Peer-reviewed version available at Pharmaceutics 2018, 10, 12; doi:10.3390/pharmaceutics10010012

Type of the Paper (Article)

Solvation and Aggregation of Meta-aminobenzoic Acid in Water: Density Functional Theory and Molecular Dynamics Study

Etienne Gaines and Devis Di Tommaso *

School of Biological and Chemical Sciences, Materials Research Institute, Queen Mary University of London, Mile End Road, E1 4NS London, United Kingdom; e.gaines@qmul.ac.uk

* Correspondence: d.ditommaso@qmul.ac.uk; Tel.: +44-207-882-6226

Abstract: Meta-aminobenzoic acid, an important model system in the study of polymorphism and crystallization of active pharmaceutical ingredients, exist in water in both the nonionic (mABA) and zwitterionic (mABA±) forms. However, the constituent molecules of the polymorph that crystallizes from aqueous solutions are zwitterionic. This study reports atomistic simulations of the events surrounding the early stage of crystal nucleation of meta-aminobenzoic acid from aqueous solutions. *Ab initio* molecular dynamics was used to simulate the hydration of mABA± and mABA, and to quantify the interaction of these molecules with the surrounding water molecules. Density functional theory calculations were conducted to determine the low-lying energy conformers of meta-aminobenzoic acid dimers and compute the Gibbs free energies in water of nonionic, $(mABA)_2$, zwitterionic, $(mABA^{\pm})_2$ and nonionic-zwitterionic, $(mABA)(mABA^{\pm})$, species. Classical molecular dynamics simulations of mixed mABA–mABA± aqueous solutions were carried out to examine the aggregation of meta-aminobenzoic acid. According to these simulations the selective crystallization of the polymorph which constituent molecules are zwitterionic is driven by the formation of zwitterionic dimers in solution, which are thermodynamically more stable than $(mABA)_2$ and $(mABA)(mABA^{\pm})$ pairs. This work represents a paradigm of the role of molecular processes during the early stages of crystal nucleation in affecting polymorph selection during crystallization from solution.

Keywords: meta-aminobenzoic acid; solvation; aggregation; polymorphism; atomistic simulations

1. Introduction

The substance meta-aminobenzoic acid is of considerable importance in the pharmaceutical industry, widely used in the synthesis of analgesics, antihypertensives, vasodilators and other drugs.[1] This molecules also represents a fascinating model system for polymorphic research because it can crystallize in five different crystal structures (I-V).[2] The very strong polymorphic character of meta-aminobenzoic acid can be related to the manifold of inter-molecular interactions between meta-aminobenzoic acid molecules (hydrogen (H) bonding, π-π interactions and H-π interactions) but also to the ability of this molecule to exist in either of both the nonionic (mABA) and zwitterionic (mABA±) forms (Figure 1).[3] In fact, in the polymorphs denoted I, III and V the meta-aminobenzoic acid molecules are zwitterionic, and in the polymorphs II and V they are nonionic.[2,4]

The nature of the solvent can significantly influence the thermodynamics and kinetics of crystal growth,[5–7] and control the formation of one specific polymorph over another. [8–10] In the case of meta-aminobenzoic acid, Form II preferentially crystallizes from dimethyl sulfoxide (DMSO),[4] where meta-aminobenzoic acid only exist in the nonionic form in this solvent. Hughes and co-workers [11] monitored the crystallization of meta-aminobenzoic acid from organosulfur solutions using a combined liquid- and solid-state *in-situ* NMR apparatus and proposed the existence of nonionic mABA aggregates linked by H bonds, but the authors could not however uniquely determine the identity of these species. A recent theoretical study conducted in our group showed however that

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 January 2018 doi:10.20944/preprints201801.0069.v1

Peer-reviewed version available at Pharmaceutics 2018, 10, 12; doi:10.3390/pharmaceutics10010012

mABA molecules in DMSO aggregate to form thermodynamically stable dimers and tetramers which structure is consistent with the classic carboxylic dimer π−π stacking synthon found in this polymorph.[12]

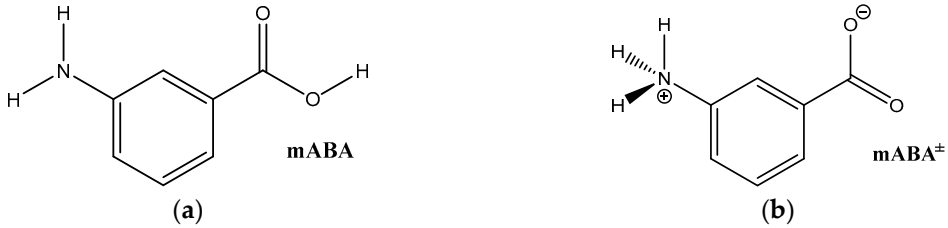


Figure 1. Schematic picture of the two tautomeric forms of meta-aminobenzoic acid: (**a**) nonionic mABA; (**a**) zwitterion mABA±.

On the other hand, the constituent molecules of crystal form I of meta-aminobenzoic acid are zwitterions. Despite it has been reported that the values of the equilibrium constant Kz = [mABA±]/[mABA] for aminobenzoic acids are of the order of unity in water,[13–15] implying a comparable distribution of mABA± and mABA molecules, Form I preferentially crystallizes from aqueous environments. The fundamental details of factors controlling the selection between zwitterionic and nonionic forms of meta-aminobenzoic acid during crystal nucleation from aqueous solution are not known yet.[16] This work aims to solve this conundrum by applying a combination of atomistic methods to follow the events surrounding the crystal nucleation of meta-aminobenzoic acid from aqueous solutions: *ab initio* molecular dynamics (MD) simulations of the hydration of mABA± and mABA in water; density functional theory (DFT) calculations of the structure and energetics of formation in water of (mABA)$_2$, (mABA)(mABA±) and (mABA±)$_2$ dimers; classical MD simulations of mixed mABA–mABA± aqueous solutions to quantify the aggregation of meta-aminobenzoic acid.

2. Computational Methods

2.1. Density functional theory calculations

Electronic structure calculations were carried out with the NWChem (version 6.3)[17] and Gaussian09 [18] codes. The Grimme's density functional (B97-D)[19] and the Minnesota 06 global hybrid functional with 54% HF exchange (M06-2X)[20] were used together with the Gaussian 6-31+G(d,p) basis set. Free energies of solvation were calculated using the SMD solvation model.[21]

The free energies of formation of nonionic, (mABA)$_2$, nonionic-zwitterionic, (mABA)(mABA±), and zwitterionic, (mABA±)$_2$, dimers were computed according to the following equation:

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

In Eq. 1, G^*_X is the total Gibbs free energy of the species X (X = AB, A or B) in the liquid. This quantity was evaluated using two different approaches. The first one follows the recommendation by Ho et al. that free energies of molecules in solution should be obtained from separate gas- and solution-phase calculations [22] and the application of the following expression:

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

In Eq. 2, $E_{e,gas}$ is the gas-phase total electronic energy of the gas-phase optimized geometry of the species X, $\delta G^{\circ}_{VRT,gas}$ is the vibrational-rotational-translational contribution to the gas-phase Gibbs free energy at T = 298 K under a standard-state partial pressure of 1 atm, ΔG^*_{solv} is 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}, and the last term is the free energy change of 1 mol of an ideal gas from 1 atm to 1 mol L^{-1} ($RTln[\tilde{R}T]$ = 1.89 kcal mol^{-1} at 298 K, $\tilde{R}$ = 0.082 K^{-1}).[23] However, the gas-phase optimization of zwitterionic dimers, (mABA±)$_2$, and nonionic-zwitterionic dimers, (mABA)(mABA±), caused the H-transfer between the molecular units (e.g. (mABA±)$_2$ → (mABA)$_2$). In these instances, stationary points in the solution do not correspond to stationary points

in the gas-phase, making it impossible to compute relevant gas-phase vibrational, translational and rotational contributions ($\delta G^{\circ}_{VRT,gas}$). Therefore, the other approach adopted was to optimize the structure of $(mABA^{\pm})_2$, $(mABA)(mABA^{\pm})$, and of the monomers mABA and $mABA^{\pm}$, in the aqueous phase. The total free energy in the solution of these species was then obtained from the expression:

$$G^{*}_{X} = E^{Tot}_{soln} + \delta G^{*}_{VRT,soln} \quad (3)$$

where $\delta G^{*}_{VRT,soln}$ is the vibrational-rotational-translational contribution to the liquid-phase, and E^{Tot}_{soln} is given by the sum of the liquid-phase expectation value of the gas-phase Hamiltonian ($E_{e,soln}$), the electronic polarization contribution to the solvation free energy based on bulk electrostatic (ΔG_{EP}), and the contribution from cavitation, dispersion and solvent structural effects (G_{CDS}):

$$E^{Tot}_{soln} = E_{e,soln} + \Delta G_{EP} + G_{CDS} \quad (4)$$

Since the potential energy surface (PES) of molecular clusters is characterized multiple low-lying energy isomers, the free energy of the dimers $(mABA)_2$, $(mABA)(mABA^{\pm})$ and $(mABA^{\pm})_2$ was determined from the Boltzmann ensemble average

$$\langle G(X) \rangle = \sum_{i=1}^{N} f_i G(X_i) \quad (5)$$

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 isomers. The Boltzmann factor was determined according to

$$f_i = \frac{e^{-G(X_i)/RT}}{\sum_j e^{-G(X_j)/RT}} \quad (6)$$

where R is the universal gas constant, T is the absolute temperature (T = 298 K) and the index j runs over all isomers. The low-lying energy structures of the meta-aminobenzoic acid dimers were located using the following computational protocol. (1) For each type of dimer [$(mABA)_2$, $(mABA)(mABA^{\pm})$ and $(mABA^{\pm})_2$] *hundreds of thousands* of candidate structures were generated using Granada,[24,25] a code designed to distribute randomly one or more molecules around a central unit (a monomer, dimer, trimer etc.) placed at the centre of a cube of defined side length. (2) Configurations satisfying the condition 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

$$f_i = \frac{e^{-(E_i - E_0)/RT}}{\sum_j e^{-(E_j - E_0)/RT}} \quad (7)$$

where E_i was the energy of the i^{th} candidate structure and E_0 was the energy of the most stable candidate structure. (4) The candidate structures with a Boltzmann factor $f_i \geq 0.01$ and ten to fifteen randomly selected structure such that $3 \leq E_i - E_0 \leq 15$ kcal mol^{-1} were selected. (4) Geometry optimization, thermochemical properties and solvation energies of the selected configurations were computed at the M06-2X/6-31+G(d,p) level of theory.

2.2. Molecular dynamics simulations

Ab initio (Born-Oppenheimer) molecular dynamics (AIMD) simulations were conducted with the electronic structure code CP2K/Quickstep code, version 4.1.[26] CP2K implements density functional theory (DFT) based on a hybrid Gaussian plane wave. We used the PBE [27] generalized gradient approximation for the exchange and correlation terms together with the general dispersion correction termed DFT-D3. Goedecker-Teter-Hutter pseudopotentials [28] were used to describe the core–valence interactions. All atomic species were represented using a double-zeta valence polarized basis set. The plane wave kinetic energy cut off was set to 1000 Ry. k-sampling was restricted to the Γ point of the Brillouin zone. Simulations were carried out with a wave function optimization

tolerance of 10^{-6} au that allows for 1.0 fs time steps with reasonable energy conservation. Periodic boundary conditions were applied throughout. Simulations were carried out in the NVT ensemble using a Nosé-Hoover chain thermostat to maintain the average temperature at T = 300 K.

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 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. The same cutoff was used for the van der Waals forces. The LINCS algorithm was applied at each step to preserve the bond lengths. The general AMBER forcefield (GAFF) [29] was used to model the nonionic and zwitterionic ($mABA^{\pm}$) forms of meta-aminobenzoic acid; this family of forcefields has been previously used to compute the aggregation and crystal growth of organic molecules.[12,30–32] Water molecules were modelled using the SPC/E potential.[33] The interactions between mABA and $mABA^{\pm}$ molecules and between meta-aminobenzoic acid and water were described using the GAFF potential. To generate the GAFF parameters for mABA and $mABA^{\pm}$ the structure and molecular electrostatic potential of these molecules were computed using the Hartree-Fock method and the 6-31G* basis set, and then using the *Antechamber* package was then used to compute partial charges according to the restrained electrostatic potential formalism. The GAFF forcefields and partial charges of mABA and $mABA^{\pm}$ are reported in Supporting Information (SI, Table SI.1.1 and SI.1.2).

Aqueous solutions of a single nonionic and zwitterionic meta-aminobenzoic acid molecule were carried out by embedding, respectively, one mABA and one $mABA^{\pm}$ in a box of 215 water molecules. Classical MD simulations were first conducted for approximately 5 ns and the last snapshot was used to conduct 20 ps of *ab initio* MD simulations.

Simulations of mixed mABA–$mABA^{\pm}$ aqueous solutions containing an equal amount of mABA and $mABA^{\pm}$ molecules were generated with the `insert-molecules` and `solvate` utilities of the GROMACS package to insert the required mABA and $mABA^{\pm}$ molecules in an empty cubic box of size 5 nm, and solvate them with water, respectively. 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 200 ns of NPT simulations; analysis were conducted on the last 40 ns of simulation. 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).

3. Results

3.1. Intermolecular properties and hydration structure

The stability of the nonionic (mABA) and zwitterionic ($mABA^{\pm}$) in aqueous solution and the interaction of these molecules with the surrounding water molecules are discussed in this section. Hereafter, the oxygen and nitrogen atoms of mABA or $mABA^{\pm}$ are denoted by O_m and N_m, the hydrogen of amino group are denoted by H_a, the hydrogen atoms of carboxylic group are denoted by H_c, and oxygen and hydrogen of water are denoted by O_w and H_w, respectively.

Figure 1 reports the time evolution of the intra- (O_m–H_c and N_m–H_a) and inter-molecular ($O_m\cdots H_w$ and $N_m\cdots H_w$) distances during the AIMD simulations of the mABA and $mABA^{\pm}$ species in water. Taking 1 Å as the average intramolecular X–H (X = N, O) bond distance, then mABA and $mABA^{\pm}$ are not involved in any proton (H) transfer reactions with the surrounding water molecules. Both mABA and $mABA^{\pm}$ molecules are therefore stable in water and should be considered when modelling the aggregation of meta-aminobenzoic acid in aqueous solution. The insets in Figures 2(a) and 2(b) indicate that the lifetime of the H-bond between mABA and the surrounding water molecules is less than 5 ps.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 January 2018 doi:10.20944/preprints201801.0069.v1

Peer-reviewed version available at Pharmaceutics 2018, 10, 12; doi:10.3390/pharmaceutics10010012

