Supplementary Information

Can formation of pharmaceutical cocrystals be computationally predicted?

II. Crystal structure prediction

Panagiotis G. Karamertzanis,[1,2†] Andrei V. Kazantsev,[1] Nizar Issa,[2] Gareth W.A. Welch,[2] Claire S. Adjiman,[1] Constantinos C. Pantelides[1] and Sarah L. Price[2]

[1] Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College, London, SW7 2AZ, United Kingdom

[2] Department of Chemistry, University College London, 20 Gordon Street, London, WC1H 0AJ, United Kingdom

Table of contents

† Author to whom all correspondence should be addressed: p.karamertzanis@imperial.ac.uk

S1 The intermolecular potential model

Section S1.1 provides details on the sensitivity of the computed lattice energies and crystal structures to the model for the intermolecular forces. Section S1.2 discusses the modifications of the Williams repulsion-dispersion parameterization that proved necessary for the modelling of 4-aminobenzoic acid - 2,2′-bipyridine cocrystal structures.

S1.1 Sensitivity of results to intermolecular potential model

The first part of this subsection investigates the effect of repulsion-dispersion parameterization and the quality of the wavefunction to derive the multipole moments on the predicted relative stability of the two 4-aminobenzoic acid polymorphs. The second part deals with the effect of wavefunction quality on the energy ranking of hypothetical 4-aminobenzoic acid and 4-nitrophenylacetic acid polymorphs.

S1.1.1 4-aminobenzoic acid polymorphs

We repeated the DMAflex-Quick lattice energy minimization of the two 4-aminobenzoic acid polymorphs using the Williams[1] and FIT[2] repulsion-dispersion parameterization and three levels of theory to compute the isolated-molecule charge density for the HF/6-31G(d,p) gas phase conformation. Ideally, these minimizations should have been done using DMAflex that updates the multipole moments as a function of conformation. However, this was not computationally feasible for large basis sets. Moreover, Table 1 in the manuscript shows that the predicted relative stability of the 4-aminobenzoic acid polymorphs with DMAflex and DMAflex-Quick differs by less than 3 kJ mol^{-1}, whilst the structural reproduction is also comparable. Hence, we expect that similar effects would have been found with the more accurate DMAflex minimization algorithm.

The DMAflex-Quick results are shown in Table S1. The wave function quality has a limited effect on the reproduction accuracy. However, the Williams repulsion-dispersion parameterization models the crystal structures more accurately and particularly the α polymorph, with the exception being the underestimation of hydrogen bond lengths that is not evident with FIT (see also section S3).

Table S1 Comparison of DMAflex-Quick (stage 2) reproductions of 4-aminobenzoic acid polymorphs using multipole moments derived from the MP2/6-31G(d,p) and MP2/aug-cc-pVTZ charge density of the HF/6-31G(d,p) gas phase conformation. The results for both FIT and Williams repulsion-dispersion parameterizations are reported. The highligthed rows refer to the model used for all results reported in the manuscript.

	Crystal Energy Partitioning[a] (kJ mol^{-1})		Lattice Details					
	E_{latt}	ΔE_{intra}	a (Å)	b (Å)	c (Å)	β (°)	Density (g cm^{-3})	RMSD$_{cs}$[b] (Å)
4-aminobenzoic acid polymorph α, Z´=2, $P2_1/n$ (CSD ref. AMBNAC01[3], R=7.03%, RT)								
Experimental		2.57	2.08	18.551	3.860	18.642	93.6	1.367	
FIT potential								
MP2/6-31G(d,p)	**-101.10**	2.31	2.30	19.369	3.691	18.426	98.6	1.399	0.465
MP2/aug-cc-pVTZ	**-107.40**	2.74	2.09	19.277	3.713	18.375	99.2	1.403	0.485
PBE0/aug-cc-pVTZ	**-109.35**	2.61	1.97	19.247	3.712	18.376	99.0	1.405	0.466
WILLIAMS potential								
MP2/6-31G(d,p)	**-107.21**	0.93	1.28	18.467	3.747	18.788	93.2	1.403	0.171
MP2/aug-cc-pVTZ	**-118.91**	0.71	0.88	18.368	3.778	18.590	93.5	1.415	failed due to short carboxyl dimer distances (O···O 2.3 Å)
PBE0/aug-cc-pVTZ	**-120.57**	0.65	0.75	18.332	3.783	18.600	93.3	1.415	
4-aminobenzoic acid polymorph β, Z´=1, $P2_1/n$ (CSD ref. AMBNAC04[4], R=4.99%, RT)								
Experimental		2.68		6.278	8.583	12.365	100.1	1.389	
FIT potential								
MP2/6-31G(d,p)	**-118.39**	1.14		6.144	8.214	12.700	99.5	1.441	0.229
MP2/aug-cc-pVTZ	**-121.21**	1.61		6.108	8.208	12.746	99.5	1.445	0.257
PBE0/aug-cc-pVTZ	**-117.47**	1.42		6.131	8.207	12.775	99.8	1.438	0.252
WILLIAMS potential								
MP2/6-31G(d,p)	**-109.78**	1.31		6.265	8.551	12.182	99.1	1.413	0.135
MP2/aug-cc-pVTZ	**-111.82**	1.77		6.233	8.510	12.274	99.3	1.418	0.137
PBE0/aug-cc-pVTZ	**-107.82**	1.63		6.254	8.523	12.309	99.6	1.408	0.120

[a] The experimental ΔE_{intra} refers to the HF/6-31G(d,p) molecular energy of the HF/6-31G(d,p) gas-phase optimized molecule with the flexible torsion angles identified in Scheme 1 constrained to their experimental values; it provides an estimate on how strained the molecular conformation is in the experimentally determined crystal. The reported values are only indicative because ΔE_{intra} depends on the X-ray uncertainty of the amino hydrogen atom positions.

[b] Hydrogen atoms were omitted for the coordination sphere (RMSD$_{cs}$) comparison,[5] that was performed using 15 and 20 molecules for α (Z′ = 1) and β polymorph (Z′ =2) respectively.

Although the lattice energies are very sensitive to the model potential, it is the stability difference E^{α}_{latt} - E^{β}_{latt} (in kJ mol^{-1}) of the two 4-aminobenzoic acid polymorphs that is the most important quantity for crystal structure prediction:

	FIT	WILLIAMS
MP2/6-31G(d,p):	+17.29	+2.57
MP2/aug-cc-pVTZ:	+13.81	-7.09
PBE0/aug-cc-pVTZ:	+8.12	-12.75

Hence, both repulsion-dispersion parameterization and charge density quality have a pronounced effect on E^{α}_{latt} - E^{β}_{latt} and exemplify the challenges in accurately predicting the relative stability of polymorphs. The reasonable stability difference of +2.57 kJ mol^{-1} with the WILLIAMS potential

