CrystEngComm

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: E. Gaines, K. Maisuria and D. Di Tommaso, *CrystEngComm*, 2016, DOI: 10.1039/C6CE00130K.

This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/crystengcomm

View Article Online
DOI: 10.1039/C6CE00130K

CrystEngComm

ARTICLE

The role of solvent in the self-assembly of m-aminobenzoic acid: a density functional theory and molecular dynamics study

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Etienne Gaines, Krina Maisuria and Devis Di Tommaso*

Solvent can have significant effects on the solution thermodynamics and crystallisation kinetics of organic compounds from solution. In the present work, the early stages of aggregation of the organic molecule m-aminobenzoic (mABA) in two different solvents, dimethyl sulfoxide (DMSO) and water were studied using a combination of quantum chemistry, molecular dynamics and metadynamics simulations. Density functional theory (B97-D and M06-2X) calculations with the continuum solvation SMD model were used to probe the potential energy surface of molecular clusters of m-aminobenzoic acid, $(mABA)_n$ (n = 2-4), locate their low-lying energy structures, and compute the Gibbs free energies of $(mABA)_n$ in solution. Starting from a large number of randomly generated candidate structures and by imposing the condition of minimum free energy in solution for the isomers of $(mABA)_n$, we proved that the most stable dimers and tetramers in solution correspond to the classic carboxylic dimer π–π stacking synthon found in the crystalline form-II of mABA. Molecular dynamics simulations of mABA solutions at different concentrations reveal a significant solvent-dependent aggregation behaviour for mABA: in water, even at low concentrations, mABA molecules spontaneously form H-bonded π-π stacking molecular clusters whereas in organosulfur solutions the molecules of mABA are in a more solvated state. Metadynamics simulations and microsolvation density functional theory calculations rationalize the low level of mABA aggregation in DMSO in terms of the energetic barrier for the desolvation of mABA molecules and formation of dimers, and the strength of mABA-solvent interactions, which are both larger in DMSO compared with water. This work shows that the solvent and its specific interaction with the organic solute molecules influence both the thermodynamics and kinetics of the molecular self-assembly process.

Introduction

In the preparation of pharmaceutical materials that require a well-defined crystal form, the phenomenon of polymorphism (the ability of a given type of molecule to crystallize in different structures) can be a threat: polymorphs can have significantly different solid-state properties and therefore different performances in materials applications, or bioavailability and stability as a drug substance.[1,2] Consequently, one of the major challenges in the synthesis of pharmaceuticals is the selection of a particular molecular crystal polymorph during solution crystallisation.[3]

It is well known from experimental evidence that the nature of the solvent, or the presence in solution of additives and impurities, can determine the formation of one specific polymorph over another.[4] In particular, several studies have suggested that during the first stages of crystallization (clustering and nucleation) the solute-solvent interactions can control the polymorphic outcome.[5,6] As the process of molecular self-assembly to form the crystallization growth unit is significantly influenced by the interaction between the solute and solvent molecules, i.e. by the properties of the solvation environment, there is considerable interest, both fundamental and technological, in understanding how the processes of clustering, nucleation and subsequent growth of a molecular crystal are affected by the solution chemistry.[7,8] However, even advanced spectroscopic techniques such as synchrotron radiation x-ray scattering or atomic force microscopy cannot give a complete understanding of these processes.[9] The interpretation of experimental measurements and, most importantly, the ability to predict how a new crystal phase nucleates and then grows increasingly depends on the use of computer simulations.[10]

Significant work has been done in the development and application of simulation techniques to model the nucleation and growth from solution.[11-14]

In particular, a recent theoretical investigation has revealed a direct relation between the computed strength of the solute-solvent interaction and the experimental rate of nucleation of organic molecules.[16] This study, however, was based on static density functional theory (DFT) gas-phase calculations of microsolvated organic molecules, which do not

School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, UK. E-mail: d.ditommaso@qmul.ac.uk; Tel: +44 (0)20 7882 6226

Electronic Supplementary Information (ESI) available: GAFF parameters for mABA and DMSO. Structures and energies of mABA clusters. Snapshot of the last configuration of molecular dynamics simulations of mABA solutions in DMSO and water. Hydrogen bond analysis of molecular dynamics simulations. Free energy profile for the mABA dimerization. See DOI: 10.1039/x0xx00000x

View Article Online
DOI: 10.1039/C6CE00130K

fully account for the thermodynamic and kinetic factors affecting the aggregation of organic molecules from solution.

Aiming at obtaining a better understanding of the role of solvent in the kinetics and thermodynamics of the early stages crystallisation from solution, we have conducted a comprehensive investigation of the aggregation of m-aminobenzoic acid (mABA) in two different solvents, dimethyl sulfoxide (DMSO) and water, using a combination of computational methods based on quantum chemistry solvation continuum, molecular dynamics and free energy techniques.

The molecule mABA is of considerable importance in the pharmaceutical industry for the synthesis of analgesics, antihypertensives and vasodilators, but it is also a fascinating model for polymorphic research due to its ability to crystallize in five different crystal forms, whose nucleation depends chiefly on the solvent.[17] The clusterization of mABA molecules is also of interest because it can form complex mABA/mABA inter-molecular interactions, i.e. XH/O and XH/π (X = O, N) hydrogen bonds and π–π interaction, all of which are important generally during polymorph selection.

Having introduced in Section 2 the details of the simulations, we next report DFT calculations of the thermodynamics of formation of the oligomers of m-aminobenzoic acid $(mABA)_n$ (n = 2-4. Results from molecular dynamics simulations of the aggregation of mABA in water and in DMSO are then reported, and the observed solvent-dependent aggregation behaviour of mABA is rationalized in terms of the free energy profile of mABA dimerization and the strength of mABA-solvent interactions.

Methods

Density functional theory calculations

Electronic structure calculations were carried out with the NWChem (version 6.3) [18] and Gaussian09 [19] codes. We used the Grimme's density functional including dispersion (B97-D)[20] and the Minnesota 06 global hybrid functional with 54% HF exchange (M06-2X).[21] The Gaussian 6-31+G(d,p) basis set was used throughout these simulations as this provides a good compromise between accuracy and computational cost.[21-23] Thermal contributions were calculated at the optimised geometries using the gas-phase harmonic frequencies, which were scaled by a factor of 0.979 in order to account for systematic errors in the density functional and for anharmonicity.[24] Free energies of solvation were calculated using the SMD solvation model,[25] and the gas-phase optimized geometries.

Free energies of association in solution. The free energy of formation of carboxylic acid clusters were computed according to the following equation:

$$\Delta G^*_{ass} = G^*_{AB} - G^*_A - G^*_B \quad (1)$$

where G^*_X is the total Gibbs free energy of the species X (X = AB, A or B) in the liquid at 298 K, which was determined (see equation 2) by the addition of the gas-phase total electronic energy of the solute ($E_{e,gas}$), the vibrational-rotational-translational contribution to the gas-phase Gibbs free energy ($\delta G^{\circ}_{VRT,gas}$) at T = 298 K under a standard-state partial pressure of 1 atm, the solvation free energy of the solute corresponding to transfer from an ideal gas at a concentration of 1 mol L^{-1} to an ideal solution at a liquid-phase concentration of 1 mol L^{-1} (ΔG^*_{solv}), and the free energy change of 1 mol of an ideal gas from 1 atm to 1 mol L^{-1} ($RTln[\widetilde{R}T]$ = 1.89 kcal mol^{-1} at 298.15 K where $\widetilde{R}$ = 0.082052 K^{-1}):[26-30]

$$G^*_X = E_{e,gas} + \delta G^{\circ}_{VRT,gas} + \Delta G^*_{solv} + RTln[\tilde{R}T] \quad (2)$$

For multiple stationary points of a molecular cluster, the free energy was determined from the Boltzmann ensemble average

$$\langle G(X) \rangle = \sum_{i=1}^{N} f_i G(X_i) \quad (3)$$

where f_i is the Boltzmann factor corresponding to the i^{th} configuration, $G(X_i)$ is the corresponding free energy and N is the number of low-lying energy structures located on the potential energy surface (PES) of the molecular cluster. The Boltzmann factor was determined according to

$$f_i = \frac{e^{-G(X_i)/RT}}{\sum_j e^{-G(X_j)/RT}} \quad (4)$$

where R is the universal gas constant, T is the absolute temperature (T = 298 K) and the index j runs over all isomers.

To locate the most stable isomers of the molecular clusters $(mABA)_n$ (n = 2-4) we have adopted a modified version of a very effective computational protocol developed in our group to locate the most stable low-lying structures of weakly interacting molecular complexes:[31]

1) For each m-aminobenzoic cluster $(mABA)_n$, several *hundreds of thousands* of candidate structures were generated using Granada,[32,33] a code designed to randomly distribute one or more molecules around a central unit (a monomer, dimer, trimer etc.) placed at the centre of a cube of defined side length. For example, we generated candidate structures for the trimer $(mABA)_3$ by considering the random distribution of one mABA molecule around the most stable dimer (the cyclic dimeric structure), and from the random distribution of two mABA molecules around a central mABA molecule. The molecular units used to generate the candidate structures were fully optimized at the M06-2X/6-31+G(d,p) level of theory.

