The Integrated DL_POLY/DL_FIELD/DL_ANALYSER Software Platform for Molecular Dynamics Simulations for Exploration of the Synthonic Interactions in Saturated Benzoic Acid/Hexane Solutions

I. Rosbottom*[a], C. W. Yong[b,c], D. Geatches[b], R.B. Hammond[a], I. T. Todorov[b] and K.J. Roberts[a]

[a]*Centre for Digital Design of Drug Products, School of Chemical and Process Engineering, University of Leeds, UK, LS2 9JT*

[b]*Scientific Computing Department, Science and Technology Facilities Council, Daresbury Laboratory, Sci-Tech Daresbury, Warrington WA4 4AD, UK*

[c]*Manchester Pharmacy School, Faculty of Medical and Human Sciences, University of Manchester, Manchester M13 9NT, UK*

* Author to whom correspondence should be addressed. Email: i.rosbottom@leeds.ac.uk

Abstract

Three separately developed software Molecular Dynamics packages at Daresbury Laboratory, namely DL_FIELD (DL_F), DL_POLY and DL_ANALYSER, have been integrated to form an efficient computational infrastructure to investigate the detailed solution chemistry of saturated benzoic acid in hexane solutions. These software capabilities are demonstrated, in combination with the Synthonic Engineering tools and density functional theory (DFT) calculations, to assess the extent that the solute-solute intermolecular synthonic interactions in solution mirrors the synthons in the crystal structure. The results show that the majority of the COOH groups are forming OH…O H-bonds, which are a combination of classic OH…O homo-dimers and three membered H-bonding clusters. The formation of pi-pi stacking interactions is observed, but in far fewer numbers than observed for the OH…O interactions. The DFT simulations of the IR spectra of the multiple benzoic acid aggregates extracted from the MD trajectories provides further in-depth analysis of previously published IR data, by matching simulated peaks to the experimental peaks, hence identifying the exact bonding modes that are responsible for such peaks. This study demonstrates the value of a multi-scale and multi-technique approach to exploring the molecular transition pathway from solution to crystal structure.

Keywords: atomic interactions, hydrogen bonds, benzoic acid, DL_FIELD, DL_POLY, DL_ANALYSER, DL_F Notation, DANAI, DFT

List of Abbreviations:

DL_F: Daresbury Laboratory Field

OPLS: Optimised Potential for Liquid Systems

CVFF: Consistent Valence Forcefield

PCFF: Polymer Consistent Forcefield

AMBER: Assisted Model Building with Energy Refinement

GAFF: Generalised AMBER Forcefield

CHARMM: Chemistry at Harvard Molecular Mechanics

List of Terms:

DANAI: The language to examine the intermolecular interactions in the MD simulations, based on the DL_F notation

DL_F Notation: The generalised notation for atoms, based on the chemistry of the atoms in the molecules. This can be mapped onto the forcefields in DL_FIELD

Synthons: Intermolecular interactions

Atom_Type: Human readable atom typing based on the chemistry of the atom

Atom_Key: The notation that the forcefield reads to identify the atom typing

1. Introduction

A molecular scale understanding of the transition state between stable solutions and solid-state molecular crystals signifies a grand challenge for theoretical and experimental scientists[1]. A fundamental understanding of this process could lay the foundations for a quality-by-design approach to industrial solution crystallisation, which is widely utilised in the pharmaceutical and fine chemical industries to achieve the required purities for their high value products. Atomistic and molecular scale modelling can be used to predict experimental conditions which will produce high quality crystalline products with tailored physical properties, optimal for downstream unit processes.

The making and breaking of covalent bonds to bring together chemicals or functional groups to build molecules is the basis of retrosynthesis in the organic synthesis community. Similarly, this concept has been expanded to view the making and breaking of intermolecular forces as the basis of bringing together the building blocks of molecular crystals, and as such the intermolecular interactions (synthons) between these building blocks are seen crucial for defining the physical properties of molecular crystals[2-7].

Molecular modelling of the intermolecular interactions (synthons) which are present in solid-state crystal structures has been utilised to rationalise the physical properties of molecular crystals[6-13], and in addition density functional theory (DFT) has been used to probe the finer electronic structure detail and chemical robustness of the these synthons. Theoretical and experimental studies have suggested that solute molecules pre-cluster (form transient associations) prior to crystal nucleation, which may impact upon solid-form selection, i.e. the polymorphs expressed via solution crystallisation and, ultimately the physical properties of the resultant crystals, such as morphology[14-19]. Indeed it has even been suggested that solute can aggregate in undersaturated homogenous solutions[20, 21], suggesting that solution structure can impact the solid form even before it has entered

into the supersaturated region. Molecular modelling studies can provide insights into solution chemistry behaviour, for example indicating that the formation of H-bonding synthons, such as the classic OH…O carboxylic acid dimer, can template the nucleation of a crystalline form from specific solutions[22, 23].

Recent scientific software developments at Daresbury Laboratory have resulted in a suite of molecular dynamics (MD) tools for the model setup, simulations and finally the results analysis by using the software packages DL_FIELD, DL_POLY and DL_ANALYSER, respectively. These packages form part of the DL_Software[24], a collective term for the scientific software packages that are developed at Daresbury Laboratory (DL).

The development of synthonic (intermolecular) modelling techniques have demonstrated capability to deconstruct 3-D crystallographic structures into their constituent intermolecular interactions (synthons) and to apply them in the comparative analysis of bulk (intrinsic) and surface (extrinsic) interactions. However such calculations do not provide detailed molecular-scale information about the intermolecular organisation within the solution state. This can, however, be extracted from MD simulations, although, the implementation of the MD simulations and subsequent analysis of trajectory files for multiple systems can be both challenging and time-consuming due to the lack of user-friendly methods for setting-up and analysing large MD simulations. The DL suite of MD software tools addresses these issues and when combined with both synthonic engineering methods and the finer-scale electronic structure method of density functional theory (DFT), produces a complementary collection of tools for the exploration of material systems.

2. The DL_POLY/DL_FIELD/DL_ANALYSER Molecular Dynamics Platform

2.1. Overview of Component Programs

DL_POLY[25] is a versatile and powerful MD program suite that can run efficiently on a variety of computer platforms, from a single-processor PC to massively parallel supercomputers. DL_FIELD[26] is a support application tool for DL_POLY, especially in setting up complex forcefield models. DL_ANALYSER[27] is a robust software tool to carry out post analysis work on the simulation outputs produced by DL_POLY.

DL_POLY is developed to be highly agnostic in nature, whereby it can handle molecular models of varying size and complexities: from inorganic materials such as minerals, zeolites, to complex topological structure models such as graphenes, organic cages, biomolecules and even mixed component systems such as the bio-inorganic models.

In the case of DL_FIELD, the philosophy behind the software development is to minimise the requirement for users to have a detailed knowledge of the inner workings of complicated forcefield descriptions and preparation procedures. It is intended to serve as a user-friendly tool that automatically processes the molecular information with minimum user's intervention. DL_FIELD is one of the few software packages that make available some of the widest range of force field (FF) schemes for model setup: CHARMM[28], AMBER[29], OPLS[30], Dreiding[31], PCFF[32], CVFF[33] and also united atom Gromos variants such as the G54A7[34]. In addition, DL_FIELD also implements inorganic force fields that apply to a range of non-biological materials, from simple binary oxides to clay minerals and glass. Perhaps unique to DL_FIELD is the consistent expression of these FF schemes in a single, unified format implemented within the DL_FIELD framework. This means that no additional scripting is necessary when one wishes to convert a molecular configuration from one FF model to another. In addition, DL_FIELD

also allows seamless setting up of multiple potential schemes within a molecular model. For instance, combination of any organic FF schemes with the inorganic models for complex bio-inorganic system setup.

On the other hand, DL_ANALYSER contains a collection of unified analysis tools that can produce results in a single-read through a collection of DL_POLY's HISTORY trajectory files. More importantly, unique to DL_ANALYSER is the implementation of the atomic interaction analysis module that can identify, quantify and analyse various modes of atomic interactions from the HISTORY files.

Drawing upon the above, the user-friendly and intuitive work flow that results from these new software tools, firstly for constructing MD simulation cells, is demonstrated via a case study in which benzoic acid molecules, as solutes, are tracked in 0.8M solutions of hexane, providing information about the solute-solute solution synthons via analysis of the HISTORY files using DL_ANALYSER. Benzoic acid at a 0.8M concentration was selected as a case study due to its well characterised crystallisation behaviour and physical properties, for example, the crystal structure is known to contain OH…O H-bonding 'classic' homo-dimers. Indeed, the experimental infra-red solution data has suggested that such H-bonding interactions are present in solutions prior to crystal nucleation in several solvents, including hexane[35, 36]. The solid-state chemistry, encapsulated in terms of intrinsic synthons, is characterised in detail and compared to the solution synthons, along with the experimental data, to present a potential unified simulation approach for examining and understanding the solution crystallisation transition-pathway for multiple solute molecules, solvents and solution concentrations.

MD, in combination with Synthonic Engineering tools and DFT results in a workflow to assess the similarities between the solution and solid-state synthonic chemistry, shown in Figure 1.

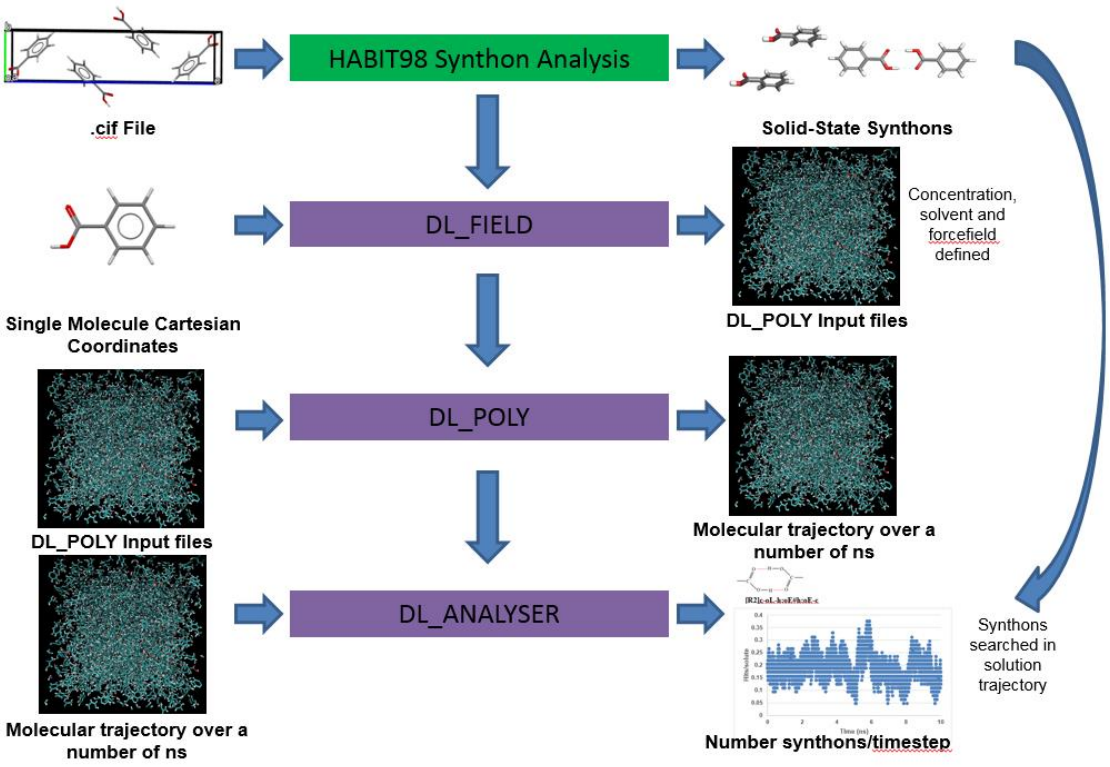


Figure 1: This workflow shows how HABIT98 can identify the most important synthons in the crystal structure, set up and run the molecular dynamics simulations of organic solutions using DL_Software and then identifying if the synthons from the crystal structures appear in the solution simulations

2.2. Software Integration within the Platform

To ensure smooth software-integration and data transitions between different packages, the standard DL_F Notation[26] is implemented within DL_FIELD. This is a universal notation scheme that standardises the expression of atom types in a molecular system, of which the atom typing procedures are essential to ensure correct setup of simulation models. The DL_F Notation is designed to provide a common solution that harmonises the atom typing which varies wildly from one FF scheme to the other. Quite often, the conversion of one force field scheme to the other is not a trivial task. The DL-F Notation removes layers of complexities involving data structure conversions and ensures smooth data transition among various force field schemes. The latest DL_FIELD version 4.2

applies the DL_F Notation to the force field schemes such as the OPLS_2005[37], CVFF and PCFF. The DL_F Notation is also partially implemented for CHARMM and AMBER and effort is currently underway to extend the notation and apply it to other FF schemes that are already implemented in DL_FIELD.