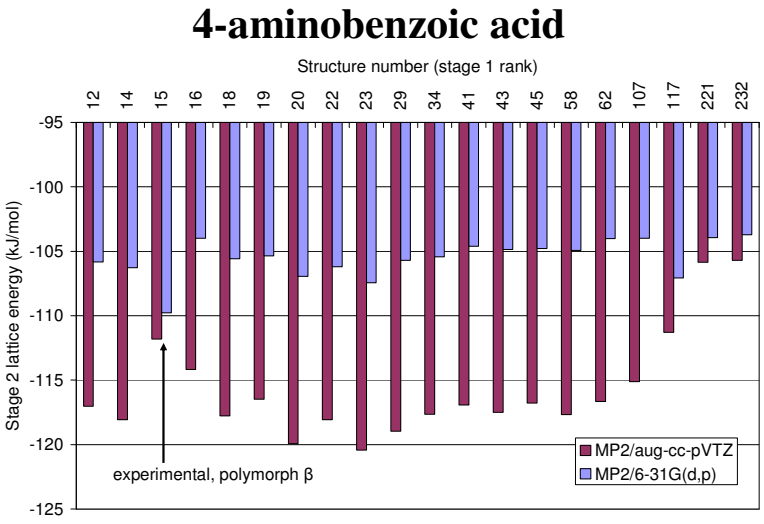
and MP2/6-31G(d,p) multipoles, i.e. the model used for all crystal structure prediction calculations in part II of this study, agrees well with experimental measurements (see manuscript), although this agreement may be due to fortuitous cancellation of errors. With the FIT parameterization and MP2/6-31G(d,p) multipoles E^{α}_{latt} - E^{β}_{latt} is unrealistically large and equal to +17.29 kJ mol^{-1} and is consistent with the DMAflex value of +20.1 kJ mol^{-1} reported in part I.

S1.1.2 Relative energies of hypothetical crystal structures

This section emphasizes further the sensitivity of the energy ranking of hypothetical 4-aminobenzoic acid and 4-nitrophenylacetic acid crystal structures on the charge density quality to compute the multipole moments. The 20 most stable DMAflex-Quick minima of each acid (computed with Williams, MP2/6-31G(d,p) charge density) were re-minimized using an aug-cc-pVTZ basis set with the rest of the computational model identical.

Figure S1 shows that increasing the basis set stabilizes all lattices energies but not to an equal extent. The large basis set deteriorates the rank of the β polymorph of 4-aminobenzoic acid. Figure 7 in the manuscript shows that basis set size mainly affects the electrostatic potential around the carbonyl acceptor of 4-aminobenzoic acid. This explains why the charge density quality alters the relative stability of carboxylic dimers and other hydrogen bond motifs, such as the COOH···NH2 hydrogen bonds found in the β polymorph.

The effect of basis set size is significantly less important for 4-nitrophenylacetic acid, probably because the majority of low energy structures contain similar hydrogen bond dimers.



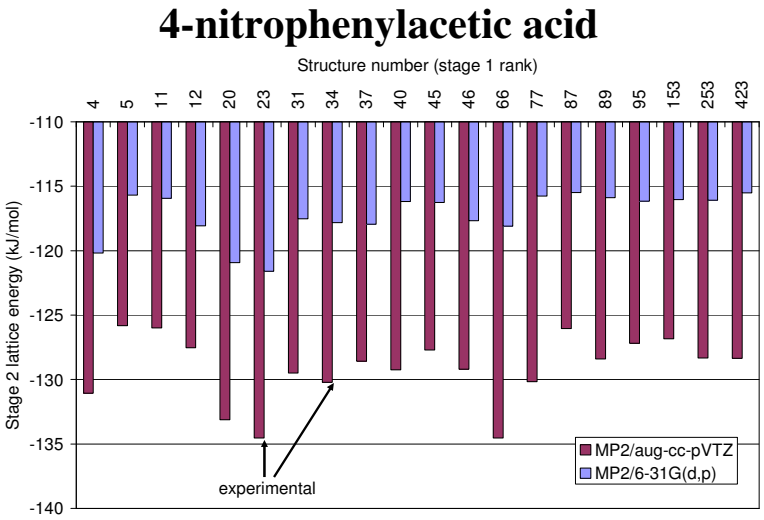


Figure S1 DMAflex-Quick lattice energies of the 20 most stable, unique (differ more that 0.4 Å in 15-molecule coordination sphere or 2 kJ mol^{-1} in lattice energy), stage 2 crystal structures of 4-aminobenzoic acid (top) and 4-nitrophenyalacetic acid (bottom) contrasting two different wavefunction qualities for the calculation of the isolated-molecule charge density (computed for the HF/6-31G(d,p) gas phase conformation).

The lower part of Figure S1 shows that there are two stage-2 4-nitrophenylacetic acid search structures that correspond to the experimentally determined crystal structure. These $P2_1/c$ structures have very similar 15-molecule coordination spheres but were not clustered as similar at stage 2 because their lattice energies differ by more than 2 kJ mol^{-1}. We note that the presence of these two closely related minima is an artifact because DMAflex-Quick uses the "rigid" degrees of freedom and reference multipole moments that have been computed for an arbitrarily chosen gas phase enantiomer. These "rigid" degrees of freedom and reference multipoles can be used to model either enantiomer of a racemic (e.g. structures belonging to $P2_1/c$) crystal structure producing two DMAflex-Quick minima that typically differ in lattice energy by a few kJ mol^{-1}. These DMAflex-Quick minima become identical when minimized with DMAflex. It would have been more appropriate to chose for the DMAflex-Quick minimization the enantiomer that closer matched the gas phase conformation. This was done for the DMAflex-Quick minimization of the experimental crystal structures in Table 2 but not during the search that relied on the DMAflex minimization to remove such fictitious duplicate structures that were erroneously not clustered as similar at stage 2. The error if the wrong enantiomer was used was small for the molecules in this study so few duplicates survived the clustering, and these were corrected at stage 3 by DMAflex, so the final results are unaffected.

S1.2 Modification of the Williams repulsion-dispersion potential

All lattice energy minimizations reported in the manuscript used the Williams repulsion-dispersion parameterization[1] because of the improved estimate of the relative stability of 4-aminobenzoic polymorphs that is grossly overestimated by FIT (see part I).

The Williams potential reproduces well all experimentally determined structures when used in conjunction with an intermolecular electrostatic model that is consistent with the model for the intermolecular forces used in its parameterization, i.e. atomic charges computed with a medium sized basis set such as 6-31G(d,p). The experimental structures are also generally well reproduced when the Williams potential is combined with atomic multipoles, although hydrogen bonded distances involving carboxylic protons are underestimated (see also section S3). The problem is particularly severe for the pyridine nitrogen - carboxylic proton hydrogen bond, which is only present in the 4-aminobenzoic acid - 2,2′-bipyridine cocrystal. Crystal structure prediction calculations for this system produced several hypothetical crystal structures that had N(pyr)···O(carbox) distances significantly shorter than 2.5 Å. These distances are much smaller than the typical N(pyr)···O(carbox) distances in the CSD as shown in Figure S2. For several of the hypothetical crystal structures the short distances caused numerical issues during lattice energy minimization.