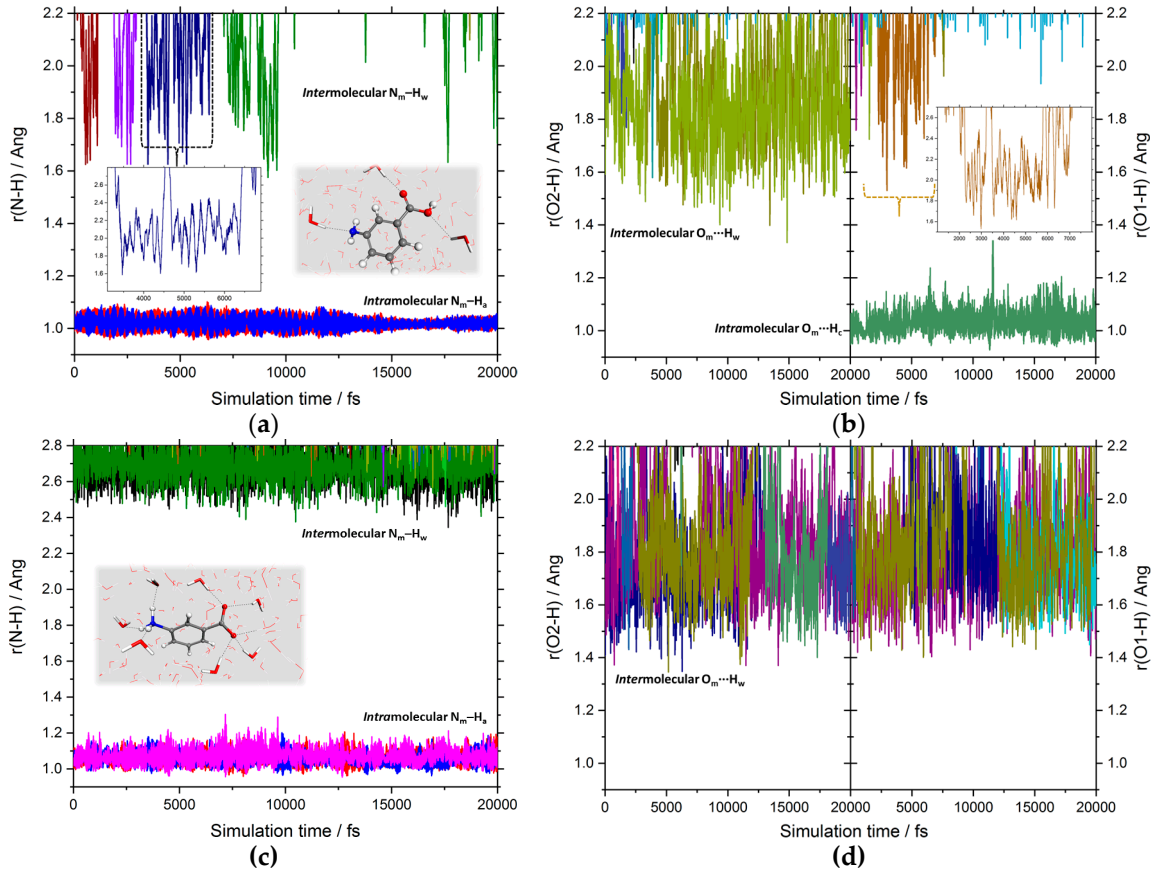


Figure 2. Time evolution of the X–H (X = N or O) distances during the AIMD simulation of the nonionic (mABA) and zwitterionic (mABA±) forms of meta-aminobenzoic acid in water: (**a**) Intramolecular (N–H) and intermolecular (N···H) distances of the mABA molecule; (**b**) Intramolecular (O–H) and intermolecular (O···H) distances of the mABA molecule; (**c**) Intramolecular (N–H) and intermolecular (N···H) distances of the mABA± molecule; (**d**) Intramolecular (O–H) and intermolecular (O···H) distances of the mABA± molecule.

A detailed characterization of this H-bonding interaction can be obtained from the analysis of the radial distribution function (RDF), $g_{\alpha\beta}(r)$, which represents the probability relative to a random distribution of finding an atom of type β at a distance r from an atom of type α. Figure 3 reports the O_m–H_w and N_m–H_w RDFs together with the running coordination number, $n(r) = (4\pi N/V)\int_0^r g(r')dr'$, where N is the number of hydrogen or oxygen atoms and V is the volume of the simulation cell. In the X_m–H_w (X = N or O) RDFs, a maximum in the [1.5–2.0] Å region and a minimum at around 2.5 Å indicate the presence of a H-bond with the surrounding water molecules.[34] On average, less than one water molecule is coordinated to each oxygen atom of the –COOH and to the nitrogen atom of the $–NH_2$ groups. On the other hand, approximately four water molecules are coordinated to the $–COO^-$ group mABA± and no water molecule is H-bonded to the nitrogen atom of the $–NH_3^+$ group. Table 1 summarizes the positions (r_{max} and r_{min}) and amplitudes (g_{max} and g_{min}) of the maxima and minima of the X_m–H_w RDFs together with the ratios $g_{max}^{X_m-H_w}/g_{min}^{X_m-H_w}$, which values can be used as a proxy for the strength of the H-bonding interactions between the X_m–H_w pairs (X = O and N).[34,35] For mABA, the $g_{max}^{O_m-H_w}/g_{min}^{O_m-H_w}$ ratio of the carboxyl oxygen atoms (9.0) is higher than nitrogen (4.5) but lower than the value obtained of $g_{max}^{O_w-H_w}/g_{min}^{O_w-H_w}$ = 19.6 obtained from AIMD simulations of pure water. Similar behavior is observed for mABA± where but the interaction of the the COO^- group ($g_{max}^{O_w-H_w}/g_{min}^{O_w-H_w}$ = 14.0) is significantly stronger than mABA.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 January 2018 doi:10.20944/preprints201801.0069.v1

Peer-reviewed version available at Pharmaceutics 2018, 10, 12; doi:10.3390/pharmaceutics10010012

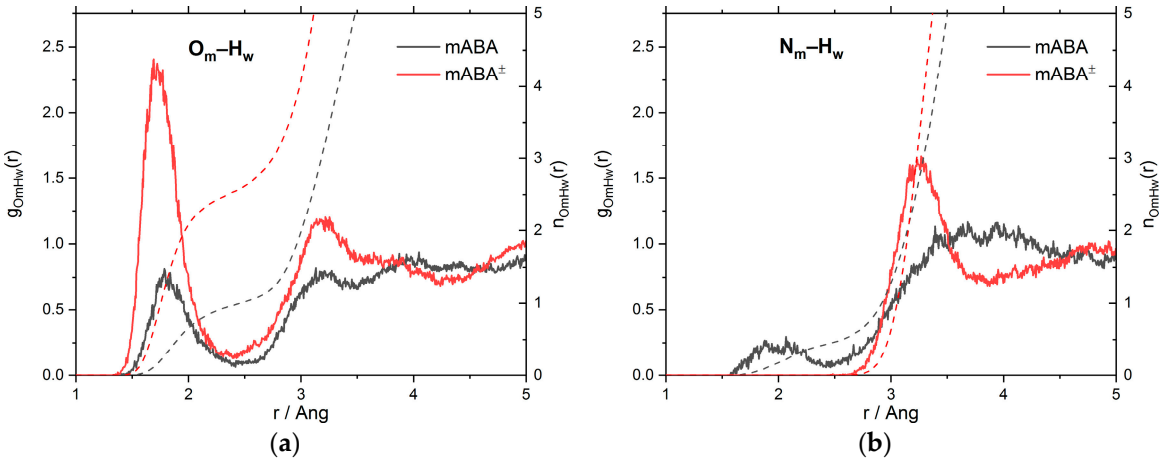


Figure 3. The radial distribution functions, $g(r)$, and running coordination numbers, $n(r)$, of mABA and mABA± with water obtained from AIMD simulations: (**a**) O_m–H_w RDFs (O_m = oxygen atoms of meta-aminobenzoic acid; H_w = hydrogen atoms of water); (**b**) N_m–H_w RDFs (N_m = nitrogen atoms of meta-aminobenzoic acid; H_w = oxygen atoms of water.

