# Supporting Information for

# "Investigating the latent polymorphism of maleic acid"

**Supporting Information contains:**
1) information on attempts at obtaining maleic acid form II;
2) further details of the computational methods;
3) unit cell parameters and lattice energies of the lowest energy predicted crystal structures.

## Information on repeated attempts at obtaining maleic acid form II

Attempts to repeat the crystallization of maleic acid form II were performed by dissolving the starting material (2:1 caffeine:maleic acid cocrystal material) in chloroform at initial concentrations of maleic acid ranging from 6.4 to 14.4 mM, with and without gentle heating to aid dissolution. In each experiment, the product obtained from slow solvent evaporation at ambient temperature was predominantly cocrystal material, as judged by its PXRD pattern (see cocrystal PXRD pattern in reference S.1), with some minor evidence of unreacted caffeine or maleic acid form I. In no instances was evidence of maleic acid form II apparent by PXRD. In the course of investigating the crystallization behavior of the caffeine:maleic acid cocrystal system, cocrystal synthesis was routinely obtained from a chloroform:methanol solvent system (reference S.1). Maleic acid form II crystals were not obtained during the course of these experiments.

Table S1. Unit cells and energies of the experimentally determined, lattice energy minimised and lowest energy predicted crystal structures of maleic acid. Structures are Z'=1 unless otherwise stated.

| form | space group | density ($g/cm^3$) | a (Å) | b (Å) | c (Å) | β (degrees) | lattice energy (kJ/mol) |
|---|---|---|---|---|---|---|---|
| **Observed** | | | | | | | |
| form I, Observed (room temperature) | $P2_1/c$ | 1.608 | 7.473 | 10.098 | 7.627 | 123.59 | - |
| form I, Observed (T = 180K) | $P2_1/c$ | 1.643 | 7.475 | 10.085 | 7.546 | 124.42 | - |
| form I, *Minimised* | $P2_1/c$ | 1.601 (-2.56%) | 7.428 (-0.63%) | 10.536 (+4.47%) | 9.403 (+24.6%) | 139.13 (+11.8%) | -93.44 |
| form II, Observed (T = 180K) | $Pc$ | 1.661 | 3.693 | 7.484 | 8.593 | 102.22 | - |
| form II, *Minimised* | $Pc$ | 1.636 (-1.51%) | 3.811 (+3.20%) | 7.493 (+0.12%) | 8.350 (-2.83%) | 98.89 (-3.26%) | -93.59 |
| **Lowest energy predictions** | | | | | | | |
| Global minimum (= form II, AAAA) | $Pc$ | 1.638 | 3.805 | 7.496 | 8.352 | 98.85 | -93.59 |
| 2nd (= form I, ABAB) | $P2_1/c$ | 1.601 | 7.428 | 10.537 | 9.402 | 139.13 | -93.44 |
| 3rd (ABAB) | $P2_1/c$ | 1.612 | 7.983 | 7.4677 | 10.4736 | 129.99 | -93.43 |
| 4th (ABAB) | $P\bar{1}$ (Z'=2) | 1.610 | 7.448 | 8.264 | 8.348 | α = 78.29<br>β = 76.12<br>γ = 76.38 | -93.35 |
| 5th (ABAB) | $P2_1/c$ | 1.614 | 7.451 | 10.440 | 6.924 | 117.51 | -93.29 |
| 6th (AABB) | $Pna2_1$ (Z'=2) | 1.618 | 10.443 | 7.461 | 12.235 | 90 | -93.26 |
| 7th (AABB) | $Pca2_1$ (Z'=2) | 1.601 | 10.466 | 7.463 | 12.334 | 90 | -92.93 |
| 8th (ABAB) | $P2_1/c$ | 1.620 | 6.223 | 10.356 | 8.861 | 123.52 | -92.79 |
| 9th (ABAB) | $P2_1/m$ (Z'=2) | 1.609 | 7.472 | 6.144 | 10.442 | 90.56 | -92.72 |
| 10th (AABB) | $Pca2_1$ (Z'=2) | 1.605 | 10.422 | 7.510 | 12.272 | 90 | -92.51 |

## Further details of the computational methods

The molecular structure was optimized within the density functional theory program DMol3 (S.2), using the VWN-BP functional and DNP basis set. Atomic charges were fitted to the resulting molecular electrostatic potential and used as the electrostatic component of the model potential in the generation of crystal structures. The W99 (S.3) empirically parameterised *exp-6* model was used for repulsion-dispersion interactions.

Each of the following 20 space groups were considered: $P2_1/c$; $P\overline{1}$; $P2_12_12_1$; $C2/c$; $P2_1$; $Pbca$; $Pna2_1$; $Pnma$; $P1$; $Pbcn$; $Cc$; $C2$; $Pca2_1$; $P2_1/m$; $P2_12_12$; $C2/m$; $P2/c$; $R\overline{3}$; $Pc$; and $Pccn$, each with Z'=1. Each space group was searched using the Monte Carlo simulated annealing algorithm in the Polymorph Predictor module of Cerius2.(S.4) The first 10 space groups were each searched 4 times to ensure a complete sampling of possible structures and the less common space groups (the final 10) were searched twice each.