In addition, the notation completely describes the actual chemical identity of an atom within a molecule in a way which can be easily interpreted by both computational modellers and experimentalists. This allows correlation of both cognitive and computational assessment of the simulation results and directly relates the analysis to the actual chemistry of the test structures over a range of different force field schemes. In addition, the ease of interpreting the DL_F Notation means DL_POLY's trajectory files can be archived for analysis using DL_ANALYSER making them accessible for assessment by other, future researchers with only minimal training required.

The atomic-interaction detection feature that is implemented within DL_ANALYSER is a unique program module to identify and quantify various modes of atomic interactions in a molecular system. The precise atomic interactions are annotated by using the DANAI notation syntax[27]. This is a standard notation scheme of atomic expression that can be easily interpreted by modellers and experimentalists. By making use of the DL_F Notation, a DANAI expression contains the actual chemical information and annotates a given set of localised atomic interactions that can be assessed by means of data analytics. Until now, and in virtually all cases, the atomic interactions are largely described by using some arbitrary diagrammatic representations, textual annotations or even symbolic representations of certain molecular motifs[38]. Although such approaches are easily interpreted by human cognition, they are not directly accessible to data query or other computational means due to their inherently low data-discoverabilities.

The universalities of DL_F Notation and DANAI means DL_FIELD and DL_ANALYSER form a flexible and powerful software infrastructure with DL_POLY to give a seamless molecular simulation engine, irrespective of the type of FF schemes, which can be applied to a wide range of atomic systems, including both crystals and condensed-phase systems.

2.3. Universal Notations for Atom and Forcefield Identification

To circumvent the issue of different atom labelling in forcefield schemes, the software platform utilises a set of universal atom keys. The atom keys are the atom or particle labels that are defined in DL_POLY's FIELD and CONFIG files. In DL_FIELD's context, the atom typing process is given a human-readable label called the *ATOM_TYPE* and the corresponding atom key is called the *ATOM_KEY*. The *ATOM_TYPE* spells out the actual chemical identity of the atoms' human-readable forms and all FF schemes share the same *ATOM_TYPE*s. There is no need to keep a separate reference list between *ATOM_TYPE*s and *ATOM_KEY*s, as these can be interconverted *in situ* within the DL_FIELD program if needed. Instead, the library files for each FF scheme keeps a reference of *ATOM_TYPE*s to the atom keys that are expressed in the same original forms as those in the respective FF schemes. From such, correct sets of interaction parameters can be determined for the FIELD file.

Figure 2 shows the ball and stick representation of ethanoic acid molecules, labelled with the original atom keys for four different FF schemes: OPLS, PCFF, CHARMM CGenFF[39] and AMBER GAFF[40]. The molecule consists of three types of elements (C, O and H) but six different atom types: two different types of carbon (methyl carbon and carbonyl carbon); two different types of hydrogen (non-polar alkyl hydrogen and polar hydroxyl hydrogen); and two different types of oxygen (the hydroxyl oxygen and the

carbonyl oxygen). Note that these atom keys are expressed via some cryptic labels that do not clearly indicate the chemical identity of the atoms.

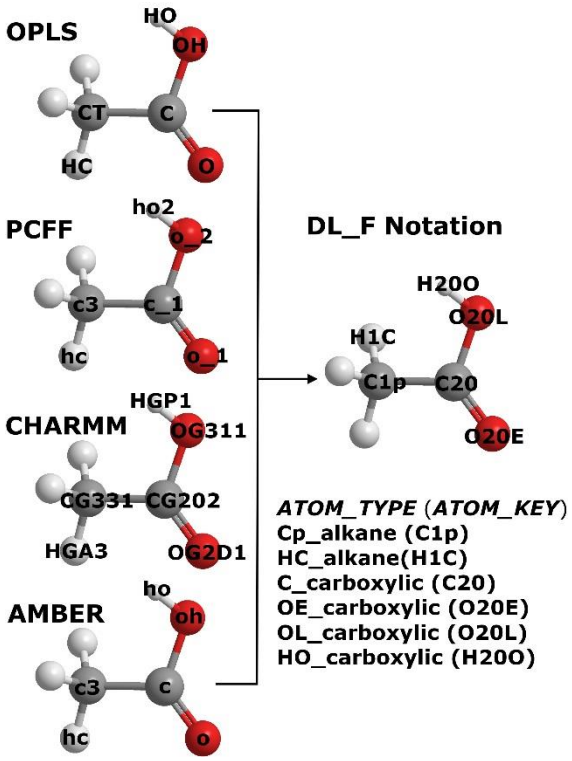


Figure 2: Ball and stick representation of ethanoic acid molecules. On the left, the molecules are labelled with the standard atom keys for OPLS, PCFF, CHARMM CGenFF and AMBER GAFF. These atom keys can be converted into a universal set of keys (ATOM_KEYs) expressed in the DL_F Notation. The corresponding ATOM_TYPE is also shown, indicating the chemical nature of every atom in the molecule.

However, by using the DL_F Notation, it was possible to produce a universal set of human-readable *ATOM_TYPE*s (Figure 2 on the right) for the molecules and the corresponding *ATOM_KEY*s were produced *in-situ* using the DL_FIELD program. The notation essentially assigned every atom to a member group with a characteristic chemical behaviour called the *Chemical Group* (CG), which is broadly similar to the functional group in chemistry. The 'alkane' and 'carboxylic' as shown in Figure 2 are the *CG*s identified in the ethanoic acid molecules. Each *CG* is referenced to the unique index called the *Chemical Group Index* (*CGI*). For instance, the *CGI* value for the 'alkane' *CG*

and ‘carboxylic’ *CG* are 1 and 20, respectively. The letters ‘p’, ‘E’ and ‘L’ are the supporting tokens that refer to *primary*, *linked atom* and *end atom*, respectively. More details about the notation syntax can be found in reference[26].

The chemical-sensitive feature of DL_F Notation means that the *ATOM_KEY*s can potentially be used as descriptors for data analysis. For instance, by making use of the DL_F Notation, a new notation syntax called DANAI is implemented in DL_ANALYSER to identify and describe specific atomic interactions in molecular systems. This also means that DANAI shares the same characteristic features of DL_F Notation of being non-FF specific and the HISTORY trajectory files can be directly analysed by DL_ANALYSER without further data modification.

Figure 3 Shows examples of hydrogen bond (HB) interactions between two benzoic acid molecules, or more specifically, between two carboxylic groups, as indicated by the general DANAI expression called the *macro-interaction*, HB_20_20. The number 20 is the *CGI* for the ‘carboxylic’ *CG*. There are various ways the HB can form between the *CG*s. These are called the *micro-interactions* for the HB_20_20 and can involve various combinations of two or more of these *CG*s. Two such interactions are shown in Figure 3.

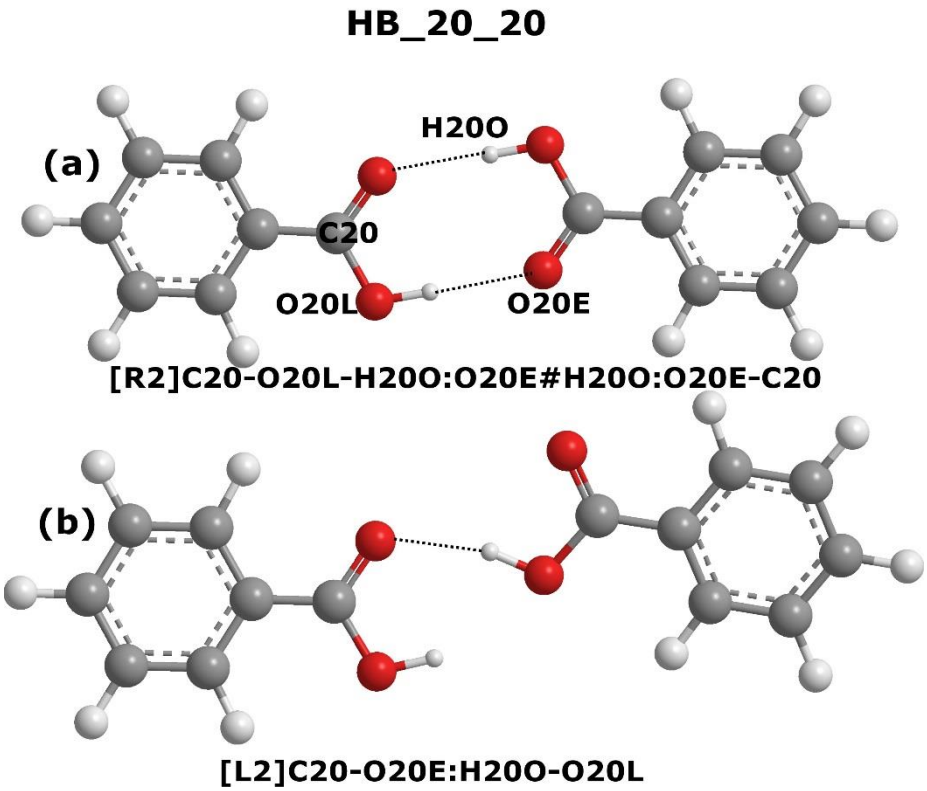


Figure 3: Diagrammatic illustration of the HB interactions between the carboxylic acid groups in benzoic acid, along with the corresponding annotations in DANAI notation. The dotted line in the diagram and the colon (:) in the DANAI expression indicate the HB interactions in question. (a) The dimeric interactions involving two hydrogen bonds. (b) The catermetic interaction involving one hydrogen bond.

In Fig. 3(a), The DANAI expression for the micro-interaction [R2]C20-O20L-H20O:O20E#H20O:O20E-C20 specifically annotates the dimeric conformation involving two hydrogen bonds, represented as the dotted lines in the diagram, or colons (:) in the DANAI expression, between the polar hydrogen atoms from one 'carboxylic' *CG*, H20O, and the carbonyl oxygen atoms, O20E, of the other 'carboxylic' *CG*. The symbols within the square bracket indicate the general overall topological structure of the interaction. In this case, [R2] means interactions between the two *CG*s that results in a ring-like topological structure. The carbonyl atom labels, C20 as shown at the beginning

and end of the DANAI expression, are referred to the same atom as shown in Fig. 3(a) to indicate the extent of the ring enclosure.