Table 1. Positions (r_{max}^{X-H} and r_{min}^{X-H} in Å, X = O and N) and amplitudes (g_{max}^{X-H} and g_{min}^{X-H}) of the maxima and minima of the first peak of the O_m–H_w and O_m–H_w RDFs, and first shell hydration number (n_w) obtained from the AIMD simulations of mABA and mABA± in water.

	mABA	**mABA±**
$r_{max}^{O_m-H_w}$	1.79	1.72
$g_{max}^{O_m-H_w}$	0.81	2.38
$r_{min}^{O_m-H_w}$	2.50	2.52
$g_{min}^{O_m-H_w}$	0.09	0.17
$g_{max}^{O_m-H_w}/g_{min}^{O_m-H_w}$	9.00	14.00
$n_w^{O_m}$	1.0	2.6
$r_{max}^{N_m-H_w}$	1.88	–
$g_{max}^{N_m-H_w}$	0.27	–
$r_{min}^{N_m-H_w}$	2.46	–
$g_{min}^{N_m-H_w}$	0.06	–
$g_{max}^{N_m-H_w}/g_{min}^{N_m-H_w}$	4.50	–
$n_w^{N_m}$	0.5	0

The RDFs and structural data of the H_c–O_w and H_a–O_w intermolecular interactions are reported in Figure 4 and Table 2. For the carboxylic group of mABA, the H_c–O_w RDF has a very well defined maximum at 1.51 Å and the running coordination number ($n_w^{H_c}$) is characterized by a clear plateau at the first RDF minimum (Figure 4(a)). The value of $g_{max}^{H_c-O_w}/g_{min}^{H_c-O_w}$ is significantly larger than $g_{max}^{O_w-H_w}/g_{min}^{O_w-H_w}$ of pure water (19.6) and consequently the H_c–O_w interaction is stronger than the intermolecular H-bonding in bulk water. The hydrogen of –COOH is therefore stably coordinated to one water molecule. For the amino group of mABA, as the H_a–O_w RDF in the [1.5–2.0] Å is not characterized by a well-defined peak, the hydrogen atoms of the $–NH_2$ group do not interact significantly with the surrounding water molecules (Figure 4(b)). On the other hand, the H_a–O_w RDF of the $–NH_3^+$ in mABA± is characterized by a distinct peak at 1.77 Å.

To summarize, the analysis of the X_m–H_w (X = N or O), H_c–O_w and H_a–O_w RDFs indicates that in aqueous solution the mABA±–water interaction is stronger than mABA–water, and for both species the interaction with the surrounding water molecules is stronger around the carboxylic acid than the amino group.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 January 2018 doi:10.20944/preprints201801.0069.v1

Peer-reviewed version available at Pharmaceutics 2018, 10, 12; doi:10.3390/pharmaceutics10010012

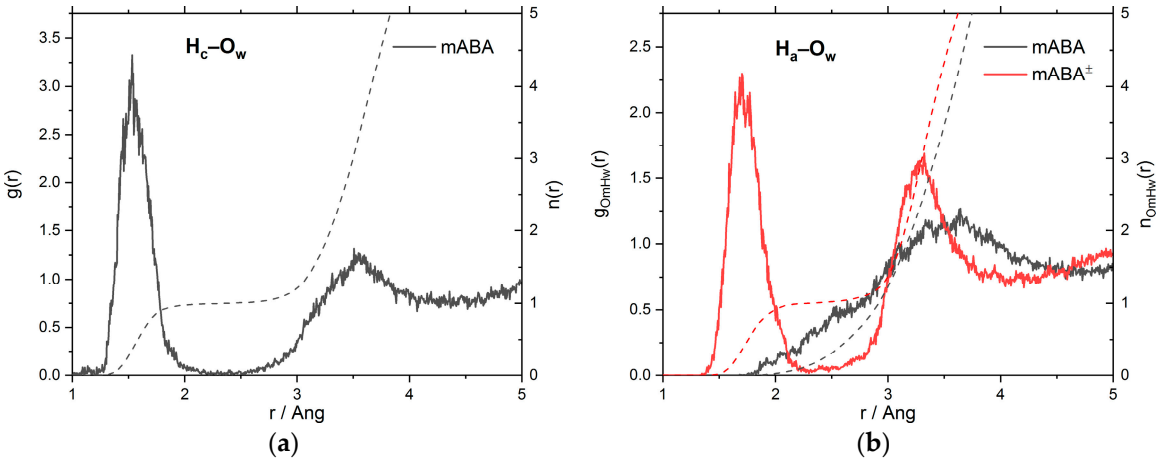


Figure 4. The radial distribution functions, $g(r)$, and running coordination numbers, $n(r)$, of mABA and mABA$^{\pm}$ with water obtained from AIMD simulations: (**a**) H_c–O_w RDFs (O_c = oxygen atoms of the carboxylic group of mABA; H_w = hydrogen atoms of water); (**b**) H_a–O_w RDFs N_m = nitrogen atoms of the amino group of mABA and mABA$^{\pm}$; O_w = oxygen atoms of water).

Table 2. Positions (r_{max}^{H-O} and r_{min}^{H-O} in Å) and amplitudes (g_{max}^{H-O} and g_{min}^{H-O}) of the maxima and minima of the first peak of the H_a–O_w and H_c–O_w RDFs, and first shell hydration number (n_w) obtained from the AIMD simulations of mABA and mABA$^{\pm}$ in water.

	mABA	**mABA$^{\pm}$**
$r_{max}^{H_c-O_w}$	1.51	–
$g_{max}^{H_c-O_w}$	3.06	–
$r_{min}^{H_c-O_w}$	2.31	–
$g_{min}^{H_c-O_w}$	0.01	–
$g_{max}^{H_c-O_w}/g_{min}^{H_c-O_w}$	306.00	–
$n_w^{H_c}$	1.0	–
$r_{max}^{H_a-O_w}$	–	1.77
$g_{max}^{H_a-O_w}$	–	2.15
$r_{min}^{H_a-O_w}$	–	2.23
$g_{min}^{H_a-O_w}$	–	0.03
$g_{max}^{H_a-O_w}/g_{min}^{H_a-O_w}$	–	71.7
$n_w^{H_a}$	0	1.0

The probability distribution of the number of water molecules in the first hydration shell (HS) of mABA and mABA was determined by computing the pair correlation functions between the center-of-masses (COM) of meta-aminobenzoic acid and water molecules (Figure 5). The position of the first HS was approximated by the first minimum in the COM(mABA)–COM(H_2O) RDFs [insets of Figure 5]. Although a hydration shell can be located for both molecules, the probability distributions of the number of water molecules surrounding mABA and mABA$^{\pm}$ show that these species display a flexible first coordination shell, where the flexibility increases on going from mABA$^{\pm}$ to mABA. There are an average of 24 water molecules in the HS of mABA with a Mean Absolute Deviation (MAD) of 1.4, and 27 water molecules in the HS of mABA$^{\pm}$ with a MAD of 1.0.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 January 2018 doi:10.20944/preprints201801.0069.v1

Peer-reviewed version available at Pharmaceutics 2018, 10, 12; doi:10.3390/pharmaceutics10010012