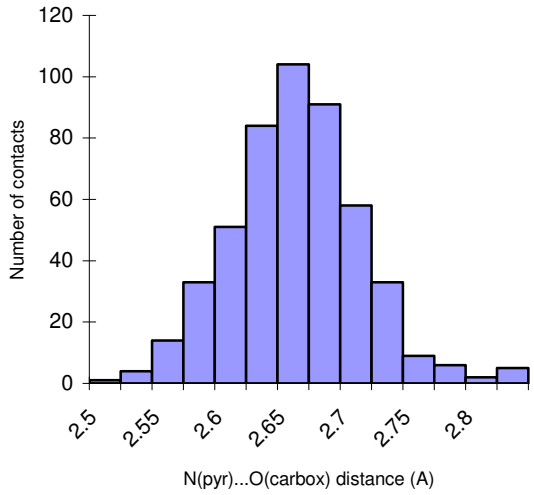


Figure S2 Distribution of N(pyr)···O(carbox) distances in the CSD (version 5.29, November 2007). The histogram was constructed by locating structures that contained this close contact and had 3D coordinates determined, R factor 7.5% or lower, were not polymeric, did not contain ions and had no disorder and no errors. This search produced 334 structures that contained 495 N(pyr)···O(carbox) close contacts.

The underestimation of the hydrogen bond distances can be attributed to the greater electrostatic attraction of the hydrogen bond donor and acceptor with the multipole model compared with the simplistic atomic charge model that was used for the Williams potential parameterization.

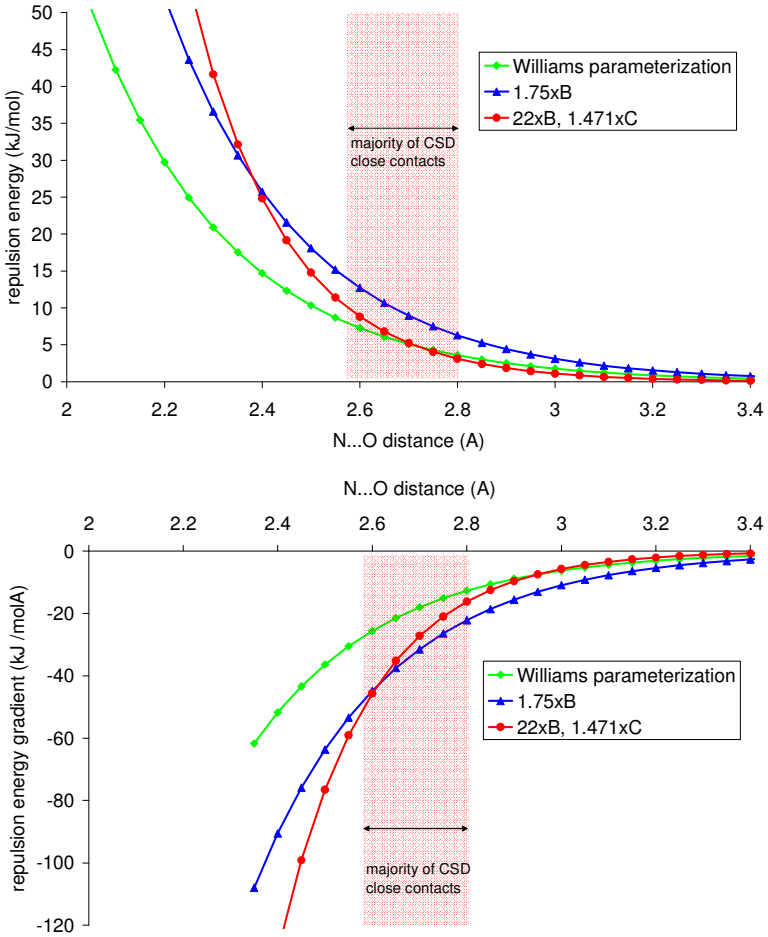


Figure S3 Modification of the Williams H3···N2 repulsion parameterization. The top graph shows the repulsion energy and the bottom graph shows the repulsion energy gradient with respect to the N···O distance assuming a perfectly linear O-H···N hydrogen bond and a 0.85 Å O-H bond length (foreshortened according to Williams parameterization). The green curves correspond to the original Williams parameterization and the red curves to the one used for all 4-aminobenzoic acid - 2,2′-bipyridine cocrystal lattice energy minimizations using multipoles. The modification that corresponds to the blue curves reproduces the N(pyr)···O(carbox) experimental distance satisfactorily but would significantly penalize the relative thermodynamic stability of the cocrystal compared with the component crystal structures.

We performed a limited modification of the Williams parameterization to avoid the numerical issues due to the underestimation of the N(pyr)···O(carbox) hydrogen bond distances in the experimental and theoretically generated 4-aminobenzoic acid - 2,2′-bipyridine cocrystals. This only involves a modification of the heteroatomic interaction term between the Williams atom types H3 and N2. Hence, for the stage 2 and stage 3 lattice minimizations this heteroatomic

interaction term is not computed with mixing rules but defined explicitly following the analysis given below. For all CrystalPredictor minimizations we used the unmodified Williams parameterization because the hydrogen bond lengths were more realistic using atomic charges.

The H3···N2 interaction term comprises an exponential term and no dispersion contribution, i.e. it has the functional form $U_{\text{H3-N2}} = Be^{-CR_{\text{H3-N2}}}$. Our objective was to modify the potential in such a way that the energy of crystal structures with N(pyr)···O(carbox) close contacts greater than 2.6 Å, which includes the majority of close contacts in the CSD, is not affected compared with the original Williams parameterization. At the same time we wanted to prevent unrealistically short N(pyr)···O(carbox) distances. This is mathematically possible if we consider that the energy of a crystal structure is determined by U but the arrangement of the atoms in the minimized crystal structure is determined by the gradient $\partial U / \partial R$. The combination 22xB and 1.471xC shown in Figure S3 is overall satisfactory and gives $R_{\text{H3-N2}}$ distances in the majority of minimized crystal structures that is within the range 2.500 Å < $R_{\text{H3-N2}}$ < 2.700 Å. This modified H3···N2 interaction term (Table S2) was adopted for all 4-aminobenzoic acid - 2,2′-bipyridine cocrystal lattice energy calculations using multipoles.

As a final confirmation, we report the lattice energy minimization results for the experimentally determined 4-aminobenzoic acid - 2,2′-bipyridine cocrystal with the original and modified H3···N2 interaction term. For these runs the molecular conformations were optimized at the HF/6-31G(d,p) level of theory with the torsion angles shown in Scheme 1 constrained to their values in the experimental crystal. These molecular conformations were held rigid for the lattice energy minimization. The multipole moments were computed from the MP2/6-31G(d,p) isolated molecule charge density for consistency with the crystal structure prediction calculations in the manuscript. The results are shown in Table S2 and demonstrate that the potential modification reproduces the cocrystal well without significantly affecting the predicted lattice energy compared with the original Williams parameterization, although the experimental N(pyr)···O(carbox) distance remains underestimated. This shows the limitations of the potential, although some of the discrepancy is due to the thermal expansion of the experimental crystal. We made no attempt to modify the H3···N2 interaction term to achieve better agreement with the N(pyr)···O(carbox) distance in the experimentally determined 4-aminobenzoic acid - 2,2′-bipyridine cocrystal, as this would require modification of other interaction terms. Section S3 shows that with the H3···N2 modification all hydrogen bonds involving carboxylic protons are underestimated to comparable extent: we hope that the underestimation of hydrogen bond lengths has a smaller effect on the modeling of the relative stabilities of the crystal structures.

