The Structural Pathway from its Solvated Molecular State to the Solution Crystallisation of the α- and β-Polymorphic Forms of Para Amino Benzoic Acid

Ian Rosbottom [a,#], *Thomas D Turner* [a,&], *Cai Y Ma* [a], *Robert B Hammond* [a] *and Kevin J Roberts* [a,*]

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

Chin W Yong [b,c], *Ilian T Todorov* [b]

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

[c] *Division of Pharmacy and Optometry, University of Manchester, Manchester, UK*

Current addresses:

[#] *GlaxoSmithKline, Gunnels Wood Rd, Stevenage, SG1 2NY, UK*

[&] *School of Mechanical Engineering, University of Leeds, Leeds, LS2 9JT, UK*

* Corresponding Author

Supplementary Information

S1. Para Amino Benzoic Acid

This supplementary section provides a selective review of some key research related to the solution crystallisation of para amino benzoic acid (PABA).

S1.1 Molecular Properties and Crystallographic Structures

Para-aminobenzoic acid (PABA), an organic compound, also known as vitamin B10, is a white crystalline substance with molecular formula $(NH_2)(C_6H_4)(COOH)$, consisting of three molecular moieties: an aromatic benzene ring substituted with carboxylic acid and amino groups which are para with respect to each other (**Figure S1**). PABA has UV absorption and antifibrotic properties, hence once being widely used for sunscreen.

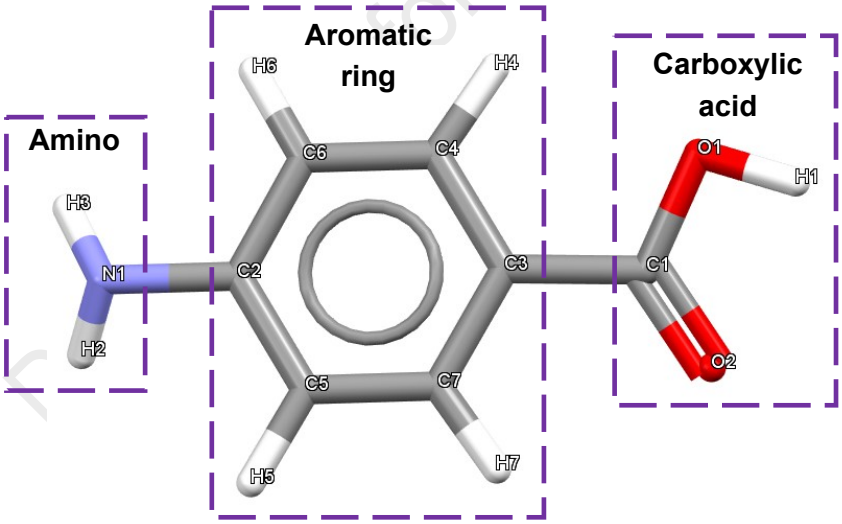

Figure S1. The molecular structure of para-aminobenzoic acid with the atom notations and the three molecular moieties (functional groups): aromatic ring, carboxylic acid and amino group as indicated in the boxes.

The crystal structures of the α- and β-forms of para-aminobenzoic acid were first solved

over 50 years ago (AMBNAC [1] and AMBNAC01 [2]) with further refinement and determination (e.g., AMBNAC04 [3], AMBNAC06 [4], AMBNAC10/11 [5]). The α- and β-forms are enantiotropic with the α-form being more stable at higher temperature than the β-form. The solubilities of these two forms are very similar which has led the exact enantiotropic transition temperature being in some doubt but studies to date seem to place this in the range of ca. 13 – 25 ^{o}C [6-9]. The crystal structures of other polymorphs including the δ-form (AMBNAC14-18 [10]) crystallised at high pressure and γ-form (AMBNAC09 [11]) have also been determined in the recent years. Recently, Cruz-Cabeza et al. [12] highlighted and reviewed polymorphism in PABA including the α-, β-, γ- and δ-forms. The comparison studies found that γ-form was found to be very similar to α-form with the δ-form having some similarities with α- and β-forms. Overall, analysis of the PABA energy landscape demonstrated the comparative rarity of the β-form.

Characteristic molecular descriptors and associated crystallographic structural data for PABA is summarised in **Table S1**. The crystal structures of PABA in both the α- and β-forms are monoclinic with a $P2_1/n$ space group. Interestingly, the β-form is more close packed with a higher density and less void space than the α-form. Although the α-form has two molecules in its asymmetric unit, all three PABA molecular structures (2 from α-form and 1 from β-form) conformers are very similar. Analysis of the hydrogen bonds (H-bonds), as shown in **Table S2**, reveals that they are stronger in the β-form structure than in the α-form, as evidenced by the OH…O and NH…O bond lengths in the β-form being ca. 5% and 7% respectively shorter than in the α-form.

In comparison to representative pharmaceutical materials [13, 14], analysis of the molecular descriptors for PABA (**Table S1**) reveals that PABA has a relatively low molecular weight (137.14 g/mol) with lower than average numbers of potential H-bond donors (3) and acceptors (2) and aromatic rings (1), with an average number of rotatable bonds (3) in terms of molecular flexibility, when compared to the distribution of the approved pharmaceuticals [13, 14]. This is well-consistent with that expected for a small molecule pharmaceutical compound.

