds, having both peaks centered at more negative energies (-86 and $-48\ kJ\ mol^{-1}$) compared to 4-Br BA (-60 and $-35\ kJ\ mol^{-1}$) and 4-COOMe BA (-72 and $-38\ kJ\ mol^{-1}$). 4-COOMe BA, which is the one with the strongest stacking interactions, lies in the middle between BA and 4-Br BA. This indicates that the strength of stacking interactions does not correlate with the one of hydrogen bonds. An increase of temperature only has a minor effect on BA, while it leads to lower peaks for the other compounds.

Weaker interactions, including aromatic stacking interactions, are almost completely separated from the single hydrogen bond peak in BA, indicating a large

difference in energy between the two interaction modes. However, 4-Br BA and 4-COOMe BA have weaker hydrogen bonds and stronger stacking interactions, leading to a partial overlap between the single hydrogen bond peak and the weaker interaction feature. Moreover, this latter feature has a width that is in trend with the strength of stacking interactions, being narrower for BA and broader for 4-COOMe, extending at more negative energies.

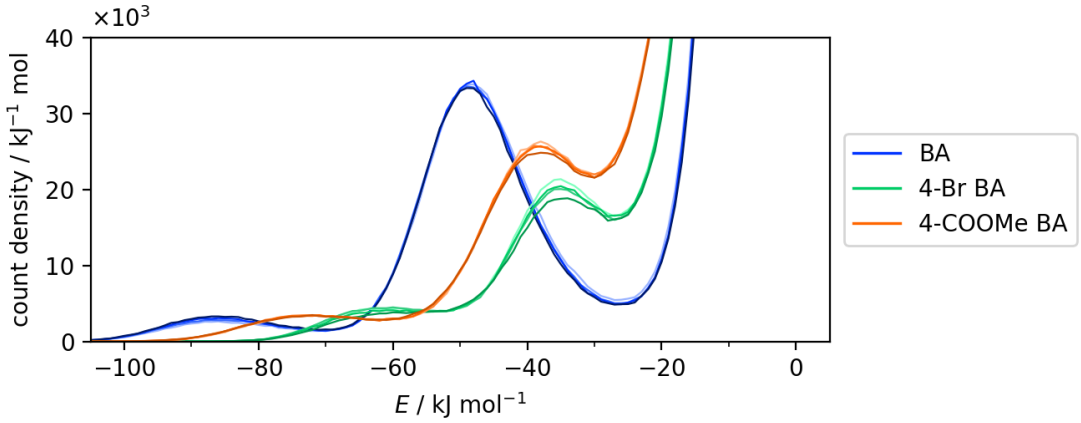

Figure 5.7 Density distribution of interaction energies in the simulations of BA, 4-Br BA, and 4-COOMe BA. Darker shades correspond to higher temperatures.

Radial distribution functions

Radial distribution functions (RDFs) $g(r)$ were exploited to study intermolecular interactions from a geometric point of view. The center of mass of 4-Br and 4-COOMe BA is significantly shifted towards the substituents compared to BA. The resulting RDFs (Figure 5.8a) show peaks related to the same feature at different distances. For example, cyclic hydrogen-bonded dimers generate a peak at 6.95 Å (BA), 10.25 Å (4-Br BA), and 9.55 Å (4-COOMe BA), although the interaction mode is almost identical. For this reason, RDFs were also calculated using ring centroids to make a better comparison between the three compounds (Figure 5.8b).

All of them share a peak at 9.25 Å that corresponds to molecules interacting through hydrogen bonds like cyclic dimers. Stacking interactions are found at lower distances. In particular, the strongest stacking interactions are around 4 Å, regardless of the compound (see Section 5.4.2). In this region, it is possible to detect some shoulders and the $g(r)$ curve is the highest for 4-COOMe BA, the compound with the

strongest stacking interactions, and the lowest for BA, which has the weakest stacking interactions. In addition, BA presents a broad peak with maximum at 5.75 Å that accounts for several interaction modes sharing similar distances, including displaced stacking and C-H···π interactions between aromatic rings. This latter interaction is less frequent in 4-Br BA which has one hydrogen atom less than BA and one more bromine atom, that is much bulkier than hydrogen, hindering this interaction mode. This is reflected in a broader peak. 4-COOMe BA has an even bulkier methyl ester group that makes the feature almost disappear. The effect of temperature on the RDFs is almost negligible.

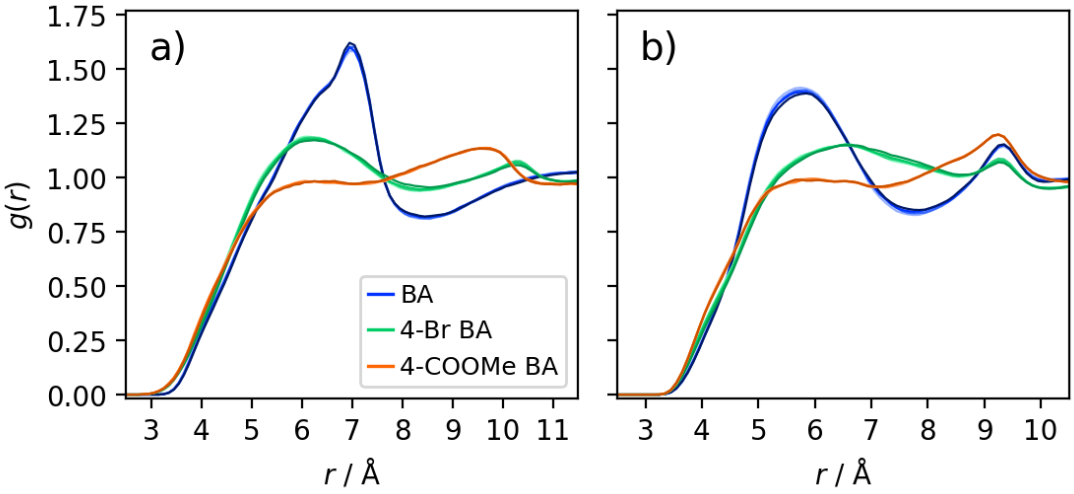

Figure 5.8 Center-of-mass RDFs (a) and ring-centroid RDFs (b) of BA, 4-Br BA, and 4-COOMe BA. Darker shades correspond to higher temperatures.

5.4.2 Analysis of stacking interactions

The specific analysis of aromatic stacking interactions was performed using the *Stacking* program described in Section 5.3. To account for proper aromatic stacking interactions, only those characterized by at least 4 overlapped atoms after projecting one molecule on the other and vice versa are considered (see Section 2.1.4). Aromatic stacking propensity based on energy, frequency, persistence of stacking interactions along with the size of stacked aggregates is also defined.

Relationship between energy and geometry

The energy of stacking interactions was studied as a function of specific geometric parameters (Figure 5.9a-d). These include the distance between ring centroids, ring centroid and ring plane, and carbonyl carbon and ring centroid, calculated as described in Section 2.1.4, along with the angle between vectors $\boldsymbol{v}_1$, in the ring plane (see Figure 3.8). The angle between $\boldsymbol{v}_2$ vectors is not shown since it is imposed to be 0 or 180° within a threshold of 30°. Also in this case, temperature only has a minor effect and only results of simulations at the lowest temperature are shown for clarity.

There is a clear relationship between the interaction energy and the distance between the ring centroids, with shorter distances corresponding to more negative interaction energies (Figure 5.9a). This is explained by a higher level of overlap between molecules that maximizes dispersion interactions. The optimal distance is around 4 Å for all the compounds but 4-COOMe BA has the most negative interaction energy while BA the least negative one. As for the distance between ring centroid and ring plane, the optimal value is at around 3.75 Å (Figure 5.9b). Shorter distances indicate either that molecules are too close to one another, making the repulsion energy increase in magnitude, or that they are not perfectly parallel and thus do not maximize dispersion interactions. Longer distances as well imply weaker interactions, and consequently less negative energies. Contrarily, interaction energy does not correlate strongly with the distance between carbonyl carbon and ring centroid for BA and 4-Br BA while a stronger correlation is found in 4-COOMe BA, that also has a narrower distance distribution (Figure 5.9c). This is explained by the presence of two carbonyl carbon atoms that increase the chances of having a shorter distance. Finally, all angles between $\boldsymbol{v}_1$ vectors (see Figure 3.8) are adopted with similar energies (Figure 5.9d). However, there is a slight preference towards 0° and 180° since molecules in these two orientations maximize the overlap compared to the 90° orientation that is slightly disfavored.

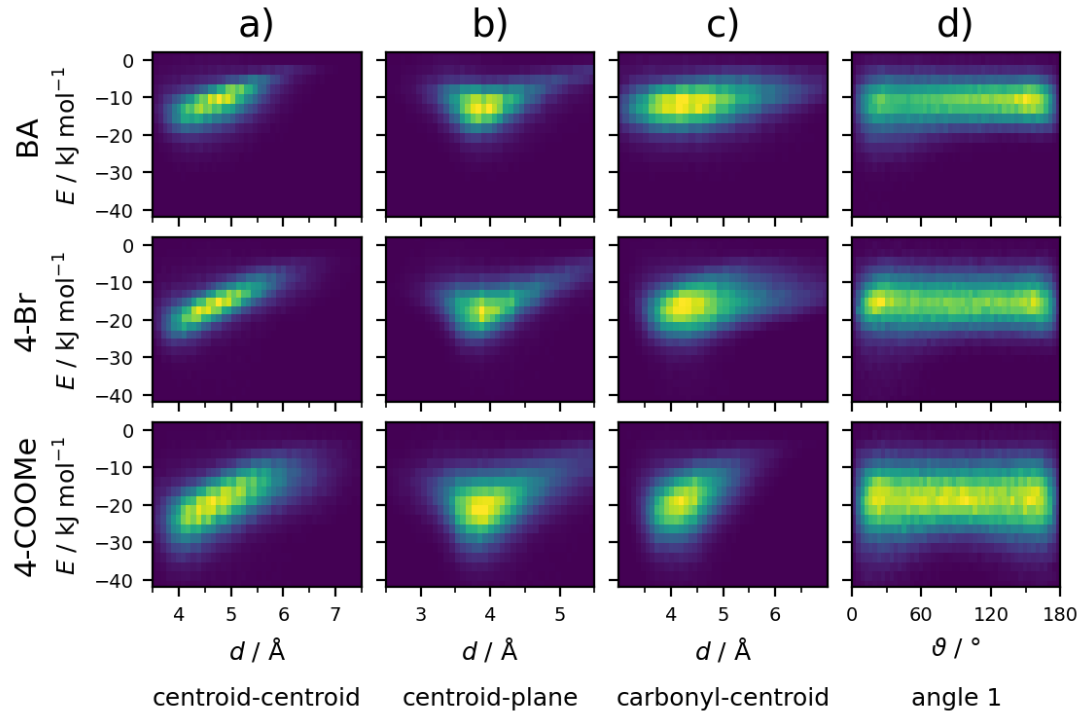

Figure 5.9 Relationship between energy of stacking interactions and distance between ring centroids (a), ring centroid and ring plane (b), carbonyl carbon and ring centroid (c), along with the angle between $\boldsymbol{v}_1$ vectors (d). Lighter colors indicate higher counts.

Energy

PIXEL stacking energy can be used as an estimate of aromatic stacking propensity. According to PIXEL, the strength of stacking interactions is BA < 4-Br BA < 4-COOMe BA with values of -11.8, -15.9, and $-18.5\ kJ\ mol^{-1}$. The average stacking energy derived from molecular dynamics simulations follows the same trend, despite some minor differences (Figure 5.10a). In particular, MD and PIXEL stacking energies differs by $\sim 0.3\ kJ\ mol^{-1}$, showing a remarkably good quantitative agreement. However, it is important to consider that PIXEL stacking energy refers to molecules frozen in their static crystal geometries and therefore does not account for dynamical effects or structural rearrangements, while the energies coming from MD simulations do. The effect of temperature on stacking energy in the range under consideration is negligible. Thus, another estimate of aromatic stacking propensity was obtained by averaging the MD stacking energy over all the temperature range, obtaining -11.5, -15.6, and $-18.9\ kJ\ mol^{-1}$ for BA, 4-Br BA, and 4-COOMe BA.

Frequency

The frequency of stacking interactions, calculated as the number of interactions per frame, has a strong dependence on temperature (Figure 5.10b). For each compound, the higher the temperature, the lower the number of stacking interactions, as expected, since high temperature promotes the disruption of weak intermolecular interactions like aromatic stacking interactions. A direct comparison between different compounds is not straightforward, since the temperatures at which they are simulated are different. However, it is emblematic how 4-COOMe BA, that is the compound with the strongest aromatic stacking interactions, has a higher frequency of stacking interactions compared to BA, that is the one with the weakest stacking interactions, although the simulation temperatures of 4-COOMe are all higher than those of BA.

Similarly to what was done for nucleation propensity (Section 5.2.2), it is possible to define aromatic stacking propensity based on the frequency of stacking interactions looking at the temperature at which a specific number of interactions is reached. As a reference, the average number of interactions across the entire dataset was used ($N_{ave} = 116.4$ interactions). This process, however, requires both interpolation and extrapolation but it is nonetheless useful for a relative comparison. The temperatures corresponding to the average number of interactions for each compound are denoted as T_{ave} and are 363.9, 499.9, and 571.4 K for BA, 4-Br BA, and 4-COOMe BA.

Persistence

Also in the case of time persistence, t_{life}, of aromatic stacking interactions, a higher temperature is associated with a higher breakage frequency of weak interactions that leads to a shorter persistence (Figure 5.10c). As in the frequency of stacking interactions, 4-COOMe BA has higher values compared to BA, despite the higher temperature of the simulations. Aromatic stacking propensity based on time persistence was estimated with the method described above for the frequency of stacking interactions, using an average value of $t_{life} = 2.71\ ps$. The temperatures T_{ave} corresponding to this value are 336.4, 507.7, and 574.8 K. Despite the short average

t_{life}, it is important to note that some of the stacking interactions persist for up to 16-28 ps.