Table S2 Rigid-body lattice energy minimization of the experimentally determined 4-aminobenzoic acid - 2,2′-bipyridine cocrystal with the original and modified Williams repulsion-dispersion parameterization. The electrostatic interactions are modeled using distributed multipoles up to hexadecapole derived from the MP2/6-31G(d,p) isolated-molecule charge density.

	H3···N2 repulsion interaction term ($U_{\mathrm{H3\text{-}N2}} = Be^{-CR_{\mathrm{H3\text{-}N2}}}$)		Energetic and geometric results		
	B (kJ mol^{-1})	C (Å^{-1})	Lattice energy (kJ mol^{-1})	N(pyr)···O(carbox) distance (Å)	RMSD$_{cs}$ (Å)
Experimental	-	-		2.825	
Williams parameterization	3441.52	3.52	-309.09	2.471	0.278
1.75×B	6022.67	3.52	-297.42	2.599	0.264
22×B, 1.471×C	75713.44	5.17792	-308.35	2.591	0.309

S2 Reproduction of flexible degrees of freedom of experimentally determined crystal structures

Table 2 in the manuscript gives the RMS$_{mol}$ overlap of the molecular conformation in the experimental and in the lattice energy minimized crystal structures. Table S3 gives the reproduction of all flexible degrees of freedom shown in Scheme 1, including the torsion angles that involve hydrogen atoms that were excluded in the RMS$_{mol}$ calculations.

Table S3 Reproduction of 4-aminobenzoic acid, 2,2′-bipyridine and 4-nitrophenylacetic acid flexible torsion angles (identified in Scheme 1) during the minimization of the experimentally determined crystal structures with the models used in the three minimization stages. Deviations that are larger than 10 ° are shown in bold.

	ξ_1		ξ_2		ξ_3		ξ_4	
4-aminobenzoic acid								
HF/6-31G(d,p) gas phase	-0.1 °		+0.2 °		+22.4 °		+137.5 °	
4-aminobenzoic acid polymorph α, Z´=2, $P2_1/n$ (CSD ref. AMBNAC01[3])								
Experimental	-11.8 °	-8.1 °	-3.4 °	+1.9 °	+24.3 °	+27.8 °	+144.6 °	+142.9 °
Stage 1	-3.2 °	-5.3 °	-6.5 °	-5.4 °	+31.0 °	+32.2 °	**+128.3** °	**+129.3** °
Stage 2	-5.3 °	-6.7 °	-7.4 °	-6.0 °	+27.8 °	+30.6 °	**+133.6** °	+135.2 °
Stage 3	-2.6 °	-3.5 °	-6.3 °	-3.8 °	+29.1 °	+30.4 °	**+132.0** °	+135.5 °
4-aminobenzoic acid polymorph β, Z´=1, $P2_1/n$ (CSD ref. AMBNAC04[4])								
Experimental	-3.8 °		+9.3 °		+26.8 °		+124.6 °	
Stage 1	-2.5 °		+7.6 °		+31.6 °		+126.3 °	
Stage 2	-4.7 °		+9.5 °		+22.4 °		**+139.3** °	
Stage 3	-6.2 °		+10.3 °		+28.7 °		+126.3 °	
4-aminobenzoic acid & 2,2′-bipyridine, Z´=3/2, $P2_1/c$ (CSD ref. DAQYUQ[6])[a]								
Experimental	+4.8 °		-0.7 °		+20.8 °		+141.6 °	
Stage 1[a]	+2.6 °		+0.7 °		+25.2 °		**+129.2** °	
Stage 2[a]	+8.5 °		**-12.9** °		+22.5 °		+140.8 °	
Stage 3[a]	+6.4 °		-6.1 °		+25.2 °		+133.6 °	
4-aminobenzoic acid & 4-nitrophenylacetic acid, Z´=2, $P2_1/a$ (CSD ref. RILJOL[7])								
Experimental	+7.7 °		+4.5 °		+13.9 °		+164.5 °	
Stage 1	+3.3 °		+3.7 °		**+24.9** °		**+138.5** °	
Stage 2	+2.2 °		+7.6 °		**+28.3** °		**+136.4** °	
Stage 3	+1.4 °		+3.7 °		+11.7 °		+167.9 °	
2,2′-bipyridine								
HF/6-31G(d,p) gas phase			-		-		-	
2,2′-bipyridine, Z´=1/2, $P2_1/n$ (CSD ref. BIPYRL04[8])[a]								
Experimental	+180.0 °		-		-		-	
Stage 1[a]	+180.0 °		-		-		-	
Stage 2[a]	+180.0 °		-		-		-	
Stage 3[a]	+180.0 °		-		-		-	
4-aminobenzoic acid & 2,2′-bipyridine, Z´=3/2, $P2_1/c$ (CSD ref. DAQYUQ[6])[a]								
Experimental	+180.0 °		-		-		-	
Stage 1[a]	+180.0 °		-		-		-	
Stage 2[a]	+180.0 °		-		-		-	
Stage 3[a]	+180.0 °		-		-		-	
4-nitrophenylacetic acid								
HF/6-31G(d,p) gas phase	+0.39 °		+65.70 °		+166.6 °		-0.3 °	
4-nitrophenylacetic acid, Z´=1, *P*bca (CSD ref. SEMTAF01; unpublished results)								
Experimental	-10.4 °		+72.3 °		+169.2 °		-3.1 °	
Stage 1	-7.5 °		+73.3 °		+166.6 °		-4.1 °	
Stage 2	-10.8 °		+71.4 °		+168.3 °		-4.0 °	
Stage 3	-14.1 °		+71.4 °		+165.7 °		-2.3 °	
4-aminobenzoic acid & 4-nitrophenylacetic acid, Z´=2, $P2_1/a$ (CSD ref. RILJOL[7])								
Experimental	-4.3 °		+43.6 °		-154.8 °		-2.4 °	
Stage 1	-0.3 °		+41.8 °		-158.2 °		+2.1 °	
Stage 2	-2.6 °		+40.0 °		-149.1 °		+0.2 °	
Stage 3	-1.8 °		+40.9 °		-154.1 °		-0.1 °	

[a] We minimized the crystal structures by reducing symmetry to $P2_1$ to have complete molecules in the asymmetric unit. We confirmed, using PLATON, that the symmetry of the minimized structures was $P2_1/c$ and hence symmetry was not lost during lattice energy minimization. The values reported in the table correspond to the flexible degrees of freedom in these PLATON-processed, minimized crystal structures.

With the most accurate, stage 3 model all flexible torsion angles are reproduced with errors that do not exceed 10 °. The only exception is the rotation of the amino hydrogen atoms for one of the

4-aminobenzoic acid molecules in the asymmetric unit of the α polymorph, which could well reflect experimental errors in the X-ray determination of the hydrogen atom positions.

