

Testing a Variety of Electronic-Structure-Based Methods for the Relative Energies of 5-Formyluracil Crystals

Matthew Habgood, * Sarah L. Price, * Gustavo Portalone, * and Simona Irrera**, *, \$

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

[‡]Dipartimento di Chimica, "Sapienza" Università di Roma, P.le Aldo Moro 5, 00185 Rome, Italy



ABSTRACT: The lattice energies of the experimental and several hypothetical crystal structures of the RNA base uracil derivative 5-formyluracil are calculated with a range of methods, based either on the electronic structure of the molecule or the lattice. The explicit modeling of the polarization within the crystal in the model intermolecular potential and the inclusion of an empirical dispersion correction to the periodic density functional energy (DFT-D2) were the only methods able to calculate the energy balance between different conformations, hydrogen bonding, and $\pi-\pi$ stacking possibilities sufficiently accurately to give the observed structure as the most stable. Even these two methods underestimated the density of the room temperature structure, showing the need for improvement in the modeling of organic crystal structures.

The physical properties of organic molecular materials in the solid state are critically dependent upon the crystal structure that they adopt. This has led to an explosion of interest in modeling molecular crystals, driven by relevance to dyes, explosives, optics, electronics, and particularly to the pharmaceutical industry. The propensity for a molecule to adopt multiple crystalline forms both threatens the manufacturer's control of the quality of their products and provides the ability to select the solid state properties of a molecule. Accurate computational modeling of crystal structures and their relative energetic differences is essential to understanding existing crystalline forms and designing new ones. The control of the quality of their products of a molecule.

A wide variety of methods are available for evaluating the zerotemperature potential energy of crystal structures (lattice energy), ranging from computationally cheap but highly approximate analytical functions (e.g., atom-centered charges plus Lennard-Jones interactions)⁸ through more sophisticated treatments of intermolecular forces (e.g., distributed multipole electrostatic models)⁹ and density functional theory with analytical van der Waal's corrections^{10–12} to sophisticated *ab initio* methods.^{13–15} Different methods need to be used in a complementary fashion, according to their relative accuracy and computational cost, 16,17 as reliable crystal structure prediction will require both accurately ranking the energies of many thousands of plausible crystal structures and the ability to simulate the transformations that can occur with temperature and pressure. 18,19 It is notable that both of the methods that have been generally successful in the international blind tests of organic crystal structure prediction²⁰ rely on molecule-specific quantum mechanical calculations and the empirical fitting of part of the model to a range of organic crystal structures. The periodic DFT method has an empirically fitted damped C₆ atom—atom dispersion model.²¹ The models that use a distributed multipole model of the ab initio molecular charge density for the electrostatic forces usually combine this with an empirically fitted exp-6 atom—atom potential

for the other contributions to the intermolecular lattice energy.²² Since there has been considerable progress in the performance of electronic structure calculations for large molecular systems, it seemed timely to test these density-functional methods for modeling the balance of interatomic forces in organic crystal structures.

The chosen test molecule is 5-formyluracil ($C_5N_2O_3H_4$, Figure 1), a major oxidation product of thymine, which is fundamental to the investigation of DNA/RNA pairing and mispairing because it is known to be a major source for transition mutations. ^{23,24} The crystal structure ²⁵ (Figure 2) is determined by the compromises between different hydrogen bonding, base stacking, and other intermolecular and intramolecular interactions, as the *syn* conformer is the only one to have been found in the crystalline phase, but the *anti* conformer is more stable in isolated molecule calculations (Figure 1).

A small set of the most stable hypothetical crystal structures for syn and anti 5-formyluracil was selected from a group generated using the program MOLPAK, 26 whose intermolecular lattice energies were within 4 kJ mol⁻¹ of the most stable when optimized using the IMP model intermolecular potential. This comprised a distributed multipole²⁷ model (derived from the optimized MP2 6-31G(d,p) molecular charge density computed using GAUSSIAN²⁸), for the electrostatic contribution to the intermolecular lattice energy $U_{\rm inter}$, and an empirical exp-6 potential for all other contributions. These, plus the corresponding experimental structure, cover a wide range of crystal packings, as has been found in the experimental and energetically competitive structures for other 5-substituted uracils. 29 The hypothetical 5-formyluracil structures have different hydrogen bonds (Supporting Information) involving the uracil and formyl groups from the observed structure (Figure 2).

Received: May 27, 2011 Published: July 13, 2011

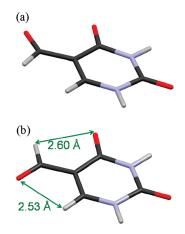


Figure 1. 5-Formyluracil in (a) the syn conformation found in the crystal structure and (b) the anti conformation showing the stabilizing intramolecular $O\cdots H$ close contacts.