Size of stacked aggregates

The size of stacked aggregates depends on temperature as well, with higher temperatures corresponding to lower sizes (Figure 5.10d). Regardless of the compound and the temperature, the size of stacked aggregates ranges from 2.3 to 2.5 molecules, with overall average value of 2.43 molecules. The differences among the three compounds are small, considering that the vast majority of stacked aggregates (~70%) are made up by only 2 molecules and another consistent part (~20%) by 3 molecules, but are nevertheless significant. There is a clear trend with respect to the strength of stacking interactions that sees 4-COOMe BA in first place as regards the size of stacked aggregates, despite the higher temperature of simulations, and BA in last place.

As shown in Section 4.2, the average information is not enough when studying processes like nucleation that stem from nanoscale inhomogeneities. That's why it is important to study also extreme cases even though they occur with a low statistical frequency. Except for dimers, which are more abundant in BA, the absolute number of stacked aggregates of 4-COOMe BA is greater than that of BA for each cluster size. In particular, the larger the cluster size, the higher the difference in favor of 4-COOMe BA. To make some examples, the number of trimers (3-mers) and tetramers (4-mers) is 4-11% higher for 4-COOMe BA, the number of 5-mers, 6-mers, and 7-mers is 11-28% higher, the number of 8-mers, 9-mers, and 10-mers is 47-136% higher, whereas larger aggregates are too underrepresented to allow for meaningful statistics. The maximum size is 17-18 molecules for all of the three compounds. Large aggregates are similar to one another, sharing a pillar-like structure reminiscent of the one found in the respective crystal structures (Figure 5.10e). However, a certain degree of disorder is still present.

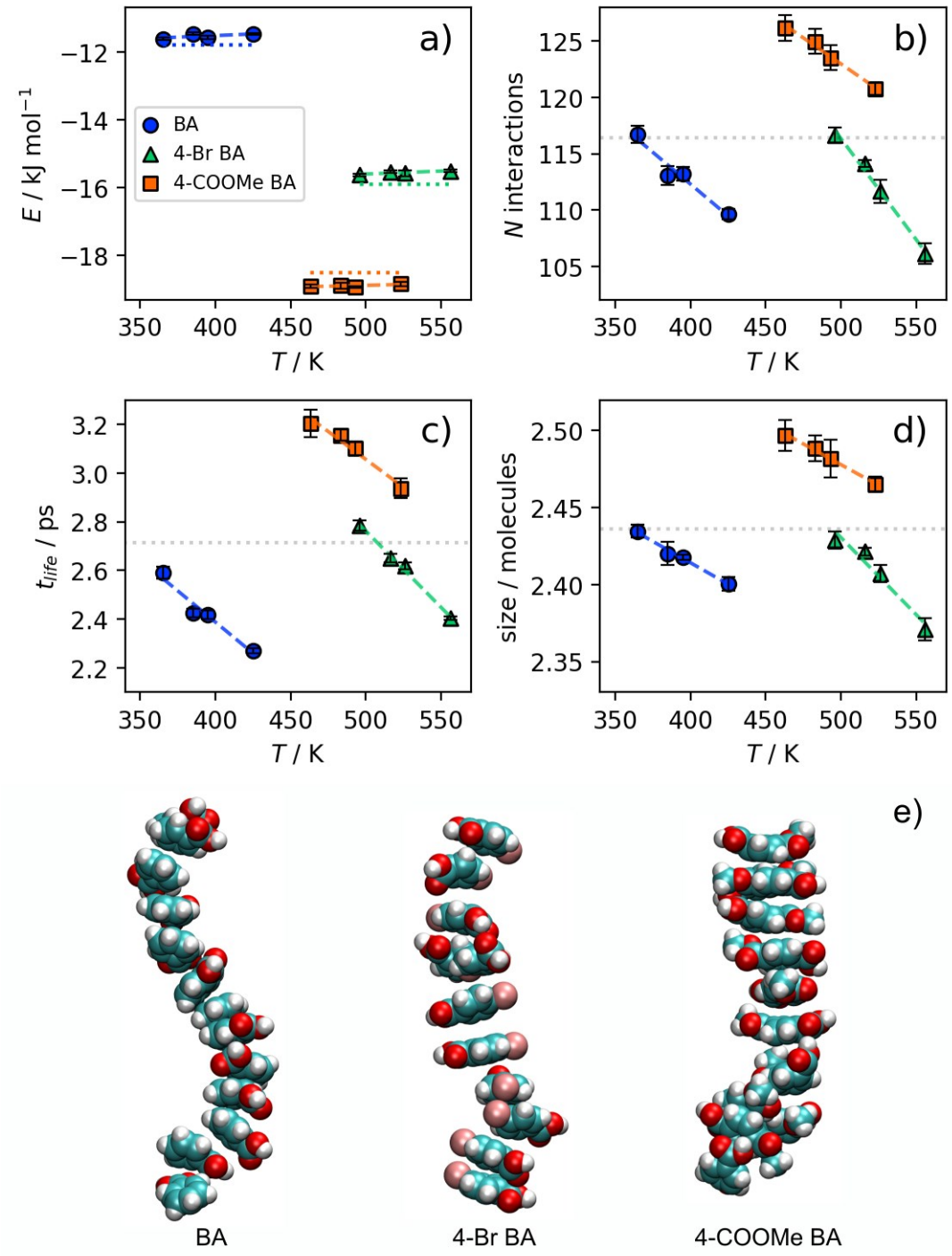

Figure 5.10 Stacking energy derived from MD simulations (a) compared with PIXEL calculations (dotted lines), number of aromatic stacking interactions (b), and their lifetime (c), along with the size of stacked aggregates (d) as a function of temperature. Horizontal dotted lines indicate the average value of the entire dataset, used as a reference for interpolation/extrapolation. Examples of stacked aggregates are also reported (e).

Aromatic stacking propensity based on the average size of stacked aggregates was estimated as described above for the frequency and persistence of stacking interactions, with an average size of 2.43 molecules for the entire dataset and corresponding temperatures T_{ave} of 360.4, 494.5, and 577.6 K for BA, 4-Br BA, and 4-COOMe BA.

Correlations with nucleation propensity

Nucleation propensity was introduced in Section 5.2.2 through the parameter S_{100}, defined as the supersaturation required to reach a nucleation rate $J = 100\ m^{-3}s^{-1}$. By definition, a higher S_{100} corresponds to a lower nucleation propensity and vice versa, showing an inverse correlation. For convenience, this relationship can be expressed directly by using the reciprocal value $1/S_{100}$, which increases with increasing nucleation propensity.

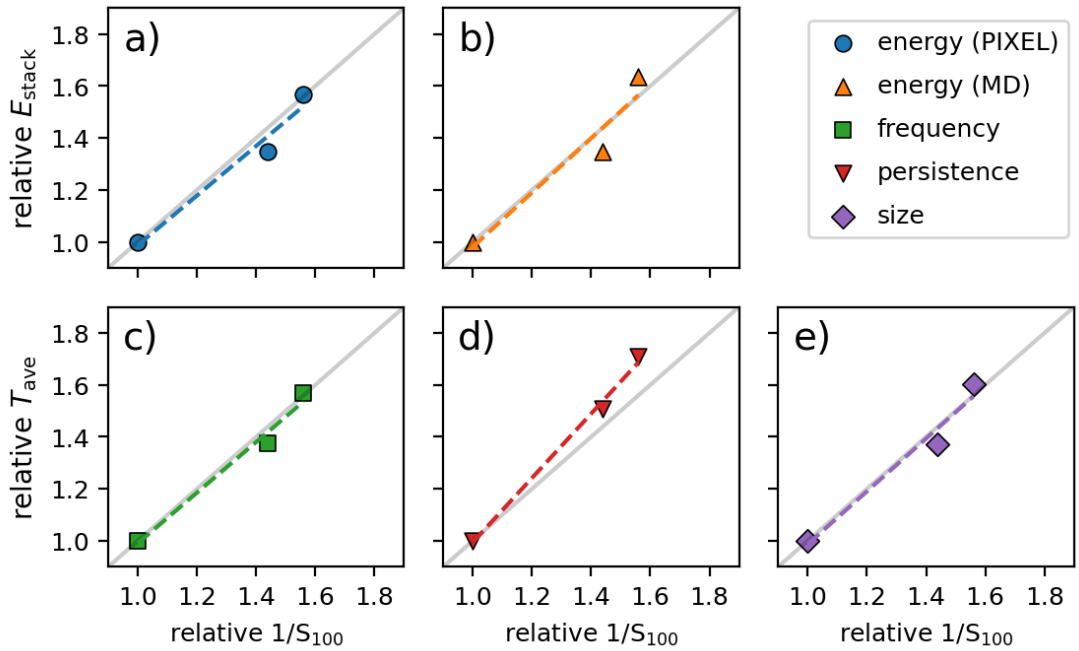

Figure 5.11 Correlation between aromatic stacking propensity and nucleation propensity. Aromatic stacking propensity is expressed in terms of energy as MD stacking energy (a) and PIXEL stacking energy (b), and also in terms of temperature to reach the average value of frequency (c) and persistence of stacking interactions (d) along with the average size of stacked aggregates (e). Nucleation rate propensity is expressed in terms of $1/S_{100}$. All values are reported relative to BA, used as a reference.

In this Section, aromatic stacking propensity was introduced in two different ways, one that is simply based on stacking energy E_{stack}, and the other one on T_{ave},

defined as the temperature at which the average value of the dataset is reached for a specific property (frequency, persistence, or size). To put everything on the same footing, $1/S_{100}$, E_{stack}, and T_{ave} have been normalized by their respective values for BA, allowing a relative comparison (Figure 5.11a-e). The results show a clear correlation between aromatic stacking propensity, derived in several ways from computational calculations, and nucleation propensity, determined experimentally. Most likely, all these correlations arise from a common underlying factor, stacking energy, which leads to a higher frequency and persistence of stacking interactions and larger stacked aggregates.

Conclusion

This work allowed to dive into the realm of molecular liquids and crystals, with the ambitious aim of improving the basic understanding of nucleation, a process of fundamental importance for many fields, including pharmaceuticals and materials science, that is elusive and hard to study experimentally. Computational techniques like molecular dynamics, PIXEL, DFT, and Monte Carlo, were combined with experimental ones, such as crystallizations, solubility measurements, nucleation rate determination, liquid-assisted grinding, and X-ray diffraction.

It should be mentioned that all the simulations performed in this work were unbiased, with longest duration of ~5 ns and largest system size of ~60 Å, reflecting the current MiCMoS limitations. Under these conditions, the direct observation of nucleation events is highly unlikely, as nucleation is a rare stochastic process that typically occurs after long and variable induction times. Nevertheless, the present computational setup is well suited to the objectives of this work, which focus on characterizing the system in the pre-nucleation regime. Within this regime, the system explores almost randomly its accessible phase space, and extending the simulations moderately, while still not observing nucleation, is not expected to lead to qualitative changes in the observed behavior. Observing nucleation directly would require simulations over much longer timescales and substantially larger systems in order to

increase the probability of sampling such rare events; however, this lies beyond the current computational feasibility. For these reasons, the results presented here remain physically meaningful and appropriate for addressing the aims of this work.

In Section 3, a new module of MiCMoS software was developed to simulate liquid systems in confined spaces such as nanolayers, nanotubes, and nanocavities. Compared to existing approaches, this setup enables simulations in both the *NVT* and *NpT* ensembles by employing infinitely extended barriers (with the exception of the nanocavity, which is non-periodic in all directions), rather than finite barriers that would require the explicit simulation of adjacent unconfined regions acting as molecular reservoirs, potentially introducing boundary artifacts. Moreover, the degree of confinement can be tuned to achieve a desired packing efficiency. Confining barriers rely on pseudoatoms that can be fine-tuned to simulate environments with different surface properties. For the present study, hydrophobic barriers of carbon atoms were employed, but the Lennard-Jones parameters can be chosen among any atom type. In addition, the stiffness of force constants can be tuned to simulate soft or hard confinement, and in future studies point charges could be added to generate electric fields, and the barriers could potentially be tailored adding functional groups able to make hydrogen bonds or any other chemical interaction.

Small nanocavities were exploited to accelerate self-aggregation and study the interaction modes of benzoic acid molecules. Small aggregates were studied in detail, with particular attention to cyclic aggregates, that are the ones with the most negative cohesive energies since they have all hydrogen bond donors and acceptors satisfied and are inherently compact, maximizing intermolecular interactions.

The effects of different confinement geometries on the dynamical and structural properties of confined liquid benzoic acid were investigated. The orientation of molecules with respect to the confining barriers is not random. Instead, molecules adsorb onto the barriers forming multiple stacked layers, indicating partial ordering, likely promoted by the hydrostatic pressure of the surrounding liquid in a system close to its maximum theoretical packing efficiency.

In addition, when confining barriers with stiff force constants are employed, a reversible liquid-liquid transition between low density and high-density states is detected in the nanocavity simulation, potentially related to the high-density liquid droplet predicted by non-classical nucleation theories. This is a major result, which demonstrates that the method can unequivocally detect elusive liquid-liquid transitions at the nanoscale, with possible applications in microfluidics.

Moreover, extreme confinement in nanolayers was exploited to study the potential formation of new interaction modes that might lead to new polymorphs of benzoic acid, that is still monomorphic to date. Long, catemeric structures constituted by up to 26 hydrogen-bonded molecules were found, resembling a top ranked crystal structure predicted using CSP methods.

In Section 4, the focus is shifted towards bulk liquid benzoic acid and, particularly, its inhomogeneities like subcritical clusters. A precise definition of these elusive molecular aggregates, which are present even before nucleation occurs and are the main character of several non-classical nucleation theories, has been lacking to date. This is particularly true for liquids containing a single molecular species in which simple geometric criteria applied to solute molecules are not sufficient.

In this work, nanoscale density inhomogeneities were further confirmed to be present even above the melting point but are more pronounced in undercooled systems. It was shown that averages should be handled carefully to avoid losing precious information in these kind of studies.