Fig. 3(b) shows a catermetic interaction between the atoms O20E and H20O with the corresponding expression [L2]C20-O20E:H20O-O20L. In this case, the [L2] represent a linear topological structure.

The capital letters C, O and H are the element symbols for the carbon, oxygen and hydrogen, respectively. However, in DANAI, the atomic symbols can be expressed in either uppercase or lowercase, which indicates the extent of the HB interactions on the atom. If an atom is specified in uppercase, it means there is no other HB interaction with the atom apart from that which is indicated in the DANAI expression. On the other hand, if an atom is specified in lowercase, it means there may be any other HB interactions with the atom but one of the hydrogen bonds must be that one indicated in the DANAI expression. Therefore, changing letter cases indicates the strictness of criteria to identify a given DANAI expression by DL_ANALYSER. For instance, by changing the DANAI expression in Figure 3(b) to [L2]C20-o20E:h20O-O20L, this means there may be more than one HB interaction between other atoms and the o20E or h20O atoms, but one of the HB interactions must be the one between the particular o20E and h20O atoms as indicated in the expression.

Since the only HB interactions mentioned in this paper are those involving carboxylic groups, the *CGI* value of 20 will be omitted in the DANAI expressions. For instance, the dimeric and catermetic interactions will be expressed as [R2]C-OL-H:OE#H:OE-C and [L2]C-OE:H-OL, respectively. More details about the DANAI expression syntax can be found in reference[27].

The software platform developed allows a user to set up a MD simulation box of a bulk solution at a defined concentration without the challenge of formatting the often complex

forcefield files for multiple component systems, needing only the Cartesian coordinates of a single solute molecule. In particular, this work flow provides the flexibility for a user to seamlessly switch between a range of forcefields and solvents. The platform then facilitates post MD simulation analysis where the detailed intermolecular interactions at each timeframe in the simulation can be characterised, without the need for any further scripting. This platform is demonstrated using the benzoic acid case study.

3. Materials and Methods

3.1. Crystallographic Structure of Benzoic Acid

Only a single crystalline form of benzoic acid has been identified experimentally which crystallises as a monoclinic $P2_1/c$ structure with a tetra-molecular unit cell, shown in Figure 4, with associated crystallographic information provided in Table 1.

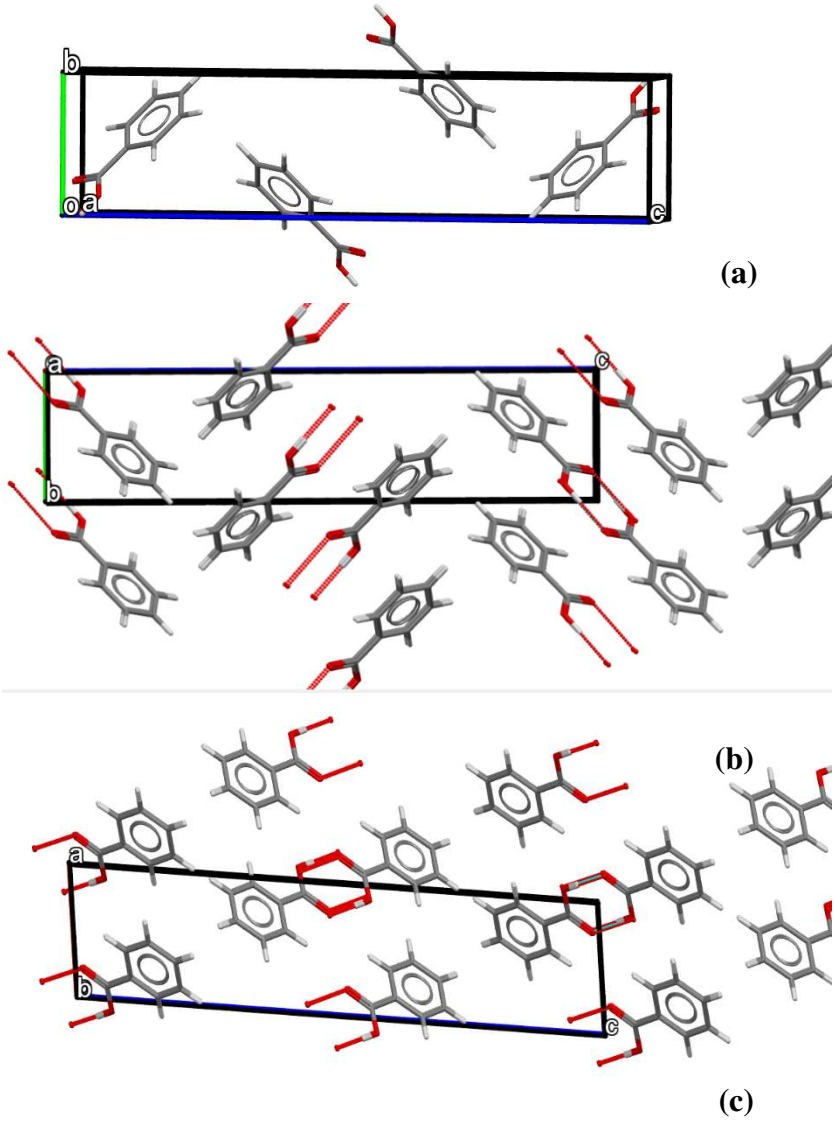


Figure 4: (a) Unit cell of benzoic acid crystalline structure; (b) packing diagram with the a-axis of the unit cell coming up out of the page; (c) packing diagram with the b-axis of the unit cell coming up out of the page

Table 1: Crystallographic information for benzoic acid crystal structure (CSD refcode BENZAC01)

a (Å)	b (Å)	(c) Å	β (°)	Z/Z'	Cell Volume (Å^3)	Space Group
5.51	5.16	21.97	97.41	4/1	619.15	$P2_1/c$ (14)

Figure 4 shows that the packing of the benzoic acid crystal structure consists of a combination of H-bonding and apolar interactions, such as pi-pi stacking interactions.

The H-bonding dimers extend approximately along the long c-axis, whereby the pi-pi stacking interactions are more approximately directed along the shorter a- and c-axes. Benzoic acid often presents as a needle or lathe-like morphology and can crystallise from both apolar and polar solutions, including hexane, ethanol, methanol, acetonitrile and chloroform[35, 36, 41-44].

3.2. Solid-State Synthon Analysis

The strongest intermolecular interactions (synthons) in the crystal structure were calculated using the HABIT98 software. The mechanics and operations of this software are detailed in multiple publications[6-12, 45, 46]. The synthon strengths were calculated using the force field published by Momany and co-workers[47], using fractional atomistic charges calculated using the Gasteiger and Marsili method[48, 49]. The interatomic interactions searched for in the MD simulations were based on the nature of the synthons found in the crystal structure.

3.3. Solution Structure using Molecular Dynamics Methods

The Cartesian coordinates of the atoms in a single molecule of benzoic acid were inputted into DL_FIELD, where the concentration, force field and solvent were specified and the CONTROL, FIELD and CONFIG files for input into DL_POLY were constructed. The simulations in this study were carried out in a 60Å^3 periodic box with a 0.1M solution of benzoic acid in hexane, containing 107 solute and 760 solvent molecules. All forces and energies were calculated using the OPLS 2005 force field[37].

A constant energy simulation was run with constant scaling at 400K for approximately 800ps to remove any pre-existing solution chemistry from the solution-box generation step. The solution was then cooled to the sampling temperature of 293K using a constant energy ensemble with constant scaling. The scaling was removed and if the temperature

was observed to remain around the sampling temperature then the solution was deemed acceptably equilibrated. The simulation was then continued for a short time using the Berensden NVT ensemble[50], before a further continuation employing the Berensden NPT ensemble[50] until the simulation box size no longer fluctuated. The sampling simulations were run for 2.5ns each with a time step of 0.002 ps using the Berensden NPT ensemble with thermostat and barostat relaxation times of 4.0ps and 5.0ps respectively[51]. The solution was sampled for 10ns in total.

The HISTORY trajectory files were inputted into DL_ANALYSER where the solute-solute H:O HB interactions and pi-pi stacking interactions were examined. Any HB interactions less than 2.5Å in distance and greater than 120° in angle were counted, whilst the pi-pi stacking interactions had to be closer than 4.5Å and with a less than 20° angle offset between the planes of the interacting rings.

Snapshots from the trajectory file were also created in DL_ANALYSER by outputting the Cartesian coordinates of a 15 X 15 X 15Å cube created around a single of molecule of PABA throughout the simulation. These snapshots which displayed solute/solute interactions were used as the input files for the DFT simulations.

3.1. Density Functional Theory Models and Method

A cif file of the molecular crystal of benzoic acid provided as a structure within Materials Studio[52] was used as the initial input structure for a DFT calculation. The remaining DFT models were extracted from a MD snapshot of the 60 $Å^3$ periodic box with a 0.8M solution of benzoic acid in hexane, described in the previous section. They include: a dimer of benzoic acid in a hexane solution, a single, dimer, half-dimer and a catamer of benzoic acid all extracted and represented in the gas-phase. As previously stated the origin of all of these systems (except the molecular crystal) derive from the MD calculations,

which implies that a dimer found stable in the hexane solution can be represented as the same dimer in the gas-phase without prejudice. This allowed a considerable speed-up of the DFT calculations, thereby increasing the feasibility of DFT being used as an additional tool to the DL_Software suite and HABIT98 combination, where electronic structure detail is required.

The dimensions of the crystal structure were the same as those mentioned in Table 1 to within 1%. All other systems were extracted from the MD snapshot and placed inside a 20 $Å^3$ box. In the case of the dimer in a hexane solution a core region was identified and extraneous hexane molecules were deleted. For the gas-phase systems all hexane molecules were deleted.

These systems were built in Materials Studio[52] and prepared for calculation using the planewave pseudopotential code, CASTEP[53] within the DFT formalism[54-56]. The 'fine' and 'ultra-fine' settings available were used, giving a kinetic energy cut-off of 990 eV. The electron-ion interactions were described by norm-conserving, on-the-fly-generated pseudopotentials, corresponding to exchange-correlation interactions represented by the generalized gradient approximation density functional, i.e. PBE-GGA[57]. Additional long-range dispersion forces were implemented using the semi-empirical dispersion interaction correction module[58], specifically the pairwise form of Tkatchenko *et al.*[59] The Brillouin zone of the molecular crystal was sampled on a Monkhorst-Pack grid[60] corresponding to the 'fine' setting of Materials Studio, i.e. 3 x 3 x 1; all gas-phase systems were sampled with a single k-point corresponding to the gamma point. The geometry was optimised using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm[61] and the electronic method used in the self-consistent field (SCF) calculations was density mixing. The following convergence criteria were also applied: electronic energy for the SCF cycles 1

x 10^{-10} eV; total energy following geometry optimisation 5.0 x 10^{-6} eV/atom; interatomic forces 0.01 eV/Å.

Following relaxation of the geometry of the molecular crystal and gas-phase dimer, density functional perturbation theory[62] (DFPT) calculations were carried out to generate IR spectra. The results are discussed in Sections 5.3.1 and 5.3.2.

4. Results and Discussion

4.1. Synthon Analysis in the Solid-State

The lattice energy was calculated to be -19.77kcal/mol, which is in reasonable agreement with a recently published sublimation enthalpy of 21.8kcal/[63], suggesting the force field acceptably reproduces the intermolecular interaction energies. The energetics of the strongest intermolecular synthons identified in the benzoic acid crystal structure are shown in Table 2, along with the molecular geometries of these synthons.