2) Using an in-house code, only those configurations such that at least one atom of each mobile molecule was within 4 Å from at least one atom of the central unit were selected as potential low-lying energy structures.

3) The energies of these structures were evaluated at the B97-D/6-31+G(d,p) level of theory and the Boltzmann factor f_i corresponding to the i^{th} configuration was determined as

$$f_i = \frac{e^{-(E_i - E_0)/RT}}{\sum_j e^{-(E_j - E_0)/RT}} \quad (5)$$

View Article Online
DOI: 10.1039/C6CE00130K

where E_i was the energy of the i^{th} candidate structure and E_0 was the energy of the most stable candidate structure. We then selected the candidate structures with a Boltzmann factor $f_i \geq 0.01$ and to increase our sampling we also selected ten to fifteen randomly selected structure such that $3 \leq E_i - E_0 \leq 15$ kcal mol^{-1}.

4) The full optimized geometries, thermochemical properties and solvation energies of the configurations selected at point 3) were computed at the M06-2X/6-31+G(d,p) level of theory.

We chose M06-2X because its assessment against representative databases showed that this method is one of the most accurate density functionals for applications involving a combination of main-group thermochemistry and noncovalent interactions.[34,35] Moreover, the application of the SMD solvation model with M06-2X to predict the free energies of aqueous solvation for 61 drug-like molecules in the SAMPL1 test set gave a mean unsigned error of only 2.0 kcal mol^{-1}.[36] However, M06-2X has shown small energy oscillations at distances away from the equilibrium.[37] For this reason, in the protocol discusses above the energies of the candidate structures were therefore evaluated using the B97-D functional, but the equilibrium geometries, thermochemical properties and solvation energies of the molecular clusters were computed with the M06-2X method. We have previously shown that the computational methodology used in the present study provides accurate predictions of the thermodynamic stability in solution of carboxylic acid clusters.[31]

DFT microsolvation modelling. The low-lying energy minima on the PES of the microsolvated zwitterionic, mABA$^{\pm}$(S), and non-zwitterionic, mABA(S)$_n$, forms of m-aminobenzoic acid (S = DMSO or H_2O; n = 1-3), were located using the following strategy:

1) Several *tens of thousands* of candidate conformations of the mABA(S)$_n$ and mABA$^{\pm}$(S)$_n$ clusters were generated using the Granada code.

2) The initial set was then reduced with a distance-based selection similar to that used for the generation of mABA molecular clusters.

3) The energy of the candidate structures and associated Boltzmann factors of each configuration were computed at the B97-D 6-31+G(d,p) level of theory.

4) For the conformations with a Boltzmann factor $f_i \geq 0.01$ the optimised geometries, thermochemical properties and solvation contributions were determined at the M06-2X/6-31+G(d,p) level of theory.

Cluster continuum approach to compute solvation free energies. The SMD continuum model can provide accurate solvation free energies for the non-zwitterionic form of mABA: using the SMD/M06-2X/6-31+G** method the mean unsigned error in the calculated solvation free energies of carboxylic acids was only 0.25 kcal mol^{-1} in water and 0.55 kcal mol^{-1} in organic solvents.[25] In contrast, due to strong electrostatic effects arising from unbalanced charges, solvation modeling of the zwitterionic form (mABA$^{\pm}$) is particularly challenging for continuum approaches.[38] To correct for these deficiencies, the solvation free energy of mABA$^{\pm}$ was determined using the cluster-continuum model:[39] the solvated molecule (mABA$^{\pm}$)(S)$_n$ (n = 0 – 3) was treated quantum mechanically (M06-2X/6-31+G**) and the bulk solvent was described with the SMD continuum model. According to this approach the solvation of mABA$^{\pm}$ is given by

$$\Delta G^*_{solv}(mABA^{\pm}) = \Delta G^o_{clust}(mABA^{\pm}(S)_n) + \Delta G^*_{solv}(mABA^{\pm}(S)_n) + n\Delta G_{vap}(S) \quad (6)$$

where $\Delta G^o_{clust}(mABA^{\pm}(S)_n)$ is the gas-phase free energy of cluster formation at 1 atm, $\Delta G^*_{solv}(mABA^{\pm}(S)_n)$ is the solvation free energy of the solvated cluster, and $\Delta G_{vap}(S)$ is the free energy of vaporization free energy of one solvent molecule. The last term was computed using the following equation:

$$\Delta G_{vap}(S) = -\Delta G^*_{solv}(S) - RTln[\tilde{R}T] - RTln[S] \quad (7)$$

where $\Delta G^*_{solv}(S)$ is the solvation energy of a solvent molecule and the last term is related to the density number of the solvent, which was computed by taking [DMSO] = 14.1 mol L^{-1} and [H_2O] = 55.5 mol L^{-1},[28] respectively. For the solvation free energy of DMSO we used the value computed at the SMD/MP2/def2-TZVPP level of theory (-7.75 kcal mol^{-1}).[40]

Classical molecular dynamics

Forcefield. The general AMBER forcefield (GAFF)[41] was used to model the mABA and DMSO molecules. The GAFF potential is well developed and it was previously used to compute processes of aggregation and crystal growth of organic molecules,[11,12,14,30,31] including *p*-aminobenzoic acid.[32,33] Water molecules were modelled using the SPC/E potential.[42] The interactions between mABA and DMSO molecules and between mABA and water molecules were described using the GAFF potential. To derive the forcefield parameters within the framework of the GAFF, the structures and molecular electrostatic potential (MEP) of mABA and DMSO were obtained using the Gaussian09 code. The *Antechamber* package was then used to compute partial charges according to the restrained electrostatic potential (RESP) formalism.[43] The selection of the quantum chemistry method and basis set to compute MEP is a key aspect in RESP derivation.[44] Here, we used the HF/6-31G* method, which was the level of theory applied for the derivation of RESP charges in the Cornel *et al.* forcefield,[45] and successive modifications of the AMBER potential.[29,38-40] Using this procedure, we obtained partial charges for DMSO consistent with the values reported by Dupradeau *et al*.[44] The GAFF forcefields and partial charges used to model mABA and DMSO are reported in Supporting Information (SI, Table SI.1.1 and SI.1.2).

Simulation details. Classical molecular dynamics (MD) simulations were performed using version 5.0.4 of the GROMACS molecular dynamics package.[46,47] The leapfrog algorithm with a time step of 2 fs was used to integrate the equations of motion. The isothermal-isobaric (constant NPT) ensemble was used to maintain a temperature of 300 K and a pressure of 1 bar. The velocity rescale thermostat and the isotropic Parrinello-Rahman barostat were used with 0.4 ps

View Article Online
DOI: 10.1039/C6CE00130K

and 2.0 ps as the thermostat and barostat relaxation times, respectively. The electrostatic forces were calculated by means of the particle-mesh Edwald approach with a cutoff of 1.2 nm. A 1.2 nm cutoff was also used for the van der Waals forces. The LINCS algorithm was applied at each step to preserve the bond lengths.

Simulation protocol. We performed MD simulations of mABA solutions in DMSO and water at different concentrations (see Table SI.2 in Supporting Information). Molecular models of mABA solutions were generated using the *insert-molecules* and *solvate* GROMACS utilities to insert the required number of mABA molecules in an empty cubic box of size 5 nm, and solvate them with DMSO or water. Each solution was at first minimized using the conjugate-gradient algorithm with a tolerance on the maximum force of 200 kJ mol^{-1}, and the temperature and volume of each system were equilibrated by running 100 ps of constant volume, constant temperature (NVT) simulation followed by 1 ns of NPT simulations. Production runs in the NPT ensemble were then conducted for tens to hundreds of ns. Details of the simulation times, number of solute and solvent molecules, and equilibrated values of the average cell length are reported in Supporting Information (Table SI.2).

Metadynamics simulations

The free energy profile for the formation of a carboxylic acid dimer (product) from two fully solvated *m*ABA molecules (reactant) was computed by means of the well-tempered metadynamics method.[43,44] We generated solutions containing two mABA molecules in 945 DMSO and 4058 water molecules, respectively. The free energy calculations were performed with GROMACS 5.0.4 equipped with the PLUMED 2.2b plugin.[48] Simulations were conducted in the NPT ensemble for 10 ns. The order parameter (OP) used to study the dimerization reaction was defined as follows:

$$OP = \frac{d_1 + d_2}{2} \quad (8)$$

where d_1 and d_2 are the two distances between the hydroxyl hydrogen (O–H) and the carbonyl oxygen (C=O) of a pair of mABA molecules.