A new definition of subcritical clusters was proposed, based on interaction energy between molecular pairs, along with time persistence and excess energy of the aggregate. Excess energy was introduced as an estimate of how an aggregate can be considered as a separate entity from the surrounding environment. This definition of subcritical clusters relies on an energy threshold that may appear arbitrary or guided by chemical intuition. Benzoic acid is a simple molecule and the choice of a reasonable threshold was easy, while more complex molecules might require trying different thresholds. The comparison of these results could even give a deeper insight. In addition, the choice to consider an aggregate destroyed when even a single molecule

detach is very harsh, but it helps to get rid of transient aggregates, keeping only persistent ones.

It was shown that subcritical clusters need to have a critical size before gaining a negative excess energy and are usually composed of hydrogen-bonded chains, folded in a compact arrangement, with aromatic stacking interactions contributing to the stability of the aggregate. Simulations on larger systems increase the chance of forming larger aggregates, as it should be for inhomogeneities, but the general conclusions remain unchanged.

The coalescence of multiple catemeric subcritical clusters may give rise to a larger subcritical cluster with a sufficiently long lifetime to allow the system to explore configurational space and eventually nucleate the dimeric form of benzoic acid. Aromatic stacking interactions could facilitate the transition from the catemeric to the dimeric structural motif by stabilizing the subcritical cluster in the correct configuration and inducing translational correlations among molecules. This transformation, involving the conversion from single to double hydrogen bonds, may occur through a concerted molecular rearrangement triggered by low-frequency fluctuations in the system. To test this non-classical mechanistic hypothesis, longer simulations combined with enhanced-sampling techniques, such as metadynamics, would be required.

Future work will be extending the scope of this study to different classes of organic molecules like simple hydrocarbons, alcohols, amines, and amides, both aromatic and aliphatic. This would allow to determine which kind of molecules are more prone to form inhomogeneous systems characterized by the presence of large semi-ordered aggregates enabling non-classical nucleation pathways. Moreover, multicomponent systems could be modelled explicitly including solvent molecules that might lead to the formation of different molecular aggregates and, potentially, to different polymorphs. In this case, highly supersaturated solutions should be simulated to favor aggregation; however, depletion effects might become important.

In Section 5, the importance of aromatic stacking interactions in nucleation was studied in detail from both an experimental and computational point of view. Although this interaction mode is significantly weaker than the double hydrogen bond in benzoic

acid dimers, it is crucial for the formation of a periodic crystal structure, since it extends infinitely in one direction contrarily to dimers that are 0-periodical entities. To confirm the key role of aromatic stacking in nucleation, the nucleation rate of benzoic acid derivatives having different stacking energy was determined to define nucleation propensity and molecular dynamics simulations were performed to determine aromatic stacking propensity. The experimental results show that the compounds characterized by the strongest stacking energy correspond to the ones with highest nucleation propensity, i.e. they require a lower relative supersaturation to have the same nucleation rate. Accordingly, molecular dynamics simulations show that the compounds with the strongest stacking energy are characterized by a higher number of stacking interactions, with a longer persistence and form larger stacked aggregates. The largest aggregates have a pillar-like structure resembling the one found in the crystal structure but still disordered. Although only larger and some orders of magnitude longer simulations might show the time evolution of these structures, they can be though as precursors of crystal nuclei.

As future work, it would be interesting to test more benzoic acid derivatives to confirm the trends found in this study and create a database of compounds to build accurate predictive models. Moreover, the influence of solvent and temperature on the relative nucleation rate should also be examined, although solubility might be a major problem for some compounds. Even though absolute nucleation rate can significantly change depending on the solvent choice, relative nucleation rate should not change much and stacking energy is expected to have a greater effect.

Overall, the results of this work contribute to the fundamental understanding of the complex nucleation process in the context of molecular crystals, with implications in the design of new materials.

Bibliography

(1) Datta, S.; Grant, D. J. W. Crystal Structures of Drugs: Advances in Determination, Prediction and Engineering. *Nat. Rev. Drug Discov.* **2004**, *3* (1), 42–57. https://doi.org/10.1038/nrd1280.

(2) Simone, E. Crystal Engineering Approaches for the Food Industry. In *Physics in Food Manufacturing: Case studies in fundamental and applied research*; IOP Publishing Bristol, UK, 2020; pp 1–7.

(3) Jia, G.; Chen, Y.; Sun, A.; Orlien, V. Control of Ice Crystal Nucleation and Growth during the Food Freezing Process. *Compr. Rev. Food Sci. Food Saf.* **2022**, *21* (3), 2433–2454. https://doi.org/10.1111/1541-4337.12950.

(4) Zhang, X.; Dong, H.; Hu, W. Organic Semiconductor Single Crystals for Electronics and Photonics. *Adv. Mater.* **2018**, *30* (44). https://doi.org/10.1002/adma.201801048.

(5) Wang, C.; Dong, H.; Jiang, L.; Hu, W. Organic Semiconductor Crystals. *Chem. Soc. Rev.* **2018**, *47* (2), 422–500. https://doi.org/10.1039/C7CS00490G.

(6) Jin, H.; Gao, X.; Ren, K.; Liu, J.; Qiao, L.; Liu, M.; Chen, W.; He, Y.; Dong, S.; Xu, Z.; Li, F. Review on Piezoelectric Actuators Based on High-Performance Piezoelectric Materials. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2022**, *69* (11), 3057–3069. https://doi.org/10.1109/TUFFC.2022.3175853.

(7) Furukawa, H.; Gándara, F.; Zhang, Y.-B.; Jiang, J.; Queen, W. L.; Hudson, M. R.; Yaghi, O. M. Water Adsorption in Porous Metal–Organic Frameworks and Related Materials. *J. Am. Chem. Soc.* **2014**, *136* (11), 4369–4381. https://doi.org/10.1021/ja500330a.

(8) Millward, A. R.; Yaghi, O. M. Metal−Organic Frameworks with Exceptionally High Capacity for Storage of Carbon Dioxide at Room Temperature. *J. Am. Chem. Soc.* **2005**, *127* (51), 17998–17999. https://doi.org/10.1021/ja0570032.

(9) Green, M. A.; Ho-Baillie, A.; Snaith, H. J. The Emergence of Perovskite Solar Cells. *Nat. Photonics* **2014**, *8* (7), 506–514. https://doi.org/10.1038/nphoton.2014.134.
(10) *Online Dictionary of Crystallography*. https://dictionary.iucr.org/Crystal (accessed 2025-12-12).
(11) Cruz-Cabeza, A. J.; Feeder, N.; Davey, R. J. Open Questions in Organic Crystal Polymorphism. *Commun. Chem.* **2020**, *3* (1), 142. https://doi.org/10.1038/s42004-020-00388-9.
(12) Firaha, D.; Liu, Y. M.; Streek, J. van de; Sasikumar, K.; Dietrich, H.; Helfferich, J.; Aerts, L.; Braun, D. E.; Broo, A.; DiPasquale, A. G.; Lee, A. Y.; Meur, S. L.; Lill, S. O. N.; Lunsmann, W. J.; Mattei, A.; Muglia, P.; Putra, O. D.; Raoui, M.; Reutzel-Edens, S. M.; Rome, S.; Sheikh, A. Y.; Tkatchenko, A.; Woollam, G. R.; Neumann, M. A. Predicting Crystal Form Stability under Real-World Conditions. *Nature* **2023**, *623* (7986), 324–328. https://doi.org/10.1038/s41586-023-06587-3.
(13) Macetti, G.; Sironi, L.; Vacchini, M.; Lo Presti, L. Symmetry-Constrained Monte Carlo for the Crystal Structure Prediction of Small Organic Molecules. *Cryst. Growth Des.* **2025**, *25* (20), 8382–8392. https://doi.org/10.1021/acs.cgd.5c00537.
(14) Li, J.; Deepak, F. L. *In Situ* Kinetic Observations on Crystal Nucleation and Growth. *Chem. Rev.* **2022**, *122* (23), 16911–16982. https://doi.org/10.1021/acs.chemrev.1c01067.
(15) 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. https://doi.org/10.1021/ar800217x.
(16) Vekilov, P. G. Two-Step Mechanism for the Nucleation of Crystals from Solution. *J. Cryst. Growth* **2005**, *275* (1–2), 65–76. https://doi.org/10.1016/j.jcrysgro.2004.10.068.
(17) Gebauer, D.; Völkel, A.; Cölfen, H. Stable Prenucleation Calcium Carbonate Clusters. *Science* **2008**, *322* (5909), 1819–1822. https://doi.org/10.1126/science.1164271.
(18) Finney, A. R.; Salvalaglio, M. Multiple Pathways in NaCl Homogeneous Crystal Nucleation. *Faraday Discuss.* **2022**, *235*, 56–80. https://doi.org/10.1039/D1FD00089F.
(19) Chattopadhyay, S.; Erdemir, D.; Evans, J. M. B.; Ilavsky, J.; Amenitsch, H.; Segre, C. U.; Myerson, A. S. SAXS Study of the Nucleation of Glycine Crystals from a Supersaturated Solution. *Cryst. Growth Des.* **2005**, *5* (2), 523–527. https://doi.org/10.1021/cg0497344.
(20) Du, J. S.; Bae, Y.; Yoreo, J. J. D. Non-Classical Crystallization in Soft and Organic Materials. *Nat. Rev. Mater.* **2024**, *9* (4), 229–248. https://doi.org/10.1038/s41578-023-00637-y.
(21) Binder, K.; Horbach, J.; Kob, W.; Paul, W.; Varnik, F. Molecular Dynamics Simulations. *J. Phys. Condens. Matter* **2004**, *16* (5), S429–S453. https://doi.org/10.1088/0953-8984/16/5/006.
(22) Anwar, J.; Zahn, D. Uncovering Molecular Processes in Crystal Nucleation and Growth by Using Molecular Simulation. *Angew. Chem. Int. Ed.* **2011**, *50* (9), 1996–2013. https://doi.org/10.1002/anie.201000463.

(23) 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. https://doi.org/10.1021/acs.chemrev.5b00744.
(24) Car, R.; Parrinello, M. Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* **1985**, *55* (22), 2471–2474. https://doi.org/10.1103/PhysRevLett.55.2471.
(25) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci.* **2002**, *99* (20), 12562–12566. https://doi.org/10.1073/pnas.202427399.
(26) Giberti, F.; Salvalaglio, M.; Parrinello, M. Metadynamics Studies of Crystal Nucleation. *IUCrJ* **2015**, *2* (2), 256–266. https://doi.org/10.1107/S2052252514027626.
(27) Perego, C.; Salvalaglio, M.; Parrinello, M. Molecular Dynamics Simulations of Solutions at Constant Chemical Potential. *J. Chem. Phys.* **2015**, *142* (14). https://doi.org/10.1063/1.4917200.
(28) Karmakar, T.; Piaggi, P. M.; Parrinello, M. Molecular Dynamics Simulations of Crystal Nucleation from Solution at Constant Chemical Potential. *J. Chem. Theory Comput.* **2019**, *15* (12), 6923–6930. https://doi.org/10.1021/acs.jctc.9b00795.
(29) Karmakar, T.; Finney, A. R.; Salvalaglio, M.; Yazaydin, A. O.; Perego, C. Non-Equilibrium Modeling of Concentration-Driven Processes with Constant Chemical Potential Molecular Dynamics Simulations. *Acc. Chem. Res.* **2023**, *56* (10), 1156–1167. https://doi.org/10.1021/acs.accounts.2c00811.
(30) Matsumoto, M.; Saito, S.; Ohmine, I. Molecular Dynamics Simulation of the Ice Nucleation and Growth Process Leading to Water Freezing. *Nature* **2002**, *416* (6879), 409–413. https://doi.org/10.1038/416409a.
(31) Indri, S. S.; Dietrich, F. M.; Wagner, A.; Hartstein, M.; Nativ-Roth, E.; Pavan, M. J.; Kronik, L.; Salvalaglio, M.; Palmer, B. A. Guanine Crystallization by Particle Attachment. *J. Am. Chem. Soc.* **2025**, *147* (22), 19139–19147. https://doi.org/10.1021/jacs.5c04543.
(32) Gavezzotti, A.; Lo Presti, L.; Rizzato, S. Molecular Dynamics Simulation of Organic Materials: Structure, Potentials and the MiCMoS Computer Platform. *CrystEngComm* **2022**, *24* (5), 922–930. https://doi.org/10.1039/D1CE01360B.
(33) Gavezzotti, A. Efficient Computer Modeling of Organic Materials. The Atom–Atom, Coulomb–London–Pauli (AA-CLP) Model for Intermolecular Electrostatic-Polarization, Dispersion and Repulsion Energies. *New J. Chem.* **2011**, *35* (7), 1360. https://doi.org/10.1039/c0nj00982b.
(34) Gavezzotti, A.; Lo Presti, L. Molecular Dynamics Simulation of Organic Crystals: Introducing the *CLP-Dyncry* Environment. *J. Appl. Crystallogr.* **2019**, *52* (6), 1253–1263. https://doi.org/10.1107/S1600576719012238.
(35) 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. https://doi.org/10.1039/D0CE00334D.