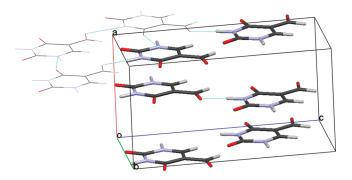


Figure 2. Layered experimental crystal structure of 5-formyluracil²⁵ (Cambridge Structural Database code GIMREA), showing the hydrogen-bonded network of the top layer with the molecules outside the conventional unit cell in wireframe. Hydrogen bonds are shown in light blue.

Three models based on separate evaluation of intermolecular and intramolecular energies were used to calculate the lattice energies of these 5-formyluracil crystal structures. To calculate the lattice energy, E_{latt} , of these crystal structures for the intermolecular potential model used in their derivation (IMP model), relative to infinitely separated molecules in their lowest energy conformation, the conformational energy penalty for the syn conformer, ΔE_{intra} , has to be added to the intermolecular lattice energy, U_{inter} , for the *syn* conformation crystal structures. The isolated molecule MP2 6-31G(d,p) calculations estimate ΔE_{intra} as 19.8 kJ mol⁻¹, so the loss of the O···H intramolecular interactions (Figure 1) significantly destabilizes the observed crystal structure. An improved lattice energy estimate (IMP+ WSM) includes the intermolecular induction energy evaluated with an explicit polarization model, using distributed anisotropic dipole—dipole polarizabilities calculated using the Williams—Stone—Misquitta (WSM) scheme, 30-32 as implemented in the CamCASP suite of programs, 33 with the induced dipoles iterated to consistency using DMACRYS.³⁴ A more approximate method of simulating the average polarization of the molecule within the crystal structure³⁵ is to calculate the distributed multipoles and relative conformational energies using the polarized continuum model (PCM)³⁶ implemented in GAUSSIAN with $\varepsilon = 3$, a value typical of organic crystals. This model (IMP+PCM) provides a

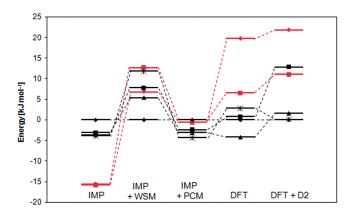


Figure 3. Lattice energies of the test set of four *syn* conformer structures (black) and two *anti* conformer structures (red) relative to the experimental structure. IMP = initial intermolecular potential model; IMP +WSM = intermolecular potential model with the addition of the induction energy; IMP+PCM = relaxed structures with average polarization from the PCM polarization model. DFT and DFT-D2 are the density functional theory relaxations without and with a dispersion correction. Some tie lines have been added to show significant changes in relative energy ordering between crystal structures with the same molecular conformation.

polarization effect that is not specific to the crystal structure, though the effect of the change in the charge distribution on the lattice energy minimized structure could be determined. Thus, these three models for the lattice energy which are based on *ab initio* calculations on a single molecule (to parametrize the electrostatic (and induction for WSM), model intermolecular potential, and provide the conformational energy difference $(\Delta E_{\rm intra})$) differ fundamentally in the treatment of the polarization of the molecule within the crystal structure.

Periodic electronic structure calculations have the advantage of not requiring separation into intermolecular and intramolecular interactions. The crystal structures were relaxed using periodic GGA-PW91 density functional theory (DFT) with the addition of an empirical dispersion correction, implemented using Grimme's method³⁷ (DFT-D2), within the QUANTUM ESPRESSO program package.³⁸ Problems of convergence of three of the structures led us to investigate the effect of annealing by DFT molecular dynamics. For three of the structures, the DFT-MD annealing led to a qualitative difference in the structures (shown in the Supporting Information) in terms of the atoms involved in hydrogen bonds and other close contacts, with the layers in one structure separating. The failure of MD using DFT without a dispersion correction to maintain the density of these structures is consistent with the considerable body of evidence pointing to the inadequate handling of long-range dispersion interactions by DFT. The excessive expansion in DFT annealing allowed sufficient reorientation of the molecules in two cases to simulate a hypothetical phase transition to an alternative low energy structure.