Results and discussion

Equilibrium of zwitterionic and non-zwitterionic forms of mABA

The molecule m-aminonbenzoic acid can exist in both non-zwitterionic (mABA) and zwitterionic ($mABA^{\pm}$) forms, which distribution in solution depends chiefly on the solvent.[49] To interpret measurements of solid- and liquid- NMR spectra of mABA, it was hypothesized that in dimethyl sulfoxide solutions, molecules of mABA are predominantly non-zwitterionic.[50] This is not an obvious conclusion because the polarity index (P') of DMSO (P' = 7.2) is higher than most other solvents, including polar protic solvents such as methanol (P' = 5.1).[51]. In order to verify this hypothesis and determine which form of mABA should be modelled in this study, we computed the microscopic constants pertaining to the zwitterionic equilibrium:

$$K_z = \frac{[mABA^{\pm}]}{[mABA]} = \mathrm{e}^{-\frac{\Delta G_z^*}{RT}} \quad (9)$$

where ΔG_z^* is the free energy for the zwitterionic equilibrium mABA ⇄ $mABA^{\pm}$. According to Eq. 2, ΔG_z^* is given by

$$\Delta G_z^* = \Delta E_{e,gas} + \Delta\delta G^{\circ}_{VRT,gas} + \Delta\Delta G^*_{solv} \quad (10)$$

Reliable estimates of the free energy of this reaction depend on the accurate determination of each term in Eq. 10, but when the species involved in the equilibrium are ionic then the accurate evaluation of the solvation free energy term ($\Delta\Delta G^*_{solv}$) becomes particularly critical.[28]

Table 1 reports the solvation free energies of $mABA^{\pm}$ computed according the cluster continuum approach (see Eq. 6) using up to three explicit DMSO solvent molecules in the microsolvated $mABA^{\pm}(S)_n$ model. The most stable structures of the $mABA^{\pm}(DMSO)_n$ clusters are displayed in Figure 1.

Table 1. Solvation free energies in DMSO of the zwitterionic ($mABA^{\pm}$) form of *m*-aminobenzoic acid obtained using the cluster-continuum approach. Gas-phase free energy contributions computed at the M06-2X/6-31+G(d,p) level and solvation free energies computed at the SMD/M06-2X/6-31+G(d,p) level. Values in kcal mol^{-1}.

	$\Delta G^{o}_{clust}(mABA^{\pm}(S)_n)$	$\Delta G^{*}_{solv}(mABA^{\pm}(S)_n)$	$\Delta G^{*}_{solv}(mABA^{\pm})$
0			-56.55
1	-26.16	-19.34	-59.03
2	-39.53	-37.15	-62.20
3	-50.72	-31.25	**-62.72**

According to Pliego and Riveros,[39] a variational principle can be established in the cluster continuum approach for the choice of the number of explicit solvent molecules *n*: the value that minimises ΔG^*_{solv} is the ideal number of explicit solvent molecules and the solvation free energy for this value is the best calculated value for ΔG^*_{solv}. For $mABA^{\pm}(S)_n$ this corresponds to *n* = 3 and the best estimate of the free energy of solvation for $mABA^{\pm}$ is -62.72 kcal mol^{-1}. This high value for the solvation free energy of $mABA^{\pm}$ is related to the localization of the negative and positive charges in the carboxylic ($–COO^-$) and amino ($–NH_3^+$) groups, which strongly interact with the polar DMSO solvent molecules. In comparison, the experimental hydration free energies of the $HCOO^-$ and $CH_3NH_3^+$ ions are –74.6 kcal mol^{-1} and –75.2 kcal mol^{-1}, respectively.[39]

View Article Online
DOI: 10.1039/C6CE00130K

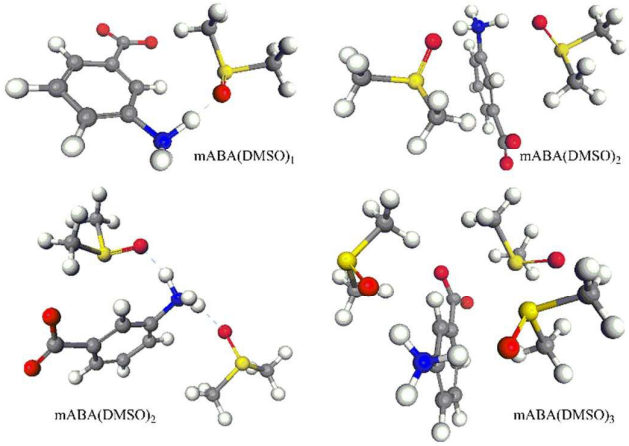


Figure 1. Optimized gas phase structures of the most stable solvated mABA±(DMSO)n (n = 1-3) clusters as computed at the M06-2X/6-31+G(d,p).

The energetics of the zwitterionic equilibrium mABA ⇄ $mABA^{\pm}$ (see Eq.10) are listed in Table 2. The solvation term $\Delta\Delta G^{*}_{solv}$ is given by the difference between the solvation free energies of $mABA^{\pm}$ (-62.72 kcal mol-1) and mABA (-11.29 kcal mol-1, SMD/M06-2X/6-31+G(d,p) level of theory), whereas the gas-phase energy contribution was computed using different density functionals (PBE, B3LYP, M06-2X) and at the ab initio MP2 level . All methods agree that the free energy of the zwitterionic equilibrium mABA ⇄ $mABA^{\pm}$ is positive (between 7.2 and 9.5 kcal mol-1), supporting the hypothesis made by Hughes and co-workers that in DMSO solutions molecules of mABA principally exist in the non-zwitterionic form.50 In water and at room temperature conditions mABA has been reported 50% in zwitterionic form (mABA±).52,53 However, since in the present work we are mostly concerned with the role of solvent in the aggregation of organic molecules, we have reduced the complexity of the system by focusing our attention to the aggregation of the non-zwitterionic form mABA from water and DMSO, and we have omitted from our models the presence of $mABA^{\pm}$ in aqueous solutions. Computer modelling of mixed mABA/$mABA^{\pm}$ solutions will be subject of future work.

Table 2. Energetics of the zwitterionic equilibrium $mABA \leftrightarrow mABA^{\pm}$. Gas-phase energies ($\Delta E_{e,gas}$) and standard state (1 atm) gas-phase free energies computed with different quantum chemistry methods. Geometries, frequencies and solvation energies computed at the M06-2X/6-31+G(d,p) level of theory. Values in kcal mol^{-1}.

Method	$\Delta E_{e,gas}$	$\Delta G^{\circ}_{z,gas}$	$\Delta\Delta G^{*}_{solv}$	$\Delta G^{*}_{z,dmso}$
M06-2X/6-31+G(d,p)	61.78	60.90	-51.43	9.47
M06-2X/aug-cc-pVTZ	61.57	60.69	-51.43	9.26
PBE/aug-cc-pVTZ	57.91	57.03	-51.43	5.60
B3LYP/aug-cc-pVTZ	60.00	59.12	-51.43	7.69
MP2/aug-cc-pVTZ	59.55	58.67	-51.43	7.24

Role of solvent in the energetics of mABA clusterization

Table 3 reports the energies and free energies for the formation, in the gas phase and in solution, of mABA clusters consisting of two (dimer), three (trimer) and four (tetramer) molecular units. Molecules of mABA can interact via H-bonding, π-π interactions, and NH/π and OH/π interactions. Consequently, for each $(mABA)_n$ there can be several low-lying energy isomers, which would be difficult to locate only by means of chemical intuition. For each $(mABA)_n$ cluster, the values in Table 3 represent the Boltzmann average of the energies, or free energies, of the most stable isomers located using the computational protocol outlined in the "Methods" section.

Dimers. Dimers are the first species that could form in solution during the molecular aggregation process, and there has been much debate on the link between the most populated dimers in solution and the structural motifs in polymorphs.[5,13,23,54] We determined the structures and free energies of 32 stable $(mABA)_2$ structures (SI, Tables SI.3.1 SI.3.2 and Figure SI.2.1), which were classified into 12 possible isomers on the PES of $(mABA)_2$ (see Figure 2). The energetics of formation of each type of dimer are reported in Supporting Information (Table SI.3.2). The classic carboxylic acid dimer I is the most stable dimer in both the gas-phase and solution. In particular, the formation of the cyclic dimeric structure I is highly exoergonic (ΔG^{*}_{ass}=-3.40 kcal mol^{-1}) in DMSO compared with water (ΔG^{*}_{ass}=-0.18 kcal mol^{-1}). For all other dimers (II-XII in Figure 2), where the mABA molecules interact via other types of H-bonding (e.g., NH_2···OH) or weaker π–π, OH/π and NH_2/π interactions, the energetics of formation is positive in both water and DMSO (see (Table SI.3.2). Therefore, according to our calculations, m-aminobenzoic acid molecules form thermodynamically stable cyclic carboxylic dimers in DMSO. These species correspond to the structural synthon found in the crystal structures II and V, the polymorphs where mABA exist in the non-zwiterionic form.[16,55]

View Article Online
DOI: 10.1039/C6CE00130K

Table 3. Energetics of formation of *m*-aminobenzoic acid clusters in the gas-phase and solution as computed at the M06-2X/6-31+G(d,p) level of theory. Values obtained from the from the Boltzmann average of the energies or free energies of the low-lying $(mABA)_n$ isomers. Values in kcal mol^{-1}.

species	reaction	$\Delta E_{e,gas}$	ΔG°_{ass}	ΔG^{*}_{ass}	
				H_2O	DMSO
dimer	mABA + mABA → $(mABA)_2$	-18.33	-6.64	-0.07	-3.40
trimer	$(mABA)_2$ + mABA → $(mABA)_3$ [a]	-13.59	2.12	1.22	5.34
	3 mABA → $(mABA)_3$ [a]	-31.95	-5.03	0.98	1.39
	3 mABA → $(mABA)_3$ [b]	-19.98	5.08	7.68	7.96
tetramer	$(mABA)_2$ + $(mABA)_2$ → $(mABA)_4$ [c]	-20.77	-0.78	-2.78	2.06
	$(mABA)_2$ + 2 mABA → $(mABA)_4$ [c]	-39.09	-7.41	-2.99	-1.36

[a] $(mABA)_3$ isomers generated from the condensation of the cyclic dimer I with a mABA molecule. [b] $(mABA)_3$ isomers generated from the condensation of three isolated mABA molecules. [c] $(mABA)_4$ isomers generated from the condensation of two cyclic dimers.

