Supplementary Information

Can solvated intermediates inform us about nucleation pathways? The case of β-*p*ABA

A. J. Cruz-Cabeza, E. Taylor, I. J. Sugden, D. H. Bowskill, S. E. Wright, H. Abdullahi, D. Tulegenov, G. Sadiq, R. J. Davey

1 Crystallographic Data

	*p*ABA solvates		*p*F-pABA solvates	
Solvate	Acetone	Dioxane	Acetone	Dioxane
Stoichiometry	1:1	1:1	1:1	1:1
CCDC Deposition No.	2009903	2009898	2009893	2009900
Chemical Formula	$2(C_7\ H_7\ N\ O_2)$, $C_3\ H_6\ O$	$2(C_7\ H_7\ N\ O_2)$, $C_4\ H_8\ O_2$	$C_7\ H_3\ F_4\ N\ O_2$, $C_3\ H_6\ O$	$C_7\ H_3\ F_4\ H\ O_2$, $2(C_4\ H_8\ O_2)$
Mr	332.35	362.37	267.18	385.31
Crystal System	triclinic	monoclinic	monoclinic	monoclinic
Space Group	P-1	$P2_1/c$	I2/m	$P2_1/c$
a (Å)	5.0395(2)	22.4397(9)	6.5218(10)	7.0886(2)
b (Å)	8.2374(3)	20.5326(7)	17.636(2)	18.8246(5)
c (Å)	20.4205(7)	11.9083(4)	10.0639(14)	12.8623(4)
α (°)	88.805(2)	90	90	90
β (°)	83.313(2)	105.194(3)	107.754(14)	92.882(2)
γ (°)	84.487(2)	90	90	90
Volume (Å^3)	838.00(5)	5294.9(3)	1102.4(3)	1714.18(9)
Z	2	12	4	4
Density (g/cm^3)	1.317	1.364	1.61	1.493
Abbsorption Coeffient μ (mm^{-1})	0.813	0.862	0.161	1.248
$\theta_{min,}\ \theta_{max}$	6.55, 72.23	4.31, 71.91	4.042, 27.905	4.17, 71.73
Theta max	72.23	71.91	27.905	71.73
Temperature (K)	150(2)	150(2)	150.02(10)	150(2)
Measured Reflections	8371	10342	3925	13651
Independent Reflections	3161	10342	1363	3275
Observed Reflections [$I > 2\sigma(I)$]	2878	6217	978	2790
Parameters	237	722	98	244
Tmin, Tmax	0.859, 0.938	0.777, 0.942	0.739, 1.000	0.761, 0.855
$R[F^2 > 2\sigma(F^2)]$	0.0381	0.0801	0.0672	0.0306
$wR(F^2)$	0.1114	0.2386	0.203	0.0811
GooF	1.067	1.04	1.103	1.021
$\Delta\rho_{max,}\ \Delta\rho_{min}$ (e Å^{-3})	0.227, -0.285	0.625, -0.319	0.873, -0.44	0.225, -0.204