Table S1. Characteristic molecular descriptors and crystallographic structural data for the para-aminobenzoic acid polymorphs.

Material Descriptor	α-form	β-form
Refcode	AMBNAC06 [4]	AMBNAC04 [3]
Molecular weight (g/mol)	137.14 / 137.14	137.14
Molecular volume (Å^3)	117.32 / 118.34	124.86
Molecular surface area (Å^2)	138.40 / 138.30	142.94
Melting point (°C)	187.3 [6]	140.0 [8]
H-bond Donors	3 / 3 (O_1, O_2, N_1 / O_3, O_4, N_2)	3 (O_1, O_2, N_1)
H-bond Acceptors	2 / 2 (O_2, N_1 / O_4, N_2)	2 (O_1, N_1)
Rotatable Bonds	3 / 3	3
Space Group	$P2_1/n$ [4]	$P2_1/n$ [3]
Z / Z'	8 / 2 [4]	4 / 1 [3]
***a* (Å)**	18.571 [4]	6.278 [3]
***b* (Å)**	3.843 [4]	8.583 [3]
***c* (Å)**	18.632 [4]	12.365 [3]
β (°)	93.67 [4]	100.13 [3]
Cell volume (Å^3)	1327.06 [4]	655.91 [3]
Packing coefficient	0.692	0.720
Void space (%)	26.33	22.91
Density (g/cc)	1.37	1.39

Note: As the α- and β-forms of PABA are enantiotropically interrelated, the melting point quoted in the literature [6,8] with respect to the β-form may be referring to the transformation temperature from the β-form to the α-form in the solid-state rather than

being a melting point from the β-form into its molten state.

Table S2. The H-bond geometrical details of the α- and β-form PABA molecules with the contribution donor (D) and acceptor (A) sites together with their respective polarizability.

Polymorph	H-bond	H···A /Å	D···A /Å	D-H /Å	D-H···A /°
α-form (1st molecule)	O2H1…O1	1.836	2.650	0.819	172.23
	N1H4…O3	2.574	3.368	0.860	154.05
α-form (2nd molecule)	O4H8…O3	1.813	2.616	0.820	166.05
	N2H11...O1	2.131	2.969	0.860	164.65
β-form	O1H1…N1	1.730	2.754	1.065	159.95
	N1H2…O2	2.186	3.045	0.884	163.98

S1.2 Crystal Chemistry and Intermolecular Interactions

The examinations of the solid-state chemistry of α- and β-form crystal structures [15] found that the α-form has strong H-bonding carboxylic acid OH…O dimer interactions, π-π stacking and NH…O interactions (**Figure S2(a)**) and that it crystallises with a needle- or lath-like morphology at higher temperatures. In contrast, the low temperature β-form has a characteristic 4-membered H-bonding ring comprising identical pairs of alternating OH…N and NH…O interactions. The β-form crystallises with a prismatic equant morphology (**Figure S2(b)**). It has been found that β-form can be quite difficult to crystallise [8, 16] even at below transition temperature and indeed it can only be formed from aqueous slurry conversion from the α-form at ca. 5 °C using an agitated batch crystalliser over a ca. two week period [17].

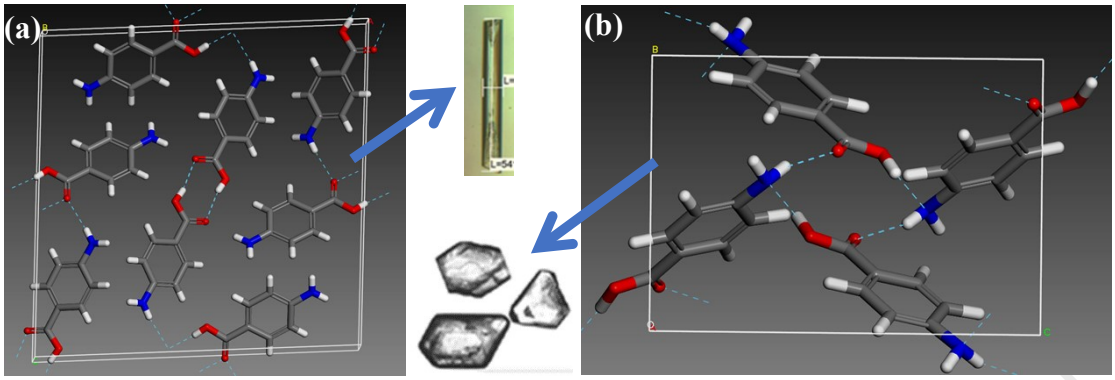

Figure S2. The crystal packing structures of para-aminobenzoic acid: (a) α-form; (b) β-form with their H-bonding networks and also observed morphologies (data derived from [15]).