Figure 3 represents the coexistence profile of the free acid monomer, mABA, and dimer, $(mABA)_2$, in DMSO computed using the monomer-dimer model of Krishnan and co-workers to quantify the stoichiometric concentration of monomers and dimers in solution:[55]

$$[(mABA)_2] = \frac{\{-1 + (1 + 8K_D[C])^{1/2}\}^2}{16K_D} \quad (11)$$

In Eq. 11, K_D is the dimerization constant, $K_D = [(mABA)_2]/[mABA]^2$, [C] is the stoichiometric concentration of the solute, [C] = [mABA] + 2[$(mABA)_2$], and [mABA] and [$(mABA)_2$] are the monomer and dimer concentrations, respectively (mol L^{-1}). The value of K_D (304.9) has been computed from the Boltzmann average value of the free energies of dimerization in DMSO (ΔG^{*}_{ass}= −3.40 kcal mol^{-1}, Table 2). Figure 3 shows that for [C] beyond 0.01 mol L^{-1} (-log[C] = 2) dimers of mABA, and in particular the classic carboxylic dimers I, is the predominant species in solution. However, this result is in contrast to what was reported by Hughes and co-workers, who, to explain changes in the ^{13}C NMR chemical shifts during the crystallization process of mABA, suggested the formation in supersaturated solution of non-zwitterionic dimers linked by O-H···N hydrogen bonds.[50] According to our static DFT calculations, none of the other H-bonded clusters is stable in DMSO. However, higher-order molecular clusters than dimers could also be stable in solution.

Trimers. We modelled the formation of trimers $(mABA)_3$ in solution from the condensation reaction of the cyclic dimer I with a mABA molecule [$(mABA)_2$ + mABA → $(mABA)_3$] and the condensation of three isolated mABA molecules [3 mABA → $(mABA)_3$] (see Table 3). The optimised geometries and energetics of the 10 located low-lying $(mABA)_3$ structures are listed in Supporting Information (Table SI.3.1-SI.3.2) and a schematic representation of the most stable trimer is reported in Figure 2, which consist of a classical carboxylic dimer interacting via two NH···O and NH–π hydrogen bonds with another mABA molecule. In solution, the Boltzmann averaged values of the free energy of the trimerization reaction [3 mABA → $(mABA)_3$] is positive (0.98 kcal mol^{-1} in H_2O and 1.39 kcal mol^{-1} in DMSO). However, these values are relatively small and we cannot exclude that traces of trimeric mABA species could be experimentally found in DMSO and account for the ^{13}C NMR chemical shifts observed during the crystallization of mABA.[50]

Tetramers. For the tetramers, we determined the structures and free energies of the most stable low-lying isomers of $(mABA)_4$ (Table SI.4.1-SI.4.2) from approximately 1500 candidate structures of type $(mABA)_{4,(D+D)}$, that is, tetramers generated from the aggregation of two cyclic dimers I. In fact, we verified that tetramers of type $(mABA)_{4,(D+D)}$ are significantly lower in energy than tetramers of type $(mABA)_{4,(D+2M)}$, that is, tetramers generated from the aggregation of dimer I with two isolated mABA molecules. A schematic representation of the most stable tetramer is reported in Figure 2. $(mABA)_4$ consist of two dimeric units arranged in a stacked configuration, which promotes the stabilization of the tetrameric structure by allowing the formation of both H-bonding and π–π interactions between pairs of m-aminobenzoic acid molecules. The energetics of formation of $(mABA)_4$ in Table 3 indicate that in DMSO, where mABA molecules form stable classic carboxylic species, the aggregation of two dimers [$(mABA)_2$ + $(mABA)_2$ ⟶ $(mABA)_4$] is endoergonic (+2.06 kcal mol^{-1}). On the other hand, the formation of $(mABA)_4$ could result from the condensation of carboxylic dimer with two mABA monomers [$(mABA)_2$ + 2 mABA ⟶ $(mABA)_4$, -1.36 kcal mol^{-1}], and in water, where both aggregation free energies are negative (see Table 3).

View Article Online
DOI: 10.1039/C6CE00130K

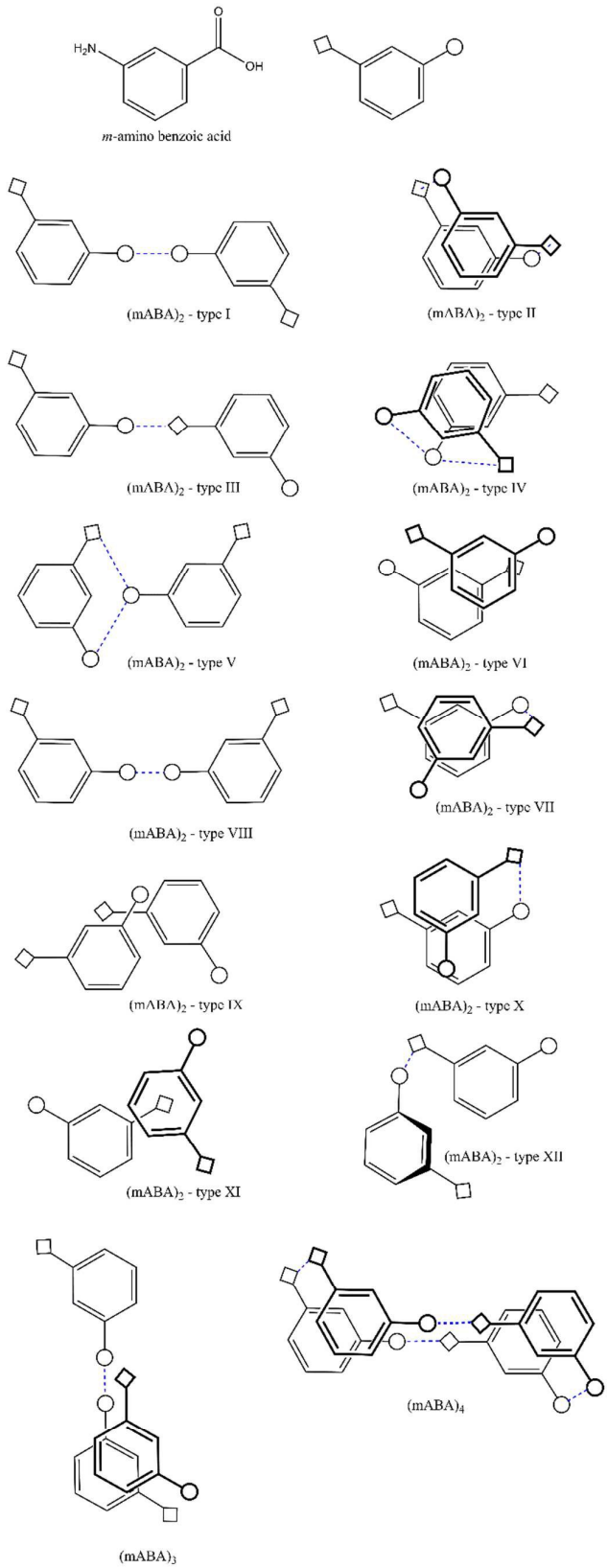


Figure 2. Schematic representation of m-aminobenzoic acid clusters, $(mABA)_n$ (n = 2-4).

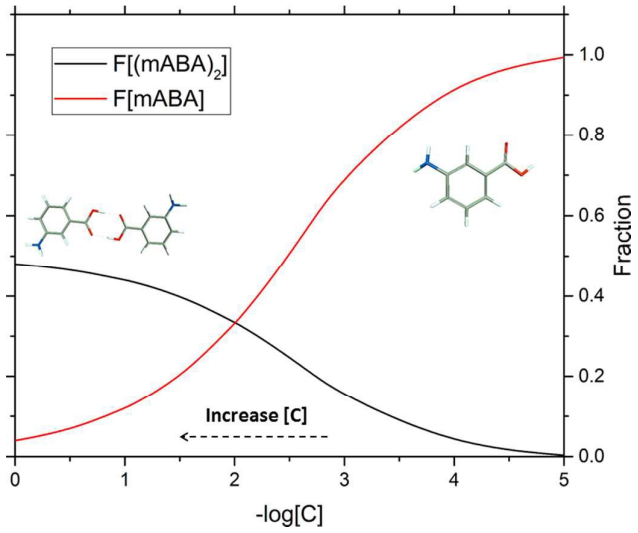


Figure 3. Fraction of the stoichiometric mABA concentration present as monomers and dimers in DMSO. Curve obtained by applying Eq. (10) with K_D = 304.9 L/mol.