(36) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1). https://doi.org/10.1063/1.2408420.
(37) Lo Presti, L.; Sironi, L.; Macetti, G. Milano Chemistry Molecular Simulation (MiCMoS) v2.4, 2025. https://doi.org/10.13140/RG.2.2.33377.68964.
(38) Sironi, L.; Macetti, G.; Lo Presti, L. Molecular Dynamics Investigation of Benzoic Acid in Confined Spaces. *Phys. Chem. Chem. Phys.* **2023**, *25* (41), 28006–28019. https://doi.org/10.1039/D3CP02886K.
(39) Zaccone, A. Explicit Analytical Solution for Random Close Packing in D=2 and D=3. *Phys. Rev. Lett.* **2022**, *128* (2), 028002. https://doi.org/10.1103/PhysRevLett.128.028002.
(40) 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), 2077–2084. https://doi.org/10.1039/C8NJ05825C.
(41) Gavezzotti, A. Calculation of Intermolecular Interaction Energies by Direct Numerical Integration over Electron Densities. I. Electrostatic and Polarization Energies in Molecular Crystals. *J. Phys. Chem. B* **2002**, *106* (16), 4145–4154. https://doi.org/10.1021/jp0144202.
(42) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Jr, J. A. M.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian 16 Revision C.01, 2016.
(43) Feld, R.; Lehmann, M. S.; Muir, K. 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. *Z. Für Krist. - Cryst. Mater.* **1981**, *157* (3–4), 215–231. https://doi.org/10.1524/zkri.1981.157.3-4.215.
(44) Fronczek, F. R. CSD Communication, 2003.
(45) Li, Y.; Wang, J. The Crystal Structure of 4-(Methoxycarbonyl)Benzoic Acid, $C_9H_8O_4$. *Z. Für Krist. - New Cryst. Struct.* **2019**, *234* (2), 349–350. https://doi.org/10.1515/ncrs-2018-0408.
(46) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.

(47) Malde, A. K.; Zuo, L.; Breeze, M.; Stroet, M.; Poger, D.; Nair, P. C.; Oostenbrink, C.; Mark, A. E. An Automated Force Field Topology Builder (ATB) and Repository: Version 1.0. *J. Chem. Theory Comput.* **2011**, *7* (12), 4026–4037. https://doi.org/10.1021/ct200196m.
(48) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate *Ab Initio* Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132* (15). https://doi.org/10.1063/1.3382344.
(49) Boys, S. F.; Bernardi, F. The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors. *Mol. Phys.* **1970**, *19* (4), 553–566. https://doi.org/10.1080/00268977000101561.
(50) Simon, S.; Duran, M.; Dannenberg, J. J. How Does Basis Set Superposition Error Change the Potential Surfaces for Hydrogen-Bonded Dimers? *J. Chem. Phys.* **1996**, *105* (24), 11024–11031. https://doi.org/10.1063/1.472902.
(51) Xiao, Y.; Tang, S. K.; Hao, H.; Davey, R. J.; Vetter, T. Quantifying the Inherent Uncertainty Associated with Nucleation Rates Estimated from Induction Time Data Measured in Small Volumes. *Cryst. Growth Des.* **2017**, *17* (5), 2852–2863. https://doi.org/10.1021/acs.cgd.7b00372.
(52) Grommet, A. B.; Feller, M.; Klajn, R. Chemical Reactivity under Nanoconfinement. *Nat. Nanotechnol.* **2020**, *15* (4), 256–271. https://doi.org/10.1038/s41565-020-0652-2.
(53) Meldrum, F. C.; O'Shaughnessy, C. Crystallization in Confinement. *Adv. Mater.* **2020**, *32* (31). https://doi.org/10.1002/adma.202001068.
(54) Ha, J.-M.; Wolf, J. H.; Hillmyer, M. A.; Ward, M. D. Polymorph Selectivity under Nanoscopic Confinement. *J. Am. Chem. Soc.* **2004**, *126* (11), 3382–3383. https://doi.org/10.1021/ja049724r.
(55) Hamilton, B. D.; Ha, J.-M.; Hillmyer, M. A.; Ward, M. D. Manipulating Crystal Growth and Polymorphism by Confinement in Nanoscale Crystallization Chambers. *Acc. Chem. Res.* **2012**, *45* (3), 414–423. https://doi.org/10.1021/ar200147v.
(56) Beiner, M.; Rengarajan; Pankaj, S.; Enke, D.; Steinhart, M. Manipulating the Crystalline State of Pharmaceuticals by Nanoconfinement. *Nano Lett.* **2007**, *7* (5), 1381–1385. https://doi.org/10.1021/nl0705081.
(57) Hamilton, B. D.; Hillmyer, M. A.; Ward, M. D. Glycine Polymorphism in Nanoscale Crystallization Chambers. *Cryst. Growth Des.* **2008**, *8* (9), 3368–3375. https://doi.org/10.1021/cg800326a.
(58) Su, Y.; Liu, G.; Xie, B.; Fu, D.; Wang, D. Crystallization Features of Normal Alkanes in Confined Geometry. *Acc. Chem. Res.* **2014**, *47* (1), 192–201. https://doi.org/10.1021/ar400116c.
(59) Adawy, A.; Amghouz, Z.; Hest, J. C. M. van; Wilson, D. A. Sub-Micron Polymeric Stomatocytes as Promising Templates for Confined Crystallization and Diffraction Experiments. *Small* **2017**, *13* (28), 1700642. https://doi.org/10.1002/smll.201700642.
(60) Liu, Y.; Wu, Y.; Yao, J.; Yin, J.; Lu, J.; Mao, J.; Yao, M.; Luo, F. Confined Crystallization and Melting Behaviors of 3-Pentadecylphenol in Anodic Alumina

Oxide Nanopores. *ACS Omega* **2021**, *6* (28), 18235–18247. https://doi.org/10.1021/acsomega.1c02112.

(61) Stephens, C. J.; Kim, Y.-Y.; Evans, S. D.; Meldrum, F. C.; Christenson, H. K. Early Stages of Crystallization of Calcium Carbonate Revealed in Picoliter Droplets. *J. Am. Chem. Soc.* **2011**, *133* (14), 5210–5213. https://doi.org/10.1021/ja200309m.

(62) Sun, Y. S.; Chung, T. M.; Li, Y. J.; Ho, R. M.; Ko, B. T.; Jeng, U. S.; Lotz, B. Crystalline Polymers in Nanoscale 1D Spatial Confinement. *Macromolecules* **2006**, *39* (17), 5782–5788. https://doi.org/10.1021/ma0608121.

(63) Mizukami, M.; Yanagimachi, T.; Ohta, N.; Shibuya, Y.; Yagi, N.; Kurihara, K. Structures of Nanoconfined Liquids Determined by Synchrotron X-Ray Diffraction. *Langmuir* **2022**, *38* (17), 5248–5256. https://doi.org/10.1021/acs.langmuir.1c02621.

(64) Weiss, H.; Cheng, H.-W.; Mars, J.; Li, H.; Merola, C.; Renner, F. U.; Honkimäki, V.; Valtiner, M.; Mezger, M. Structure and Dynamics of Confined Liquids: Challenges and Perspectives for the X-Ray Surface Forces Apparatus. *Langmuir* **2019**, *35* (51), 16679–16692. https://doi.org/10.1021/acs.langmuir.9b01215.

(65) Alba-Simionesco, C.; Dosseh, G.; Dumont, E.; Frick, B.; Geil, B.; Morineau, D.; Teboul, V.; Xia, Y. Confinement of Molecular Liquids: Consequences on Thermodynamic, Static and Dynamical Properties of Benzene and Toluene. *Eur. Phys. J. E* **2003**, *12* (1), 19–28. https://doi.org/10.1140/epje/i2003-10055-1.

(66) Buntkowsky, G.; Breitzke, H.; Adamczyk, A.; Roelofs, F.; Emmler, T.; Gedat, E.; Grünberg, B.; Xu, Y.; Limbach, H.-H.; Shenderovich, I.; Vyalikh, A.; Findenegg, G. Structural and Dynamical Properties of Guest Molecules Confined in Mesoporous Silica Materials Revealed by NMR. *Phys. Chem. Chem. Phys.* **2007**, *9* (35), 4843. https://doi.org/10.1039/b707322d.

(67) Nakamuro, T.; Sakakibara, M.; Nada, H.; Harano, K.; Nakamura, E. Capturing the Moment of Emergence of Crystal Nucleus from Disorder. *J. Am. Chem. Soc.* **2021**, *143* (4), 1763–1767. https://doi.org/10.1021/jacs.0c12100.

(68) Liao, Z.; Wynne, K. A Metastable Amorphous Intermediate Is Responsible for Laser-Induced Nucleation of Glycine. *J. Am. Chem. Soc.* **2022**, *144* (15), 6727–6733. https://doi.org/10.1021/jacs.1c11154.

(69) Urquidi, O.; Brazard, J.; LeMessurier, N.; Simine, L.; Adachi, T. B. M. In Situ Optical Spectroscopy of Crystallization: One Crystal Nucleation at a Time. *Proc. Natl. Acad. Sci.* **2022**, *119* (16). https://doi.org/10.1073/pnas.2122990119.

(70) Giri, A. K.; Teixeira, F.; Cordeiro, M. N. D. S. Structure and Kinetics of Water in Highly Confined Conditions: A Molecular Dynamics Simulation Study. *J. Mol. Liq.* **2018**, *268*, 625–636. https://doi.org/10.1016/j.molliq.2018.07.083.

(71) Zhu, X.; Vandamme, M.; Jiang, Z.; Brochard, L. Molecular Simulation of the Confined Crystallization of Ice in Cement Nanopore. *J. Chem. Phys.* **2023**, *159* (15). https://doi.org/10.1063/5.0169783.

(72) Desiraju, G. R. Supramolecular Synthons in Crystal Engineering—A New Organic Synthesis. *Angew. Chem. Int. Ed. Engl.* **1995**, *34* (21), 2311–2327. https://doi.org/10.1002/anie.199523111.

(73) Uchida, A.; Hasegawa, M.; Yamaguchi, S.; Takezawa, E.; Ishikawa, A.; Kagayama, T. (1 *S* *,2 *S* *,4 *R* *,5 *R* *)-Cyclohexane-1,2,4,5-Tetracarboxylic Acid. *Acta Crystallogr. Sect. E Struct. Rep. Online* **2014**, *70* (1), o75–o75. https://doi.org/10.1107/S1600536813033795.
(74) *CSD Space Group Statistics - Space Group Frequency Ordering*. https://www.ccdc.cam.ac.uk/ (accessed 2025-12-12).
(75) *Chemeo*. https://www.chemeo.com/ (accessed 2025-12-12).
(76) Sun, T.; Teja, A. S. Density, Viscosity, and Thermal Conductivity of Aqueous Benzoic Acid Mixtures between 375 K and 465 K. *J. Chem. Eng. Data* **2004**, *49* (6), 1843–1846. https://doi.org/10.1021/je0497247.
(77) Kohlrausch, R. Ueber Das Dellmann'sche Elektrometer. *Ann. Phys.* **1847**, *148* (11), 353–405. https://doi.org/10.1002/andp.18471481102.
(78) Williams, G.; Watts, D. C. Non-Symmetrical Dielectric Relaxation Behaviour Arising from a Simple Empirical Decay Function. *Trans. Faraday Soc.* **1970**, *66*, 80. https://doi.org/10.1039/tf9706600080.
(79) Phillips, J. C. Stretched Exponential Relaxation in Molecular and Electronic Glasses. *Rep. Prog. Phys.* **1996**, *59* (9), 1133–1207. https://doi.org/10.1088/0034-4885/59/9/003.
(80) Ngai, K. L. Alternative Explanation of the Difference between Translational Diffusion and Rotational Diffusion in Supercooled Liquids. *J. Phys. Chem. B* **1999**, *103* (48), 10684–10694. https://doi.org/10.1021/jp990554s.
(81) Wojdyr, M. *Fityk* : A General-Purpose Peak Fitting Program. *J. Appl. Crystallogr.* **2010**, *43* (5), 1126–1128. https://doi.org/10.1107/S0021889810030499.
(82) Granick, S. Motions and Relaxations of Confined Liquids. *Science* **1991**, *253* (5026), 1374–1379. https://doi.org/10.1126/science.253.5026.1374.
(83) Gebauer, D.; Kellermeier, M.; Gale, J. D.; Bergström, L.; Cölfen, H. Pre-Nucleation Clusters as Solute Precursors in Crystallisation. *Chem Soc Rev* **2014**, *43* (7), 2348–2371. https://doi.org/10.1039/C3CS60451A.
(84) Sedlák, M.; Rak, D. Large-Scale Inhomogeneities in Solutions of Low Molar Mass Compounds and Mixtures of Liquids: Supramolecular Structures or Nanobubbles? *J. Phys. Chem. B* **2013**, *117* (8), 2495–2504. https://doi.org/10.1021/jp4002093.
(85) Karthika, S.; Radhakrishnan, T. K.; Kalaichelvi, P. A Review of Classical and Nonclassical Nucleation Theories. *Cryst. Growth Des.* **2016**, *16* (11), 6663–6681. https://doi.org/10.1021/acs.cgd.6b00794.
(86) Zhang, X. *Crystallization via Nonclassical Pathways Volume 1: Nucleation, Assembly, Observation & Application*; ACS Symposium Series; American Chemical Society: Washington, 2020.
(87) Pouget, E. M.; Bomans, P. H. H.; Goos, J. A. C. M.; Frederik, P. M.; With, G. de; Sommerdijk, N. A. J. M. The Initial Stages of Template-Controlled $CaCO_3$ Formation Revealed by Cryo-TEM. *Science* **2009**, *323* (5920), 1455–1458. https://doi.org/10.1126/science.1169434.
(88) Habraken, W. J. E. M.; Tao, J.; Brylka, L. J.; Friedrich, H.; Bertinetti, L.; Schenk, A. S.; Verch, A.; Dmitrovic, V.; Bomans, P. H. H.; Frederik, P. M.; Laven, J.; Schoot, P. van der; Aichmayer, B.; With, G. de; DeYoreo, J. J.; Sommerdijk, N.

A. J. M. Ion-Association Complexes Unite Classical and Non-Classical Theories for the Biomimetic Nucleation of Calcium Phosphate. *Nat. Commun.* **2013**, *4* (1), 1507. https://doi.org/10.1038/ncomms2490.

(89) Rimer, J. D.; Tsapatsis, M. Nucleation of Open Framework Materials: Navigating the Voids. *MRS Bull.* **2016**, *41* (5), 393–398. https://doi.org/10.1557/mrs.2016.89.

(90) Kellermeier, M.; Rosenberg, R.; Moise, A.; Anders, U.; Przybylski, M.; Cölfen, H. Amino Acids Form Prenucleation Clusters: ESI-MS as a Fast Detection Method in Comparison to Analytical Ultracentrifugation. *Faraday Discuss.* **2012**, *159*, 23. https://doi.org/10.1039/c2fd20060k.