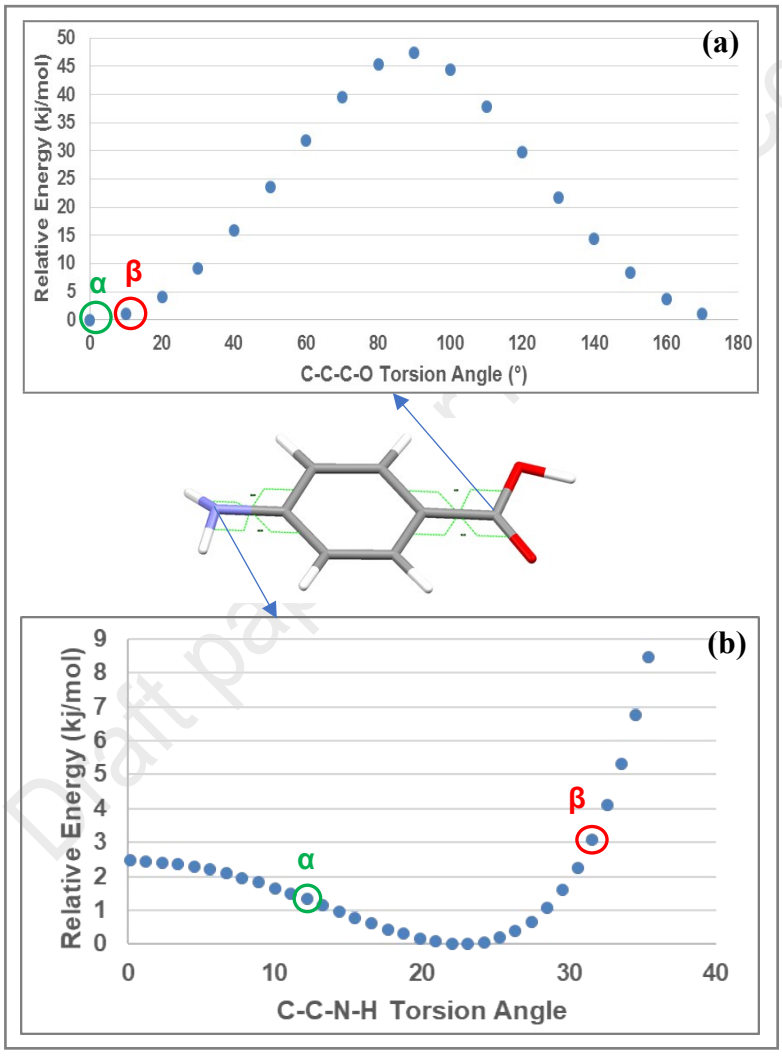

Figure S3. The conformational DFT analysis of α- and β-form PABA molecular structures using (a) rigid carboxyl acid (COOH) rotation and (b) pyramidal bend of

amino (NH_2) groups (data derived from [18]).

Detailed molecular conformational analysis for these two polymorphic structures reveals some interesting differences. **Figure S3** shows results from an analysis of the relative conformational stabilities [18] of both α- and β-form molecules with respect to the carbonyl (a) and amino (b) groups. This reveals the β-form conformation to be slightly distorted with respect to the more stable α-form. In the α-form conformer, the carboxyl acid group was found to be very close to the planar conformational minima and the slight pyramidal bend of amino group, also being found to be close to the pseudopyramidal conformational minima. For the β-form molecule, the carboxyl acid group was found to be slightly rotated away from a planar conformation, whilst the amino group was found to be more pyramidal. This conformational distortion reflects the need to complete the β-form's tetra-molecular H-bonded ring structure. As the α-form PABA conformation is closer to the minimum energy conformation (**Figure S3**), this would present a lower energy barrier to the crystallization of α-form when compared to the β-form, consistent with the known challenges in crystallising the β-form. The rigid planar COOH dimers formed in the α-form crystal structure strongly hold the COOH group planar (**Figure S2(a)**). The NH…O H-bond is also relatively planar, leading overall to the α-form's tetra-molecular core building block remaining planar. On the contrast, the 4-membered H-bond ring in β-form structure distorts conformation of both functional groups as the OH…N H-bond pulls N atom towards OH group and the NH…O H-bond pulls the H atom towards the C=O group (**Figure**

S2(b)). That said, whilst the conformation deformation penalty could imply lower crystallisability of the β-form, once formed the strength of its core 4 member H-bonding ring structure would suggest a strong and stable arrangement for its subsequent development and growth (**Table S2**).