Therefore, starting from a large number of randomly generated candidate structures and by imposing a minimum free energy condition for the isomers of $(mABA)_4$, we were able to locate the most stable tetrameric species in solution. Figure 4 highlights the correspondence between the most stable tetramer and the π–π stacking crystal synthon found in form II of mABA.[17] Consequently, the transfer of structural information from the solution- to the solid state-phase is not only related with the presence in solution of stable carboxylic acid dimers,[5,13,23] but the higher-order prenucleation clusters $(mABA)_4$ (Figure 2) could also direct the nucleation of mABA towards the formation of the polymorphic form II.

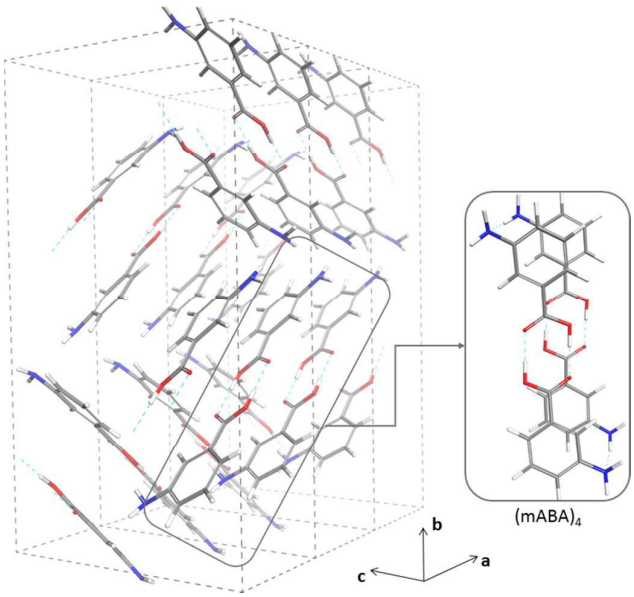


Figure 4. Correspondence between the most stable *m*-aminobenzoic acid tetramer, $(mABA)_4$, in solution and the synthon found in the (1×3×1) unit cell of form II of mABA.

View Article Online
DOI: 10.1039/C6CE00130K

Role of solvent in the kinetics of mABA aggregation

The quantum mechanical continuum calculations reported above describe the solution thermodynamics of the process of mABA cluster formation with respect to *infinitely* separated mABA molecules, but ignore intermediate process such as separation of solute-solvent molecules and diffusion of solute molecules across the solution, which come before the aggregation step.

Figures 5 and Figure SI.5.1 in Supporting Information display the last configurations of the MD simulations of mABA solutions in water (20 ns) and DMSO (200 ns). The results indicate a clear solvent-dependent behaviour for the aggregation of mABA molecules. In water, even at low concentrations (0.3 mol L^{-1}), we observed the spontaneous formation of π–π associated clusters, whereas in DMSO the molecules of mABA are present in a more solvated state; the level of clusterisation in water is therefore significantly larger than in DMSO. For the DMSO solutions, extending the simulation period to values larger than 200 ns did not results in a significant increase of the level of aggregation and we could not observe the formation of the thermodynamically stable cyclic dimer I that was predicted by DFT solvation continuum calculations (see Table 3).

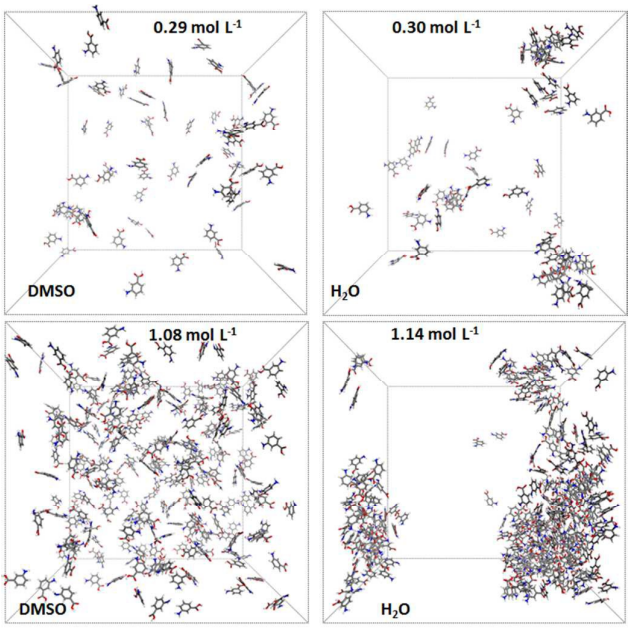


Figure 5. Last configuration of MD simulations of 0.3 mol L^{-1} and 1.1 mol L^{-1} mABA solutions in DMSO (200 ns) and water (30 ns). For clarity, solvent molecules were removed.

We have quantified the level of aggregation of mABA in water and DMSO by computing, every 10 ps, the number of mABA pairs within 3.5 Å (Figure 6) and the number of hydrogen bonds between mABA molecules (Figure SI.5.2.). As expected, the number of hydrogen-bonded clusters increases with the concentration of the solute, but the role of solvent in assisting, or hindering, the process of clusterisation is substantial. For example, in the 0.3 mol L^{-1} and 1.1 mol L^{-1} aqueous solutions the average number of pairs in water are approximately 50 % and 130 % higher than in DMSO, respectively (see Figure 6). Notice also the small temporal fluctuations of the number of mABA pairs in DMSO (after 200ns) with its equivalent concentration in water (after 10 ns), which indicates that the clustering of mABA in water is still progressing and is likely to be considerably higher than the one reported in Figure 6 and Figure SI.5.2 in Supporting Information.

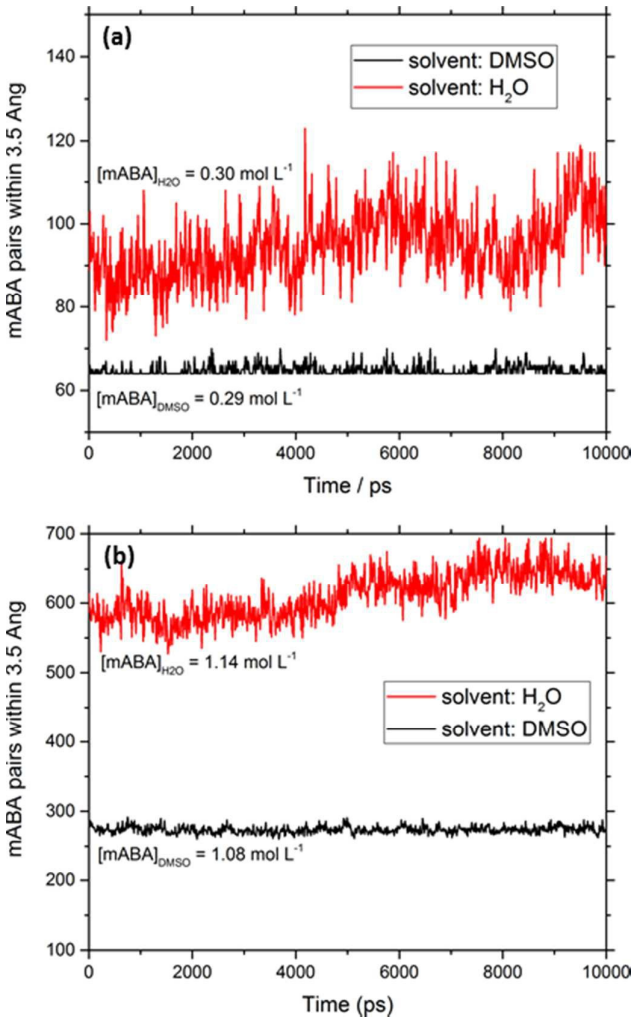


Figure 6. Time evolution of the number of pairs between mABA molecules computed during the last 10 ns of the MD simulations of (a) 0.3 mol L^{-1} and (b) 1.1 mol L^{-1} solutions.

Free energy profile of mABA dimerization in DMSO and water

Figure 7 shows free energy profile for the dimerization of mABA in DMSO and water. In DMSO, the carboxylic dimer I (**A** in Figure 7) is stable with respect to highly separated (40 Å) mABA molecules and thermodynamically more stable than in water, in agreement with the DFT values for free energies of dimerization (see Table 3. As these two units are brought together during the metadynamics simulations, the free energy profiles in DMSO and water have significant differences

View Article Online
DOI: 10.1039/C6CE00130K

even at large distances between the mABA molecules, which indicates a different desolvation behaviour of mABA in the two solvents. In particular, the formation of $(mABA)_2$ goes through two highly solvated intermediate states, **D** and **C**, and an activated state **B**.

In aqueous solution, the formation of $(mABA)_2$ is less thermodynamically favourable than in DMSO, in agreement with the quantum mechanical continuum calculations (see Table 2), but the free energy profile of mABA dimerization has a significantly lower energy barrier. On the other hand, in DMSO the activation free energy required for the formation of the mABA classic carboxylic dimer (3 kcal/mol) is approximately half the activation free energy in water and significantly larger than kT (0.593 kcal mol^{-1} at 300K), which rationalizes the absence of classic carboxylic acid dimers during unbiased MD simulations in DMSO solutions (see Figure 5).