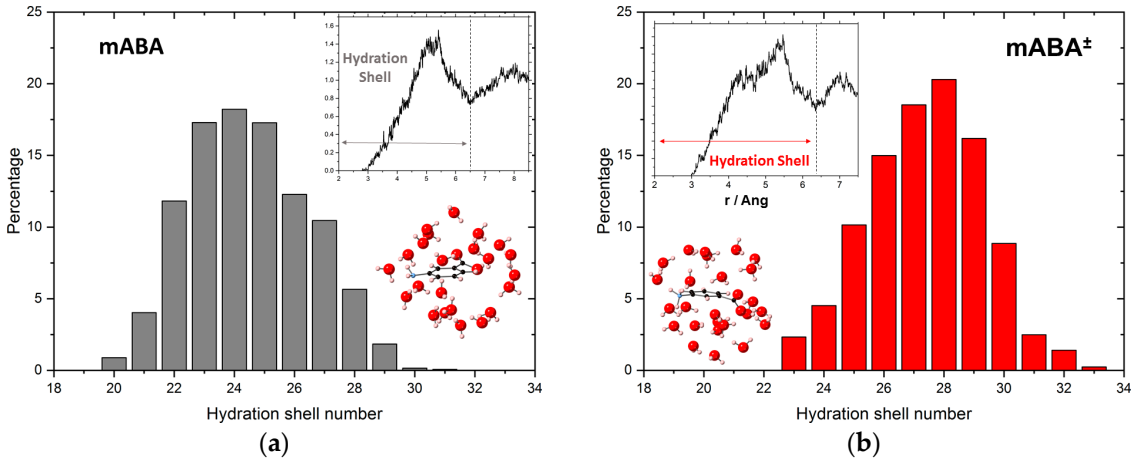


Figure 5. (**a**) Probability distribution of the coordination number in the hydration shell of mABA, together with the mABA–H_2O radial distribution function of the center-of-masses of mABA and water, and the optimized structure of the hydration shell of mABA. (**b**) Probability distributions of the coordination number in the first hydration shell of mABA±, together with the mABA±–H_2O radial distribution function of the center-of-masses of mABA± and water, and the optimized structure of the hydration shell of mABA±.

3.2. Dimerization of meta-aminobenzoic acid

Stable dimers in solution have often been linked to the structural synthon found in the crystal polymorph that crystallizes from solution.[36,37] This section reports therefore results from extensive DFT calculations to determine the structure and the thermodynamic stability in water of dimers of meta-aminobenzoic acid. The Boltzmann averaged energetics of formation of the nonionic, $(mABA)_2$, zwitterionic, $(mABA^{\pm})_2$ and nonionic-zwitterionic, [(mABA)(mABA±)], dimers are reported in Table 3. The free energy of formation $(mABA)_2$ ranges from –0.1 to 2.4 kcal mol mol^{-1}, depending on the method used to compute the total free energies of the dimers and monomer in water. The formation of (mABA)(mABA±) (2.4 kJ mol^{-1}) is also endergonic. On the other hand, the dimerization free energy of the zwitterionic aggregate $(mABA^{\pm})_2$ is large and negative (–5.8 kJ mol^{-1}).

Table 3. Energetics of dimerization of meta-aminobenzoic acid: $\Delta E_{e,gas}$ is the gas phase interaction energy; ΔG°_{ass} is the standard state (1 atm) gas-phase association free energy at 298 K; ΔG^{*}_{ass} is the standard state (1 mol/L) free energy of reactions in the liquid-phase. Calculations conducted at the M06-2X/6-31+G(d,p) level of theory using the SMD solvation model. Values obtained from the Boltzmann average of the energies, or free energies, of the isomers of nonionic $(mABA)_2$, zwitterionic $(mABA^{\pm})_2$ and mixed (mABA)(mABA±) dimers. Values in kcal mol^{-1}.

Reaction	$\Delta E_{e,gas}$	ΔG°_{ass}	ΔG^{*}_{ass}
2 mABA → $(mABA)_2$	-18.3	–6.6	–0.1 [1]
			2.4 [2]
mABA + mABA± → (mABA)(mABA±)			1.3 [2]
2 mABA± → $(mABA^{\pm})_2$	–	–	–5.8 [2]

[1] Gas-phase optimized geometries and free energies in water obtained using Eq. 2. [2] Solution-phase optimized geometries and free energies in water obtained using Eq. 3.

Figure 6 reports the structures of the two thermodynamically most stable $(mABA)_2$ and (mABA)(mABA±) species in water. The $(mABA)_2$ dimer corresponds to the structural synthon found in Form II,[2] where the two nonionic meta-aminobenzoic acid molecules interact through a double H-bond to form a classic carboxylic dimer (Figure 6(a). In the (mABA)(mABA±) dimer the two monomers are arranged to maximize the concomitant H-bonding and π-π interactions (Figure 6(b)). All other $(mABA)_2$ and (mABA)(mABA±) dimeric structure have significantly higher free energies of formation in water (2.5 kcal mol^{-1} < ΔG^{*}_{ass} < 10 kcal mol^{-1}) and consequently they are very unstable in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 January 2018 doi:10.20944/preprints201801.0069.v1

Peer-reviewed version available at Pharmaceutics 2018, 10, 12; doi:10.3390/pharmaceutics10010012

aqueous solution. On the other hand, several very stable zwitterionic dimers, $(mABA^{\pm})_2$, were found in solution (Figure 6(c)). Therefore, despite the distribution between zwitterions and nonionic molecules in water is close to unity,[13–15] the selective crystallization of the polymorphs that only contain zwitterionic molecules (Form I, III and V) could be driven by the higher stability in water of zwitterionic $(mABA^{\pm})_2$ dimers.

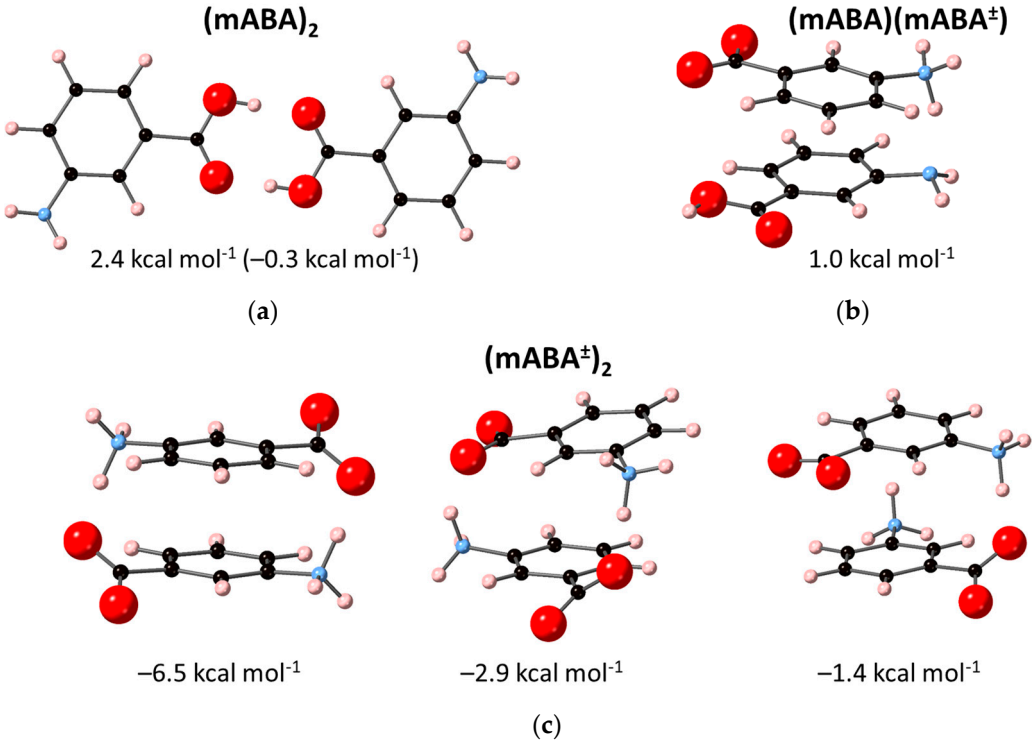


Figure 6. Optimized structures of most stable meta-aminobenzoic acid dimers in water: (**a**) nonionic $(mABA)_2$ dimer (in parenthesis value obtained using the gas-phase optimized geometries of $(mABA)_2$ and mABA); (**b**) nonionic–zwitterionic (mABA)($mABA^{\pm}$) dimer; (**c**) zwitterionic $(mABA)_2$ dimer. Beneath the structure is reported free energy of dimer formation in water.