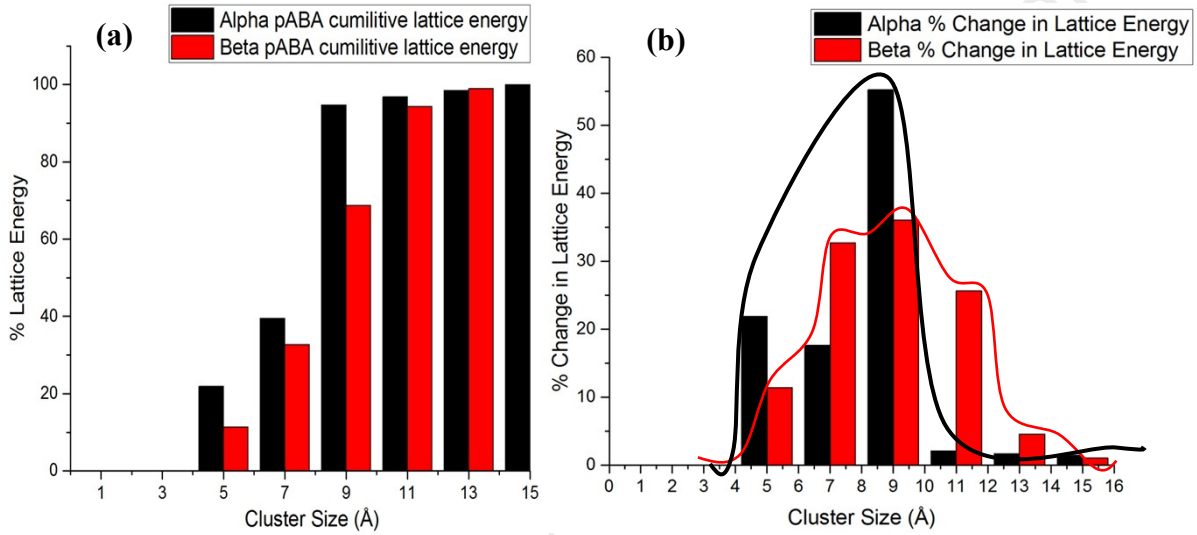

Figure S4. Cumulative (a) and discretized (b) lattice energy distributions as a function of the cluster size (Black – α-form PABA; Red – β-form PABA) (data derived from [15]).

Calculation of the lattice energy for the α- and β-form structures of PABA reveals that the α-form achieves more cohesive energy from its near inter-molecular neighbours than the β-form (**Figure S4**). This is consistent with the α-form needing a smaller cluster size for its structural stability at nucleation when compared to the β-form which implies, in turn, that high supersaturation would be expected to favour the formation of the α-form and vice versa with respect to the β-form.

As shown in **Table S3**, the functional group analysis for these two polymorphs indicates that the α-form might be expected to dominate crystallisation due to the impact of its

dominant carboxylic dimer interaction (ca. 50% contribution of lattice energy), even at the lower temperatures where β-form is expected to be more thermodynamically stable [19].

Table S3. Lattice energy together with the percentage of contributions from three functional groups (amino, aromatic ring and carboxyl acid) (data derived from [19]).

α-form		β-form
-24.45	**Lattice Energy (kcal/mol)**	-22.73
15.3 (donor only)	**NH_2 (%)**	23.8 (both donor and acceptor)
39.8	**C_6H_6 (%)**	42.5
44.7	**COOH (%)**	33.7

S1.3 Solution Properties, Solvation and Solubility

Synthon propensity studies [15, 18], using the COSMO-RS approach, was ultilised to compare the relative populations of the energetically top-ranked synthons (Aα, Bα and Aβ, Bβ) in solution for a range of solvent systems (**Figure S5**). Synthon Aα associated with OH…O H-bonding interactions was found to be dominant for all solvents studied with the exception of water where the concentration of this synthon was found to be much lower and where π-π interactions were much more prevalent notably the π-π head-to-head (H2H) (synthon Bα) and head-to-tail (H2T) (synthon Aβ) interactions which were found to be the most stable. In this, calculation of the surface charge distributions for the synthons reveals that the π-π interaction synthons (Bα, Aβ) have greater polar surface area when compared to the OH…O H-bonding synthons (Aα, Bβ). Overall, whilst the data supports the preferential crystallisation of the α-form from most

solvents (except water), it does suggest that the β-form would be more likely to be crystallised from aqueous solutions.