pABA derivatives									
Compound	2-F-pABA	pF-pABA	3-Cl-pABA	2-Me-pABA	3-Me-pABA	3-OH-pABA	2-OMe-pABA	2-OMe-5-Cl-pABA	2-OEt-5-Cl-pABA
CCDC Deposition No.	2009895	2009899	2009894	2009902	2009901	2009891	2009892	2009897	2009896
Chemical Formula	$C_7 H_6 F N O_2$	$C_7 H_3 F_4 N O_2$	$C_7 H_6 Cl N O_2$	$C_8 H_9 N O_2$	$C_8 H_9 N O_2$	$C_7 H_7 N O_3$	$C_8 H_9 N O_3$	$C_8 H_8 Cl N O_3$	$C_9 H_{10} Cl N O_3$
Mr	155.13	209.1	171.58	151.16	302.32	153.14	167.16	201.6	215.63
Crystal System	monoclinic	monoclinic	monoclinic	orthorhombic	monoclinic	orthorhombic	monoclinic	triclinic	monoclinic
Space Group	C2/c	$P2_1/n$	$P2_1/c$	$P2_12_12_1$	$P2_1/c$	Pbca	$P2_1/c$	P-1	$P2_1/c$
a (Å)	14.7711(8)	4.1957(2)	3.7229(3)	7.74940(14)	10.7033(4)	11.6926(9)	8.1846(7)	7.2780(6)	9.7014(12)
b (Å)	3.7877(2)	14.9867(6)	11.1634(8)	8.12614(16)	8.7280(3)	8.1121(7)	13.5952(10)	7.5330(7)	14.9141(12)
c (Å)	24.3334(12)	11.5145(5)	16.9454(11)	22.7285(4)	16.1405(7)	14.1796(10)	7.2087(6)	16.2004(11)	6.9582(9)
α (°)	90	90	90	90	90	90	90	94.713(6)	90
β (°)	107.378(4)	97.566(4)	95.218(7)	90	95.802(4)	90	109.909(10)	91.340(6)	109.984(14)
γ (°)	90	90	90	90	90	90	90	106.926(7)	90
Volume ($Å^3$)	1299.28(12)	717.73(6)	701.34(8)	1431.28(5)	1500.10(10)	1344.95(17)	754.18(11)	845.80(12)	946.1(2)
Z	8	4	4	8	4	8	4	4	4
Density (g/cm^3)	1.586	1.935	1.625	1.403	1.339	1.513	1.472	1.583	1.514
Abbsorption Coeffient μ (mm^{-1})	0.136	0.208	0.483	0.842	0.097	0.12	0.114	0.422	0.383
$\theta_{min,}\ \theta_{max}$	2.873, 28.686	2.239,	4.067	6.018	1.882	4.159	3.981	3.935	4.136
Theta max	28.686	26.307	28.522	76.525	27.739	28.971	27.109	28.06	26.216
Temperature (K)	150.00(10)	100.12(10)	150	150.00(10)	150.01(10)	150	150	150	150
Measured Reflections	7607	8394	5117	6832	18127	6923	5879	7197	4839
Independent Reflections	1554	1458	1728	2901	3603	1669	1877	3951	2250
Observed Reflections [I > 2σ(I)]	1386	1306	1342	2813	2552	1150	1219	2562	1420
Parameters	100	139	102	271	201	103	111	243	129
Tmin, Tmax	0.64803, 1	0.3095, 1	0.35418, 1	0.87873, 1	0.40453, 1	0.84492, 1	0.79079, 1	0.84804, 1	0.91675, 1
$R[F^2 > 2\sigma(F^2)]$	0.0366	0.0297	0.0427	0.0277	0.0506	0.0495	0.0532	0.0559	0.0634
$wR(F^2)$	0.1098	0.0871	0.1027	0.0737	0.1437	0.1562	0.1382	0.1794	0.19
GooF	1.081	1.092	1.059	1.058	1.035	0.928	1.03	0.935	0.958
$\Delta\rho_{max,}\ \Delta\rho_{min}$ (e $Å^{-3}$)	0.403, -0.274	0.345, -0.204	0.314, -0.336	0.159, -0.147	0.26, -0.278	0.267, -0.286	0.227, -0.262	0.312, -0.321	0.499, -0.521

2 Solubility Measurements

pF-*p*ABA in water. Using gravimetric analysis, the solubilities of the pF-*p*ABA and 2F-*p*ABA samples were determined between 5^0 and 40^0C in aqueous solutions. Saturated solutions of the pure solids were prepared in a 100mL parafilm insulated jacketed vessel. The temperature of the vessel was controlled using a water bath and the saturated solutions of each compound were constantly stirred at 300rpm for a period of 24h to allow for equilibration. After stirring, the excess undissolved solute was left to settle for 1h without agitation. 4ml samples of the saturated solutions were taken using a 5ml syringe and filtered using a 0.22µm membrane filter to remove smaller undissolved solute. The remaining solution was transferred into 5 sample vials of known weight. The samples and vials were weighed and allowed to evaporate at 55°C in a drying oven. The masses of the residues were then calculated. The mean solubility values at each temperature were calculated using a mass balance and are given below:

	Solubility ± Standard Deviation (g/kg water)*			
Molecule	**5 °C**	**10 °C**	**25 °C**	**40 °C**
2-F-pABA	0.464 ± 0.093	0.700 ± 0.031	1.261 ± 0.089	1.453 ± 0.073
2,3,4,5-tetraF-pABA	7.641 ± 0.640	7.872 ± 0.330	13.440 ± 0.336	23.260 ± 0.083
2-Cl-pABA	0.1004 ± 0.037	0.2325 ± 0.035	0.3141 ± 0.091	1.0104 ± 0.078
3-Cl-pABA	0.5236 ± 0.631	0.2165 ± 0.067	0.5843 ± 0.595	0.7206 ± 0.126