3.3. Molecular aggregation in mixed mABA–mABA± aqueous solutions

Extensive classical MD simulations (≥ 200 ns) of mixed mABA–$mABA^{\pm}$ aqueous solutions were conducted to examine the aggregation behaviour of meta-aminobenzoic acid as a function of concentration. Four solution were considered: 0.04 mol L^{-1}, which corresponds to conditions below the limit of aqueous solubility of meta-aminobenzoic acid (5.9 g L^{-1});[38] 0.08 mol L^{-1}, 0.16 mol L^{-1} and 0.31 mol L^{-1}, which correspond to increasingly supersaturated conditions. Representative configurations of these solutions are reported in Figure 7, where the number of molecular aggregates that form in solution increases as a function of solute concentration. This aggregation process has been quantified in terms of the number of (mABA···mABA), ($mABA^{\pm}$···$mABA^{\pm}$) and (mABA···$mABA^{\pm}$) pairs within 4.0 Å (Figure 8 and Figure SI.3.1, ESI). The number of molecular pairs increases with the concentration but the number of nonionic clusters is significantly higher than mixed and zwitterionic species. As the dehydration of the molecules of solute is a crucial step during crystal nucleation from solution,[39] the stronger interaction of $mABA^{\pm}$ with the surrounding water molecules discussed in Section 3.1 could explain the observed different level of aggregation of nonionic and zwitterionic species in water.

Moreover, a close view of the clusters formed during the MD simulations reveals that meta-aminobenzoic acid interact via a manifold of inter-molecular interactions: H-bonding X–H···X (X = O or N) between amino (NH_2 and NH_3^+) and carboxylic (COOH and COO^-) groups; π-π interactions between benzine (C_6H_4) groups; X–H···π interactions.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 January 2018 doi:10.20944/preprints201801.0069.v1

Peer-reviewed version available at Pharmaceutics 2018, 10, 12; doi:10.3390/pharmaceutics10010012

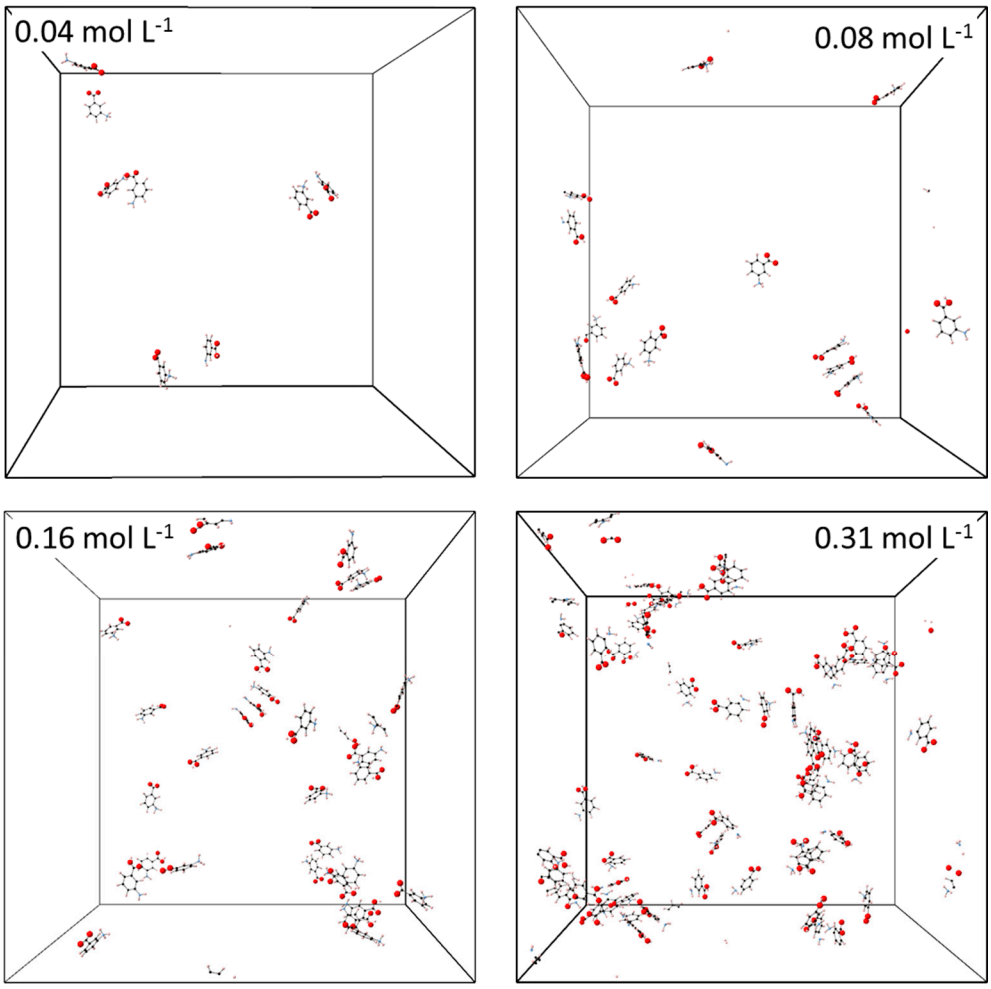


Figure 7. Configuration at 200 ns of mixed mABA–mABA± aqueous solutions. Water removed.

To obtain a characterization of these interactions during the aggregation process, a three-body simplified representation of the nonionic mABA (A–B–C) and zwitterionic mABA± (A*–B*–C*) molecules has been adopted (Figure 8), where A and A* represent the center-of-masses of $–NH_2$ and $–NH_3^+$; B and B* represent the center-of-masses of the benzine (C_6H_4) groups; C and C* represent the center-of-masses of–COOH and $–COO^-$.