Table 2: Strongest intermolecular synthons found in the BENZAC01 crystal structure of benzoic acid. Multiplicity defined as the amount of intermolecular synthons one molecule in the structure can form of each type. Lattice energy contribution calculated as (multiplicity*intermolecular energy)/total lattice energy

Synthon	Multiplicity	Intermolecular Energy	Lattice Energy Contribution (%)	Interaction Type
A	1	-5.3	26.8	H-bonding dimers
B	2	-2.0	19.8	Dispersive head-head pi-pi stack
C	1	-1.5	7.6	Dispersive head-tail polar stack
D	2	-1.4	14.2	Dispersive offset head-head stack

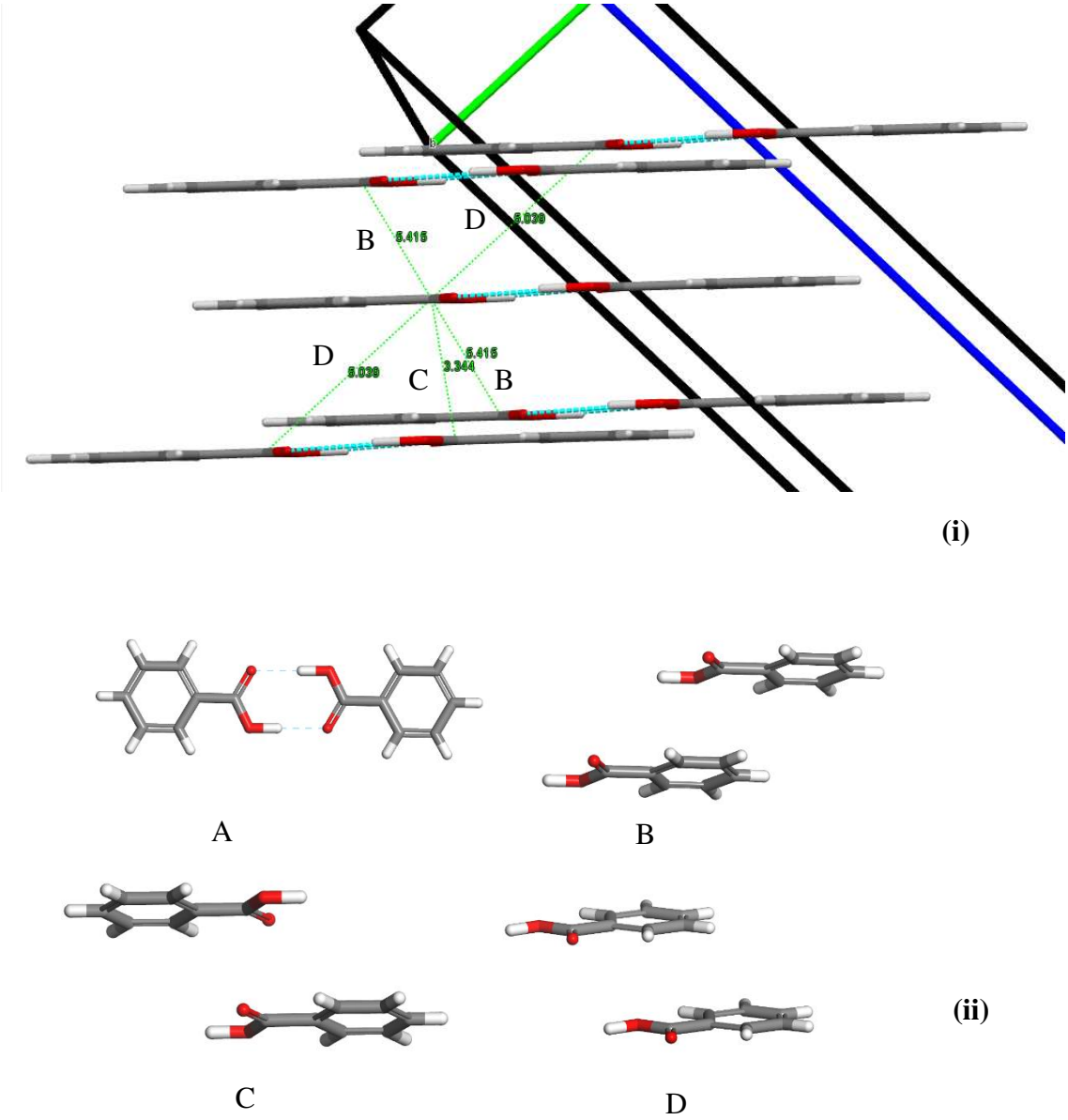


Figure 5: (i) The packing of the benzoic acid crystal structure showing how the dispersive stacking interactions B, C and D align in the structure. The head to head interactions B and D can form one above and one below the central molecule, giving a multiplicity of 2, whilst the head to tail interaction C only forms an interaction above the central molecule giving a multiplicity of 1. (ii) Molecular geometry found from the strongest intermolecular synthons shown in Table 2

A

Table 2 shows that synthon A has by far the strongest intermolecular HB interaction energy found in the crystal structure, which is shown to be the classic carboxylic acid dimer in Figure 5. Synthons B and D are more characterised by dispersive interactions, whereby the molecules can form offset π-π stacking interactions, whilst, energetically, the greatest contribution to the strength of synthon C is from polar (electrostatic) interactions between the COOH groups.

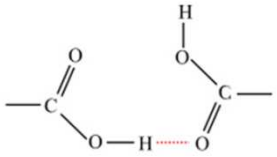
It has been hypothesised that both COOH H-bonding dimers and π-π stacking interactions can be influential in directing the nucleation of molecular crystals[23, 64-66]. In addition to this, there have been significant experimental data reported which suggest that benzoic acid molecules pre-cluster in solution prior to nucleation, particularly evidence supporting the hypothesis that HB interactions form in solution[35, 36]. There is particularly strong evidence that such interactions form in supersaturated hexane solutions[35].

4.2. Synthon Analysis in Solution

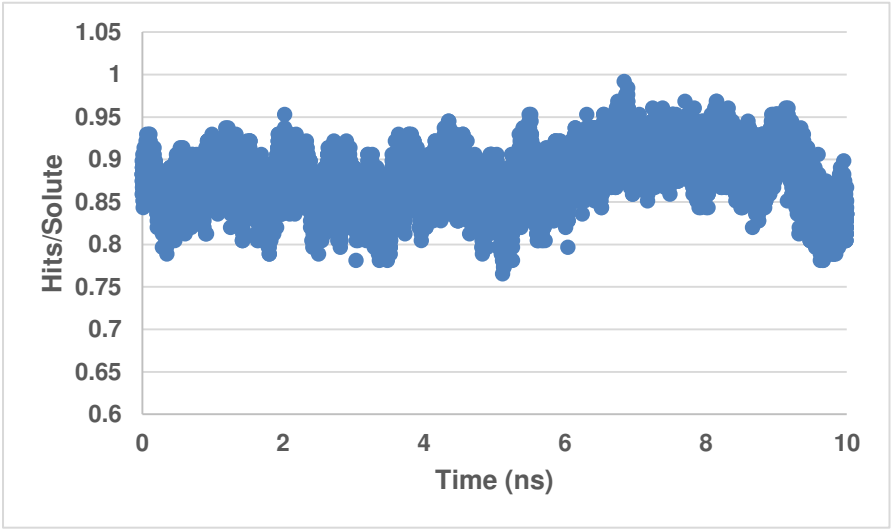
Based on the synthon analysis in the solid-state, any occurrences of the HB dimers or pi-pi stacking interactions in a 10ns MD simulation of a 0.8M solution of benzoic acid in hexane were examined.

4.2.1. **COOH Synthons in Solution**

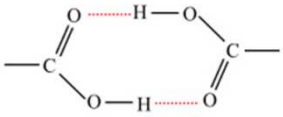
Figure 6 shows the number of singular H:O and the number of de-solvated H:O dimers, normalised per solute, throughout the 10ns simulation.



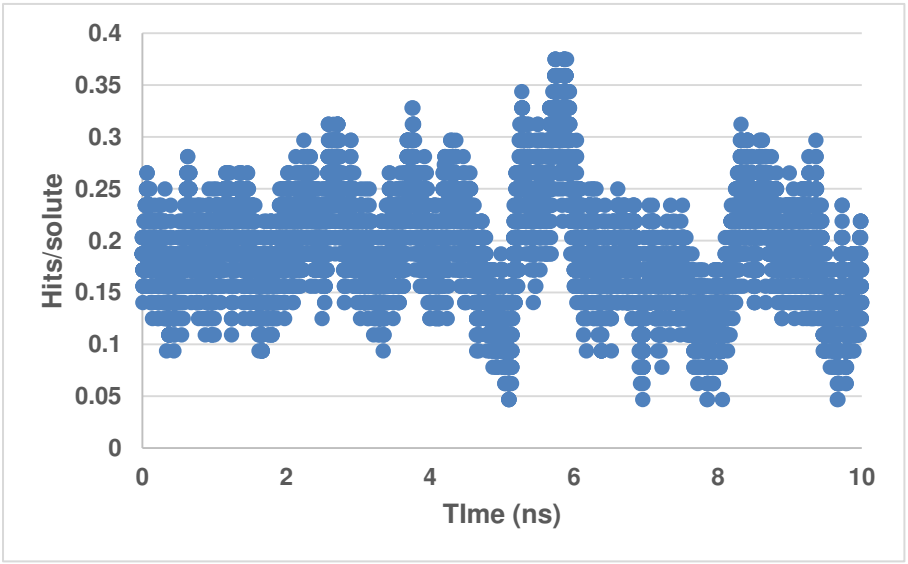
[L2]oL-h:oE-c



(a)



[R2]c-oL-h:oE#h:oE-c



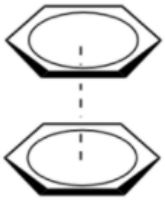
(b)

Figure 6: The hits/solute found from a 10ns simulation for: (a) an OH…O catermetic interaction which can also have further interactions with surrounding solute and (b) a fully de-solvated OH…O H-bonding dimer. DANAI expressions shown for the interactions

Figure 6(a) shows that over the 10ns of the simulations, around 85% of the COOH groups in benzoic acid were found to be forming interactions with each other, in the [L2]oL-h:oE-c mode. Figure 6(b) shows that around 20% of the COOH groups had in fact formed the 'classic' H:O dimer ([R2]c-oL-h:oE#h:oE-c) in the solution. The fact that the vast majority of the benzoic acid molecules are forming HB interactions agrees with the experimental IR studies of the solution, which observed a single carbonyl stretch and OH wag at very similar wavenumbers to the solid-state spectra[35]. The lack of multiple C=O environments for these peaks in the experimental IR spectra suggest that there are none, or very few, free COOH groups in the solution. This was also shown in the MD results.

4.2.2. **Pi-pi Stacking Synthons in Solution**

Figure 7 shows the propensity of pi-pi 'sandwich' stacking synthons to form in the MD simulations. The corresponding macro-interaction for this type of interaction is PS_6_6, where the value 6 is the *CGI* for benzene rings.