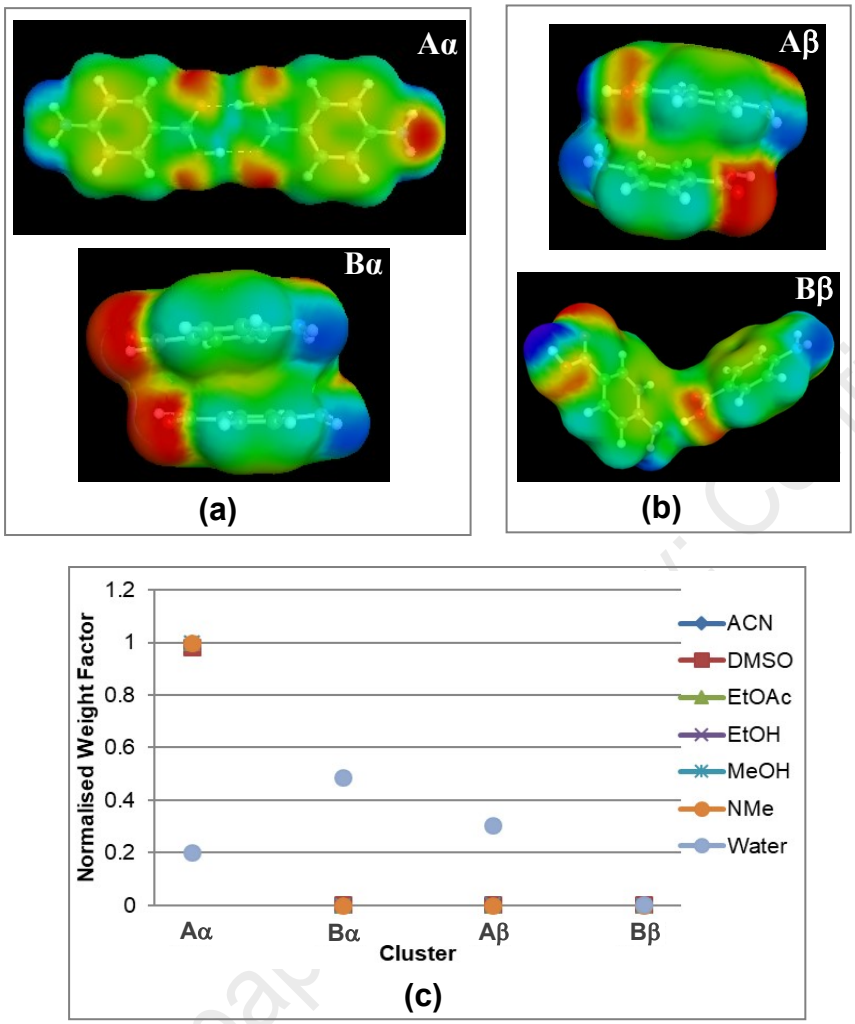

Figure S5. The surface charge density distributions of top two dimers building blocks derived from α-form (a) and β-form (b) PABA crystal structures, calculated using COSMO-RS based on DFT modelling within a solvent continuum environment, and (c) the calculated populations of dimers (Aα, Bα, Aβ, Bβ) in seven solvents (AcN, DMSO, EtOAc, EtOH, MeOH, NMe and water) (data derived from [15, 18]).

A more detailed analysis of the intermolecular chemistry associated with the solvation of PABA molecules has been provided using intermolecular grid-search methods [20-22] which reveals (**Figure S6**) the predicted solvation shell structures for 3 solvent systems,

EtOH, AcN and water. The calculated solvation energies of -29.86, -26.3 and -19.63 kcal/mol, respectively, were found to be consistent with the respective order associated with their known solubilities [22]. In this, the data confirms PABA to be well-solvated by the protic solvent EtOH which can interact strongly with both the aromatic ring through van der Waals interactions with the alkyl group and with the carboxylate through H-bonding. In contrast, the aprotic AcN solvates to a much lesser degree only forming only quite limited interactions with the COOH group compared to EtOH and water. This limited solvation would perhaps be consistent with PABA in the solution-phase preferably forming OH…O H-bonding dimers (synthon Aα) hence rationalising its easier crystallisability. Water, in contrast, fails to solvate the whole surface area of the PABA molecule apart from COOH group, particularly the amino group and the hydrophobic phenyl ring. This would also be consistent with its much lower solubility.

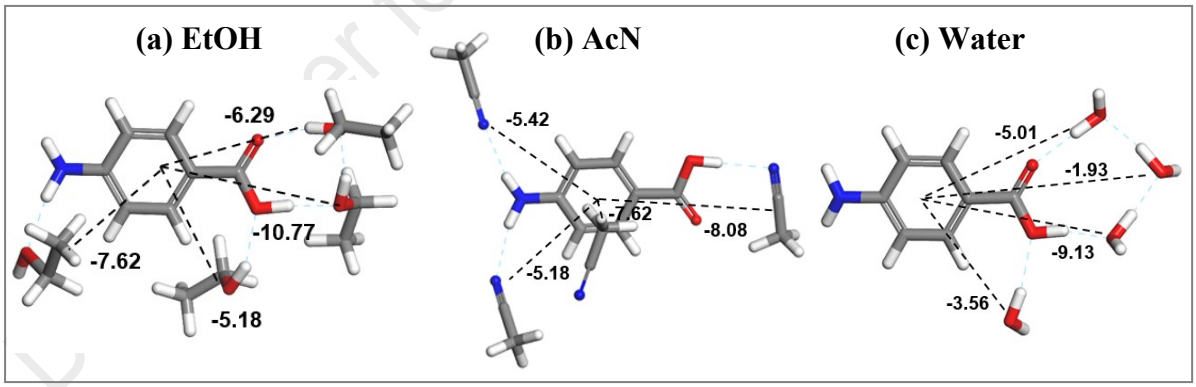

Figure S6. The solvation shell structures of PABA molecule with three solvents: (a) EtOH, (b) AcN and (c) water, calculated using molecule-molecule grid searching (data derived from [22]).

Solubility data [23], obtained from solution dissolution measurements (**Figure S7(a-c)**, upper plots) was analysed using van't Hoff plots for EtOH, AcN and water solvents

(**Figure S7(d)**). These reveal less than ideal solubility (activities -0.91, 0.43, 0.02, respectively) for all three solvents, consistent with a solution structure with an enhanced degree of solute/solute interactions and associated solute clustering.