(91) Carpenter, J. E.; Grünwald, M. Pre-Nucleation Clusters Predict Crystal Structures in Models of Chiral Molecules. *J. Am. Chem. Soc.* **2021**, *143* (51), 21580–21593. https://doi.org/10.1021/jacs.1c09321.

(92) Sarupria, S.; Debenedetti, P. G. Homogeneous Nucleation of Methane Hydrate in Microsecond Molecular Dynamics Simulations. *J. Phys. Chem. Lett.* **2012**, *3* (20), 2942–2947. https://doi.org/10.1021/jz3012113.

(93) Demichelis, R.; Raiteri, P.; Gale, J. D.; Quigley, D.; Gebauer, D. Stable Prenucleation Mineral Clusters Are Liquid-like Ionic Polymers. *Nat. Commun.* **2011**, *2* (1), 590. https://doi.org/10.1038/ncomms1604.

(94) Velzen, D. van; Cardozo, R. L.; Langenkamp, H. Liquid Viscosity and Chemical Constitution of Organic Compounds: A New Correlation and a Compilation of Literature Data. EUR 4735.; 1972.

(95) Kramers, H. A. Brownian Motion in a Field of Force and the Diffusion Model of Chemical Reactions. *Physica* **1940**, *7* (4), 284–304. https://doi.org/10.1016/S0031-8914(40)90098-2.

(96) Moynihan, C. T.; Schroeder, J. Non-Exponential Structural Relaxation, Anomalous Light Scattering and Nanoscale Inhomogeneities in Glass-Forming Liquids. *J. Non-Cryst. Solids* **1993**, *160* (1–2), 52–59. https://doi.org/10.1016/0022-3093(93)90283-4.

(97) Pálinkás, G.; Jedlovszky, P. Network of Strongly Interacting Atoms in Liquid Argon. *Chem. Phys.* **1994**, *185* (2), 173–178. https://doi.org/10.1016/0301-0104(94)00128-6.

(98) Sironi, L.; Macetti, G.; Lo Presti, L. Nanoscale Inhomogeneities in Undercooled Benzoic Acid: A Molecular Dynamics Study. *J. Mol. Liq.* **2024**, *414*, 126141. https://doi.org/10.1016/j.molliq.2024.126141.

(99) Hunter, C. A.; Lawson, K. R.; Perkins, J.; Urch, C. J. Aromatic Interactions. *J. Chem. Soc. Perkin Trans. 2* **2001**, No. 5, 651–669. https://doi.org/10.1039/b008495f.

(100) Cruz-Cabeza, A. J.; Davey, R. J.; Sachithananthan, S. S.; Smith, R.; Tang, S. K.; Vetter, T.; Xiao, Y. Aromatic Stacking – a Key Step in Nucleation. *Chem. Commun.* **2017**, *53* (56), 7905–7908. https://doi.org/10.1039/C7CC02423A.

(101) Hwang, J.; Li, P.; Carroll, W. R.; Smith, M. D.; Pellechia, P. J.; Shimizu, K. D. Additivity of Substituent Effects in Aromatic Stacking Interactions. *J. Am. Chem. Soc.* **2014**, *136* (40), 14060–14067. https://doi.org/10.1021/ja504378p.

(102) Martinez, C. R.; Iverson, B. L. Rethinking the Term "Pi-Stacking." *Chem. Sci.* **2012**, *3* (7), 2191. https://doi.org/10.1039/c2sc20045g.

(103) Beyer, T.; Price, S. L. Dimer or Catemer? Low-Energy Crystal Packings for Small Carboxylic Acids. *J. Phys. Chem. B* **2000**, *104* (12), 2647–2655. https://doi.org/10.1021/jp9941413.

(104) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **2016**, *72* (2), 171–179. https://doi.org/10.1107/S2052520616003954.

(105) 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.; Wood, P. A. *Mercury 4.0* : From Visualization to Analysis, Design and Prediction. *J. Appl. Crystallogr.* **2020**, *53* (1), 226–235. https://doi.org/10.1107/S1600576719014092.

(106) Nauha, E.; Bernstein, J. "Predicting" Polymorphs of Pharmaceuticals Using Hydrogen Bond Propensities: Probenecid and Its Two Single-Crystal-to-Single-Crystal Phase Transitions. *J. Pharm. Sci.* **2015**, *104* (6), 2056–2061. https://doi.org/10.1002/jps.24449.

(107) Tang, S. K.; Davey, R. J.; Sacchi, P.; Cruz-Cabeza, A. J. Can Molecular Flexibility Control Crystallization? The Case of *Para* Substituted Benzoic Acids. *Chem. Sci.* **2021**, *12* (3), 993–1000. https://doi.org/10.1039/D0SC05424K.

Appendix A

PIXEL calculations

Pixpar.par file

```
0.00  3.000   150.0   4800.0   1200.0  0
   4 0 0.000000  0.0
   4 0 0.000000  0.0
 3.0  0
```

Input.inp file

```
#BRBZAP01 'P 21/n'
   1   0  15   0
   0.000
   1  12  0.0000  0.0000    C
   2  12  0.0000  0.0000    C
   3  12  0.0000  0.0000    C
   4  12  0.0000  0.0000    C
   5  12  0.0000  0.0000    C
   6  12  0.0000  0.0000    C
   7   2  0.0000  0.0000    H
   8   2  0.0000  0.0000    H
   9   2  0.0000  0.0000    H
  10   2  0.0000  0.0000    H
  11  43  0.0000  0.0000   Br
  12  10  0.0000  0.0000    C
  13  27  0.0000  0.0000    O
  14  28  0.0000  0.0000    O
  15   6  0.0000  0.0000    H
     0.000    18.000
```

```
   3.8539    6.0370   29.3380   90.0000   93.1620   90.0000
 1.000  0.000  0.000  0.000  1.000  0.000  0.000  0.000  1.000
 0.000  0.000  0.000
   0.885979 -0.385026  0.258449
   0.461314  0.675033 -0.575778
   0.047228  0.629353  0.775683
  -0.279853    6.093112   10.516928
  4
  1.00   0.00   0.00   0.00   1.00   0.00   0.00   0.00   1.00
 0.000000  0.000000  0.000000
  1.00   0.00   0.00   0.00  -1.00   0.00   0.00   0.00   1.00
 0.500000  0.500000  0.500000
 -1.00   0.00   0.00   0.00  -1.00   0.00   0.00   0.00  -1.00
 0.000000  0.000000  0.000000
 -1.00   0.00   0.00   0.00   1.00   0.00   0.00   0.00  -1.00
-0.500000 -0.500000 -0.500000
```

Gaussian calculations

ESP fitting and charge density

Input.gjf file

```
#MP2/6-31G** guess=core nosym density=MP2 pop=esp cube=cards cube=frozencore

 #BENZAC02 'P 21/c'   fragm   1

 0 1
 C   -0.000127   -0.004966    0.176928
 C    0.007273    1.200758   -0.507189
 C    0.007904    1.198267   -1.890537
 C   -0.002003    0.023434   -2.586193
 C   -0.007616   -1.177703   -1.906811
 C   -0.004263   -1.193650   -0.529042
 H    0.012518    2.134533    0.035492
 H    0.016317    2.136016   -2.426264
 H   -0.005398    0.034243   -3.666061
 H   -0.014716   -2.107479   -2.456373
 H   -0.004906   -2.135358   -0.000303
 C   -0.001162   -0.026847    1.661119
 O   -0.013013    1.067403    2.282855
 O    0.013671   -1.155658    2.236837
 H   -0.014331    1.109134    3.281940

/path/BENZAC02a.den
    0   -3.240000   -4.760000   -5.880000
```

```
  80    0.080000    0.000000    0.000000
 120    0.000000    0.080000    0.000000
 140    0.000000    0.000000    0.080000
```

Energy of molecular aggregates

Dimer.gjf file

```
%chk=benzac02.chk
%nprocshared=2
# B3LYP/6-311G** nosym opt freq EmpiricalDispersion=GD3

 #BENZAC02   fragm   1

 0 1
C101    7.49785    -4.77529   7.42846
C102    6.43931    -5.69229   7.57649
C103    6.77447    -7.04298   7.48023
C104    8.09360    -7.48119   7.37563
C105    9.06028    -6.58810   7.02398
C106    8.74933    -5.22934   7.03416
H107    5.38954    -5.34439   7.61630
H108    5.98765    -7.73941   7.74023
H109    8.38957    -8.48056   7.32178
H110   10.14096    -6.88991   7.29569
H111    9.58205    -4.58644   7.04481
C112    7.21472    -3.38576   7.23848
O113    6.01614    -2.95843   7.26911
O114    8.21902    -2.59842   7.25352
H115    5.89808    -1.92096   7.17779
C201    6.99407     1.79612   7.19090
C202    8.03447     2.65054   7.27164
C203    7.74405     3.95816   7.53106
C204    6.49013     4.40449   7.75204
C205    5.42392     3.48193   7.71823
C206    5.66799     2.22833   7.39474
H207    8.99114     2.25393   7.42972
H208    8.53541     4.67016   7.46399
H209    6.24934     5.45254   7.84691
H210    4.45322     3.84183   7.91357
H211    4.88750     1.52355   7.61320
C212    7.35505     0.30089   7.25431
O213    8.58458    -0.03111   7.43847
O214    6.37137    -0.46762   7.17989
H215    8.78566    -1.06340   7.53559
```

BSSE correction via counterpoise method

Dimer.gjf file

```
%chk=benzac02.chk
%nprocshared=2
# B3LYP/6-311G** nosym EmpiricalDispersion=GD3 counterpoise=2

 #BENZAC02   fragm   1

 0 1 0 1 0 1
C101(fragment=1)   7.648419   -4.904476    7.336437
C102(fragment=1)   6.552748   -5.770580    7.422834
C103(fragment=1)   6.753838   -7.146457    7.387545
C104(fragment=1)   8.044268   -7.661263    7.266371
C105(fragment=1)   9.137740   -6.799090    7.180178
C106(fragment=1)   8.941791   -5.423603    7.214983
H107(fragment=1)   5.556676   -5.358589    7.516460
H108(fragment=1)   5.905422   -7.817629    7.454387
H109(fragment=1)   8.197938   -8.734296    7.239092
H110(fragment=1)  10.140144   -7.200366    7.086058
H111(fragment=1)   9.774826   -4.735164    7.149618
C112(fragment=1)   7.473184   -3.428246    7.370385
O113(fragment=1)   6.221025   -3.029302    7.487343
O114(fragment=1)   8.429303   -2.658624    7.294312
H115(fragment=1)   6.175085   -2.029664    7.503226
C201(fragment=2)   6.886214    1.856959    7.490826
C202(fragment=2)   7.982140    2.722840    7.405829
C203(fragment=2)   7.781307    4.098777    7.441261
C204(fragment=2)   6.490880    4.613805    7.561230
C205(fragment=2)   5.397123    3.751824    7.646074
C206(fragment=2)   5.592804    2.376330    7.611077
H207(fragment=2)   8.978238    2.310682    7.313170
H208(fragment=2)   8.629952    4.769759    7.375427
H209(fragment=2)   6.337365    5.686857    7.588644
H210(fragment=2)   4.394718    4.153315    7.739313
H211(fragment=2)   4.759572    1.688021    7.675283
C212(fragment=2)   7.061205    0.380744    7.456600
O213(fragment=2)   8.313395   -0.018589    7.341227
O214(fragment=2)   6.104884   -0.388800    7.531059
H215(fragment=2)   8.358267   -1.018303    7.324672
```

MiCMoS MD simulations

Topology files

Benzac02.top

```
#BENZAC02 'P 21/c'   topology  MD LJC
   15
    1  -0.00013  -0.00497   0.17693  12    0.1028
    2   0.00727   1.20076  -0.50719  12   -0.1934
    3   0.00790   1.19827  -1.89054  12   -0.0681
    4  -0.00200   0.02343  -2.58619  12   -0.1361
    5  -0.00762  -1.17770  -1.90681  12   -0.0449
    6  -0.00426  -1.19365  -0.52904  12   -0.2206
    7   0.01252   2.13453   0.03549   2    0.1289
    8   0.01632   2.13602  -2.42626   2    0.1074
    9  -0.00540   0.03424  -3.66606   2    0.1134
   10  -0.01472  -2.10748  -2.45637   2    0.1055
   11  -0.00491  -2.13536  -0.00030   2    0.1482
   12  -0.00116  -0.02685   1.66112  10    0.6016
   13  -0.01301   1.06740   2.28285  27   -0.5857
   14   0.01367  -1.15566   2.23684  27   -0.5303
   15  -0.01433   1.10913   3.28194   6    0.4714
    0  nslav-u
    0  ncore-v
    0  nslav-v
  112.0    0.0 volu-u,volu-v
   15  nstr-u
    1    2   1.386  4958.6   C- C
    1    6   1.383  5038.1   C- C
    1   12   1.484  2894.0   C- C
    2    3   1.383  5020.8   C- C
    2    7   1.080  3600.0   C- H
    3    4   1.365  5399.2   C- C
    3    8   1.080  3600.0   C- H
    4    5   1.380  5091.9   C- C
    4    9   1.080  3600.0   C- H
    5    6   1.378  5136.3   C- C
    5   10   1.080  3600.0   C- H
    6   11   1.080  3600.0   C- H
   12   13   1.259  7254.9   C- O
   12   14   1.267  7061.6   C- O
   13   15   1.000  4250.0   O- H
    0  nstr-v
   22  nbend-u
    1    2    3  119.00   576.7   C- C- C
    1    2    7  120.00   505.0   C- C- H
    1    6    5  120.00   583.8   C- C- C
    1    6   11  120.00   505.0   C- C- H
```

```
   1   12   13  119.00   655.6   C- C- O
   1   12   14  118.00   637.5   C- C- O
   2    1    6  120.00   583.8   C- C- C
   2    1   12  120.00   583.8   C- C- C
   2    3    4  121.00   590.8   C- C- C
   2    3    8  120.00   505.0   C- C- H
   3    2    7  120.00   505.0   C- C- H
   3    4    5  120.00   583.8   C- C- C
   3    4    9  120.00   505.0   C- C- H
   4    3    8  120.00   505.0   C- C- H
   4    5    6  120.00   583.8   C- C- C
   4    5   10  120.00   505.0   C- C- H
   5    4    9  120.00   505.0   C- C- H
   5    6   11  120.00   505.0   C- C- H
   6    1   12  120.00   583.8   C- C- C
   6    5   10  120.00   505.0   C- C- H
  12   13   15  122.00   450.0   C- O- H
  13   12   14  123.00   727.7   O- C- O
   0  nbend-v
  15  ntors-u
  6   1   2   3  50.00   -1.0    1.0   C- C- C- C
  2   1   6   5  50.00   -1.0    1.0   C- C- C- C
  2   1  12  13  14.00   -1.0    2.0   C- C- C- O
  1   2   6  12 100.00   -1.0    1.0   C- C- C- C
  1   2   3   4  50.00   -1.0    1.0   C- C- C- C
  2   1   3   7 100.00   -1.0    1.0   C- C- C- H
  2   3   4   5  50.00   -1.0    1.0   C- C- C- C
  3   2   4   8 100.00   -1.0    1.0   C- C- C- H
  3   4   5   6  50.00   -1.0    1.0   C- C- C- C
  4   3   5   9 100.00   -1.0    1.0   C- C- C- H
  4   5   6   1  50.00   -1.0    1.0   C- C- C- C
  5   4   6  10 100.00   -1.0    1.0   C- C- C- H
  6   1   5  11 100.00   -1.0    1.0   C- C- C- H
  1  12  13  15  35.00   -1.0    2.0   C- C- O- H
 12   1  13  14 100.00   -1.0    1.0   C- C- O- O
   0  ntors-v
   2  nlist-u
 7 15 11 15
   0  nlist-v
  0.410   235.0   650.0   77000.0
   0  nextr
```