* Five measurements

***p*ABA in water at various pHs.**

Solubilities (α form) were measured as a function of pH (1.5 – 6.5) at 10 and 25°C. Saturated solutions were prepared by the addition of excess pABA (the calculated ideal values acted as a guide) to a stoppered 50ml flask containing 30 ml of water and fitted with a stirring bar. Known volumes of HCl or NaOH were added in small increments to alter the pH to a desired value. If visual observation indicated the presence of excess solids, the solution was allowed to stir for at least 3 hours in order to reach equilibrium.

pH was measured using a Mettler Toledo AG8603 pH meter (±0.01). 1 mL samples (five repeats) of these saturated pABA solutions were subjected to gravimetric analysis by evaporating to dryness at 65°C.From the measured mass of residue and the known volumes of water and acid or base solution added, compositions of saturated solutions in ML^{-1} were calculated.

25°C		**10°C**	
pH	**Concentration mol L^{-1}**	pH	**Concentration mol L^{-1}**
0.77	0.675	1.25	0.622
1	0.788	1.67	0.288
1.5	0.473	1.87	0.136
1.85	0.2	2.51	0.098
2.46	0.071	3.8	0.029
3.8	0.035	5.01	0.097
4.85	0.054	5.65	0.201
5.72	0.197	6.3	0.789
6.16	0.716		

3 Computational methods

3.1 Crystal structure prediction (CSP)

CSP calculations for all systems ((1) *p*ABA ((a) Z'=1, (b) Z'=2), (2) water, (3) *p*ABA:water 1:1 ratio, (4) *p*ABA:water 1:2 ratio) were carried out using the code CrystalPredictor (version 2.4.3). Flexibility was determined using gas phase finite difference perturbations around degrees of freedom indicated as potentially flexible by second derivatives at the gas phase minimum; this meant the carboxylic acid group torsion. The level of theory used was B3LYP, with the Aug-CC-pVTZ basis set, in Gaussian 09. A uniform local approximate model (LAM) grid was set up with LAMs evaluated at 0.0 and 180.0 degrees; a pass of the adaptive LAM algorithm indicated that this was sufficient to accurately describe the flexibility within the molecules. The potential parameters for C, H-C (hydrogen attached to carbon), N, O, and H-n (hydrogen attached to a polar atom) from the work of Williams and co-workers were used to describe the exchange-repulsion and dispersion interactions. The structure

generation stage sampled the 59 most common space groups. 500k and 1 million structural minimisations were run for the $Z' = 1$ (1a, 2, 3, 4 and 5) and 2 (1b) searches respectively. After the CrystalPredictor calculations were completed a final clustering of generated structures was carried out with the COMPACK algorithm.

3.2 Refinement of Lattice energies

In order to refine the calculated lattice energies, the generated structures were minimised with an improved energy model for electrostatics consisting of atomic multipoles, with extended flexibility (amine hydrogen angles, and all angles within the carboxylic acid group), together with the same FIT potential, using CrystalOptimizer. The same level of theory (B3LYP/ Aug-CC-pVTZ) was employed

3.3 Improved lattice energies: TPSS-D3

As a final step the lattice energies of the experimental form of *p*ABA (alpha) and ice (hexagonal) and the five lowest energy structures from investigations 3 and 4 were recomputed with periodic density functional theory with van der Waals corrections. For this, the TPSS functional was used, with Grimme's D3 vdW corrections, as implemented in the VASP code (version 5.4). An energy cut off of 1000 eV was used for the plane waves. The Brillouin zone was sampled using a Gamma centred Monkhorst–Pack approximation at k-point grids separated by approximately $0.025*2\pi$ $Å^{-1}$. Crystal structures were relaxed with this model allowing the unit cell volume as well as the atomic positions to optimise. Structural relaxations were halted when the calculated force on every atom was less than 0.01 eV $Å^{-1}$.

4 Computational Results

Sections 4.1 – 4.3 summarise the results of CSP of water and *p*ABA 1:1 and 1:2 hydrates. Tables 4.1 and 4.2 provide the calculated lattice energies of the *p*ABA polymorphs[3], α and β and the free energies of hydrate formation of the most stable predicted hydrates.

4.1 CSP investigation of water.

The alcohol OH FIT potential used for water hydrogens[1], as has been used in previous studies[2].