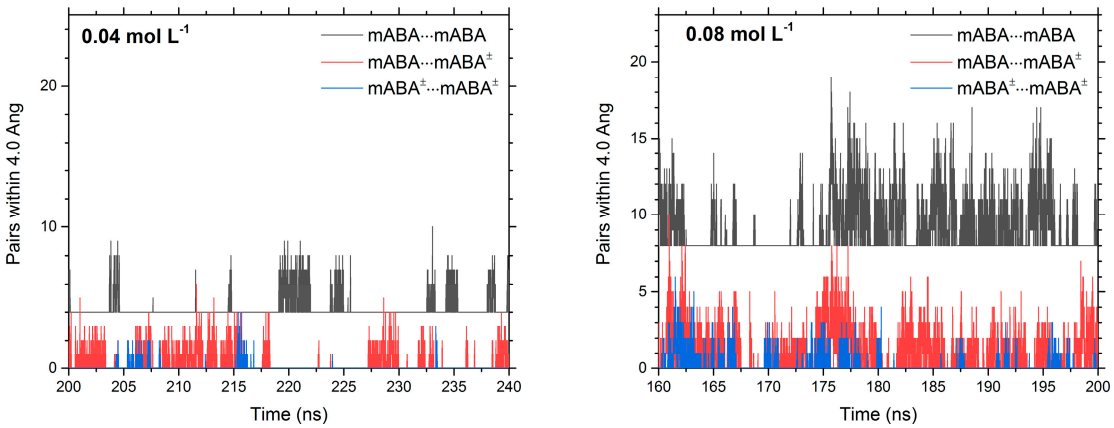
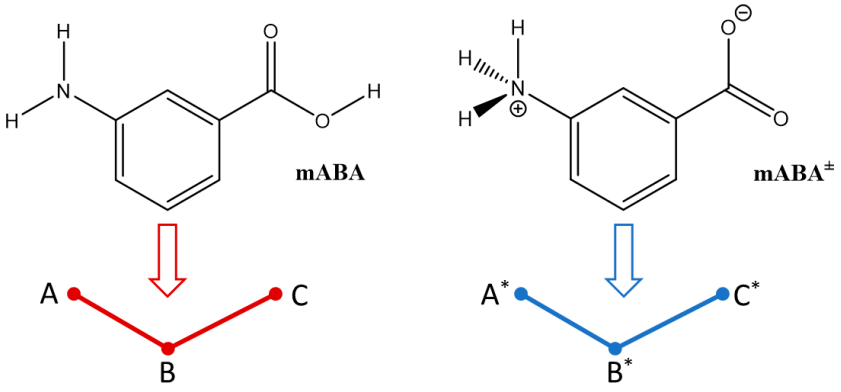


Figure 8. Time evolution of the number of pairs between meta-benzoic acid molecules in mixed mABA–mABA± aqueous solutions computed during the last 40 ns of the MD simulations.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 January 2018 doi:10.20944/preprints201801.0069.v1

Peer-reviewed version available at Pharmaceutics 2018, 10, 12; doi:10.3390/pharmaceutics10010012



(a)

Figure 8. Three-body representations (A–B–C) and (A*–B*–C*) of the nonionic, mABA, and zwitterionic, mABA± forms of meta-aminobenzoic acid: A and A* are the center-of-masses (COMs) of the –NH_2 and –NH_3^+ groups; B and B* are the COMs of the benzine (C_6H_4) group; C and C* are the COM of the –COOH and –COO^- groups.

A symmetric pairwise interaction matrix (PIM) can therefore be defined to quantify the interactions between (A–B–C) and (A*–B*–C*):

$$PIM = \begin{bmatrix} p_{A^*A^*} & p_{A^*B^*} & p_{A^*C^*} & p_{A^*A} & p_{A^*B} & p_{A^*C} \\ & p_{A^*B^*} & p_{B^*C^*} & p_{B^*A} & p_{B^*B} & p_{B^*C} \\ & & p_{C^*C^*} & p_{C^*A} & p_{C^*B} & p_{C^*C} \\ & & & p_{AA} & p_{AB} & p_{AC} \\ & & & & p_{BB} & p_{BC} \\ & & & & & p_{CC} \end{bmatrix} \quad (8)$$

In Eq. (6) the elements of the PIM matrix are defined as

$$p_{ij} = \left\langle \sum_i \sum_{i>j} f(r_{ij}) \right\rangle \quad (9)$$

where the pairwise interaction function $f(r_{ij})$ quantifies to the existence of a (i, j) pair within a cutoff distance of 4.0 Å:

$$f(r_{ij}) = \begin{cases} 0, & r_{ij} > 4.0 \text{ Å} \\ 1, & r_{ij} < 4.0 \text{ Å} \end{cases} \quad (10)$$

For example: the element p_{AA} corresponds to COOH···COOH interactions found in the classic carboxylic dimer $(mABA)_2$ (Figure 6(a)); p_{A^*A} and p_{C^*C} correspond to the COO^-···COOH and NH_3^+···NH_2 interactions in the nonionic-zwitterionic dimer (mABA)(mABA±) (Figure 6(b)); p_{B^*B} and $p_{A^*C^*}$ correspond to π···π and NH_3^+···COO^- interacting pairs in the structures of the most stable zwitterionic dimers $(mABA^{\pm})_2$ (Figure 6(c)). For the mixed 0.08 mol L^{-1} mABA–mABA± aqueous solutions, the pairwise interaction matrix in Table 4 reveals a higher proportion of NH_3^+···COO^- (A*···C* = 8.7%) and π··· π (B*···B* = 9.1%) pairs than COOH ···COOH (C···C = 6.5%), COO^-···COOH (C*···C = 6.5%) and NH_3^+···NH_2 (A*···A = 5.3%). Very similar PIM matrices were obtained from the calculation of the three-body pairwise interactions of the other systems (SI, Tables SI.4.1-3). This analysis consequently implies that aqueous solutions of meta-aminobenzoic acid contain a higher proportion of stable zwitterionic $(mABA^{\pm})_2$ pairs, in agreement with the DFT calculations of dimerization free energies.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 January 2018 doi:10.20944/preprints201801.0069.v1

Peer-reviewed version available at Pharmaceutics 2018, 10, 12; doi:10.3390/pharmaceutics10010012

Table 4. Matrix elements p_{ij} of the pairwise interaction matrix for the mixed 0.08 mol L^{-1} mABA–mABA$^{\pm}$ aqueous solutions. Values of p_{ij} expressed as percentage.

	A*	B*	C*	A	B	C
A*	0.2	0.6	8.7	5.3	3.4	3.5
B*		9.1	2.7	2.6	10.0	7.6
C*			0.1	3.6	2.3	3.6
A				4.3	4.2	5.2
B					6.1	10.6
C						6.5

5. Conclusions

Atomistic simulations of aqueous solutions of meta-aminobenzoic acid, an important model system to understand polymorphism in active pharmaceutical ingredients, were conducted to determine the solvation and aggregation of nonionic (mABA) and zwitterionic (mABA$^{\pm}$) molecules.

Ab initio molecular dynamics (AIMD) of mABA and mABA$^{\pm}$ were conducted to determine the stability, intermolecular and hydration properties of these species. AIMD were performed on simulation cells where single mABA and mABA$^{\pm}$ species were considered in combination with around 200 water molecules. A detailed analysis of the number and strength of hydrogen bonds between mABA and mABA$^{\pm}$ and the surrounding water molecules, shows that the mABA$^{\pm}$–water interaction is stronger than mABA–water, and that for both species the interaction with the surrounding water molecules is stronger around the carboxylic acid than around the $–NH_2$ (mABA) and $–NH_3^+$ (mABA$^{\pm}$) groups. Analysis of the mABA–H_2O and mABA$^{\pm}$–H_2O pair correlation functions indicate that although a hydration shell can be located for both molecules, the probability distributions of the number of water molecules surrounding mABA and mABA$^{\pm}$ show that these species display a flexible first coordination shell, where the flexibility increases on going from mABA$^{\pm}$ to mABA.

Density functional theory calculations with a polarizable continuum model to describe the aqueous environment were used to locate the low-lying energy structures and thermodynamic stability in water of nonionic, $(mABA)_2$, zwitterionic, $(mABA^{\pm})_2$ and nonionic-zwitterionic, (mABA)(mABA$^{\pm}$), dimers. Results show that the only thermodynamically dimers in solution are $(mABA^{\pm})_2$, whereas the formation of the nonionic classic carboxylic dimer $(mABA)_2$ and the π-π stacked (mABA)(mABA$^{\pm}$) dimer is endoergonic.