S1.4 Metastability, Solute Clustering and Nucleation

Analysis of the slopes of the associated crystallisation on-set points (Tc) as a function of the solution cooling rate (q) probing the kinetic balance between the relative rates of supersaturation generation and nucleation, provides a helpful indication as to the relative crystallisabilities (**Figure S7(a-c)**, lower plots)) for the three solvents. The data shows distinctly different behavior between the 3 solvents with water having a much flatter slope compared to EtOH and AcN, respectively, suggesting the rate-limiting parameters regarding crystallisability would have an order of (kinetics driven) EtOH > AcN > water (thermodynamics driven). Also as shown in (**Figure S7**(d)), the crystallisation on-sets in EtOH as a function of cooling rate shows a marked difference in slope with respect to the ideal solubility line, moving away from the equilibrium solubility which is slightly less than ideal to greater than ideal with increasing cooling rate, undercooling and hence supersaturation. Hence, qualitatively taking the slope as an indicator of the enthalpy of dissolution (ΔH_{diss}), ΔH_{diss} decreases from ideal solubility (2500 J/mol) to 488 J/mol (1.0 ^{o}C/min) as the solutions undercool. This change in the apparent crystallisability with increased cooling rate can also be seen in the data when replotted in terms of critical supersaturation at Tc (**Figure S7(e)**). This reveals that the highest cooling rates access a much wider range of supersaturations and

nucleation cluster sizes across the solute concentration range (1.2 – 1.6) when compared to lowest cooling rates (1.05 – 1.10).

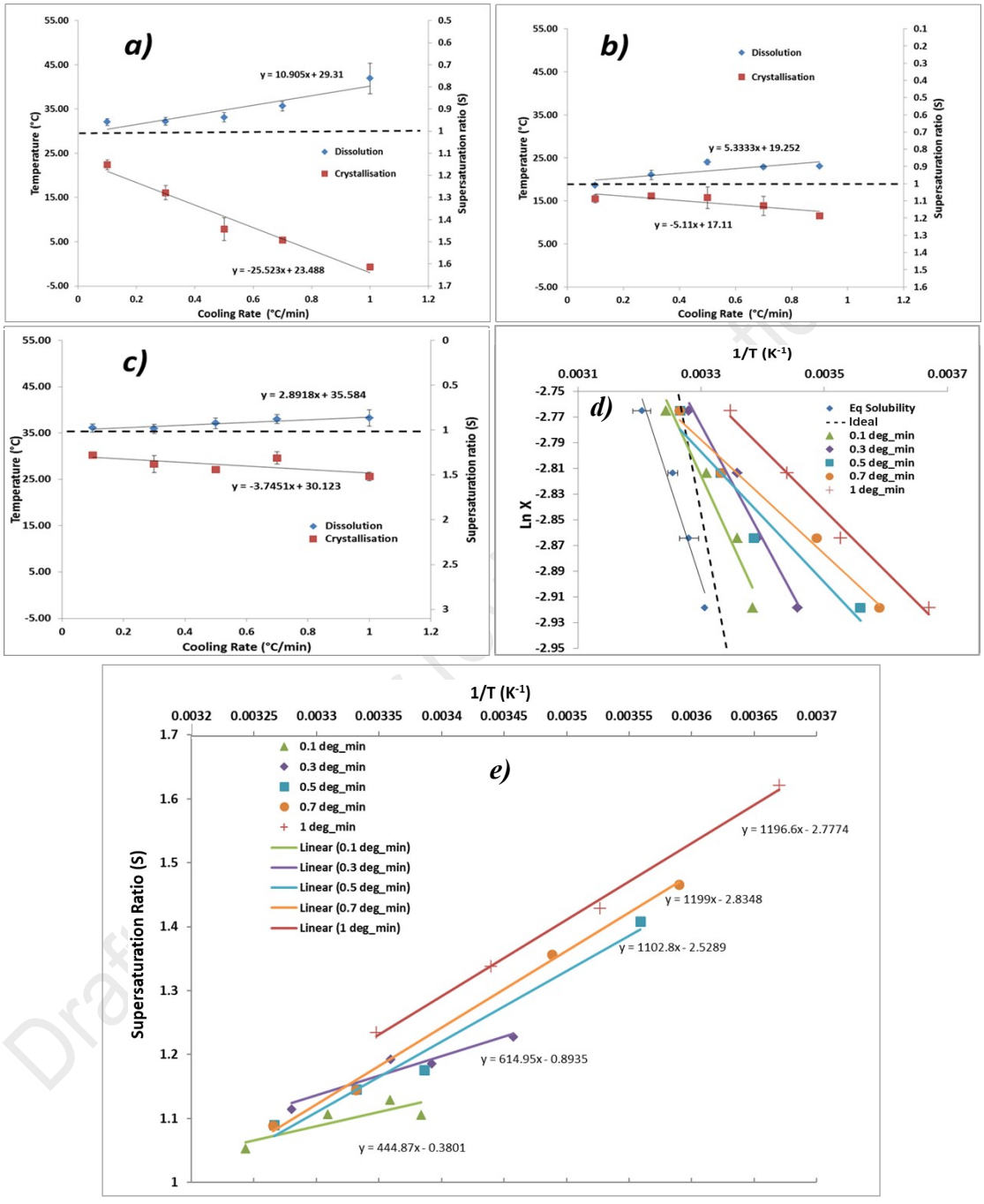

Figure S7. Dissolution and crystallisation temperatures of PABA in (a) EtOH, (b) AcN and (c) water under various cooling rate for the study of PABA crystallisability, and (d) crystallisation temperatures of PABA in EtOH with van't Hoff coordinates, showing a

marked differences in slope with respect to the ideal solubility line (---) and (e) supersaturation ratios in van't Hoff plotting (data derived from [23]).