We have also computed the free energy profile with respect to the carbon (C=O) carbon (C=O) distance of two mABA molecules (Figure SI.6.1). This order parameter leads to the formation of an H-bonded π-π dimer. In DMSO, the activation free energy is approximately 3.5 kcal mol^{-1} whereas in water the formation of a π-π dimer is almost barrier-less, in agreement with the spontaneous formation of π-π clusters observed during unbiased MD simulations.

These results confirm that explicit solute-solvent interactions, which are critical during the process of mABA desolvation, control the kinetics of molecular dimerization from solution and, consequently, also the formation of larger pre-nucleation clusters.

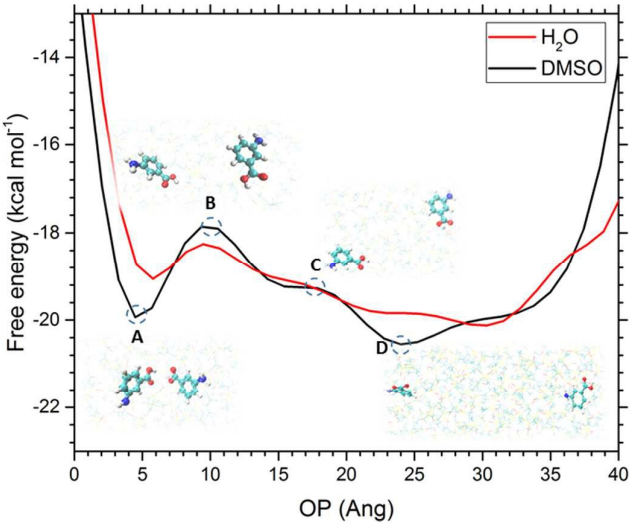


Figure 7. Free energy profiles for the *m*-aminobenzoic acid dimerization in water and DMSO. The order parameter (OP) used to study this reaction was defined by averaging two distances between the hydroxyl hydrogen and carboxyl oxygen of a pair of mABA molecules. **A**, **B**, **C** and **D** represent the structure of the microstates along the OP trajectory determined using the METAGUI computational tool.[57]

Quantification of the strength of mABA-solvent interaction

The observed differences in the aggregation of mABA clusters in aqueous and organosulfur environments can also be rationalized in terms of the strength of mABA-solvent interaction. We used as descriptors the energy ($\Delta \boldsymbol{E_{clust}}$) and free energy($\Delta \boldsymbol{G^o_{clust}}$) of solute-solvent clustering process,

$$\Delta E_{clust} = E_{[mABA(S)_n]} - \left(E_{[mABA]} + nE_{[S]}\right) \quad (13)$$

$$\Delta G_{clust} = G^o{}_{[mABA(S)_n]} - \left(G^o{}_{[mABA]} + nG^o{}_{[S]}\right) \quad (14)$$

and the successive solvent binding energy (ΔE_{inc})

$$\Delta E_{inc} = E_{[mABA(S)_n]} - E_{[mABA(S)_{n-1}]} - E_{[S]} \quad (15)$$

In Eq. 13-15, E and $\boldsymbol{G^o}$ are the gas-phase energies and gas-phase free energies of the solvated cluster [mABA(S)$_n$], solute (mABA) and solvent (H_2O or DMSO), which values were computed at the M06-2X/6-31+G(d,p). The structures of the low-lying structures of mABA(S)$_n$ (n = 1-3) were determined using the computational protocol mentioned in the computational section starting from *tens of thousands* of candidate structures. The energetics of binding and successive binding are reported in Table 4. For comparison, the solvation free energies of mABA in water and DMSO computed at the SMD/M06-2X/6-31+G(d,p) level of theory are -10.97 and -11.29 kcal mol^{-1}, respectively.

The energetics in Table 4 are in agreement with the results from metadynamics simulations: the clustering energies and free energies of mABA with DMSO are substantially larger (up to 10 kcal mol^{-1}) than with water. The energy required to desolvate mABA, which is related with the specific strength of the solute-solvent interaction, is therefore a crucial factor in determining the level of clusterization, and subsequent growth, of mABA in water and in DMSO. This result also supports what previously suggested by Rasmuson and co-workers that the strength of the solute-solvent molecular interaction is directly related with experimental rate of nucleation of organic molecules from solution.[16]

Table 4. Solute-solvent binding energy (ΔE_{clust}) and free energy (ΔG_{clust}) of solute-solvent clustering, and successive solvent binding energy (ΔE_{inc}) and solvation free energies of *m*-aminobenzoic acid in water and in DMSO. Geometries, frequencies and solvation energies computed at the M06-2X/6-31+G(d,p) level of theory. Values in kcal mol^{-1}.

	ΔE_{clust}		ΔE_{inc}		ΔG_{clust}	
species	**H_2O**	**DMSO**	**H_2O**	**DMSO**	**H_2O**	**DMSO**
mABA(S)$_1$	-7.88	-17.68	-7.88	-17.68	2.40	-5.80
mABA(S)$_2$	-22.43	-30.21	-14.55	-12.52	-0.43	-4.90
mABA(S)$_3$	-33.06	-43.20	-10.63	-12.99	0.34	-5.49

Conclusions

Extensive computational chemistry simulations based on density functional theory (DFT), molecular dynamics and metadynamics methods have revealed the role of solvent in the kinetics and thermodynamics of aggregation of m-aminobenzoic acid (mABA) molecules in aqueous and organosulfur (DMSO) solutions.

View Article Online
DOI: 10.1039/C6CE00130K

DFT calculations with a continuum model to describe the solvation environment were used to probe the potential energy surfaces of molecular clusters of m-aminobenzoic acid, $(mABA)_n$ (n = 2-4), locate the low-lying energy structures of $(mABA)_n$, and compute their Gibbs free energies in water and DMSO. Starting from a large number of randomly generated candidate structures and by imposing the condition of minimum free energy in solution for the isomers of $(mABA)_n$, we proved that the structure of the most stable dimer and tetramer in solution correspond to the classic carboxylic dimer π–π stacking synthon found in the crystalline form-II of mABA. Consequently, the transfer of structural information from the solution- to the solid state-phase of mABA is not only related with the presence in solution of stable carboxylic acid dimers, but also to higher-order prenucleation clusters $(mABA)_4$, which could also direct the nucleation process towards the formation of the polymorphic form II of mABA.

Molecular dynamics simulations of mABA solutions show a significant solvent-dependent behaviour for the aggregation of mABA molecules in solution. In aqueous solutions, even at relatively low concentrations, mABA molecules spontaneously form H-bonded π-π stacking cluster, whereas in organosulfur solutions the molecules of mABA are in a more solvated state and high concentrations of mABA are required to observe appreciable levels of solute mABA aggregation. We could not observe however the formation of the stable dimeric and tetrameric species predicted by DFT calculations.

Calculation of the free energy profile for the mABA dimerization using metadynamics simulations shows that in DMSO solutions mABA molecules have a higher barrier associated with the diffusion and desolvation of mABA, which are necessary and pre-requisite steps for the aggregation of organic molecules from solution to occur. In particular, the formation of $(mABA)_2$ in DMSO has an activation barrier that is twice the one in water, and much larger than the value of kT at 300 K. This rationalizes the observed low level of aggregation of mABA in DMSO solutions compared with water. DFT calculations of the energy and free energy of formation of microsolvated mABA clusters, $mABA(S)_n$, agree with the results of metadynamics simulations and show that mABA-DMSO interaction is substantially larger (up to 10 kcal mol^{-1}) mABA-water interaction.

It is well known that during the crystallization of inorganic crystals such as magnesite ($MgCO_3$) or calcite ($CaCO_3$), the desolvation of the building units, and particularly of the cations,[78,79] can be rate-determining. The computational results obtained in this study suggest that the early stages of crystallization of organic crystals can be significantly influenced by the desolvation of mABA molecules in solution: the solvent and its specific interaction with the organic solute molecules influence not only the thermodynamics of aggregation but also the rate of desolvation in DMSO compared with water, which limits the extend of mABA self-assembly process in DMSO. This result could have implications for the computational methods that can be employed to model the crystallization of organic molecules from solution: approaches based on a continuum description of the solvent to describe the thermodynamic stability of the molecular aggregation,[31] or iterative methods where solvent molecules are excluded during the aggregation step,[65] cannot quantify the desolvation of molecules and could therefore overlook rate-determining processes such as desolvation of the building units in homogeneous nucleation or desolvation of the surface during surface growth.

Acknowledgements

D.D.T. thanks the UK's Royal Society for the award of a Royal Society Industry Fellowship. This research utilised Queen Mary's MidPlus computational facilities, supported by QMUL Research-IT and funded by EPSRC grant EP/K000128/1. Via our membership of the UK's HEC Materials Chemistry Consortium, which is funded by EPSRC (EP/L000202), this work used the ARCHER UK National Supercomputing Service (http://www.archer.ac.uk). Dr. G. Tribello (Belfast) is thanked for guidance in the use of the Plumed code. Dr. C. Hughes and Prof. K. Harris (Cardiff) are acknowledged for useful discussions.