PS_6_6

[L2]<c6>:<c6>

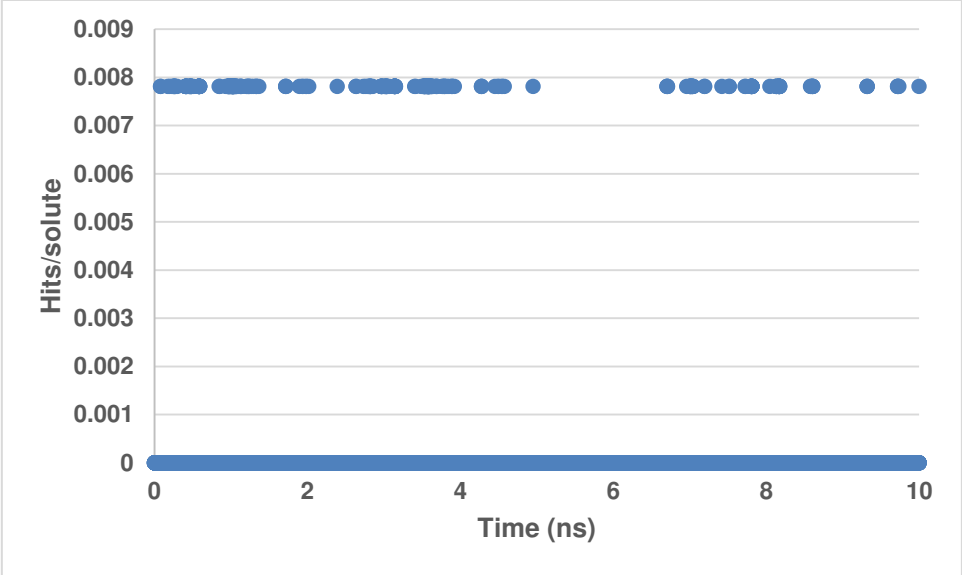
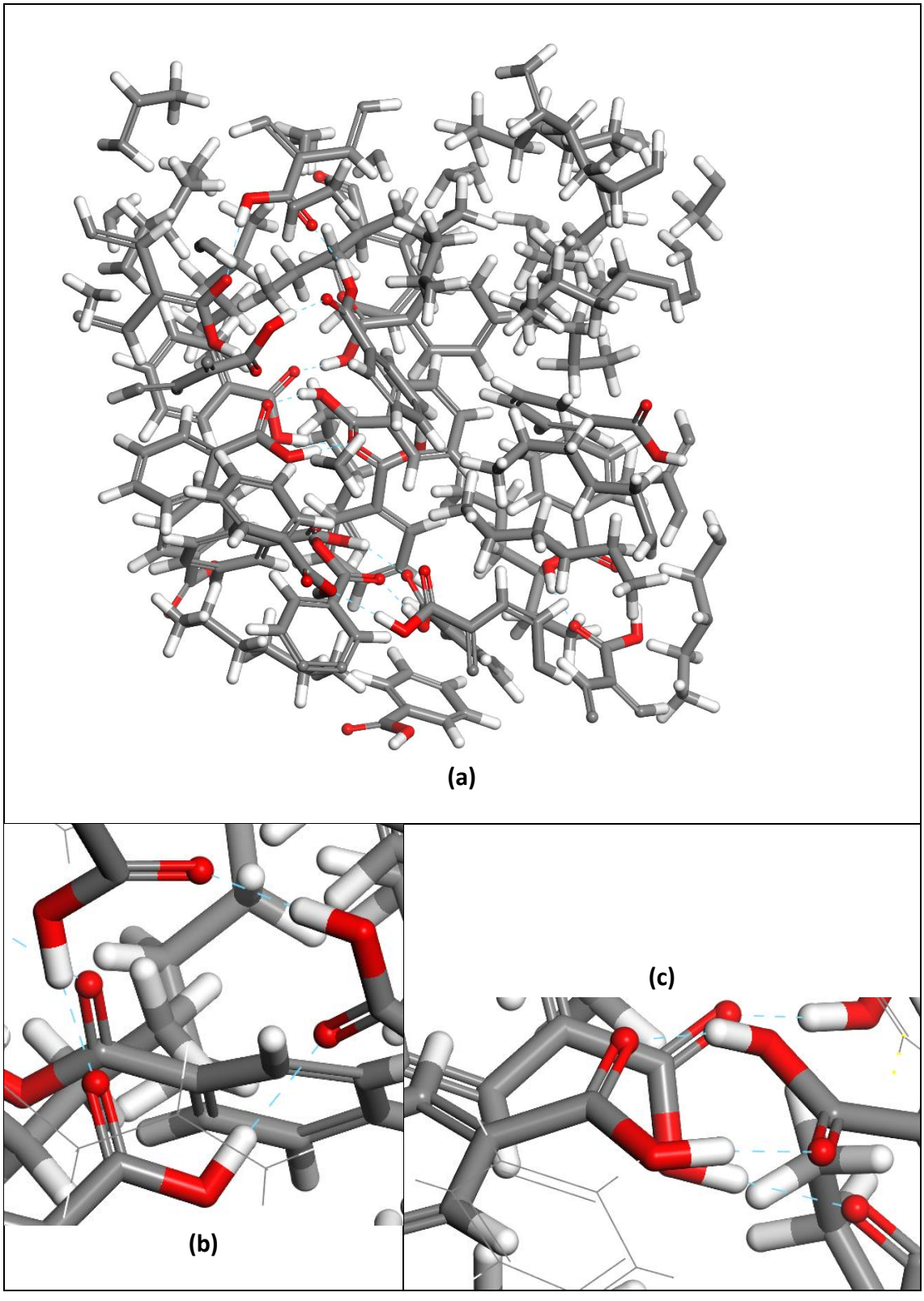


Figure 7: Number of pi-pi stacking interactions found in the 10ns simulation of benzoic acid in hexane, normalised to the amount of solute in the simulation. DANAI expressions shown for the interaction

Figure 7 shows that throughout the simulation there are pi-pi stacking interactions regularly detected, however it is found that no more than one of these interactions exists at any one time. Inspection of the trajectory files show that these interactions are extremely transient, i.e. they only exist for a particular time-step and then disappear and reappear. This is in contrast to the multiple HB interactions detected, which although the amount of these in the simulation fluctuate, their existence is consistent throughout the simulation, suggesting that this is indeed the crucial synthon for templating the nucleation of the crystalline form.

Figure 8 shows the molecular geometry of the solute and solvent at the 2.5ns step of the simulation.


(a)
(b)
(c)

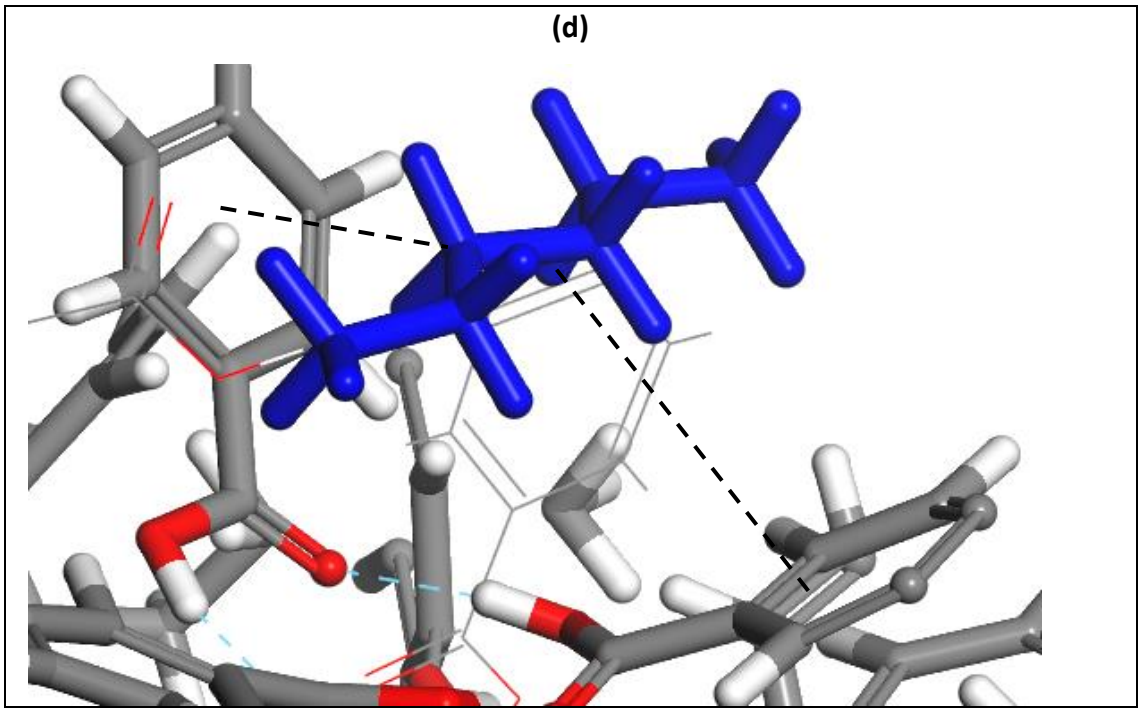


Figure 8: (a) Snapshot of a 10Å cubed area of the 0.12g/ml benzoic acid in hexane solution at approximately the 2.5ns step of the 10ns simulation; (b) zoom in on a three membered OH…O H-bonding benzoic acid ring interaction; (c) zoom in on a classic OH…O H-bonding homo-dimer; (d) zoom in on a hexane molecule (coloured blue) sitting between two benzoic acid phenyl ring structures

Figure 8 shows the molecular orientation of the different types of interactions detected, displaying how the polar carboxylic groups can form the classic dimeric interactions (Figure 8 (c)) or clusters of H:O interactions (Figure 8(b)), forming a trimeric structure with a corresponding representative DANAI expression of [R3]C-OE:H#OE:H#OE:H-OL-C. The dimer interactions are analogous to synthon A identified from the solid-state structure, shown in Table 2 and Figure 5. The simulated benzoic acid clustering through OH…O H-bonding is in agreement with the experimental data which suggests OH…O interactions are formed in solution[35].

The propensity for the apolar hexane molecule to interact with the apolar phenyl ring of benzoic acid (Figure 8(d)) probably contributes to the low number of pi-pi 'sandwich'

stacking interactions observed in the simulation. The apolar nature of hexane means that it will barely interact with the carboxylic group, leaving the formation of solute-solute HB interactions almost barrier-less. Hence, the pathway from the liquid structure, detailed in this section, to solid-state structure detailed in Section 4.1, is likely facilitated by the OH…O H-bonding Synthon A which is prominent in both the hexane mediated solution and solid-state. These interactions are the likely first step in encouraging the solute molecules to cluster in hexane, whereby the less polar parts of the benzoic acid molecule would be harder to de-solvate in hexane and hence probably play a secondary role in the formation and growth of critical nuclei sized clusters for the crystallisation of benzoic acid from hexane solutions.

4.3. DFT Refinement of MD Identified Solution

Although the molecular dynamics calculations provided atomistic configurations of the solute/solvent interactions, which were then identified and quantified using DANAI, there is no information of the fine, electronic structure of these interactions. To address this requires application of DFT, which would be computationally expensive were it to explore the whole MD simulation system. However, by extracting a smaller environment from an equilibrated MD simulation snapshot, and applying DFT to this, it is feasible to obtain more accurate detail on the fine conformational and intermolecular structure of the material, at relatively little extra computational cost.