The crystal structures in the lowest 10 kJ/mol from this search were then re-energy minimised using an improved model potential: the same *exp-6* parameters, but with the atomic charges replaced by atomic multipoles up to hexadecapole, obtained from a distributed multipole analysis of a B3LYP/6-31G(*d,p*) wavefunction, calculated with the program CADPAC.(S.5) *exp-6* interactions were summed to a 15Å cutoff, charge-charge, charge-dipole and dipole-dipole interactions by Ewald summation and all higher order electrostatic terms (up to $R^{-5}$) to a 15Å molecule-based cutoff. Energy minimisations were carried out using the program DMAREL (S.6) with space group symmetry constrained and each final structure was checked for stability by examining the calculated elastic stiffness tensor and **k**=0 phonon frequencies. Unstable structures were re-minimised with all symmetry constraints removed (sometimes resulting in Z'>1 structures). All final minima were then clustered using the COMPACK algorithm (S.7) to remove identical structures. Unit cell parameters and calculated lattice energies for the lowest 10 of these are reported in Table S1, along with unit cell parameters of the known polymorphs, before and after energy minimization with the *exp-6* + multipoles model potential.

In evaluating the relative energies of polar and non-polar crystal structures, an important consideration is that the energy of a polar crystal is dependent on the external shape of the crystal, *via* a surface correction term to the electrostatic energy. The polar crystal correction term, and its implications for crystal structure prediction, has been discussed in detail elsewhere.(S.8) This extra term has been excluded from our calculations of the relative energies of maleic acid polymorphs, for reasons explained below, and we estimate here the maximum errors that this could lead to in our results.

The energy term in question arises from the energy of the dipole of an internal unit cell in the field of charged crystal surfaces, resulting from unit cell dipoles at the external surfaces of the crystal. The term reaches a maximum for a plate-like morphology, with unit cell dipoles perpendicular to the large face of the plate:

$$U^{correction} \le 2\pi\left(p^2/V\right),$$

where $p$ and $V$ are the unit cell dipole and volume, respectively.

For example, if the *Pc* form II polymorph (unit cell dipole moment of 1.37 eÅ and volume of 235.6 Å$^3$) crystallised with such a morphology, under conditions where the surface charges are not compensated (say, by a build-up of external charges on the surface to compensate for the surface charge - "tin-foil boundary conditions"), it would be destabilized by approximately 1.7 kJ/mol relative to non-polar crystals. This would raise the energy of form II to be 1.6 kJ/mol less stable than form I. The correction would be similar for most AAAA crystals in our predicted list. The ABAB and AABB are centrosymmetric or have very small unit cell dipoles, so would be unaffected.

Our argument against including this energy term is that such a morphology is unrealistic for these crystals. Due to their internal structure (i.e. molecular sheets), the molecular and unit cell dipoles are almost certainly oriented *parallel* to the largest external face. In the limit of a thin plate, with such an orientation of dipoles, the surface correction term vanishes,(S.8a) leaving the relative energies as in Table S1.

A CIF file is also provided containing all predicted crystal structures in the lowest cluster (up to about ΔE = 5kJ/mol above the global minimum). Our suggestion for a possible form III, with AABB stacking of sheets, is the 6th structure in this list and the packing is shown in Figure 4.

References for Supporting Information

(S.1) Trask, A. V.; Motherwell, W. D. S.; Jones, W. *Cryst. Growth Des.* 2005, 5, 1013-1021.
(S.2) Delley, B. *J. Chem. Phys.* **1990**, *92*, 508-517.
(S.3) (a) Williams, D. E. *J. Mol. Struct.*, **1999**, *486*, 321-347; (b) Williams, D. E. *J. Comput. Chem.* **2001**, *22*, 1-20; (c) Williams, D. E. *J. Comput. Chem.* **2001**, *22*, 1154-1166
(S.4) Accelrys Inc.: Cerius2 version 4.6, San Diego, 2001.
(S.5) Amos, R. D. *et al*; CADPAC 6.5 ed.: Cambridge, **2001**.
(S.6) Price, S. L.; Willock, D. J.; Leslie, M.; Day, G. M.; DMAREL 3.1, 2001.
(S.7) Chisholm, J.A.; Motherwell, W. D. S. *J. Appl. Crystallogr.* **2005**, *38*: 228-231.
(S.8) (a) van Eijck, B. P. and Kroon, J. *J. Phys. Chem. B* **1997**, *101*, 1096.; (b) Pillardy, J.; Wawak, R. J.; Arnautova, Y. A.; Czaplewski, C. and Scheraga, H. A. *J. Am. Chem. Soc.* **2000**, *122*, 907; (c) van Eijck, B. P. and Kroon, J. *J. Phys. Chem.* **2000**, *104*, 8089; (d) Wedemeyer, W. J.; Arnautova, Y. A.; Pillardy, J.; Wawak, R. J.; Czaplewsky, C. and Scheraga, H. A. *J. Phys. Chem.* **2000**, *104*, 8090-8092.