References

1. H. G. Brittain and G. D. J. W., "Effect of polymorphism and solid-state solvation on solubility and dissolution rate," in *Polymorphism in Pharmaceutical Solids*, New York, Marcel Dekker, Inc., 1999, 279-330.

2. "ANDAs: Pharmaceutical Solid Polymorphism Chemistry, Manufacturing, and Controls Information," U.S. Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research, Rockville, MD, 2007.

3. D. Braga, F. Grepioni, L. Maini and M. Polito, "Crystal Polymorphism and Multiple" *Struct. Bond.,* 2009, **132**, 25-50.

4. C. Stoica, P. Verwer, H. Meekes and P. J. C. M. van Hoof, "Understanding the Effect of a Solvent on the Crystal Habit" *Cryst. Growth Des.,* 2004, **4**, 765-768.

5. S. Parveen, R. J. Davey, G. Dent and R. G. Pritchard, "Linking solution chemistry to crystal nucleation: the case of tetrolic acid" *Chem. Comm.,* 2005, **109**, 1531-1533.

6. A. Spilateri, C. A. Hunter, J. F. McCabe, M. J. Packer and S. L. Cockroft, "A 1NMR study of crystal nucleation from solution" *CrystEngComm*, 2004, **6**, 490-493.

7. R. J. Davey, K. R. Back and R. A. Sullivan, "Crystal nucleation from solutions - transition states, rate determining steps and complexity" *Faraday Discuss.,* 2015, **179**, 9-26.

View Article Online
DOI: 10.1039/C6CE00130K

8. R. A. Sullivan, R. J. Davey, G. Sadiq, G. Dent, K. R. Back, J. H. ter Horst, D. Toroz and R. B. Hammond, "Revealing the Roles of Desolvation and Molecular Self-Assembly in Crystal Nucleation from Solution: Benzoic and p-Aminobenzoic Acids" *Crystal Growth & Design,* 2014, **14**, 2689-2696.

9. E. Ruiz-Agudo and C. Putnis, "Direct observations of mineral-fluid reactions using atomic force microscopy: The specific example of calcite", *Mineralogical Magazine,* 2012, **76**, 227-253.

10. A. S. Myerson and B. L. Trout, "Nucleation from Solution" *Science,* 2013, **341**, 855-856.

11. M. Salvalaglio, T. Vetter, F. Giberti, M. Mazzotti and M. Parrinello, "Uncovering molecular details of urea crystal growth in the presence of additives" *J. Am. Chem. Soc.,* **134**, 17221-17233.

12. M. Salvalaglio, C. Perego, F. Giberti and M. Parrinello, "Molecular-dynamics simulations of urea nucleation from aqueous solution" *Proc. Natl. Acad. Sci. USA,* 2015, **112**, E6-E14.

13. J. Chen and B. L. Trout, "Computational study of solvent effects on the molecular self-assembly of tetrolic acid in solution ans implications for the polymorphism formed from crystallization" *Journal of Physical Chemistry B,* 2008, **112**, 7794-7802.

14. S. Hamad, C. E. Hughes, C. R. A. Catlow and K. D. M. Harris, "Clustering of glycine molecules in aqueous solution studies by molecular dynamics", *J. Phys. Chem. B,* 2008, **112**, 7280-7288.

15. R. J. Davey, S. L. M. Schroeder and J. H. ter Horst, "Nucleation of Organic Crystals – A Molecular Perspective" *Angewandte Chemie Int. Ed.,* 2013, **52**, 2166-2179.

16. K. Dikshitkumar, J. Zeglinski, D. Mealey and A. C. Rasmuson, "Investigating the Role of Solvent–Solute Interaction in Crystal Nucleation of Salicylic Acid from Organic Solvents", *J. Am. Chem. Soc.,* 2014, **136**, 11664–11673.

17. P. A. Williams, C. E. Hughes, G. Keat Lim, B. M. Kariuki and K. D. M. Harris, "Discovery of a new system exhibiting abundant polymorphism: m-aminobenzoic acid", *Cryst. Growth Des.,* 2012, **12**, 3104-3113.

18. M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus and W. A. de Jong, "NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations", *Comput. Phys. Commun.,* 2010, **181**, 1477–1489.

19 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, Sonnenberg, L. J., M. Hada, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, "Gaussian, Inc.," Wallingford CT, 2009.

20 S. Grimme, "Semiempirical GGA-type density functional constructed with a long-range dispersion correction", *J. Comp. Chem.,* 2006, **27**, 1787-99.

21 Y. Zhao and D. G. Truhlar, "The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent", *Theor. Chem. Acc.,* 2008, **120**, 215−241.

22 B. Lynch, Y. Zhao and D. Truhlar, "Effectiveness of Diffuse Basis Functions for Calculating Relative Energies by Density Functional Theory", *J. Phys. Chem. A,* 2003, **107**, 1384−1388.

23. D. Di Tommaso, "The molecular self-association of carboxylic acids in solution: testing the validity of the link hypothesis using a quantum mechanical continuum solvation", *CrystEngComm,* 2013, **15**, 6564-6577.

24. I. M. Alecu, J. Zheng, Y. Zhao and D. G. Truhlar, "Computational Thermochemistry: Scale Factor Databases and Scale Factors for Vibrational Frequencies Obtained from Electronic Model Chemistries", *J. Chem. Theory Comput.,* 2010, **6**, 2872-2887.

25. A. V. Marenich, C. J. Cramer and D. G. Truhlar, "Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions", 2009, **113**, 6378–6396.

26. R. F. Ribeiro, A. V. Marenich, C. J. Cramer and D. G. Truhlar, "The Solvation, Partitioning, Hydrogen Bonding, and Dimerization of Nucleotide Bases: A Multifaceted Challenge for Quantum Chemistry", *Phys. Chem. Chem. Phys.,* 2011, **13**, 10908-10922.

27. V. S. Bryantsev, M. S. Diallo and W. A. Goddard III, "Calculation of Solvation Free Energies of Charged Solutes Using Mixed Cluster/Continuum", *J. Phys. Chem. A*, 2008, **112**, 9709–9719.

View Article Online
DOI: 10.1039/C6CE00130K

28. E. Tang, D. Di Tommaso and N. H. de Leeuw, "Accuracy of the microsolvation–continuum approach in computing the pKa and the free energies of formation of phosphate species in aqueous solution", *Phys. Chem. Chem. Phys.,* 2010, **12**, 13804-13815.

29. J. Chocolousova, J. Vacek and P. Hobza, "Acetic acid dimer in the gas phase, nonpolar solvent. Microhydrated environment, and dilutre and concentrated acetic acid: Ab initio aquantum chemical and molecular dynamics study", *J. Phys. Chem. A,* 2003, **107**, 3086-3092.

30. C. J. Cramer, Essentials of Computational Chemistry. Theories and models, New York: Wiley, 2004.

31. D. Di Tommaso and K. L. Watson, "Density Functional Theory Study of the Oligomerization of Carboxylic Acids", *J. Phys. Chem. A,* 2014, **118**, 11098−11113.

32. L. A. Montero, "Generacion de Celdas con Agregados Moleculares Optimizados," [Online]. Available: http://karin.fq.uh.cu/mmh/.

33. L. Montero, A. Esteva, J. Molina, A. Zapardiel, L. Hernández, H. Márquez and A. Acosta, "A Theoretical Approach to Analytical Properties of 2,4-Diamino-5-phenylthiazole in Water Solution. Tautomerism and Dependence on pH", *J. Am. Chem. Soc.,* 1998, **120**, 12023-12033.

34. Y. Zhao and D. G. Truhlar, "Exploring the Limit of Accuracy of the Global Hybrid Meta Density Functional for Main-Group Thermochemistry, Kinetics, and Noncovalent Interactions", *J. Chem. Theory Comp.,* 2008, **4**, 1849-1868.

35. R. Peverati and D. G. Truhlar, "Quest for a universal density functional: the accuracy of density functionals across a broad spectrum of databases in chemistry and physics", *Phil. Trans. R. Soc. A,* 2014, **372**, 20120476/1-51.

36. A. V. Marenich, C. J. Cramer and D. G. Truhlar, "Performance of SM6, SM8, and SMD on the SAMPL1 test set for the prediction of small-molecule solvation free energies", *J. Phys. Chem. B,* 2009, **113**, 4538–4543.

37. E. R. Johnson, I. D. Mackie and G. A. DiLaio, "Dispersion interactions in density-functional theory", *J. Phys. Org. Chem.,* 2009, **22**, 1127-1135.

38. C. P. Kelly, C. J. Cramer and D. G. Truhlar, "Adding explicit solvent molecules to continuum solvent calculations for the calculation of aqueous acid dissociation constants", *J. Phys. Chem. A,* 2006, **110**, 2493-2499.

39. J. R. Pliego and J. M. Riveros, "The cluster-continuum model for the calculation of the solvation free energy of ionic species", *J. Phys. Chem. A,* 2001, **105**, 7241-7247.