Brbzap01.top

```
#BRBZAP01 'P 21/n'   topology
  15
   1  -0.00363  -0.02886   1.93751  12    0.1697
   2   0.03965   1.17714   1.24716  12   -0.2771
   3   0.04779   1.19314  -0.14094  12    0.1093
   4   0.00518  -0.02271  -0.81729  12   -0.2139
   5  -0.05302  -1.23799  -0.14872  12    0.1307
```

```
  6  -0.04319  -1.24075   1.23888  12   -0.3050
  7   0.06711   2.10776   1.79444   2    0.1545
  8   0.08624   2.12647  -0.68302   2    0.0874
  9  -0.10508  -2.16556  -0.69924   2    0.0849
 10  -0.06602  -2.17594   1.77850   2    0.1668
 11   0.00066   0.01314  -2.71545  43   -0.0526
 12   0.00444  -0.04786   3.42524  10    0.5599
 13   0.09459  -1.10010   4.06201  27   -0.4945
 14  -0.09493   1.12996   3.98625  28   -0.5641
 15   0.00398   1.06847   4.97952   6    0.4441
  0  nslav-u
  0  ncore-v
  0  nslav-v
132.0    0.0 volu-u,volu-v
 15  nstr-u
  1    2   1.390  4874.6   C- C
  1    6   1.399  4683.0   C- C
  1   12   1.488  2819.7   C- C
  2    3   1.388  4918.3   C- C
  2    7   1.080  3600.0   C- H
  3    4   1.392  4839.7   C- C
  3    8   1.080  3600.0   C- H
  4    5   1.388  4917.3   C- C
  4   11   1.899  2000.0   C-Br
  5    6   1.388  4930.5   C- C
  5    9   1.080  3600.0   C- H
  6   10   1.080  3600.0   C- H
 12   13   1.233  7823.4   C- O
 12   14   1.308  6140.5   C- O
 14   15   1.000  4250.0   O- H
  0  nstr-v
 22  nbend-u
  1    2    3  120.00   583.8   C- C- C
  1    2    7  120.00   505.0   C- C- H
  1    6    5  120.00   583.8   C- C- C
  1    6   10  120.00   505.0   C- C- H
  1   12   13  122.00   709.7   C- C- O
  1   12   14  115.00   583.5   C- C- O
  2    1    6  120.00   583.8   C- C- C
  2    1   12  120.00   583.8   C- C- C
  2    3    4  118.00   569.7   C- C- C
  2    3    8  121.00   492.5   C- C- H
  3    2    7  120.00   505.0   C- C- H
  3    4    5  122.00   597.8   C- C- C
  3    4   11  118.00   601.0   C- C-Br
  4    3    8  121.00   492.5   C- C- H
  4    5    6  119.00   576.7   C- C- C
  4    5    9  121.00   492.5   C- C- H
  5    4   11  120.00   601.0   C- C-Br
  5    6   10  120.00   505.0   C- C- H
  6    1   12  119.00   576.7   C- C- C
```

```
   6    5    9  121.00   492.5   C- C- H
  12   14   15  111.00   450.0   C- O- H
  13   12   14  124.00   745.7   O- C- O
   0  nbend-v
  15  ntors-u
   6   1   2   3  50.00   -1.0    1.0   C- C- C- C
   2   1   6   5  50.00   -1.0    1.0   C- C- C- C
   2   1  12  13  14.00   -1.0    2.0   C- C- C- O
   1   2   6  12 100.00   -1.0    1.0   C- C- C- C
   1   2   3   4  50.00   -1.0    1.0   C- C- C- C
   2   1   3   7 100.00   -1.0    1.0   C- C- C- H
   2   3   4   5  50.00   -1.0    1.0   C- C- C- C
   3   2   4   8 100.00   -1.0    1.0   C- C- C- H
   3   4   5   6  50.00   -1.0    1.0   C- C- C- C
   4   3   5  11 100.00   -1.0    1.0   C- C- C-Br
   4   5   6   1  50.00   -1.0    1.0   C- C- C- C
   5   4   6   9 100.00   -1.0    1.0   C- C- C- H
   6   1   5  10 100.00   -1.0    1.0   C- C- C- H
  13  12  14  15  35.00   -1.0    2.0   C- C- O- H
  12   1  13  14 100.00   -1.0    1.0   C- C- O- O
   0  ntors-v
   2  nlist-u
 15 7   15 10
   0  nlist-v
  0.410   235.0   650.0   77000.0
   0  nextr
```

Qiyruo.top

```
#QIYRUO   'P21/c'    topology
  21
   1  -0.08081  -0.16715  -0.99018  12   -0.1920
   2  -0.02501  -1.31359  -0.22065  12   -0.0057
   3   0.01386  -1.22975   1.15935  12   -0.2090
   4  -0.02103   0.01106   1.78205  12    0.0887
   5  -0.10177   1.15930   1.01241  12   -0.1592
   6  -0.12752   1.07253  -0.36766  12   -0.0697
   7   0.03950   0.09699   3.26606  10    0.5746
   8  -0.05293  -0.31208  -2.47918  10    0.7338
   9  -0.01319  -2.28910  -0.70300   2    0.1007
  10   0.07358  -2.13821   1.75726   2    0.1525
  11  -0.14462   2.13474   1.49539   2    0.1333
  12  -0.18557   1.97951  -0.96746   2    0.0996
  13  -0.10909  -1.37163  -3.05095  27   -0.5060
  14   0.07898   0.84487  -3.10544  23   -0.3243
  15   0.18596   0.78220  -4.55578  13   -0.1456
  16   0.28997   1.78941  -4.95654   3    0.1262
  17  -0.70981   0.31857  -4.96648   3    0.1091
  18   1.05778   0.19152  -4.82995   3    0.1107
  19   0.14350  -0.97884   3.92596  27   -0.5230
  20  -0.01161   1.23422   3.80436  27   -0.5432
```

```
 21   0.03895   1.13600   4.79223   6    0.4484
  0  nslav-u
  0  ncore-v
  0  nslav-v
158.9    0.0 volu-u,volu-v
 21  nstr-u
  1    2   1.382  5051.6   C- C
  1    6   1.388  4923.1   C- C
  1    8   1.496  2642.4   C- C
  2    3   1.383  5026.3   C- C
  2    9   1.080  3600.0   C- H
  3    4   1.389  4907.4   C- C
  3   10   1.080  3600.0   C- H
  4    5   1.385  4992.9   C- C
  4    7   1.488  2822.9   C- C
  5    6   1.383  5027.6   C- C
  5   11   1.080  3600.0   C- H
  6   12   1.080  3600.0   C- H
  7   19   1.266  7080.9   C- O
  7   20   1.259  7240.8   C- O
  8   13   1.205  8448.2   C- O
  8   14   1.322  5832.1   C- O
 14   15   1.456  2844.1   O- C
 15   16   1.080  3600.0   C- H
 15   17   1.080  3600.0   C- H
 15   18   1.080  3600.0   C- H
 20   21   1.000  4250.0   O- H
  0  nstr-v
 32  nbend-u
  1    2    3  120.00   583.8   C- C- C
  1    2    9  120.00   505.0   C- C- H
  1    6    5  120.00   583.8   C- C- C
  1    6   12  120.00   505.0   C- C- H
  1    8   13  124.00   745.7   C- C- O
  1    8   14  113.00   547.4   C- C- O
  2    1    6  120.00   583.8   C- C- C
  2    1    8  118.00   569.7   C- C- C
  2    3    4  120.00   583.8   C- C- C
  2    3   10  120.00   505.0   C- C- H
  3    2    9  120.00   505.0   C- C- H
  3    4    5  120.00   583.8   C- C- C
  3    4    7  120.00   583.8   C- C- C
  4    3   10  120.00   505.0   C- C- H
  4    5    6  120.00   583.8   C- C- C
  4    5   11  120.00   505.0   C- C- H
  4    7   19  118.00   637.5   C- C- O
  4    7   20  118.00   637.5   C- C- O
  5    4    7  121.00   590.8   C- C- C
  5    6   12  120.00   505.0   C- C- H
  6    1    8  122.00   597.8   C- C- C
  6    5   11  120.00   505.0   C- C- H
```

```
    7   20   21  109.00   450.0   C- O- H
    8   14   15  116.00   601.5   C- O- C
   13    8   14  123.00   727.7   O- C- O
   14   15   16  110.00   630.0   O- C- H
   14   15   17  109.00   642.5   O- C- H
   14   15   18  109.00   642.5   O- C- H
   16   15   17  109.00   470.0   H- C- H
   16   15   18  109.00   470.0   H- C- H
   17   15   18  109.00   470.0   H- C- H
   19    7   20  123.00   727.7   O- C- O
    0  nbend-v
   19  ntors-u
   6   1   2   3  50.00   -1.0    1.0   C- C- C- C
   2   1   6   5  50.00   -1.0    1.0   C- C- C- C
   2   1   8  13  14.00   -1.0    2.0   C- C- C- O
   1   2   6   8 100.00   -1.0    1.0   C- C- C- C
   1   2   3   4  50.00   -1.0    1.0   C- C- C- C
   2   1   3   9 100.00   -1.0    1.0   C- C- C- H
   2   3   4   5  50.00   -1.0    1.0   C- C- C- C
   3   2   4  10 100.00   -1.0    1.0   C- C- C- H
   3   4   5   6  50.00   -1.0    1.0   C- C- C- C
   3   4   7  19  14.00   -1.0    2.0   C- C- C- O
   4   3   5   7 100.00   -1.0    1.0   C- C- C- C
   4   5   6   1  50.00   -1.0    1.0   C- C- C- C
   5   4   6  11 100.00   -1.0    1.0   C- C- C- H
   6   1   5  12 100.00   -1.0    1.0   C- C- C- H
  19   7  20  21  35.00   -1.0    2.0   C- C- O- H
   7   4  19  20 100.00   -1.0    1.0   C- C- O- O
   1   8  14  15  40.00    1.0    1.0   C- C- O- C
   8   1  13  14 100.00   -1.0    1.0   C- C- O- O
   8  14  15  16   2.50    1.0    3.0   C- O- C- H
    0  ntors-v
   16  nlist-u
 21 10   21 11   15 12   16 12   17 12   18 12   15  9   16  9   17  9   18
9   16  6   17  6   18  6   16  2   17  2   18  2
    0  nlist-v
   0.410   235.0   650.0   77000.0
    0  nextr
```

Bulk liquids parameter files

Input.mdi file

```
Benzoic acid ljc 425 K liq unbiased
# n.steps irvel ipri ibox idstr timestep Emolim iengt ibias + Ebias Nbias
  2500000    2     0      1    0     0.001   -5.0    0     0
# cutoffu cutoffv cutoffuv factin  ipots ianh inano
   16.0    0.0     0.0       0.7      1    0    0
#  N(T)   Tset Tstart Trelax  0/1 weak/stiff
    100    425   425    0.6      2
```

```
#  N(P)   Pset  comprs  0/1ianis ipr  ww   iextstr + stra11 22, 33, 12 13
23, GPa
     50    1.0   0.4      0      0    0.0    0
#  N(com)  nwbox   nwre  npri
    100     500    500    500
```

Confined liquids parameter files

Barrier.par file

```
# iplane(XY), iplane(XZ), iplane(YZ), iattr
          1         0         0         0
# ispbar,    rvdw,    qqbar,    offset
    12    1.7700      0.000000    6.0000
# dampk(XY),  dampk(XZ),  dampk(YZ
)  0.025000  0.025000  0.025000
# zacsize,      nmolzacu,   nmolzacv(0=default)
    0.000000         0         0
```