All three minimization stages are based on the artificial partitioning of the intramolecular degrees of freedom into rigid and flexible. In the course of this study, it became evident that freezing the rigid degrees of freedom, ξ^r, to their values in the gas phase conformation at stages 1 and 2 causes up to 2-3 kJ mol^{-1} error in lattice energies. Some of this error is corrected in stage 3 by the partial relaxation of ξ^r within the isolated molecule *ab initio* optimization. However, the lattice energy minimization should ideally optimize a larger set of flexible intramolecular degrees of freedom in response to the crystal packing forces.[9] Moreover, all the molecules considered here are sufficiently small that calculating correlated intramolecular energies at the MP2(fc)/6-31G(d,p) level does not change the predicted relative stability of the generated structures by more than 2-3 kJ mol^{-1} relative to the HF/6-31G(d,p) values used in this study. Nevertheless, electron correlation will make a more significant contribution to ΔE_{intra} for larger molecules[10] which have functional groups that can be found in close proximity or much further apart in different molecular conformations.

S3 Reproduction of hydrogen bond geometries of experimentally determined crystal structures

Table S4 gives the reproduction of all hydrogen bond distances and angles with the stage 3 model. It is evident that all hydrogen bonds formed with the carboxylic proton acting as donor are underestimated.

Table S4 Reproduction of hydrogen bond geometries of the experimental crystal structures using the stage 3 model. Hydrogen bonds formed with a carboxylic proton acting as donor are shown in bold.

hydrogen bond	hydrogen bond length (X(-H)···Y)			hydrogen bond angle (X-H···Y)	
	experimental	minimized	difference	experimental	minimized
4-aminobenzoic acid polymorph α, Z´=2, $P2_1/n$ (CSD ref. AMBNAC01[3])					
O=COH···OCOH	**2.642**	**2.428**	**-0.214**	**167.1**	**174.3**
O=COH···OCOH	**2.610**	**2.399**	**-0.211**	**166.1**	**173.0**
NH···O=COH	3.351	3.273	-0.078	141.7	143.2
NH···O=COH	2.984	3.011	+0.027	173.7	167.4
4-aminobenzoic acid polymorph β, Z´=1, $P2_1/n$ (CSD ref. AMBNAC04[4])					
O=COH···NH2	**2.754**	**2.542**	**-0.212**	**160.0**	**173.5**
OHC=O···HN	3.045	3.069	+0.024	164.0	167.4
4-nitrophenylacetic acid, Z´=1, *P*bca (CSD ref. SEMTAF01; unpublished results)					
O=COH···OCOH	**2.603**	**2.409**	**-0.194**	**173.0**	**175.5**
4-aminobenzoic acid & 2,2′-bipyridine, Z´=3/2, $P2_1/c$ (CSD ref. DAQYUQ[6])[a]					
O=COH···N(pyr)	**2.825**	**2.572**	**-0.253**	**166.6**	**160.2**
NH···O=COH	3.057	3.083	+0.026	165.2	170.7
NH···NH	3.066	3.160	+0.094	165.8	160.1
4-aminobenzoic acid & 4-nitrophenylacetic acid, Z´=2, $P2_1/a$ (CSD ref. RILJOL[7])					
O=COH···OCOH	**2.593**	**2.415**	**-0.178**	**162.2**	**171.6**
O=COH···OCOH	**2.739**	**2.532**	**-0.207**	**160.5**	**177.4**

NH···O=N	3.133	3.078	-0.055	141.1	136.3
NH···O=N	3.161	3.099	-0.062	157.9	157.5
[a] The carboxylic proton - pyridine nitrogen interaction term was modified as described in section S1.3.					

S4 Extended list of hypothetical crystal structures

This section contains tables with the ten most stable crystal structures found in the crystal structure prediction searches. All crystal structures have been minimized with DMAflex as described in the manuscript and are stored on CCLRC e-Science Centre data portal. They are available from the authors on request (email: Prof. S.L. Price, s.l.price@ucl.ac.uk).

Table S5 The 10 most stable, fully DMAflex optimized crystal structures of 4-aminobenzoic acid. We fully minimized with DMAflex the 25 most stable, distinct single-point DMAflex structures. This resulted in 24 unique minima. For all structure comparisons we used a tolerance of 0.4 Å in comparing the 15-molecule coordination spheres.[5]

		molecular conformation[b]					cell geometry	
rank, space group	E_{latt} [a] (kJ mol^{-1})	ΔE_{intra} (kJ mol^{-1})	ξ_1 (°)	ξ_2 (°)	ξ_3 (°)	ξ_4 (°)	density (g cm^{-3})	conventional cell[c] a(Å), b(Å), c(Å), α(°),β(°),γ(°) and dominant hydrogen bond motif
1[d], $P2_1/c$	-112.73	2.82	-6.0	10.0	30.3	126.0	1.411	6.336,8.445,12.246,-,100.0 °,- OCOH···NH2···OCOH four member rings
2, $P2_1/c$	-106.60	0.39	2.4	2.1	18.4	135.1	1.418	3.838,15.271,10.966,-,92.3 °,- carboxylic dimers
3, $P2_1$	-106.44	6.12	-18.3	0.8	34.9	127.3	1.397	4.009,14.489,5.661,-,97.5 °,- OCOH···NH2···OCOH chain
4, $P2_1/c$	-106.35	0.50	0.9	1.0	19.4	132.2	1.401	3.707,11.100,15.979,-,98.5 °,- carboxylic dimers
5, $P2_1/c$	-106.23	0.74	-4.1	-4.5	29.3	136.8	1.375	3.849,12.977,13.514,-,101.2,- carboxylic dimers
6, $P2_1/c$	-106.00	0.44	-3.3	-2.7	27.5	132.1	1.395	3.751,13.014,13.393,-,93.3 °,- carboxylic dimers
7, $P2_1/c$	-105.89	1.79	-3.9	-13.9	26.4	136.3	1.406	6.570,13.881,7.106,-,90.1 °,- carboxylic dimers
8, $P2_1/c$	-105.59	0.63	-1.0	4.2	19.6	132.1	1.391	3.663,15.847,11.322,-,94.9 °,- carboxylic dimers
9, $P2_1/c$	-105.42	1.43	2.1	3.9	13.6	132.2	1.382	3.665,13.896,13.297,-,102.6 °,- carboxylic dimers
10, *P*bcn	-105.05	0.90	-0.9	4.5	32.2	130.3	1.378	13.339,15.074,6.573,-,-,- carboxylic dimers, NH2···NH2

[a] Lattice energy (in kJ mol^{-1}) expressed as the sum of the inter- and intra-molecular energies.
[b] Torsion angles defined in Scheme 1.
[c] Only cell angles not constrained by symmetry given.
[d] Corresponds to the experimentally determined low-temperature polymorph β.