40. N. F. Carvalho and J. R. Pliego, "Cluster-continuum quasichemical theory calculation of the lithium ion solvation in water, acetonitrile and dimethyl sulfoxide: an absolute single-ion solvation free energy scale", *Phys. Chem. Chem. Phys.,* 2015, **17**, 26745-26755.

41. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, "Development and testing of a general amber force field", *J. Comput. Chem.,* 2004, **25**, 1157-1174.

42. H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, "The missing term in effective pair potentials", *J. Phys. Chem.,* 1987, **91**, 6269-6271.

43. C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model", *J. Phys. Chem.,* 1993, **97**, 10269-10280.

44. F.-Y. Depradeau, A. Pigache, T. Zaffran, C. Savineau, R. Lelong, N. Grivel, D. Lelong, W. Rosanski and P. Cieplak, "The R.E.D. tools: advantages in RESP and ESP charge derivation and force field library building", *Phys. Chem. Chem. Phys.,* 2010, **12**, 7821-7939.

45. W. D. Cornell, P. Cieplak, C. Bayly, I. R. Gould, K. M. Merz, D. C. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules", *J. Am. Chem. Soc.,* 1995, **117**, 5179–5197.

46. B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, "GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable", *J. Chem. Theory Comput.,* 2008, **4**, 435-447.

47. D. van der Spoel, GROMACS, 2015. [Online]. Available: http://www.gromacs.org/.

48. G. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, "Plumed 2: New feathers for an old bird", *Comput. Phys. Commun.*, 2014, **185**, 604–613.

49. M. Svard, F. L. Nordstrom, T. Jasnobulka and A. C. Rasmuson, "Thermodynamics and Nucleation Kinetics of m-Aminobenzoic Acid Polymorphs", *Cryst. Growth Des.,* 2010, **10**, 195-204.

50. C. E. Hughes, P. A. Williams and K. D. E. Harris, ""CLASSIC NMR": an in-situ NMR strategy for mapping the time-evolution of crystallization processes by combined liquid-state and solid-state measurements", *Angew. Chem. Int. Ed. Engl.,* 2014, **53**,

View Article Online
DOI: 10.1039/C6CE00130K

8939-8943.

51. P. C. Sadek, Illustrated Pocket of Dictionary of Chromatography, Hoboken: Wiley, 2004.

52. W. D. Kumler, "Acidic and basic dissociation constants and structure", *J. Org. Chem.,* vol. 20, pp. 700-706, 1955.

53. N. Bjerrum, "Die Konstitution der Ampholyte, besonders der Aminos€auren, und die Dissoziationskonstanten", *Z. Phys. Chem.,* 1923, **104**, 147.

54. M. Friant-Miche and M. F. Ruiz-Lopez, "Glycine dimers: structure, stanility, and medium effects", *ChemPhysChem,* 2010, **11**, 3499-3504.

55. T. Krishnan, W. C. Carlton, S. Goldman and J. Fortier, "On the use of dilution calorimetry in the study of hydrogen-bondingt self-association reactions: benzoic-acid and benzene", 1979, *Can. J. Chem.,* 1979, **57**, 530-537.

56. E. M. Wooley and D. S. Rushforth, "Molecular association of hydrogen bonding solutes, o- m-, and p-Cresol in carbon tetrachloride", *Can. J. Chem.,* 1974, **52**, 653-660.

57. X. Biarnes, F. Pietrucci, F. Marinelli and A. Laio, "METAGUI. A VMD interface for analyzing metadynamics and molecular dynamics simulations", *Comput. Phys. Commun.,* 2012, **183**, 203-211.

58. The CP2K developers group, "CP2K version 2.7 (Development Version)" 2015. [Online]. Available: http://www.cp2k.org/.

59. J. Hutter, M. Iannuzzi, F. Schiffmann and J. VandeVondele, "CP2K: atomistic simulations of condensed matter systems", *Wiley Interdisciplinary Reviews: Computational Molecular Science,* 2014, **4**, 15-25.

60. J. P. Perdew, K. Burke and M. Ernzerhof, "Generalized gradient approximation made simple", *Physical Review Letters,* 1996, **77**, 3865-3868.

61. S. Grimme, S. Ehrlich and H. Krieg, "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**, 154104.

62 S. Goedecker, M. Teter and J. Hutter, "Separable dual-space Gaussian pseudopotentials", *Phys. Rev. B,* 1996, **54**, 1703-1710.

63. S. Grimme, S. Ehrlich and L. Goerigk, "Effect of the damping function in dispersion corrected density functional theory", *J. Comput. Chem.,* 2011, **32**, 1456.

64. D. W. Cheong and D. Y. Boon, "Comparative study of the force field for molecular dynamics simulations of alpha-Glycine crystal growth from solution", 2010, *Cryst. Growth Des.,* **10**, 5146-5158.

65. P. Ectors, P. Duchstein and D. Zhan, "Nucleation mechanisms of a polymorphic molecular crystal: solvent-dependent structural evolution of benzamide aggregates", *Cryst. Growth Des.,* 2014, **14**, 2972–2976.

66. R. A. Sullivan, R. J. Davey, G. Sadiq, G. Dent, K. R. Back, J. H. ter Horst, D. Toroz and R. B. Hammond, "Revealing the roles of desolvation and molecular self-assembly in crystal nucleation from solution: benzoic and p-aminobenzoic acids", *Cryst. Growth Des.,* 2014, **14**, 2689-2696.

67. D. Toroz, R. B. Hammond, K. J. Roberts, S. Harris and T. Ridley, "Molecular dynamics simulations of organic crystal dissolution: The lifetime and stability of the polymorphic forms of para-amino benzoic acid in aqueous environment", *J. Cryst. Growth,* 2014, **401**, 38-43.

68 B. H. Besler, K. M. Merz Jr. and P. A. Kollman, "Atomic charges derived from semiempirical methods", *J. Comp. Chem.,* 1990, **11**, 431-439.

69. T. E. Cheatham III, P. Cieplak and P. A. Kollman, "A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat", *J. Biomol. Struct. Dyn.,* 1999, **16**, 845-862.

70. J. Wang, P. Cieplak and P. A. Kollman, "How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?", *J. Comput. Chem.,* 2000, **21**, 1049-1074.

71. V. Hornak, R. Abel, B. Strockbine, A. Roitberg and C. Simmerling, "Comparison of multiple Amber force fields and development of improved protein backbone parameters", *Proteins,* 2006, **65**, 712-725.

72. A. Barducci, G. Bussi and M. Parrinello, "Well-tempered metadynamics: A smoothly converging and tunable free-energy method", *Phys. Rev. Lett.,* 2008, **100**, 020603.

73. A. Laio and F. G. Gervasio, "Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science," *Rep. Prog. Phys.,* 2008, **71**, 126601.

74. V. S. Bryantsev, M. S. Diallo, A. C. T. van Duin and G. W. A.,

View Article Online
DOI: 10.1039/C6CE00130K

"Hydration of copper(II): New insights from density functional theory and the COSMO solvation model", *J. Phys. Chem. A,* 2008, **112**, 9104-9112.

75. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, "Comparison of simple potential functions for simulating liquid water", *J. Chem. Phys.,* 1983, **79**, 926-935.

76. M. Svard, F. L. Nordstrom, T. Jasnobulka and A. C. Rasmuson, "Thermodynamics and nucleation kinetics of m-aminobenzoic acid polymorphs", *Cryst. Growth Des.,* 2010, **10**, 195-204.

77. N. Blagden, M. Song, R. Davey, L. Seton and C. S. Seaton, "Ordered Aggregation of Benzamide Crystals Induced Using a "Motif Capper" Additive", *Cryst. Growth Des.,* 2005, **5**, 467-471.

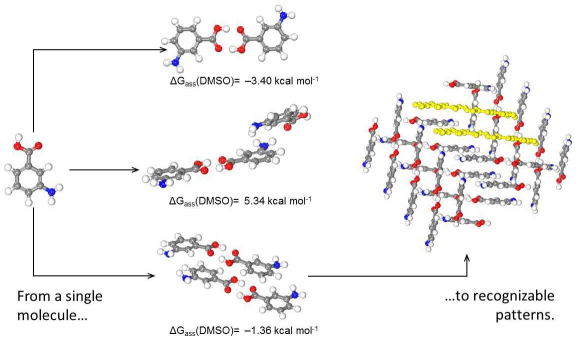
78. S. Piana, F. Jones and J. D. Gale, "Assisted desolvation as a key kinetic step for crystal growth", *J. Am. Chem. Soc.*, 2006, **128**, 13568–13574.

79. O. S. Pokrovsky and J. Schott, "Surface Chemistry and Dissolution Kinetics of Divalent Metal Carbonates", *Environ. Sci. Technol.*, 2002, **36**, 426–432.

View Article Online
DOI: 10.1039/C6CE00130K

The role of solvent in the self-assembly of m-aminobenzoic acid: a density functional theory and molecular dynamics study

Etienne Gaines, Krina Maisuria and Devis Di Tommaso



The role of solvent in the thermodynamics and kinetics of m-aminobenzoic acid aggregation from solution is revealed by computer simulations.