Analysis of the polythermal crystallisation behavior as a function of solvent type and solution concentration indicated that the nucleation mechanism was mostly instantaneous and consistent with a two-step nucleation mechanism [23] except at low solute concentrations where it was progressive. Comparatively for the instantaneous cases, it could be said that EtOH was more instantaneous than AcN and water, this perhaps reflecting their relative solubilities [23].

The strong evidence for solute clustering from the less than ideal equilibrium solubilities and instantaneous nucleation mechanism has been further examined using small angle X-ray scattering (SAXS) [17, 24]. The data confirmed the formation of solvent clusters within PABA in ethanolic solutions during cooling, highlighting at low Q an increase in cluster numbers and their size (> 40 nm) during cooling. Such clustering behavior is consistent with a number of previous studies on e.g. glycine [25], citric acid [26], and L-isoleucine [27]. Simultaneous analysis at high Q reveals solution de-supersaturation during cooling with evidence for formation of small cluster structures (ca. 1 - 2 nm) well matched to the overall dimensions of the Aα synthon.

A tentative nucleation pathway for PABA crystallisation from EtOH solutions is shown schematically in **Figure S8** which draws together the structural information obtained from the analysis of the modelling and experimental data to date. This indicates the progress of nano-scale assemblies from large disordered liquid-like PABA nano-

clusters (dimers and monomers) in under-saturated condition, then the growth of these clusters in sizes under supersaturated conditions including the increasing degree of structural ordering at cluster/solution interface increasing the mass fractal dimension with their increasing interfacial smoothness and faceting at the interface and through to the formation of the crystalline phase (nuclei) by further growth. Overall, the data supports a two-step nucleation mechanism associated with the formation and development of nano-crystallite clusters together with the growth of smaller molecular clusters.

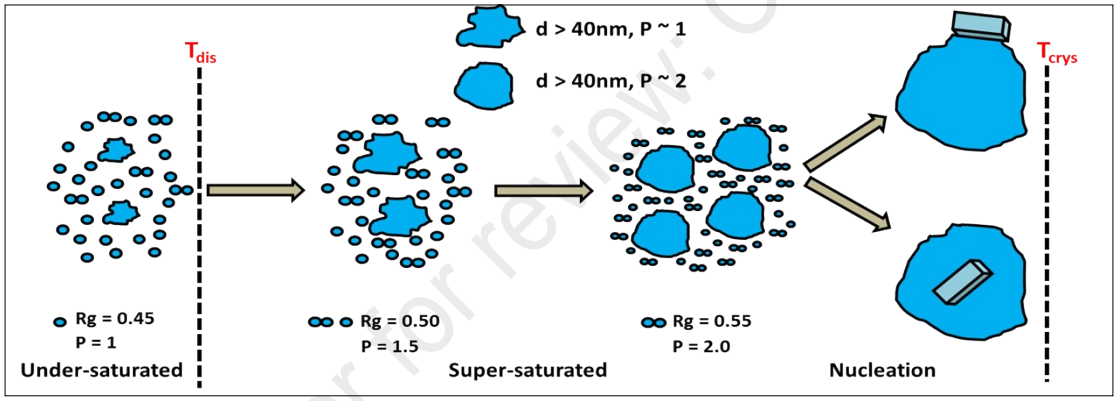

Figure S8. The tentative structural schematic of α-PABA nucleation pathway from EtOH solutions, based on the modelling and experimental evidence obtained so far, indicating large liquid-like clusters surrounded by a population of monomer and dimerised PABA molecules. Note the pathway for β-PABA nucleated in water can have different parameters. Rg is the radius of gyration which is related to cluster size (nm) and P is the fractal dimensionality and related to interface structure where a higher value infers a more ordered interface (data derived from [17]).

It is important to remark that the SAXS measurements require good contrast between the aggregated and continuum phases demanding, in turn, comparatively high solute concentrations. This aspect was linked SAXS studies to EtOH solutions (only) and in

this case only at the relatively high solute concentration typified by instantaneous nucleation. In effect, this restriction precludes SAXS studies over a range of concentrations in EtOH and on the AcN and aqueous solutions. Hence, at this stage we do not know how this model would apply more generally.