Table S6 The 10 most stable, fully DMAflex optimized crystal structures of 2,2′-bipyridine. We fully minimized with DMAflex the 25 most stable, distinct single-point DMAflex structures. This resulted in 25 unique minima. For all structure comparisons we used a tolerance of 0.4 Å in comparing the 15-molecule coordination spheres.[5] These 25 unique DMAflex-minimized structures show a varying degree of π-π stacking of the planar molecules. The molecular planes approach each other at an angle and in some structures this angle becomes close to 90° corresponding to a T-shaped configuration.

		molecular conformation[b]		cell geometry	
rank, space group	E_{latt} [a] (kJ mol^{-1})	ΔE_{intra} (kJ mol^{-1})	ξ_1 (°)	density (g cm^{-3})	conventional cell[c] a(Å), b(Å), c(Å), α(°),β(°),γ(°)
1, $P2_1/c$	-81.46	0.05	-177.1	1.301	6.017,11.249,12.951,-,114.6 °,-
2[d], $P2_1/c$, Z′=½	-81.41	0.00	-180.2	1.293	5.680,6.180,12.289,-,111.5 °,-
3, $Pna2_1$	-79.57	0.00	-180.2	1.274	13.701,5.061,11.744 °,-,-,-
4, $P2_1/c$, Z′=½	-79.47	0.00	180.0	1.293	3.787,9.735,10.902,-,93.4 °,-
5, $P2_12_12_1$	-79.44	0.08	176.0	1.272	7.419,10.155,10.821,-,-,-
6, $P2_1/c$	-79.39	0.08	176.0	1.278	6.796,11.304,11.585,-,114.2 °,-
7, $Pna2_1$	-79.06	0.02	178.1	1.300	18.041,3.771,11.730,-,-,-
8, $P2_1/c$, Z′=½	-79.03	0.00	180.0	1.308	3.735, 9.162, 11.897,-,103.07 °,-
9, $P2_1/c$	-79.01	0.01	179.1	1.274	5.198, 13.318,12.429,-,108.9 °,-
10, $P2_1/c$, Z′=½	-78.96	0.00	-180.0	1.253	5.994,11.724,7.149,-,124.5 °,-

[a] Lattice energy (in kJ mol^{-1}) expressed as the sum of the inter- and intra-molecular energies.
[b] Torsion angles defined in Scheme 1.
[c] Only cell angles not constrained by symmetry given.
[d] Corresponds to the experimentally determined polymorph.

Table S7 The 10 most stable, fully DMAflex optimized crystal structures of 4-nitrophenylacetic acid. We fully minimized with DMAflex the 25 most stable, distinct single-point DMAflex structures. This resulted in 22 unique minima. For all structure comparisons we used a tolerance of 0.4 Å in comparing the 15-molecule coordination spheres.[5] All 22 distinct DMAflex-minimized structures contain carboxylic dimers.

			molecular conformation[b]					cell geometry
rank, space group	E_{latt} [a] (kJ mol^{-1})	ΔE_{intra} (kJ mol^{-1})	ξ_1 (°)	ξ_2 (°)	ξ_3 (°)	ξ_4 (°)	density (g cm^{-3})	conventional cell[c] a(Å), b(Å), c(Å), α(°),β(°),γ(°)
1[d], *P*bca	-117.30	1.02	-13.4	71.4	165.9	-2.6	1.450	15.178,6.822,16.027,-,-,-
2, $P2_1/c$	-116.90	0.82	10.7	109.8	-172.3	3.2	1.445	8.021,15.254,6.805,-,90.4 °,-
3, $P2_1/c$	-115.65	1.22	3.0	88.9	160.1	-2.3	1.494	6.511,4.793,26.093,-,98.4 °,-
4, $P2_1/c$	-114.97	2.30	-19.6	96.8	-175.4	-1.4	1.535	4.755,15.258,10.923,-,98.5 °,-
5, $P2_1/c$	-114.94	1.72	13.3	54.4	152.2	0.0	1.538	6.543,17.517,7.444,-,113.5 °,-
6, *C*2/c	-114.40	1.71	-12.3	78.5	-170.9	-0.6	1.455	23.057,4.972,16.258,-,117.5 °,-
7, *P*bca	-114.37	1.09	10.8	121.9	-163.6	5.4	1.479	11.996,8.505,15.947,-,-,-
8, $P2_1/c$	-114.35	0.47	10.1	69.5	165.5	-0.5	1.479	8.181,7.019,14.257,-,96.6 °,-
9, $P2_1/c$	-114.14	0.04	-1.0	67.5	169.0	0.6	1.478	9.932,6.758,15.349,-,127.8 °,-
10, $P2_1/c$	-114.08	0.79	-3.2	102.2	-93.3	-4.6	1.460	7.127,10.738,11.645,-,112.4 °,-

[a] Crystal energy (in kJ mol^{-1}) expressed as the sum of the inter- and intra-molecular energies.
[b] Torsion angles defined in Scheme 1.
[c] Only cell angles not constrained by symmetry given.
[d] Corresponds to the experimentally determined polymorph.

Table S8 The 10 most stable predicted 4-aminobenzoic acid - 2,2′-bipyridine 1:1 cocrystal structures. We fully minimized with DMAflex the 25 most stable, distinct single-point DMAflex structures. This resulted in 25 unique minima. For all structure comparisons we used a tolerance of 0.4 Å in comparing the 20-molecule coordination spheres.[5]

		molecular conformation[b]						cell geometry
rank, space group	E_{latt} [a] (kJ mol^{-1})	ΔE_{intra} (kJ mol^{-1})	ξ_1 (°)	ξ_2 (°)	ξ_3 (°)	ξ_4 (°)	density (g cm^{-3})	conventional cell[c] a(Å), b(Å), c(Å), α(°),β(°),γ(°) and dominant hydrogen bond motif
1, $P2_1/c$	-198.68	1.95 2.55	0.4 158.5	-4.0	39.2	131.3	1.343	3.830, 13.066, 29.100, -, 95.2, - OCOH···N (pyr), NH2···OCOH
2, $P2_1$	-196.04	2.59 2.00	7.3 -160.9	-3.2	37.8	133.0	1.353	3.955, 12.623, 14.512, -, 96.5, - OCOH···N (pyr), NH2···OCOH
3, $P2_1/c$	-195.20	1.25 6.66	-5.1 145.9	-7.3	31.7	134.5	1.309	7.041, 30.514, 7.341, -, 109.6, - OCOH···N (pyr), NH2···OCOH
4, $P2_1$	-195.01	1.84 2.01	-6.6 173.7	-4.1	34.1	127.7	1.343	3.856, 13.283, 14.217, -, 95.2, - OCOH···N (pyr), NH2···OCOH
5, $P2_12_12_1$	-194.93	1.62 0.76	-0.5 -168.1	-3.2	37.5	133.1	1.325	3.999, 13.275, 27.706, -, -, - OCOH···N (pyr), NH2···OCOH
6, $P2_1/c$	-194.57	3.88 2.98	-11.3 -156.8	5.3	13.1	158.8	1.296	7.922 16.649 12.881 90.00 117.8 OCOH···N (pyr), NH2···N (pyr)
7, $Pna2_1$	-194.12	2.34 0.22	3.2 -173.6	-7.0	38.7	130.1	1.358	28.982, 3.849, 12.863, -, -, - OCOH···N (pyr), NH2···OCOH
8, $Pca2_1$	-194.08	1.13 3.17	6.2 156.1	-1.7	19.7	132.2	1.351	21.309, 3.836, 17.634, -, -, - OCOH···N (pyr), NH2···OCOH
9, $Pca2_1$	-193.6	2.32 0.41	3.0 171.2	-5.4	39.4	131.8	1.327	29.733, 3.858, 12.801, -, -, - OCOH···N (pyr), NH2···OCOH
10, $P2_1/c$	-193.59	1.38 3.27	0.5 155.7	-1.7	36.5	128.8	1.321	3.716, 30.152, 13.647, 105.3 OCOH···N (pyr), NH2···OCOH

[a] Lattice energy (in kJ mol^{-1}) expressed as the sum of the inter- and intra-molecular energies.
[b] Torsion angles defined in Scheme 1; the first row refers to 4-aminobenzoic acid and the second to 2,2′-bipyridine.
[c] Only cell angles not constrained by symmetry given.