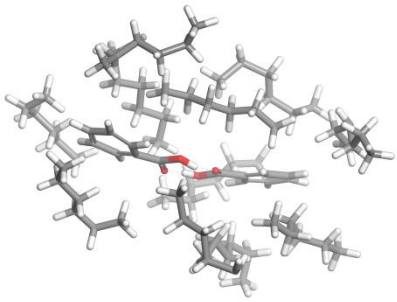
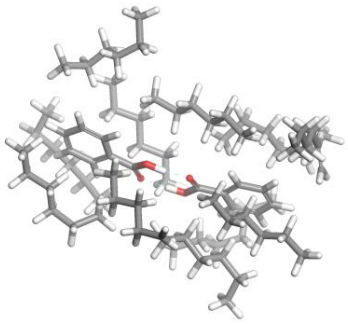
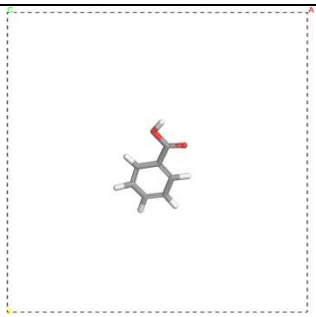
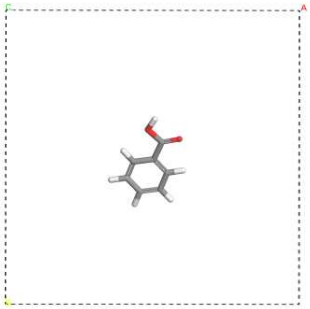
4.3.1. Crystal Structure Optimisation

The molecular crystal relaxed to lattice parameters: *a* - 5.35, *b* - 5.12, *c* - 21.54 Å, ß – 100.03°, all of which lie within 3.5% of the experimental parameters given in Table 1. This is within the standard 1 to 5% accuracy range of typical DFT calculations and

confirms the suitability of the PBE-GGA exchange-correlation functional and the TS dispersive terms.

4.3.2. Solution Cluster Optimisation

The initial pre-DFT, i.e. MD parent structures and the post-DFT optimised structures are shown in Figure 9.

 A	
H-bonds: 1.82, 1.71 Å	H-bonds: 1.62, 1.55 Å
 B	
C-OH angle: 100.07° C-C-H angle: 118.11°	C-OH angle: 104.95° C-C-H angle: 119.79°

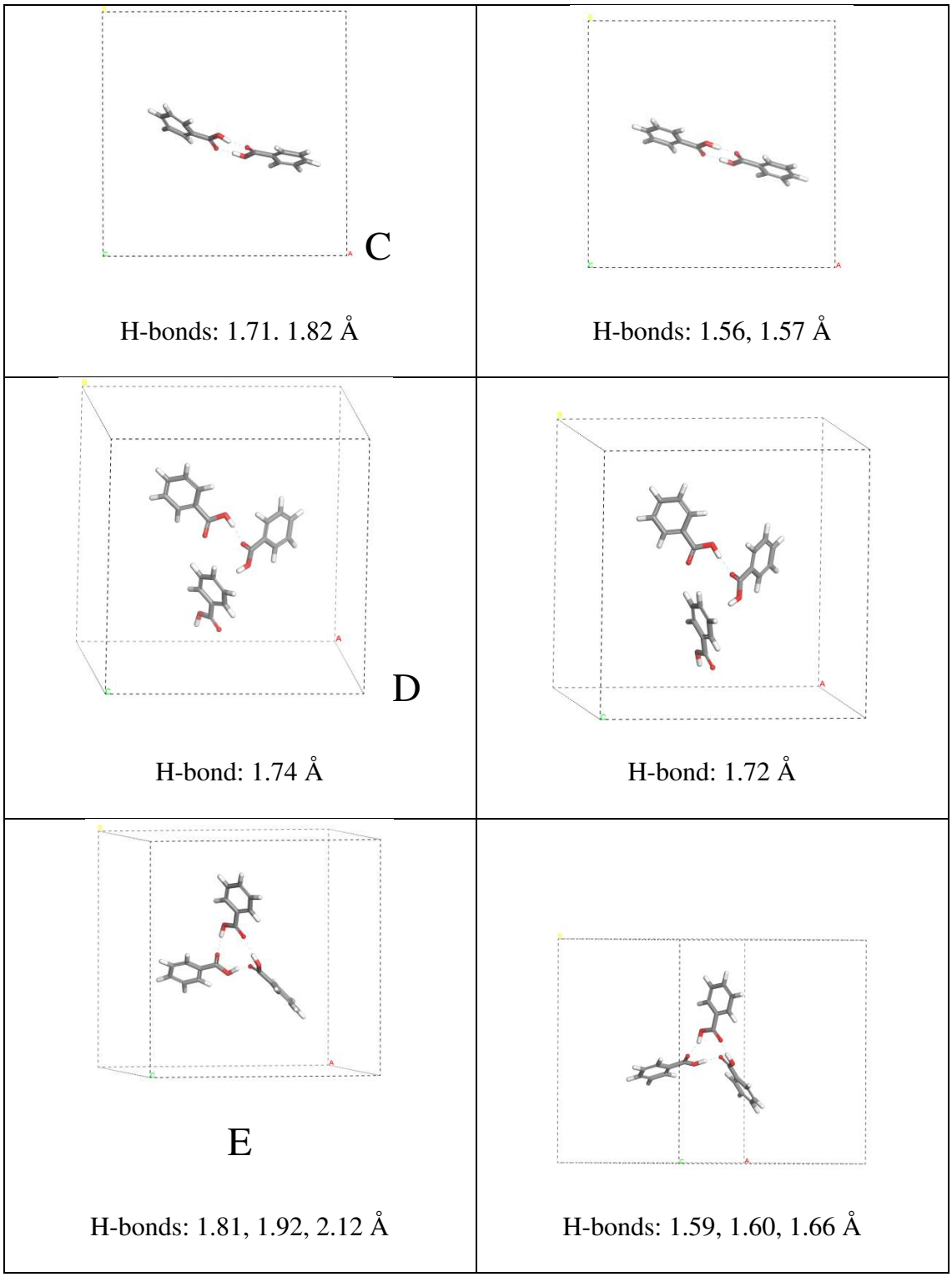


Figure 9: Pre-DFT i.e., 'MD parent structure' (LHS) and post-DFT (RHS) i.e., optimised structures of benzoic acid in hexane solution (A) and gas-phase states: single molecule (B), dimer (C), half-dimer (D) –see text for further explanation, catamer (E). All of these structures originated from a 60 Å^3 snapshot of an equilibrated MD simulation as described in the previous section.

Figure 9, row A, represents an intermolecular interaction between benzoic acid molecules, found from the MD trajectory file, which is similar to the solid-state synthon

A from Section 4.2. The DFT optimisation of this cluster shows a flattening of the plane of the solute dimer, compared to the same when directly extracted from the MD simulation. This flattening of the solute-solute H-bonding dimer is also observed when it is optimised by itself in the gas phase (Figure 9, row C). There is also a decrease in the length of both hydrogen bonds that accompanies a general contraction of the whole dimer/hexane region. Given the lack of any notable structural differences in the MD- and DFT-optimised dimers, continuing the investigation of the structures in the gas-phase is justified, with the consequent benefit of reduced computational cost.

Figure 9, row B demonstrates that the optimized single molecule of benzoic acid becomes planar upon DFT optimisation, compared to the approximately 10° torsion of the COOH group found from direct extraction from the MD structure.

The trimer gas-phase structure (Figure 9, row D) shows a 'half' H:O H-bond where a third benzoic acid molecule has inserted itself almost orthogonally and disrupted the classic homo-dimer structure, such as in row C. On relaxation the distance between H-bonding molecules and the unbonded molecule increases by 0.12Å, whereas the single H-bond contracts by an insignificant 0.02Å. The COOH torsions in the MD structure range from 4.6 to 11.7° and those in the DFT-optimised structure from 0.75 to 4.1°.

Figure 9, row E shows a H-bonded trimer structure, whereby on DFT relaxation all of the H-bonds shrink and the catamer orients itself into a more pronounced tetrahedral configuration than is apparent in the MD structure. This could be due to the absence of other catamers and hexane molecules allowing greater freedom in the gas-phase, DFT model. The COOH torsions in the MD structure range from 8.6 to 17.3°, and those in the DFT-optimised structure from 0.3 to 1.5°.

The slight changes in benzoic acid conformation and intermolecular distances from the structures extracted directly from the MD, in comparison to the same structures post DFT

gas phase optimisation, could be due to the presence of the hexane molecules in the MD simulation which are not present in the DFT simulation. However, the similarity in geometry of the benzoic acid OH…O homo-dimer when optimised in the presence of hexane molecules (Figure 9, row A) and when optimised without the hexane molecules (Figure 9, row C) suggests that this is not the case. It is more likely that the kinetic energy of the finite energy MD simulation, compared to a zero kelvin DFT simulation, results in fluctuations in the molecular conformation and intermolecular distances. These discrepancies in the exact structure between the MD and DFT simulations is reflected in previous studies of methylene blue/kaolinite clay mineral systems when described by the two different approaches[67]. There could also be a case for further examination of the forcefield description of the system, however that is outwith this study.

The lack of major structural change in the solute clusters, extracted from the MD trajectory, when optimised using DFT suggests that these can be taken forward and used to predict IR spectra.

5.3.2. Cross Validation of Simulations with Experimental IR Spectra

Figure 10 shows the calculated IR spectra for solid state benzoic acid and the gas-phase systems shown in Figure 9 (B to E), together with some highlighted frequencies, corresponding to the domain of focus in the experimental work of Davey *et al.*[35].

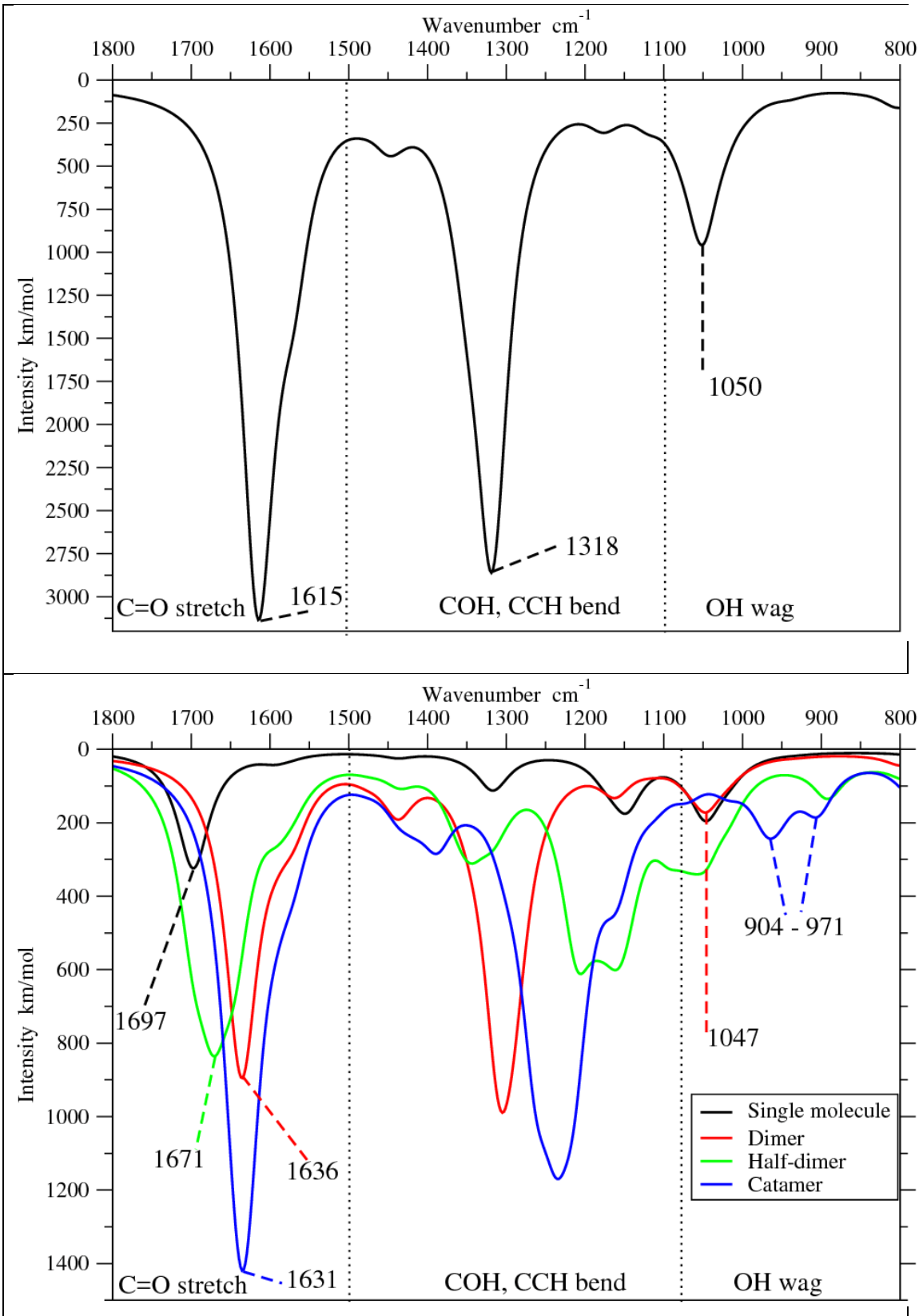


Figure 10: Top: Simulated IR spectra of solid state benzoic acid; bottom: gas-phase benzoic acid in the configurations B to E shown in Figure 9 (i.e. solution state proxy). The frequency of the peak for the single molecule at 1047 cm^{-1} is not an OH wag but COH and CCH bending