Classical molecular dynamics simulations of meta-aminobenzoic acid aqueous solutions containing an equal amount of nonionic and zwitterionic species were conducted to examine the aggregation behavior as a function of concentration of solute. Analysis of the aggregates formed during the simulation shows a higher proportion of π···π and NH_3^+···COO^-pairs, which interactions occur in the most stable zwitterionic dimers $(mABA^{\pm})_2$ located using DFT calculations.

According to these simulations the selective crystallization of the polymorph which constituent molecules are zwitterionic is driven by the formation of zwitterionic dimers in solution, which are thermodynamically more stable than (mABA)2 and (mABA)(mABA±) pairs.

The atomistic simulations reported in this work suggest therefore that the selective crystallization polymorphs which constituent molecules are zwitterionic is driven by the higher stability of zwitterionic species in solution. This work represents a paradigm of the role of molecular processes during the early stages of crystal nucleation in affecting polymorph selection during crystallization from solution.

Supplementary Materials: The following are available online at www.mdpi.com/link, Table SI.1: General AMBER forcefield parameters used to model mABA in GROMACS, Table SI.2: General AMBER forcefield parameters used to model mABA$^{\pm}$ in GROMACS, Table SI.3: Details of molecular dynamics simulation, Figure SI.3.1: Time evolution of the number of pairs between meta-benzoic acid molecules in mixed mABA–mABA$^{\pm}$

aqueous solutions computed during the last 40 ns of the MD simulations, Table SI.4.1: Matrix elements p_{ij} of the pairwise interaction matrix for the mixed 0.04 mol L^{-1} mABA–mABA$^{\pm}$ aqueous solutions.

Acknowledgments: 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). This research utilised Queen Mary's Apocrita HPC facility, supported by QMUL Research-IT. http://doi.org/10.5281/zenodo.438045

Author Contributions: E.G. and D.D.T. conceived and designed the experiments; E.G. and D.D.T. performed the experiments; E.G. and D.D.T. analyzed the data; D.D.T. wrote the paper.

Conflicts of Interest: "The authors declare no conflict of interest."

References

(1) Alcolea Palafox M., G. & J. L. N. M. *Spectrosc. Lett.* **1996**, *29:4* (January 2015), 609–629.

(2) Williams, P. A.; Hughes, C. E.; Lim, G. K.; Kariuki, B. M.; Harris, K. D. M. *Cryst. Growth Des.* **2012**, *12* (6), 3104–3113.

(3) Théorêt, A. *Spectrochim. Acta Part A Mol. Spectrosc.* **1971**, *27* (1), 11–18.

(4) Svärd, M.; Nordström, F. L.; Jasnobulka, T.; Rasmuson, Å. C. *Cryst. Growth Des.* **2010**, *10* (1), 195–204.

(5) Lahav, M.; Leiserowitz, L. *Chem. Eng. Sci.* **2001**, *56* (7), 2245–2253.

(6) ter Horst, J. H.; Geertman, R. M.; van Rosmalen, G. M. *J. Cryst. Growth* **2001**, *230* (1–2), 277–284.

(7) Gaines, E.; Maisuria, K.; Di Tommaso, D. *CrystEngComm* **2016**, *18* (16), 2937–2948.

(8) Blagden, N.; Davey, R. J. *Cryst. Growth Des.* **2003**, *3* (6), 873–885.

(9) Musumeci, D.; Hunter, C. A.; McCabe, J. F. *Cryst. Growth Des.* **2010**, *10* (4), 1661–1664.

(10) Kitamura, M.; Umeda, E.; Miki, K. *Ind. Eng. Chem. Res.* **2012**, *51* (39), 12814–12820.

(11) Hughes, C. E.; Williams, P. A.; Harris, K. D. M. *Angew. Chemie Int. Ed.* **2014**, *53* (34), 8939–8943.

(12) Gaines, E.; Maisuria, K.; Di Tommaso, D. *CrystEngComm* **2016**, *18* (16), 2937–2948.

(13) KUMLER, W. D. *J. Org. Chem.* **1955**, *20* (6), 700–706.

(14) Cohn, E. J.; Edsall, J. T. *Proteins, Amino Acids and Peptides*; Reinhold Publishing Corporation: New York, 1943.

(15) Bjerrun, N. *Z. Phys. Chem.* **1923**, *104*, 147.

(16) Hughes, C. E.; Williams, P. A.; Harris, K. D. M. *Angew. Chemie Int. Ed.* **2014**, *53* (34), 8939–8943.

(17) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J. J. Van; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. A. *Comput. Phys. Commun.* **2010**, *181* (9), 1477–1489.

(18) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Had, D. J. In *Gaussian, Inc Wallingford CT*; 2009.

(19) Grimme, S. *J. Comput. Chem.* **2006**, *27* (15), 1787–1799.

(20) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, 215–241.

(21) Marenich, A. V; Cramer, C. J.; Truhlar, D. G. *J.Phys.Chem. B* **2009**, *113*, 6378–6396.

(22) Ho, J.; Klamt, A.; Coote, M. L. *J. Phys. Chem. A* **2010**, *114* (51), 13442–13444.

(23) Ribeiro, R. F.; Marenich, A. V; Cramer, C. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2011**, *13* (23), 10908–10922.

(24) Montero, L. A.; Esteva, A. M.; Molina, J.; Zapardiel, A.; Herna, L.; Màrquez, H.; Acosta, A. *J. Am. Chem. Soc.* **1998**, *120* (6), 12023–12033.

(25) Sanchez-Garcia, E.; Studentkowski, M.; Montero Luis, A.; Sander, W. *ChemPhysChem* **2005**, *6* (4), 618–624.

(26) Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*

(1), 15–25.

(27) Perdew, J.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77* (18), 3865–3868.

(28) Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54* (3), 1703–1710.

(29) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. **2004**, *25* (9), 1157–1174.

(30) Salvalaglio, M.; Perego, C.; Giberti, F.; Mazzotti, M.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *112* (1), E6–E14.

(31) Salvalaglio, M.; Giberti, F.; Parrinello, M. *Acta Crystallogr. Sect. C Struct. Chem.* **2014**, *70*, 132–136.

(32) Toroz, D.; Hammond, R. B.; Roberts, K. J.; Harris, S.; Ridley, T. *J. Cryst. Growth* **2014**, *401*, 38–43.

(33) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91* (24), 6269–6271.

(34) Gaigeot, M.-P.; Sprik, M. *J. Phys. Chem. B* **2004**, *108* (22), 7458–7467.

(35) Tang, E.; Di Tommaso, D.; de Leeuw, N. H. *J. Chem. Phys.* **2009**, *130* (23), 234502.

(36) Davey, R. J.; Dent, G.; Mughal, R. K.; Parveen, S. *Cryst. Growth Des.* **2006**, *6* (8), 1788–1796.

(37) Du, W.; Cruz-Cabeza, A. J.; Woutersen, S.; Davey, R. J.; Yin, Q. *Chem. Sci.* **2015**, *6* (6), 3515–3524.

(38) National Center for Biotechnology Information. PubChem Compound Database. 3-AMINOBENZOIC ACID https://pubchem.ncbi.nlm.nih.gov/compound/7419 (accessed Jan 2, 2018).

(39) Khamar, D.; Zeglinski, J.; Mealey, D.; Rasmuson, Å. C. *J. Am. Chem. Soc.* **2014**, *136* (33), 11664–11673.