Input.mdi file

```
Benzoic acid ljc 350 K liq unbiased nanolayer
# n.steps irvel ipri ibox idstr timestep Emolim iengt ibias + Ebias Nbias
 500000     0      0      1    1     0.001   -5.0    0     0
# cutoffu cutoffv cutoffuv factin  ipots ianh inano
   16.0    0.0     0.0       0.7      1    0    1
#  N(T)   Tset Tstart Trelax  0/1 weak/stiff
    100    350   350    0.6      0
#  N(P) Pset comprs 0/1ianis ipr ww  iextstr+stra11 22,33,12,13,23,GPa
     50    1.0   0.4      1      0    0.0    0
#  N(com)  nwbox   nwre  npri
    100     500    500    500
```

Gromacs MD simulations

Topology file

Topology.top file

```
; Include forcefield parameters
#include "./gromos54a7_atb.ff/forcefield.itp"

[ moleculetype ]
; Name   nrexcl
6YZ7     3
[ atoms ]
```

```
;  nr  type  resnr  resid  atom  cgnr  charge    mass
    1  CAro    1    6YZ7     C1    1   -0.010  12.0110
    2  CAro    1    6YZ7     C2    2   -0.133  12.0110
    3  CAro    1    6YZ7     C3    3   -0.109  12.0110
    4  CAro    1    6YZ7     C4    4   -0.109  12.0110
    5  CAro    1    6YZ7     C5    5   -0.109  12.0110
    6  CAro    1    6YZ7     C6    6   -0.133  12.0110
    7    HC    1    6YZ7     H7    7    0.135   1.0080
    8    HC    1    6YZ7     H8    8    0.131   1.0080
    9    HC    1    6YZ7     H9    9    0.134   1.0080
   10    HC    1    6YZ7    H10   10    0.131   1.0080
   11    HC    1    6YZ7    H11   11    0.135   1.0080
   12  CPos    1    6YZ7    C12   12    0.621  12.0110
   13    OA    1    6YZ7    O13   13   -0.582  15.9994
   14 OEOpt    1    6YZ7    O14   14   -0.556  15.9994
   15  HS14    1    6YZ7    H15   15    0.454   1.0080
; total charge of the molecule:   0.000
[ bonds ]
;  ai   aj  funct   c0         c1
   15   13    2   0.0972   1.9581e+07
   13   12    2   0.1360   1.0200e+07
   12   14    2   0.1230   1.6600e+07
   12    1    2   0.1480   5.7300e+06
    1    6    2   0.1400   8.5400e+06
    1    2    2   0.1400   8.5400e+06
    6   11    2   0.1090   1.2300e+07
    6    5    2   0.1390   8.6600e+06
    5   10    2   0.1090   1.2300e+07
    5    4    2   0.1390   8.6600e+06
    4    9    2   0.1090   1.2300e+07
    4    3    2   0.1390   8.6600e+06
    3    8    2   0.1090   1.2300e+07
    3    2    2   0.1390   8.6600e+06
    2    7    2   0.1090   1.2300e+07
[ pairs ]
;  ai   aj  funct  ;  all 1-4 pairs but the ones excluded in GROMOS it
   15   14    1
   15    1    1
   13    6    1
   13    2    1
   12   11    1
   12    5    1
   12    3    1
   12    7    1
   14    6    1
   14    2    1
    1   10    1
    1    8    1
    6    9    1
    6    7    1
   11   10    1
```

```
   11    4    1
   11    2    1
    5    8    1
   10    9    1
   10    3    1
    4    7    1
    9    8    1
    9    2    1
    8    7    1
[ angles ]
;  ai   aj   ak  funct   angle     fc
   15   13   12    2    109.50   450.00
   13   12   14    2    124.00   730.00
   13   12    1    2    115.00   610.00
   14   12    1    2    121.00   685.00
   12    1    6    2    120.00   560.00
   12    1    2    2    120.00   560.00
    6    1    2    2    120.00   560.00
    1    6   11    2    120.00   505.00
    1    6    5    2    120.00   560.00
   11    6    5    2    120.00   505.00
    6    5   10    2    120.00   505.00
    6    5    4    2    120.00   560.00
   10    5    4    2    120.00   505.00
    5    4    9    2    120.00   505.00
    5    4    3    2    120.00   560.00
    9    4    3    2    120.00   505.00
    4    3    8    2    120.00   505.00
    4    3    2    2    120.00   560.00
    8    3    2    2    120.00   505.00
    1    2    3    2    120.00   560.00
    1    2    7    2    120.00   505.00
    3    2    7    2    120.00   505.00
[ dihedrals ]
; GROMOS improper dihedrals
;  ai   aj   ak   al  funct   angle     fc
    5    6   10    4    2      0.00   167.36
    6    1   11    5    2      0.00   167.36
    1   12    6    2    2      0.00   167.36
    2    1    3    7    2      0.00   167.36
    3    4    8    2    2      0.00   167.36
    4    5    9    3    2      0.00   167.36
   12   13   14    1    2      0.00   167.36
[ dihedrals ]
;  ai   aj   ak   al  funct    ph0      cp     mult
   15   13   12    1    1    180.00    16.70    2
   14   12    1    6    1    180.00     5.86    2
    1    6    5    4    1    180.00    41.80    2
    6    1    2    3    1    180.00    41.80    2
    6    5    4    3    1    180.00    41.80    2
    5    4    3    2    1    180.00    41.80    2
```

```
    4    3    2    1    1    180.00    41.80    2
    2    1    6    5    1    180.00    41.80    2
[ exclusions ]
;  ai   aj  funct  ;  GROMOS 1-4 exclusions
    1    4
    6    3
    5    2

; Include Position restraint file
#ifdef POSRES
#include "posre.itp"
#endif

[ system ]
; Name
Benzoic acid, liquid

[ molecules ]
; Compound        #mols
6YZ7     432
```

Parameter file

Parameter.mdp file

```
integrator      = md            ; leap-frog integrator
nsteps          =   1500000 ;   2*1500000 = 3000000 fs (3000 ps = 3 ns)
dt                  = 0.002     ; 2 fs
; Output control
nstxout                 = 1500          ; save coordinates every 3.0 ps
nstvout                 = 1500          ; save velocities every 3.0 ps
nstenergy               = 1500          ; save energies every 3.0 ps
nstlog                  = 1500          ; update log file every 3.0 ps
nstxout-compressed  = 1500      ; save compressed coordinates every 3.0 ps
                                ; nstxout-compressed replaces nstxtcout
compressed-x-grps   = System    ; replaces xtc-grps
; Bond parameters
continuation            = no    ; Restarting after NPT
constraint_algorithm    = lincs     ; holonomic constraints
constraints                 = all-bonds ; all bonds (even heavy atom-H
bonds) constrained
lincs_iter                  = 1             ; accuracy of LINCS
lincs_order                 = 4             ; also related to accuracy
; Neighborsearching
cutoff-scheme   = Verlet
ns_type             = grid              ; search neighboring grid cells
nstlist             = 10            ; 20 fs, largely irrelevant with Verlet
scheme
rcoulomb            = 1.6               ; short-range electrostatic cutoff
(in nm)
```

```
rvdw                = 1.6               ; short-range van der Waals cutoff
(in nm)
; Electrostatics
coulombtype         = PME               ; Particle Mesh Ewald for long-range
electrostatics
pme_order           = 4             ; cubic interpolation
fourierspacing  = 0.16          ; grid spacing for FFT
; Temperature coupling is on
tcoupl                  = V-rescale             ; modified Berendsen
thermostat
tc-grps                 = 6YZ7
tau_t                   = 0.1
ref_t                   = 350       ; reference temperature, one for each
group, in K
; Pressure coupling is on
pcoupl                  = Berendsen             ; Pressure coupling on in
NPT
pcoupltype              = isotropic                 ; uniform scaling of box
vectors
tau_p                   = 2.0            ; time constant, in ps
ref_p                   = 1.0                       ; reference pressure, in
bar
compressibility     = 2.0e-4            ; isothermal compressibility just
guessed
; Periodic boundary conditions
pbc             = xyz           ; 3-D PBC
; Dispersion correction
DispCorr        = EnerPres      ; account for cut-off vdW scheme
; Velocity generation
gen_vel         = yes   ; Velocity generation is on
gen_temp        = 350           ; temperature for Maxwell distribution
gen_seed        = -1            ; generate a random seed
```

Appendix B

PIXEL calculations

Table B1. Selection of 17 structures ranked according to PIXEL stacking energy, having space group number 14, max 25 atoms, 1 aromatic ring, 1 carboxylic group, stacked molecules related by translation symmetry only.

ID	refcode	Benzoic acid name	*a* / Å	*b* / Å	*c* / Å	*β* / °	*d* com / Å	*E* stack / kJ mol^{-1}
3	COVJIG01	3-Fluoro-	6.747	3.719	24.285	92.9	3.719	-11.5
4	PFBZAD15	4-Fluoro-	3.704	6.264	26.660	93.0	3.704	-11.7
5	BENZAC12	Benzoic acid	5.415	5.039	21.630	96.1	5.039	-11.8
8	ANISIC04	4-Methoxy-	3.876	10.930	16.620	93.4	3.876	-12.6
10	PIDGOZ	3,5-Difluoro-	3.769	13.400	14.041	93.8	3.769	-13.4
11	QUYREI	3-Acetyl-	3.820	15.648	12.928	98.5	3.820	-13.5
13	EFINEO02	3-methoxy-	3.802	15.603	11.976	90.9	3.802	-14.0
19	BRBZAP01	p-Bromo-	3.854	6.037	29.338	93.2	3.854	-15.9
24	EFUZUD	3-(trifluoromethoxy)-	4.984	5.264	30.415	90.7	5.264	-16.2
25	WUWPEM	4-amino-3-chloro-	3.723	11.163	16.945	95.2	3.723	-16.5
35	QIYRUO	4-(methoxycarbonyl)-	4.854	5.880	29.194	91.4	5.880	-18.5
39	COWYET	3,5-Dihydroxy-4-nitro-	6.837	4.963	22.565	93.9	4.963	-19.1
56	VITDEL01	3-(azidomethyl)-	3.771	6.123	34.868	93.1	3.771	-21.6
59	HEKMOZ	3,5-Dimethoxy-	10.879	4.924	16.388	104.4	4.924	-22.0
71	FISSOS	3-methoxy-4-[(prop-2-yn-1-yl)oxy]-	4.940	26.515	7.463	104.1	4.940	-25.3
109	GEFVEV	4-(1H-1,2,3-triazol-1-yl)-	4.806	5.470	31.794	90.7	5.470	-31.1
145	TIJXAN	3-(Phosphonomethyl)-	5.476	5.215	33.117	94.1	5.215	-41.7

Table B2. List of all 191 structures of benzoic acid derivatives ranked according to PIXEL stacking energy. Progressive ID number, refcode, space group (SG) number and symmetry, number of atoms, aromatic rings, and carboxylic (COOH) groups, symmetry relating stacked molecules and stacking energy are reported.

ID	refcode	SG *n*	SG sym	*n* atoms	*n* arom	*n* COOH	Stack sym	*E* stack / kJ mol^{-1}
1	KUNDAZ	14	P21/n	27	1	1	inv	-8.1
2	DAKWAP	60	Pbcn	20	1	2	transl	-8.6
3	COVJIG01	14	P21/n	15	1	1	transl	-11.5
4	PFBZAD15	14	P21/c	15	1	1	transl	-11.7
5	BENZAC12	14	P21/n	15	1	1	transl	-11.8
6	BEBTIP	15	C2/c	27	1	1	inv	-11.9
7	TARXET	14	P21/c	29	1	1	inv	-12.2
8	ANISIC04	14	P21/n	19	1	1	transl	-12.6
9	IJUMEG	15	C2/c	18	1	1	transl	-13.3
10	PIDGOZ	14	P21/c	15	1	1	transl	-13.4
11	QUYREI	14	P21/c	20	1	1	transl	-13.5
12	TUPNOI	14	P21/c	37	2	2	inv	-13.5
13	EFINEO02	14	P21/c	19	1	1	transl	-14.0
14	BENZDC11	14	P21/c	18	1	2	transl	-14.1
15	MULWIC	14	P21/n	30	1	1	inv	-14.1
16	TEJMAW	15	C2/c	25	2	1	transl	-14.2
17	NUPTAV	2	P-1	17	1	1	transl	-15.6
18	SAJFIT	14	P21/c	43	3	1	inv	-15.8
19	BRBZAP01	14	P21/n	15	1	1	transl	-15.9
20	ULEKOM	11	P21/m	18	1	2	transl	-15.9
21	NBZOAC04	15	A2/a	17	1	1	transl	-16.0
22	MEHCIN	14	P21/n	36	1	1	transl	-16.1
23	MOSCOP	15	C2/c	17	1	1	transl	-16.1
24	EFUZUD	14	P21/n	19	1	1	transl	-16.2
25	WUWPEM	14	P21/c	17	1	1	transl	-16.5
26	CLBZAP10	2	P-1	15	1	1	transl	-16.6
27	GOVZEW01	2	P-1	17	1	1	transl	-16.8
28	JOZZIH01	14	P21/c	16	1	1	inv	-16.9
29	KUNDED	2	P-1	29	1	1	inv	-17.1
30	MBBNZA01	14	P21/n	15	1	1	inv	-17.2
31	TEPHTH15	2	P-1	18	1	2	transl	-17.3
32	PUVLOK	14	P21/c	21	1	1	inv	-17.5
33	MCBZAC03	14	P21/c	15	1	1	glide	-17.6
34	VUNKEU	2	P-1	17	1	1	inv	-17.6
35	QIYRUO	14	P21/c	21	1	1	transl	-18.5
36	VODWEQ	14	P21/c	19	1	1	inv	-18.5
37	IMOJOM	14	P21/c	50	3	1	inv	-18.7
38	XUMZIP	14	P21/n	53	2	1	inv	-18.9
39	COWYET	14	P21/c	19	1	1	transl	-19.1
40	DMXBZA01	2	P-1	23	1	1	transl	-19.7
41	PTOLIC01	2	P-1	18	1	1	inv	-19.7
42	XIYKUO	14	P21/n	17	1	1	inv	-20.1
43	PEXBZA01	15	I2/a	22	1	1	inv	-20.3
44	ZZZMVY01	14	P21/c	17	1	1	inv	-20.4
45	PAMNUO	14	P21/n	48	2	2	inv	-20.6
46	SERZOI	2	P-1	24	1	1	transl	-20.7