Davey et al (Figure 3(a), reference 31) identified the asymmetric carbonyl stretch in the experimental solid state at 1684 cm^{-1} and an out-of-plane O-H wag at 935 cm^{-1}, which were deemed consistent with OH…O H-bonding homo-dimers. The solid state simulations show the asymmetric carbonyl stretch at 1615 cm^{-1} and an out-of-plane, O-H wag at 1050 cm^{-1}, in reasonable agreement with the experimental data. In solution, the experimental asymmetric carbonyl stretch shifted to 1697 cm^{-1} and the out-of-plane O–H wag to 934 cm^{-1}, whereby the minimal change in these peaks in comparison to the solid-state structure were deemed to be consistent with the OH…O homo-dimers remaining associated in solution. In the solution simulations, (here simulated by the gas-phase systems as previously justified) the asymmetric carbonyl stretch of the dimer shifted to 1636 cm^{-1} and the out-of-plane O–H wag to 1047 cm^{-1}.

The simulated solid-state spectra and the simulated cluster spectra were examined in more detail to characterise the modes that contributed to the peaks shown in Figure 10. The wavenumbers for the major peaks in each spectra are shown in Table 3, along with the identified IR mode.

Table 3: Wavenumbers (cm-1) corresponding to wagging, bending and stretching modes of OH, COH + CCH and C=O bonds respectively, of the systems shown in Figure 9 (B to E) as well as of the solid state crystal of benzoic acid.

System	OH Wagging	COH, CCH Bending	C=O Stretching
Experimental solid-state[35]	935	unassigned	1684
Solid-state (simulated)	1049 - 1052	1318 - 1452	1615
Experimental solvated-state[35]	934	unassigned	1697
Single molecule (B) (simulated)	----------	1047 - 1317	1697
Dimer (C) (simulated)	1047	1290 – 1438	1636
Half-dimer (D) (simulated)	----------	1048 – 1351	1645 – 1697
Catamer (E) (simulated)	904 – 971	1159 - 1417	1631 – 1642

The simulated IR spectra for systems B to E generally show three distinct regions between 800 and 1800 cm^{-1}, encompassing OH wagging, COH and CCH bending and C=O stretching. The experimental spectra also shows multiple peaks in these areas which are challenging to de-convolute. It is noticeable from the spectra and from Table 3 that the frequency of the peak for the single molecule (Figure 10, black) at 1047 cm^{-1} is not an

OH wag but COH and CCH bending. Indeed the single molecule and the half-dimer (green) do not display OH wagging, unlike the solid state crystal, the dimer and the catamer. The presence of such peaks in the experimental solution state spectra confirms the likelihood that the COOH groups are indeed aggregated in solution. Interestingly, the experimental spectra shows a shoulder on the identified OH wagging 934cm^{-1} peak, which could be consistent with the doublet OH wagging frequency calculated from the catamer IR spectra (blue).

Although these simulated spectra do not agree quantitatively with Davey *et al.*[35] they do agree qualitatively and serve to illustrate the additional information it is possible to extract from detailed simulations using a range of tools. Where Davey *et al.*[35] identified dimers in solution, the combined analysis presented here suggests that the benzoic acid molecules are indeed aggregated through interactions between the COOH groups in multiple cluster types. Similar peaks are calculated for classic homo-dimer interactions and the three membered catamer interactions, however the observation of the shoulder on the experimental peak at 934cm^{-1} indeed suggests that the three membered H-bonding aggregates are likely, although not previously hypothesised for benzoic acid.

There are many potential reasons for the lack of numerical agreement between experiment and simulation, not least of which is the purity of the modelled systems. There are no extraneous factors or materials apparent in the models, neither crystal defects nor impurities whose potential effect on absorbance is unclear. In addition there could be modelling artefacts exaggerating numerical differences. However, in the context of this study where an example of the application of methods to a system of interest is the focus rather than the end results themselves, the DFT-generated IR spectra show sufficient agreement with the experimental work of Davey *et al.*[35], to illustrate the complementarity of the suite of DL-methods, Synthonic Engineering and DFT.

5. Conclusions

This multi-scale and multi-technique study has yielded intimate details of how the benzoic acid COOH groups aggregate in hexane solutions, whilst the hexane molecules impede the less polar solute-solute interactions. The correlation of the synthonic solution chemistry to the synthonic solid-state chemistry reveals that it is likely the OH…O H-bonding interactions drive the crystallisation of the singular benzoic acid polymorph.

The detailed description of the interatomic interactions, provided by the DANAI approach, combined with molecular geometry snapshots from the trajectory file uncovers that the COOH interactions are not just the classic homo-dimer interaction, but also three and four membered H-bonded ring interactions. DFT simulated IR spectra of the dimeric and trimeric H-bonding clusters provided increased clarity to the experimentally published IR data by characterising the exact modes that give rise to the multiple peaks in the spectra. Such depth of analysis can assist in the identification of solute clustering from spectroscopic data.

This study demonstrates a new platform for user-friendly input file construction, molecular dynamics simulations and synthon analysis of the trajectories of organic molecular solutions. The synergy of these calculations with high level electronic structure DFT calculations provides further granulation of the solution chemistry and experimental results, demonstrating how a multi-technique approach to exploring molecular solution chemistry can provide detailed structural information on the transition pathway from solution to crystal, inaccessible from a single modelling technique.

We intend to extend this initial case study to the analysis of more complex molecules, along with multiple concentrations and solvents.

Acknowledgment

We gratefully acknowledge the support of the Advanced Manufacturing Supply Chain Initiative through the funding of the 'Advanced Digital Design of Pharmaceutical Therapeutics' (Grant No. 14060) project in terms of supporting pharmaceutical crystallisation and modelling research at Leeds. We are also grateful to the funding of the EPSRC for some of the development of DL_ANALYSER under the auspices of the EPSRC's Collaborative Computational Project No. 5 (CCP5), of grant no: EP/M022617/1.

We are also grateful to the EPSRC for the support of crystallisation research at Leeds and Manchester through the award of a Critical Mass grant 'Molecules, Clusters and Crystals' (EP/I014446/1).

References:

(1) EPSRC Grand Challenge Network in Directed Assembly, *Web http://beyondthemolecule.org.uk/d6/,accessed 12/3/14*.
(2) Desiraju, G. R., DESIGNER CRYSTALS: INTERMOLECULAR INTERACTIONS, NETWORK STRUCTURES AND SUPRAMOLECULAR SYNTHONS. *Acta Crystallographica a-Foundation and Advances* **1996,** 52, C2-C2.
(3) Thalladi, V. R.; Goud, B. S.; Hoy, V. J.; Allen, F. H.; Howard, J. A. K.; Desiraju, G. R., Supramolecular synthons in crystal engineering. Structure simplification, synthon robustness and supramolecular retrosynthesis. *Chemical Communications* **1996**, (3), 401-402.
(4) Etter, M. C., ENCODING AND DECODING HYDROGEN-BOND PATTERNS OF ORGANIC-COMPOUNDS. *Accounts of Chemical Research* **1990,** 23, (4), 120-126.
(5) Etter, M. C.; Macdonald, J. C.; Bernstein, J., GRAPH-SET ANALYSIS OF HYDROGEN-BOND PATTERNS IN ORGANIC-CRYSTALS. *Acta Crystallographica Section B-Structural Science* **1990,** 46, 256-262.
(6) Roberts, K. J.; Hammond, R. B.; Ramachandran, V.; Docherty, R., Synthonic engineering: from molecular and crystallographic structure to the rational design of pharmaceutical solid dosage forms. In *Computational Approaches in Pharmaceutical Solid State Chemistry* Abramov, Y., Ed. Wiley: New Jersey, USA, 2016.
(7) Pickering, J.; Hammond, R. B.; Ramachandran, V.; Soufian, M.; Roberts, K. J., Synthonic Engineering Modelling Tools for Product and Process Design. In *Engineering Crystallography: From Molecule to Crystal to Functional Form*, Roberts, K. J.; Docherty, R.; Tamura, R., Eds. Springer Netherlands: Dordrecht, 2017; pp 155-176.
(8) Rosbottom, I.; Roberts, K. J.; Docherty, R., The solid state, surface and morphological properties of p-aminobenzoic acid in terms of the strength and directionality of its intermolecular synthons. *CrystEngComm* **2015,** 17, (30), 5768-5788.
(9) Moldovan, A. A.; Rosbottom, I.; Ramachandran, V.; Pask, C. M.; Olomukhoro, O.; Roberts, K. J., Crystallographic Structure, Intermolecular Packing Energetics, Crystal Morphology and Surface Chemistry of Salmeterol Xinafoate (Form I). *Journal of Pharmaceutical Sciences* **2017,** 106, (3), 882-891.
(10) Nguyen, T. T. H.; Rosbottom, I.; Marziano, I.; Hammond, R. B.; Roberts, K. J., Crystal Morphology and Interfacial Stability of RS-Ibuprofen in Relation to Its Molecular and Synthonic Structure. *Crystal Growth & Design* **2017,** 17, (6), 3088-3099.
(11) Rosbottom, I.; Ma, C. Y.; Turner, T. D.; O'Connell, R. A.; Loughrey, J.; Sadiq, G.; Davey, R. J.; Roberts, K. J., Influence of Solvent Composition on the Crystal Morphology and Structure of p-Aminobenzoic Acid Crystallized from Mixed Ethanol and Nitromethane Solutions. *Crystal Growth & Design* **2017,** 17, (8), 4151-4161.
(12) Rosbottom, I.; Roberts, K. J., Crystal Growth and Morphology of Molecular Crystals. In *Engineering Crystallography: From Molecule to Crystal to Functional Form*, Roberts, K. J.; Docherty, R.; Tamura, R., Eds. Springer Netherlands: Dordrecht, 2017; pp 109-131.
(13) Ramachandran, V.; Murnane, D.; Hammond, R. B.; Pickering, J.; Roberts, K. J.; Soufian, M.; Forbes, B.; Jaffari, S.; Martin, G. P.; Collins, E.; Pencheva, K., Formulation Pre-screening of Inhalation Powders Using Computational Atom–Atom Systematic Search Method. *Molecular Pharmaceutics* **2014,** 12, (1), 18-33.
(14) Kashchiev, D.; Vekilov, P. G.; Kolomeisky, A. B., Kinetics of two-step nucleation of crystals. *The Journal of Chemical Physics* **2005,** 122, (24), 244706.
(15) Vekilov, P. G., Two-step mechanism for the nucleation of crystals from solution. *Journal of Crystal Growth* **2005,** 275, (1–2), 65-76.