Table S9 The 10 most stable predicted 4-aminobenzoic acid - 2,2′-bipyridine 2:1 cocrystal structures. We fully minimized with DMAflex the 25 most stable, distinct single-point DMAflex structures. This resulted in 25 unique minima. For all structure comparisons we used a tolerance of 0.4 Å in comparing the 20-molecule coordination spheres.[5]

		molecular conformation[b]						cell geometry
rank, space group	$U + \Delta E$ [a] (kJ mol^{-1})	ΔE (kJ mol^{-1})	ξ_1 (°)	ξ_2 (°)	ξ_3 (°)	ξ_4 (°)	density (g cm^{-3})	conventional cell[c] a(Å), b(Å), c(Å), α(°),β(°),γ(°) and dominant hydrogen bond motif
1[d], $P2_1/c$, Z′=3/2	-309.55	1.32 0.00	6.4 0.0	-5.9	25.2	133.4	1.365	16.173, 4.282, 15.732, -, 106.0, - OCOH···N (pyr), NH2···OCOH,·NH2··NH2
2, $P2_1/c$, Z′=3/2	-308.98	2.12 0.00	1.6 0.0	-5.9 0.0	38.9 0.0	131.7 0.0	1.319	3.904, 13.082, 21.311, -, 95.2, - OCOH···N (pyr), NH2···OCOH
3, $P2_1/c$, Z′=3/2	-306.44	4.17 0.00	15.4 0.0	-4.8 0.0	27.2 0.0	134.5 0.0	1.279	8.430, 4.689, 28.583, -, 98.4, - OCOH···N (pyr), NH2···OCOH,·NH2··NH2
4, $P2_1/c$, Z′=3/2	-306.02	1.25 0.00	-3.2 0.0	-2.9 0.0	34.9 0.0	129.6 0.0	1.317	3.778, 21.677, 13.509, -, 101.1, - OCOH···N (pyr), NH2···OCOH
5, $P2_1/c$, Z′=3/2	-305.98	2.28 0.00	10.7 0.0	-4.5	23.5	132.9	1.321	16.807, 4.206, 16.288, -, 109.9, - OCOH···N (pyr), NH2···OCOH,·NH2··NH2
6, $P2_1/c$, Z′=3/2	-304.59	2.68 0.00	11.2 0.00	-6.8	25.3	134.9	1.340	4.432, 30.331, 8.252, -, 105.9, - OCOH···N (pyr), NH2···OCOH
7, $P2_1$	-304.16	1.50 0.23 0.19	3.4 1.0 -174.1	-9.8 -1.1	19.1 24.7	133.8 -132.2	1.327	15.954, 4.131, 17.203, -, 108.1,- OCOH···N (pyr), NH2···OCOH,·NH2··NH2
8, $P2_1$	-300.02	2.26 1.33 2.35	6.6 8.4 -159.3	-11.4 -1.1	26.6 -18.0	137.1 -135.8	1.320	14.790, 4.834, 15.142, -, 90.6, - OCOH···N (pyr), NH2···OCOH,·NH2··NH2
9, $P2_1$	-299.76	2.26 1.33 2.35	24.0 8.0 138.6	7.4 -4.5	20.9 -26.3	132.4 -127.5	1.326	15.873, 3.768, 18.205, -, 98.1, - OCOH···N (pyr), NH2···OCOH,·NH2··NH2
10, $P2_1/c$, Z′=3/2	-299.59	1.01 0.00	-0.2 0.0	-4.9	30.9	128.5	1.307	3.833, 23.877, 11.959, -, 91.8, - OCOH···N (pyr), NH2···OCOH

[a] Lattice energy (in kJ mol^{-1}) expressed as the sum of the inter- and intra-molecular energies. Lattice energy refers to 2 mol of 4-aminobenzoic acid and 1 mol 0f 2,2′-bipyridine.

[b] Torsion angles defined in Scheme 1; for all crystal structures in which 2,2′-bipyridine lies on the inversion centre (Z′=3/2), only the torsion angles of one aminobenzoic acid molecule are given as symmetry independent; for the other cases the first and second rows refer to 4-aminobenzoic acid and the third to 2,2′-bipyridine.

[c] Only cell angles not constrained by symmetry given.

[d] Corresponds to the experimentally determined polymorph.

Table S10 The 10 most stable, fully DMAflex optimized 4-aminobenzoic acid - 4-nitrophenylacetic acid cocrystal structures. We fully minimized with DMAflex the 25 most stable, distinct single-point DMAflex structures. This resulted in 25 unique minima. For all structure comparisons we used a tolerance of 0.4 Å in comparing the 20-molecule coordination spheres.[5] The DMAflex minimized experimental crystal was found 10.4 kJ mol^{-1} above the global minimum and is given in Table 2 of the manuscript.

		molecular conformation[b]						cell geometry
rank, space group	E_{latt} [a] (kJ mol^{-1})	ΔE_{intra} (kJ mol^{-1})	ξ_1 (°)	ξ_2 (°)	ξ_3 (°)	ξ_4 (°)	density (g cm^{-3})	conventional cell[c] a(Å), b(Å), c(Å), α(°),β(°),γ(°) and dominant hydrogen bond motif
1, $P2_1/c$	-228.05	1.98 2.72	-4.5 6.1	-2.6 67.4	40.9 87.1	125.6 -6.3	1.456	7.318, 15.287, 13.954, -, 111.6, - 4-aminobenzoic acid carboxylic homo-dimers, OCOH(4-nit)···NH2, NH2···O2N
2, P-1	-226.04	3.99 1.43	-3.3 -2.7	1.3 100.6	28.9 54.6	125.9 3.6	1.437	5.102, 12.082, 13.119, -, 105.8, 101.2, 101.4, - 4-aminobenzoic acid carboxylic homo-dimers, OCOH(4-nit)···NH2, NH2···O2N
3, $P2_1/c$	-225.62	4.42 2.50	-3.8 12.6	5.0 82.7	30.2 -72.5	123.2 -1.1	1.448	13.166, 7.471, 17.660, -, 122.8, - 4-aminobenzoic acid carboxylic homo-dimers, OCOH(4-nit)···NH2, HOCO(4-amn,4-nit)···H2N
4, $P2_1/c$	-225.10	4.93 1.89	-0.9 1.1	-2.4 75.7	30.6 -38.5	123.8 2.1	1.455	4.496, 25.887, 12.482, -, 90.3, - 4-aminobenzoic carboxylic homo-dimers, OCOH(4-nit)···NH2, HOCO(4-nit)···H2N, NO2···H2N
5, $P2_1/c$	-225.10	4.94 1.67	-3.2 17.5	-2.0 104.3	20.0 -143.6	126.8 -1.2	1.430	7.787, 16.436, 13.721, -, 122.66, - 4-aminobenzoic acid carboxylic homo-dimers, OCOH(4-nit)···NH2, OCO(H)(4-nit)···H2N, NO2···H2N
6, $P2_1/c$	-224.70	5.34 1.09	0.0 4.6	-10.4 36.3	21.3 162.4	142.0 -1.6	1.486	7.711, 17.129, 12.928, -, 123.59, - carboxylic heterodimers, NO2···H2N, HOCO(4-amn)···H2N
7, $P2_1/c$	-224.56	5.48 0.84	2.5 2.0	-3.0 87.0	31.0 70.9	130.1 0.1	1.467	8.252 16.275 11.352 90.00 109.07 carboxylic heterodimers, NO2···H2N
8, $P2_1/c$	-224.10	5.93 3.35	2.0 -0.8	14.7 81.8	33.4 -90.5	125.1 -6.3	1.436	8.138, 27.410, 6.749, -, 102.04, - chains involving both COOH and NH2
9, P-1	-223.50	6.54 1.09	2.8 -0.6	9.2 115.7	25.4 -173.3	130.6 -0.8	1.462	5.377, 11.292, 12.880, 69.3, 81.4, 89.0 4-aminobenzoic acid carboxylic homo-dimers, 4-nitrophenylacetic acid carboxylic homo-dimers, NO2···H2N
10, $P2_1/c$	-223.35	6.68 0.88	-5.2 -11.6	2.6 105.0	29.9 -81.8	132.6 0.6	1.448	4.337, 24.845, 13.810, -, 101.03, - carboxylic heterodimers, NO2···H2N, OCO(H)(4-amn)···H2N