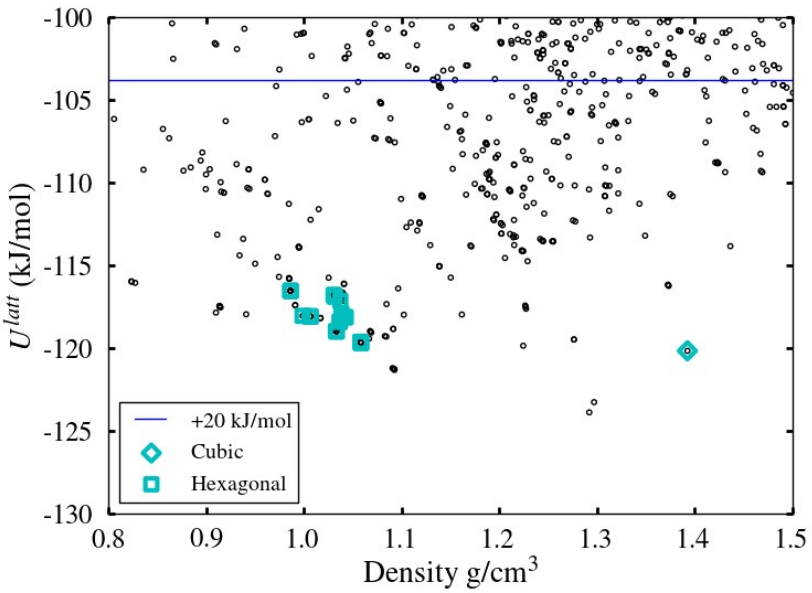

Figure 4.1 Polymorphic landscape for water, following refinement.

Several static proton fixed structures match with the disordered experimental hexagonal ice structure, consistent with an approximation of said disorder. The structures are significantly denser than experiment, consistent with previous investigations showing collapse of structure when disorder is ignored[2].

4.2 CSP investigation of *p*ABA:water 1:1 ratio.

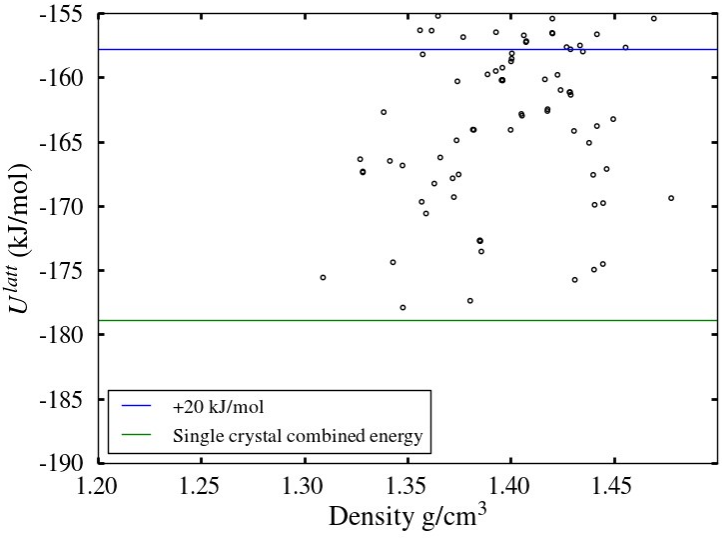

Figure 4.2 1:1 *p*ABA:hydrate polymorphic landscape. Green line is the sum of the lowest energy of neat water and the lattice energy of β *p*ABA (AMBNAC08). Blue line is +20 kJ/mol from the global minimum. Energies are reported as per mole of asymmetric unit.