List of Symbols and Abbreviations

a, b, c:	Crystal unit cell parameters (Å)
AcN:	Acetonitrile
Aα, Aβ:	Synthon A of α-form, synthon A of β-form
D···A:	Length between donor and acceptor (Å)
D-H:	Bond length between donor and hydrogen (Å)
D-H···A:	Angle of donor – hydrogen – acceptor (Å)
DFT:	Density functional theory
DMSO:	Dimethyl sulfoxide
EtOAc:	Ethyl acetate
EtOH:	Ethanol
H···A:	Hydrogen bond length between acceptor and hydrogen (Å)
H2H:	Head-to-head
H2T:	Head-to-tail
H-bond:	Hydrogen bond
MeOH:	Methanol
NMe:	Nitromethane
PABA:	Para amino benzoic acid
PXRD:	Powder X-ray diffraction
q:	Solution cooling rate (m/s)
Q:	Scattering vector (nm^{-1})
SAXS:	Small angle X-ray scattering
Tc:	Crystallisation on-set point (critical temperature) (°C)
Z / Z':	Number of molecules in the asymmetric cell / unit cell
α, β, γ:	Crystal unit cell parameters (°)
ΔH_{diss}:	Enthalpy of dissolution (kcal/mol)

References

1. M. Alleaume, G. Salascim and J. Decap, *J. Compt. Rend., Ser. C*, 1966, **262**, 416-417.
2. T. F. Lai and R. E. Marsh, *Acta Crystallographica*, 1967, **22**, 885-893.
3. S. Gracin and A. Fischer, *Acta Crystallographica Section E: Crystallographic Communications*, 2005, **E61**, o1242-o1244.
4. S. Athimoolam and S. Natarajan, *Acta Crystallographica Section C: Structural Chemistry*, 2007,

C63, o514-o517.
5. R. A. Sullivan and R. J. Davey, *CrystEngComm*, 2015, **17**, 1015-1023.
6. S. Gracin and Å. C. Rasmuson, *Crystal Growth & Design*, 2004, **4**, 1013-1023.
7. H. X. Hao, M. Barrett, Y. Hu, W. Y. Su, S. Ferguson, B. Wood and B. Glennon, *Organic Process Research & Development*, 2012, **16**, 35-41.
8. M. Svard, F. L. Nordström, E.-M. Hoffmann, B. Aziz and A. C. Rasmuson, *CrystEngComm*, 2013, **15**, 5020-5031.
9. T. D. Turner, S. Caddick, R. B. Hammond, K. J. Roberts and X. Lai, *Crystal Growth & Design*, 2018, **18**, 1117−1125.
10. M. R. Ward, S. Shatha Younis, A. J. Cruz-Cabeza, C. L. Bull, N. P. Funnell and I. D. H. Oswald, *CrystEngComm*, 2019, **21**, 2058-2066.
11. R. Benali-Cherif, R. Takouachet, E.-E. Bendeif and N. Benali-Cherif, *Acta Crystallographica Section C: Structural Chemistry*, 2014, **C70**, 323-325.
12. A. J. Cruz-Cabeza, R. J. Davey, I. D. H. Oswald, M. R. Ward and I. J. Sugden, *CrystEngComm*, 2019, **21**, 2034-2042.
13. M. J. Bryant, S. N. Black, H. Blade, R. Docherty, A. G. P. Maloney and S. C. Taylor, *J. Pharm. Sci.*, 2019, **108**, 1655-1662.
14. C. Wang, I. Rosbottom, T. D. Turner, S. Laing, A. Maloney, A. Y. Sheikh, R. Docherty, Y. Q. and K. J. Roberts, *Pharmaceutical Research*, 2021, **38**, 971-990.
15. I. Rosbottom, PhD, PhD Thesis, University of Leeds, UK, 2015.
16. R. A. Sullivan, R. J. Davey, G. Sadiq, G. Dent, K. R. Back, J. H. Ter Horst, D. Toroz and R. B. Hammond, *Crystal Growth & Design*, 2014, **14**, 2689-2696.
17. T. D. Turner, PhD, PhD Thesis, University of Leeds, UK, 2015.
18. I. Rosbottom, D. Toroz, R. B. Hammond and K. J. Roberts, *CrystEngComm*, 2018, **20**, 7543-7555.
19. I. Rosbottom, K. J. Roberts and R. Docherty, *CrystEngComm*, 2015, **17**, 5768-5788.
20. R. B. Hammond, C. Y. Ma, K. J. Roberts, P. Y. Ghi and R. K. Harris, *The Journal of Physical Chemistry B*, 2003, **107**, 11820-11826.
21. R. B. Hammond, R. S. R S Hashim, C. Y. Ma and K. J. Roberts, *J. Pharm. Sci.*, 2006, **95**, 2361-2372.
22. I. Rosbottom, J. H. Pickering, R. B. Hammond and K. J. Roberts, *Organic Process Research & Development*, 2020, **24**, 500−507.
23. T. D. Turner, D. M. C. Corzo, D. Toroz, A. Curtis, M. M. Dos Santos, R. B. Hammond, X. Lai and K. J. Roberts, *Physical Chemistry Chemical Physics*, 2016, **18**, 27507-27520.
24. D. Toroz, I. Rosbottom, T. D. Turner, D. M. C. Corzo, R. B. Hammond, X. Lai and K. J. Roberts, *Faraday Discussions*, 2015, **179**, 79-114.
25. A. S. Myerson and P. Y. Lo, *Journal of Crystal Growth*, 1990, **99**, 1048-1052.
26. H. C. Groen and K. J. Roberts, *Journal of Physical Chemistry B*, 2001, **105**, 10723-10730.
27. N. Salwani Md Azmi, N. Anuar, N. Fitrah Abu Bakar, N. Fadhilah Kamalul Aripin and K. J. Roberts, *Journal of Crystal Growth*, 2019, **519**, 91-99.