[a] Lattice energy (in kJ mol^{-1}) expressed as the sum of the inter- and intra-molecular energies.
[b] Torsion angles defined in Scheme 1; the first row refers to 4-aminobenzoic acid and the second to 4-nitrophenylacetic acid.
[c] Only cell angles not constrained by symmetry given.

Table S9 shows a scarcity of hypothetical cocrystal structures containing 4-nitrophenylacetic acid homodimers. The stronger tendency of 4-aminobenzoic acid to form dimers when in competition with 4-nitrophenylacetic acid is explained in Figure S4 that shows that the carbonyl acceptor of 4-nitrophenylacetic acid is much weaker compared with 4-aminobenzoic acid.

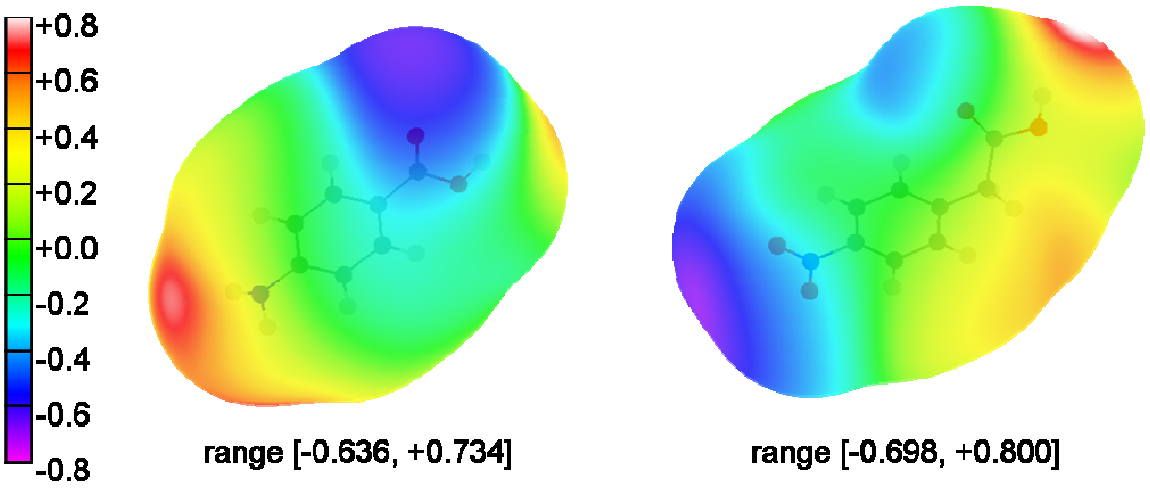


Figure S4 Electrostatic potential (in Volts) of the HF/6-31G(d,p) globally optimized molecular conformation of 4-aminobenzoic acid (left) and 4-nitrophenylacetic acid (right) on the 2×vdW surface (vdW radius for polar hydrogen atoms set to 1 au; vdw radii for other atoms taken from Bondi[11]), computed with atomic multipole moments up to rank 4 derived from the MP2(fc)/aug-cc-pVTZ isolated-molecule charge density. The maps were displayed using ORIENT.[12] The aliphatic carboxylic group of 4-nitrophenylacetic acid (right) is a much stronger donor and weaker acceptor than the aromatic carboxylic group of 4-aminobenzoic acid (left).

Reference List

1. **Williams, D. E.** ***J. Comput. Chem.*** **2001,** ***22*****, 1154-1166.**

2. **Coombes, D. S.; Nagi, G. K.; Price, S. L.** ***Chem. Phys. Lett.*** **1997,** ***265*****, 532-537.**

3. **Lai, T. F.; Marsh, R. E.** ***Acta Crystallogr.*** **1967,** ***22*****, 885-893.**

4. **Gracin, S.; Fischer, A.** ***Acta Crystallogr. , Sect. E*** **2005,** ***61*****, O1242-O1244.**

5. **Chisholm, J. A.; Motherwell, S.** ***J. Appl. Crystallogr.*** **2005,** ***38*****, 228-231.**

6. **Bowers, J. R.; Hopkins, G. W.; Yap, G. P. A.; Wheeler, K. A.** ***Cryst. Growth Des.*** **2005,** ***5*****, 727-736.**

7. **Smith, G.; Lynch, D. E.; Byriel, K. A.; Kennard, C. H. L.; Colin, H. L.** ***J. Chem. Crystallogr.*** **1997,** ***27*****, 307-317.**

8. **Kuhn, F. E.; Groarke, M.; Bencze, E.; Herdtweck, E.; Prazeres, A.; Santos, A. M.; Calhorda, M. J.; Romao, C. C.; Goncalves, I. S.; Lopes, A. D.; Pillinger, M.** ***Chem. Eur. J*** **2002,** ***8*****, 2370-2383.**

9. **Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C.** ***In preparation.*** **2009.**

10. **van Mourik, T.; Karamertzanis, P. G.; Price, S. L.** ***J. Phys. Chem. A*** **2006,** ***110*****, 8-12.**

11. **Bondi, A.** ***J. Phys. Chem.*** **1964,** ***68*****, 441-451.**

12. ***ORIENT: a program for studying interactions between molecules,*** **version 4.6; Stone, A. J.; Dullweber, A.; Engkvist, O.; Fraschini, E.; Hodges, M. P.; Meredith, A. W.; Nutt, D. R.; Popelier, P. L. A.; Wales, D. J. University of Cambridge, 2006**