4.3 CSP investigation *p*ABA:water 1:2 ratio.

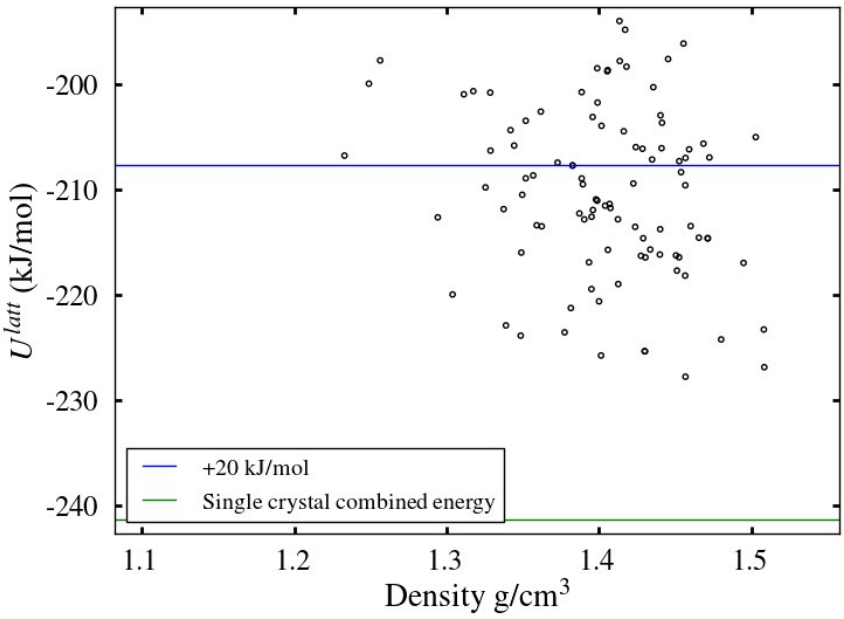

Figure 4.3. 1:2 *p*ABA:hydrate polymorphic landscape. Green line is the sum of twice the lowest energy of neat water and the lattice energy of β *p*ABA (AMBNAC08). Blue line is +20 kJ/mol from the global minimum. Energies are reported as per mole of asymmetric unit.

4.4 Improved lattice energies: TPSS-D3

The results presented in the previous sections indicate that there is no thermodynamic driving force for the formation of hydrates, in that the lowest lattice energy hydrate structures are higher in energy than the sum of the independent molecular component's (ice and *p*ABA) lowest energy crystal structure. However, given the approximations inherent in the computational method, particularly in the case of water, it was decided to perform full wavefunction calculations for the lowest 5 hydrate structures from each investigation, and compare the predicted lattice energies to the neat crystal structures at the same level of theory. Energies are reported per mole of molecules in the asymmetric unit.

The free energy of hydrate formation was calculated using the following equation[4]:

$$\Delta G_{hydrate\ formation} \approx \Delta E_{latt-gain}^{Hydrate} + 0.6$$

where by $\Delta E_{latt-gain}^{Hydrate}$ is the difference in lattice energy between the predicted hydrate minus the lattice energy of α *p*ABA minus the lattice energy of ice times the appropriate stoichiometry

(being either 1 for monohydrate or 2 for dihydrate). This confirms that none of the structures are expected to appear experimentally since all are less stable than the pure components.

	$RMSD_{15}$ (Å)	Ulatt (kJ/mol)
α *p*ABA	0.072	-131.1
Hexagonal ice	0.132	-69.7

Table 4.1 Lattice Energies (VASP) for Hexagonal Ice and α *p*ABA.

System	U_{latt} (kJ/mol)	$\Delta G_{hydrate\ formation}\ (\frac{kJ}{mol})$
*p*ABA: hydrate 1:1, rank 1	-187.7	13.7
*p*ABA: hydrate 1:1, rank 2	-194.1	7.3
*p*ABA: hydrate 1:1, rank 3	-188.5	12.9
*p*ABA: hydrate 1:1, rank 4	-186.4	15.0
*p*ABA: hydrate 1:1, rank 5	-186.9	14.5
*p*ABA: hydrate 1:2, rank 1	-221.1	50
*p*ABA: hydrate 1:2, rank 2	-18.1	253
*p*ABA: hydrate 1:2, rank 3	-252.3	18.8
*p*ABA: hydrate 1:2, rank 4	-251.8	19.3
*p*ABA: hydrate 1:2, rank 5	-256.0	15.1

Table 4.2 Lattice Energies and Free Energies of Hydrate Formation for most stable predicted *p*ABA hydrates.

References

1 S. R. Cox, L. -Y Hsu and D. E. Williams, *Acta Crystallogr. Sect. A*, 1981, **37**, 293–301.

2 A. T. Hulme and S. L. Price, *J. Chem. Theory Comput.*, 2007, **3**, 1597–1608.

3. A. J. Cruz-Cabeza, R. J. Davey, I. D. H. Oswald, M. R. Ward and I. J. Sugden, CrystEngComm, 2019, 21, 2034-2042.

4. A. J. Cruz-Cabeza, S. E. Wright and A. Bacchi, Chem. Commun., 2020 5127-5130.