ID	refcode	SG *n*	SG sym	*n* atoms	*n* arom	*n* COOH	Stack sym	*E* stack / kJ mol^{-1}
47	TEGTAD	14	P21/n	50	2	1	inv	-20.7
48	MATBER	14	P21/n	17	1	1	inv	-20.9
49	QAJPAV	15	C2/c	28	2	2	transl	-21.0
50	DAQJIR	15	C2/c	33	2	1	inv	-21.1
51	NBZOAC16	14	P21/n	17	1	1	inv	-21.3
52	BOLBOW	14	P21/n	17	1	1	inv	-21.4
53	BODBUU	2	P-1	25	1	2	inv	-21.5
54	NBZOAC17	14	P21/c	17	1	1	inv	-21.5
55	VIHPOT	15	C2/c	45	3	2	inv	-21.6
56	VITDEL01	14	P21/n	20	1	1	transl	-21.6
57	ZEMGIL	14	P21/c	54	2	1	inv	-21.6
58	JISGAW	14	P21/c	26	1	1	transl	-21.8
59	HEKMOZ	14	P21/c	23	1	1	transl	-22.0
60	KEHRUN	2	P-1	16	1	1	inv	-22.2
61	FIDJIO	14	P21/n	27	1	1	transl	-22.3
62	VITDEL	14	P21/c	20	1	1	inv	-22.5
63	FOXBUQ	2	P-1	26	2	1	transl	-23.3
64	NORDEG	14	P21/c	20	1	1	inv	-23.6
65	COQTOT	15	C2/c	23	1	1	transl	-23.7
66	FOGJOB	2	P-1	29	1	1	transl	-24.1
67	EFINOY	2	P-1	29	1	1	inv	-24.7
68	XOHFAC	14	P21/c	33	2	1	transl	-24.8
69	YUCGUZ	14	P21/c	27	1	1	transl	-25.1
70	MICCIN	14	P21/c	26	1	1	inv	-25.2
71	FISSOS	14	P21/n	25	1	1	transl	-25.3
72	ISAGEP	2	P-1	28	2	2	inv	-25.3
73	COBFUU	2	P-1	21	1	1	transl	-25.4
74	NIDPUP	15	C2/c	28	2	1	transl	-25.4
75	VITDEL02	2	P-1	20	1	1	inv	-25.5
76	LOJRUX	61	Pbca	48	4	1	glide	-26.4
77	ODAQEQ	2	P-1	18	1	1	inv	-26.4
78	AZAKAO	14	P21/c	37	3	1	transl	-26.6
79	TAPSAF01	2	P-1	29	2	2	transl	-26.9
80	XETZAB	14	P21/c	30	1	2	inv	-27.1
81	CUSDOL	15	C2/c	29	2	2	transl	-27.3
82	PUZTUB	15	C2/c	23	1	1	transl	-27.3
83	PEXLIN	14	P21/a	54	2	1	inv	-27.7
84	TONJAK	2	P-1	48	2	1	inv	-27.7
85	MUYBOA	14	P21/n	32	2	1	transl	-27.8
86	IRIBUI	14	P21/c	49	2	1	inv	-27.9
87	JESYEP	14	P21/n	37	3	1	inv	-27.9
88	ZZZDJU11	2	P-1	37	1	1	transl	-27.9
89	BILQOF	2	P-1	40	1	1	inv	-28.0
90	ZZZZCA02	14	P21/c	34	1	1	transl	-28.0
91	LOSDUT	2	P-1	25	1	1	inv	-28.3
92	QISVAT	14	P21/c	33	2	1	transl	-28.4
93	NOMBEA01	14	P21/c	22	1	1	inv	-28.7
94	EVATUR	2	P-1	25	1	1	transl	-28.8
95	CUKCAM15	15	C2/c	19	1	1	inv	-29.0
96	ULUGEN	14	P21/n	29	1	1	transl	-29.0
97	ZZZZCB02	2	P-1	40	1	1	transl	-29.0
98	CUKCAM04	14	P21/c	19	1	1	inv	-29.1
99	UCATAV	14	P21/c	37	3	1	transl	-29.1

ID	refcode	SG *n*	SG sym	*n* atoms	*n* arom	*n* COOH	Stack sym	*E* stack / kJ mol^{-1}
100	CEGFOJ	14	P21/c	44	1	1	transl	-29.3
101	RALDEQ	14	P21/c	29	2	2	transl	-29.6
102	QISWUO	14	P21/n	27	2	1	transl	-29.9
103	RIRRAO	14	P21/n	28	2	1	transl	-30.2
104	QISTIZ	2	P-1	30	2	1	transl	-30.3
105	ROMJIN	14	P21/c	23	1	1	inv	-30.3
106	ZZZDIK10	2	P-1	43	1	1	transl	-30.5
107	CUVDAZ	13	P2/n	30	2	2	transl	-31.0
108	FAYVEI	14	P21/c	51	4	1	inv	-31.1
109	GEFVEV	14	P21/c	21	1	1	transl	-31.1
110	ADAFIU	2	P-1	57	3	1	inv	-31.5
111	NEZXIB	14	P21/c	29	2	1	transl	-31.8
112	ULUGAJ	14	P21/n	35	1	1	transl	-31.9
113	BUDMOD	2	P-1	30	2	2	transl	-32.3
114	DUDGEP	15	C2/c	37	2	2	transl	-32.6
115	ODASAO	14	P21/c	33	2	1	transl	-32.8
116	COCFIM	14	P21/n	20	1	1	inv	-32.9
117	PUQFUD	14	P21/c	19	1	1	inv	-33.3
118	MOTCEE	14	P21/n	26	1	1	inv	-33.6
119	TARWES	14	P21/c	31	1	1	inv	-33.9
120	ZZZDIA11	2	P-1	46	1	1	transl	-34.0
121	JUKBEY	2	P-1	46	2	1	transl	-34.1
122	EREPAU	61	Pbca	35	2	3	inv	-34.3
123	MAQVOP	2	P-1	50	2	1	transl	-34.8
124	QEMHAS	14	P21/c	25	1	2	inv	-35.2
125	QACMOX01	2	P-1	49	1	1	transl	-35.7
126	ISAMUM	15	C2/c	34	2	4	glide	-36.0
127	GOWYID	2	P-1	31	2	1	transl	-36.2
128	RISGOQ	14	P21/c	33	2	1	transl	-36.2
129	CIGDOO	2	P-1	65	1	1	transl	-37.0
130	PAMNOI	14	P21/n	46	2	2	inv	-37.3
131	VORPOI	14	P21/c	30	2	1	transl	-37.5
132	MOPNAH01	14	P21/c	53	2	1	transl	-37.6
133	WAFQEB	2	P-1	31	2	1	transl	-37.8
134	LOSHOR	15	C2/c	36	2	4	screw	-38.3
135	BINRAU	2	P-1	30	1	1	transl	-38.4
136	ITOVAP	14	P21/c	26	1	1	inv	-38.7
137	NUCQOS	2	P-1	57	4	3	inv	-38.7
138	DAQCII	2	P-1	54	2	1	transl	-39.1
139	ODASAO01	2	P-1	33	2	1	transl	-40.1
140	XATWEV	15	C2/c	37	2	1	inv	-40.1
141	GIMVUW	14	P21/c	31	2	1	inv	-40.3
142	TINWIX	14	P21/n	38	2	1	transl	-40.5
143	ZILKUD	61	Pbca	35	2	1	inv	-41.2
144	MISWIY	14	P21/c	27	2	1	inv	-41.4
145	TIJXAN	14	P21/n	23	1	1	transl	-41.7
146	MOVREU	14	P21/n	30	1	1	inv	-41.8
147	YETJIT	2	P-1	36	2	1	transl	-41.9
148	QEPRIO	14	P21/n	49	4	1	inv	-42.1
149	WUTYAM	14	P21/n	54	2	1	inv	-42.1
150	COKLEW	4	P21	46	2	2	transl	-42.3
151	PEYBIE	14	P21/c	50	2	1	inv	-42.4
152	QOYCOZ	15	C2/c	34	2	1	transl	-42.4

ID	refcode	SG *n*	SG sym	*n* atoms	*n* arom	*n* COOH	Stack sym	*E* stack / kJ mol^{-1}
153	YIFNAD	11	P21/m	31	2	2	transl	-42.6
154	COKLUM	4	P21	46	2	2	transl	-42.8
155	MOPNAH03	14	P21/n	53	2	1	transl	-42.8
156	YADXUA	14	P21/c	32	2	1	inv	-42.8
157	YIKYOI	14	P21/n	38	2	1	transl	-42.8
158	YUDROH	14	P21/c	30	1	1	transl	-42.8
159	CEGVUH	14	P21/n	26	1	1	transl	-42.9
160	DIXFAR04	2	P-1	22	1	1	transl	-43.2
161	CEHLUY	15	C2/c	34	2	1	transl	-43.8
162	TAQMAD	2	P-1	33	1	1	inv	-43.8
163	BEVTUU	2	P-1	64	4	1	inv	-43.9
164	BINREY	15	C2/c	30	1	1	transl	-44.0
165	RUCJIL	14	P21/c	42	4	1	transl	-44.1
166	IMEQUP	14	P21/c	39	3	1	transl	-44.4
167	BEYROQ	2	P-1	32	2	1	transl	-44.6
168	JELKOE	2	P-1	56	3	1	transl	-44.6
169	ARUVAK	14	P21/n	47	2	2	inv	-45.4
170	ROLVAR	14	P21/n	33	2	1	inv	-45.4
171	KEMKOG	2	P-1	69	1	1	transl	-45.5
172	PIDSEE	2	P-1	30	2	1	inv	-45.9
173	EMEZEE	2	P-1	38	3	1	inv	-46.0
174	CAHNEH	14	P21/n	44	3	2	transl	-48.3
175	QQQBSS02	2	P-1	38	1	1	transl	-48.5
176	EJAHED	2	P-1	33	2	1	inv	-49.2
177	AZESAB	2	P-1	49	2	1	inv	-50.1
178	HUMCAW	2	P-1	34	2	1	inv	-50.1
179	YOCPUD	61	Pbca	42	2	2	glide	-53.3
180	NOXGIT	14	P21/c	34	2	2	inv	-53.9
181	EMIHAM	15	C2/c	39	3	1	rotation	-54.6
182	CUMPIK	15	C2/c	53	2	2	transl	-55.7
183	ELERUK	14	P21/n	36	2	1	transl	-56.4
184	FEDGON	14	P21/c	45	2	1	transl	-63.2
185	IWONUE	2	P-1	57	2	1	transl	-65.4
186	HACWAL	2	P-1	63	4	3	transl	-66.6
187	YAYFAH	15	C2/c	43	3	3	inv	-72.2
188	VEBKIZ	14	P21/n	51	3	1	inv	-79.8
189	YEKHII	15	C2/c	34	2	2	transl	-80.5
190	DICBEY	2	P-1	38	3	2	inv	-85.6
191	IFAXUJ	2	P-1	48	2	1	inv	-89.8

Acknowledgements

I would like to express my deepest gratitude to all the people and institutions who have supported me throughout my academic journey.

First and foremost, I wish to thank all those who supervised and assisted me in my work. I am sincerely grateful to my supervisor, Prof. Leonardo Lo Presti, for his constant availability, fairness, and for granting me the independence necessary to carry out this research. I am also profoundly thankful to Prof. Aurora Cruz-Cabeza, who supervised me during the wonderful six-month research period in her group at Durham University and welcomed me as if I were a member of her own family. My heartfelt thanks extend as well to my co-supervisor, Prof. Silvia Rizzato, and to my colleague, Dr. Giovanni Macetti, for the helpful discussions, guidance, and assistance over these years. A special thanks goes to Lorenzo, who applied the confinement algorithm to the exploratory simulations involving few molecules and extreme confinement during his BSc thesis work, and to Frank for assisting me with the experimental work at Durham University during his PhD.

I would also like to acknowledge the institutions that have supported me throughout these years. I am extremely grateful to the Italian Crystallographic Association (AIC), British Association for Crystal Growth (BACG), European Crystallographic Association (ECA), and Fondazione Zegna for the generous

scholarships, awards, bursaries, and travel grants that made possible my international research period, participation in conferences, and attendance at scientific schools. I am additionally grateful to INDACO and CINECA for providing the computational resources essential to this research.

My gratitude extends to all the people who have accompanied me on this journey and supported me with their presence, encouragement, and friendship. I wish to thank Prof. Lucia Carlucci and Prof. Davide Proserpio, who have guided me since my undergraduate years with human warmth, insight, and invaluable advice, and who sparked my passion for supramolecular chemistry, both experimental and computational. My sincere thanks also go to the numerous colleagues and friends I met during the many conferences, schools, and courses I attended in Turin, Trieste, Durham, Parma, Naples, Bologna, Loughborough, Padua, Copenhagen, Leeds, and Florence, as well as to my fellow Zegna Scholars.

I am particularly thankful to my high school classmates and to the university peers who have remained by my side throughout these years, in particular Elisa, Giulia, Nico, Daniel and Tino, for their friendship and the many memories we share. I am equally grateful to the members of my current research group, Giovanni, Margherita, and Stefano, including former thesis students Lorenzo, Fabio, Alessio, Andrea, Cristian, Gabriele, Loris, and Margherita, for their collaboration, kindness, and for fostering a stimulating and supportive environment. I would also like to thank Pavel for the collaboration and support throughout our time working together. I also wish to extend my appreciation to my colleagues in Aurora's research group in Durham, Amrita, Amy, Frank, Henry, and Marta, with whom I had the privilege to work and grow during my time at Durham University.

Finally, I am deeply thankful to my parents and my sister, who have always supported and encouraged me, even in the face of the clear challenges characterizing the path I chose to pursue. A very special thanks goes to my girlfriend, Alessandra, for her patience, understanding, and love.

To all of you, thank you for being part of this journey.