(16) Hamad, S.; Moon, C.; Catlow, C. R. A.; Hulme, A. T.; Price, S. L., Kinetic Insights into the Role of the Solvent in the Polymorphism of 5-Fluorouracil from Molecular Dynamics Simulations. *The Journal of Physical Chemistry B* **2006,** 110, (7), 3323-3329.
(17) Erdemir, D.; Lee, A. Y.; Myerson, A. S., Nucleation of Crystals from Solution: Classical and Two-Step Models. *Accounts of Chemical Research* **2009,** 42, (5), 621-629.
(18) Vekilov, P. G., Nucleation. *Crystal Growth & Design* **2010,** 10, (12), 5007-5019.
(19) Vekilov, P. G., The two-step mechanism of nucleation of crystals in solution. *Nanoscale* **2010,** 2, (11), 2346-2357.
(20) Jawor-Baczynska, A.; Moore, B. D.; Lee, H. S.; McCormick, A. V.; Sefcik, J., Population and size distribution of solute-rich mesospecies within mesostructured aqueous amino acid solutions. *Faraday Discussions* **2013,** 167, 425-440.
(21) Jawor-Baczynska, A.; Sefcik, J.; Moore, B. D., 250 nm Glycine-Rich Nanodroplets Are Formed on Dissolution of Glycine Crystals But Are Too Small To Provide Productive Nucleation Sites. *Crystal Growth & Design* **2013,** 13, (2), 470-478.
(22) Di Tommaso, D.; Watson, K. L., Density Functional Theory Study of the Oligomerization of Carboxylic Acids. *The Journal of Physical Chemistry A* **2014,** 118, (46), 11098-11113.
(23) Gaines, E.; Maisuria, K.; Di Tommaso, D., The role of solvent in the self-assembly of m-aminobenzoic acid: a density functional theory and molecular dynamics study. *CrystEngComm* **2016,** 18, (16), 2937-2948.
(24) DL_Software is the collective term for a range of scientific software developed at the Daresbury Laboratory, spanning across multi-length and -time scales. https://www.scd.stfc.ac.uk/Pages/Materials-Modelling-Software.aspx.
(25) Todorov, I. T.; Smith, W.; Trachenko, K.; Dove, M. T., DL_POLY_3: new dimensions in molecular dynamics simulations via massive parallelism. *Journal of Materials Chemistry* **2006,** 16, (20), 1911-1918.
(26) Yong, C. W., Descriptions and Implementations of DL_F Notation: A Natural Chemical Expression System of Atom Types for Molecular Simulations. *Journal of Chemical Information and Modeling* **2016,** 56, (8), 1405-1409.
(27) Yong, C.; Todorov, I., DL_ANALYSER Notation for Atomic Interactions (DANAI): A Natural Annotation System for Molecular Interactions, Using Ethanoic Acid Liquid as a Test Case. *Molecules* **2018,** 23, (1), 36.
(28) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **1983,** 4, (2), 187-217.
(29) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society* **1995,** 117, (19), 5179-5197.
(30) Damm, W.; Frontera, A.; Tirado–Rives, J.; Jorgensen, W. L., OPLS all-atom force field for carbohydrates. *Journal of Computational Chemistry* **1997,** 18, (16), 1955-1970.
(31) Mayo, S. L.; Olafson, B. D.; Goddard, W. A., Dreiding - A Generic Force-Field For Molecular Simulations. *J Phys Chem* **1990,** 94, (26), 8897-8909.
(32) Sun, H.; Mumby, S. J.; Maple, J. R.; Hagler, A. T., An ab Initio CFF93 All-Atom Force Field for Polycarbonates. *Journal of the American Chemical Society* **1994,** 116, (7), 2978-2987.
(33) Dauber-Osguthorpe, P.; Roberts, V. A.; Osguthorpe, D. J.; Wolff, J.; Genest, M.; Hagler, A. T., Structure and energetics of ligand binding to proteins: Escherichia coli

dihydrofolate reductase-trimethoprim, a drug-receptor system. *Proteins: Structure, Function, and Bioinformatics* **1988,** 4, (1), 31-47.
(34) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F., Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *European Biophysics Journal* **2011,** 40, (7), 843.
(35) Davey, R. J.; Dent, G.; Mughal, R. K.; Parveen, S., Concerning the relationship between structural and growth synthons in crystal nucleation: Solution and crystal chemistry of carboxylic acids as revealed through IR spectroscopy. *Crystal Growth & Design* **2006,** 6, (8), 1788-1796.
(36) Burton, R. C.; Ferrari, E. S.; Davey, R. J.; Finney, J. L.; Bowron, D. T., The Relationship between Solution Structure and Crystal Nucleation: A Neutron Scattering Study of Supersaturated Methanolic Solutions of Benzoic Acid. *Journal of Physical Chemistry B* **2010,** 114, (26), 8807-8816.
(37) Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M., Integrated Modeling Program, Applied Chemical Theory (IMPACT). *Journal of Computational Chemistry* **2005,** 26, (16), 1752-1780.
(38) Infantes, L.; Chisholm, J.; Motherwell, S., Extended motifs from water and chemical functional groups in organic molecular crystals. *CrystEngComm* **2003,** 5, (85), 480-486.
(39) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **2010,** 31, (4), 671-690.
(40) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004,** 25, (9), 1157-1174.
(41) Holmback, X.; Rasmuson, A. C., Size and morphology of benzoic acid crystals produced by drowning-out crystallisation. *Journal of Crystal Growth* **1999,** 198, 780-788.
(42) Sullivan, R. A.; Davey, R. J.; Sadiq, G.; Dent, G.; Back, K. R.; ter Horst, J. H.; Toroz, D.; Hammond, R. B., Revealing the Roles of Desolvation and Molecular Self-Assembly in Crystal Nucleation from Solution: Benzoic and p-Aminobenzoic Acids. *Crystal Growth & Design* **2014,** 14, (5), 2689-2696.
(43) McArdle, P.; Hu, Y.; Lyons, A.; Dark, R., Predicting and understanding crystal morphology: the morphology of benzoic acid and the polymorphs of sulfathiazole. *CrystEngComm* **2010,** 12, (10), 3119-3125.
(44) Liang, Z. Z.; Chen, J. F.; Ma, Y.; Wang, W.; Han, X. L.; Xue, C. Y.; Zhao, H., Qualitative rationalization of the crystal growth morphology of benzoic acid controlled using solvents. *Crystengcomm* **2014,** 16, (27), 5997-6002.
(45) Clydesdale, G.; Docherty, R.; Roberts, K. J., HABIT - a program for predicting the morphology of molecular crystals. *Comput Phys Commun* **1991,** 64, (2), 311-328.
(46) Docherty, R.; Clydesdale, G.; Roberts, K. J.; Bennema, P., Application Of Bravais-Friedel-Donnay-Harker, Attachment Energy And Ising-Models To Predicting And Understanding The Morphology Of Molecular-Crystals. *J Phys D Appl Phys* **1991,** 24, (2), 89-99.
(47) Momany, F. A.; Carruthers, L. M.; McGuire, R. F.; Scheraga, H. A., Intermolecular potentials from crystal data. III. Determination of empirical potentials and

application to the packing configurations and lattice energies in crystals of hydrocarbons, carboxylic acids, amines, and amides. *The Journal of Physical Chemistry* **1974,** 78, (16), 1595-1620.
(48) Gasteiger, J.; Marsili, M., New Model For Calculating Atomic Charges In Molecules. *Tetrahedron Letters* **1978**, (34), 3181-3184.
(49) Gasteiger, J.; Marsili, M., Iterative Partial Equalization Of Orbital Electronegativity - A Rapid Access To Atomic Charges. *Tetrahedron* **1980,** 36, (22), 3219-3228.
(50) Berendsen, H. J. C.; Postma, J. P. M.; Gunsteren, W. F. v.; DiNola, A.; Haak, J. R., Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **1984,** 81, (8), 3684-3690.
(51) Piana, S.; Gale, J. D., Understanding the barriers to crystal growth: Dynamical simulation of the dissolution and growth of urea from aqueous solution. *Journal of the American Chemical Society* **2005,** 127, (6), 1975-1982.
(52) Biovia, D. S. *Biovia Materials Studio 2016.0*, Dassault Systèmes: San Diego, 2016.
(53) Clark, S. J.; Segall, M. D.; Pickard, C. J.; Hasnip, P. J.; Probert, M. I. J.; Refson, K.; Payne, M. C., First principles methods using CASTEP. *Z. Kristallogr.* **2005,** 220, 567-570.
(54) Hohenberg, P.; Kohn, W., Inhomogeneous electron gas. *Phys. Rev.* **1964,** 136, 864-871.
(55) Kohn, W.; Sham, L. J., Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **1965,** 140, A1133-A1138.
(56) Payne, M. C.; Teter, M. P.; Allan, D. C.; Arias, T. A.; Joannopoulous, J. D., Iterative minimization for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Reviews of Modern Physics* **1992,** 64, 1045-1097.
(57) Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized gradient approximation made simple. *Physical Review Letters* **1996,** 77, 3865-3868.
(58) McNellis, E. R.; Meyer, J.; Reuter, K., Azobenzene at coinage metal surfaces: Role of dispersive van der Waals interactions. *Phys. Rev. B* **2009,** 80, 205414.
(59) Tkatchenko, A.; Scheffler, M., Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009,** 102, 0730051-4.
(60) Monkhorst, H. J.; J.D., P., Special points for Brillouin-zone integrations*. *Phys. Rev. B* **1976,** 13, 5188-5192.
(61) Pfrommer, B. G.; Côté, M.; Louie, S.; Cohen, M. L., Relaxation of crystals with the Quasi-Newton Method. *Journal of Computational Physics* **1997,** 131, 233-240.
(62) Refson, K.; Clark, S. J.; Tulip, P. R., Variational density functional perturbation theory for dielectrics and lattice dynamics. *Phys. Rev. B* **2006,** 73, 155114.
(63) Ribeiro da Silva, M. A. V.; Monte, M. J. S.; Santos, L. M. N. B. F., The design, construction, and testing of a new Knudsen effusion apparatus. *The Journal of Chemical Thermodynamics* **2006,** 38, (6), 778-787.
(64) Di Tommaso, D., The molecular self-association of carboxylic acids in solution: testing the validity of the link hypothesis using a quantum mechanical continuum solvation approach. *CrystEngComm* **2013,** 15, (33), 6564-6577.
(65) Sullivan., R. A.; Davey., R. J.; Sadiq., C.; ter Horst., J. H.; Dent., G.; Back., K. R.; Toroz., D.; Hammond., R. B., *Crystal Growth & Design. Revealing the Roles of Desolvation and Molecular Self-Assembly in Crystal Nucleation from Solution: Benzoic and p -Aminobenzoic Acids* **2014,** 14, 2689.

(66) 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. *Chemical Communications* **2017,** 53, (56), 7905-7908.
(67) Greathouse, J. A.; Geatches, D. l.; Pike, D. Q.; Greenwell, H. C.; Johnston, C. T.; Wilcox, J.; Cygan, R. T., Methylene Blue Adsorption on the Basal Surfaces of Kaolinite: Structure and Thermodynamics from Quantum and Classical Molecular Simulation. *Clays and Clay Minerals* **2015,** 63, (3), 185-198.