The remaining six structures, two *anti* and four *syn* (including the experimental structure), did have a DFT minimum sufficiently close to the original [judged by the capacity to overlay a cluster of 15 nearest-neighbor molecules from the two structures to within distance tolerances of 20% and angle tolerances of 20°, as determined in the Crystal Packing Similarity module of Mercury⁴¹ 2.4] for a meaningful comparison of lattice energies for all five computational methods (details in the Supporting Information, Table S1). The lattice energies relative to the

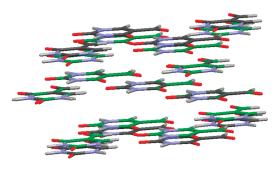


Figure 4. Overlay of three layers of the experimental structure, relaxed using uncorrected DFT (gray) and DFT-D2 (green). The closer interlayer spacing is more realistic for DFT-D2 but still overestimates the cell parameter in the stacking direction (a, Figure 2) by 9%, providing further evidence of deficiencies for π stacking. ^{11,13}

experimental structure, which can be reasonably assumed to be the most stable (Figure 3), show qualitative differences in the relative ordering and lattice energy differences in excess of the few kilojoules per mole usually associated with the experimental energy differences between polymorphs. The difference in the isolated molecule conformational energies results in the anti structures being considerably more stable than the *syn* structures with the initial intermolecular potential model (IMP), showing that this model does not balance the inter- and intramolecular interactions to give reasonable relative energies, despite giving the best match to the experimental structure. The PCM model reduces the difference between the two conformers considerably but, otherwise, results in little change in the energetic rankings of structures with the same conformer. It is perhaps not surprising that the redistribution of the isolated molecule charge densities as a response to an averaged electrostatic background does not accurately model the differences produced by the different hydrogen bonds in the crystalline environment. In contrast, lattice energies calculated using the WSM model yield a significant reranking, such that the most stable structures are those of the syn conformer, and the experimental crystal structure is the most stable overall. Hence, an intermolecular potential that includes the induction energy, representing how the charge density of the molecule adjusts to the specific crystalline environment (illustrated in the Supporting Information, Figure S1.3), is a considerable improvement over the initial model, which assumes that the isolated molecule electron density is undistorted in the crystal.

Following relaxation with DFT, the syn structures are the most stable overall, but the experimental structure is still less stable than another syn structure. With the inclusion of the empirical dispersion correction (DFT-D2), however, the experimental structure becomes the most stable. This can be related (Figure 4) to the dispersion correction having a significant effect on the interlayer distance in the lattice energy optimized experimental structure. However, the DFT-D2 method still significantly underestimates the density of the 298 K experimental structure, whereas the neglect of thermal expansion in all lattice energy models should result in an overestimate of the density by a few percent. Indeed, the density of all structures increases (Supporting Information, Table S1) in the order DFT, DFT-D2, IMP, IMP+PCM, (experiment), suggesting that improvements in the quality of the molecular or crystal charge density and the other approximations made in representing the intra- and intermolecular forces are needed before the neglect of thermal expansion becomes the dominant error.

Even without the dispersion correction, DFT resolves the apparent paradox of the syn conformer appearing in the solid state, whereas the anti conformer is the more stable in the gas phase. This is consistent with such methods treating all nuclei and electrons on the same footing, so that intra- and intermolecular polarization are both modeled well. This is in contrast to the intermolecular potential models when the polarization within the crystal is not modeled, although some average effect is absorbed in the empirically fitted exp-6 potential. Explicit modeling of the differential polarizations in the different intermolecular hydrogen bonds is needed, through using a distributed polarizability model, for the observed structure to be calculated as the most stable. The stability of the anti conformer in the gas phase is partially the result of the close internal contacts between the carbonyl oxygens and hydrogen atoms (see Figure 1), which could be viewed as weak intramolecular hydrogen bonds. Thus, this study reinforces the finding 42 that the lattice energy differences between polymorphs of oxalyl dihydrazide and o-acetamidobenzamide, which differ in the number of intra- and intermolecular hydrogen bonds, were only plausible when the induction energy was explicitly modeled, or when a dispersion corrected periodic electronic structure method was used.

The crystal structure of 5-formyluracil is not unique in having a molecular conformation that is a high energy conformer for the isolated molecule, and in having alternative crystal structures with different hydrogen bonds that are close in energy. Thus, these structures of 5-formyluracil and the conformational polymorphs of oxalyl dihydrazide and o-acetamidobenzamide form a particularly stringent test of the ability to model condensed phases and complexes of uracils and peptides, respectively. The crystal structures of other uracils²⁹ and amino acids⁴³ have been successfully predicted, demonstrating the specificity of the challenge to the range of energetically competitive crystal structures. This study has demonstrated that the relative energies of conformational polymorphs can require modeling of both the polarization and dispersion interactions, which are often inadequately represented by readily evaluated model intermolecular potentials and periodic DFT calculations, respectively. The use of distributed polarizability models to improve the model intermolecular potentials, or the addition of dispersion to periodic DFT methods, appears to have the potential for systematically improving modeling of the balance between inter- and intramolecular forces in these challenging organic systems.

■ ASSOCIATED CONTENT

Supporting Information. Further details of the computational methods, crystal energy landscape, extent of polarization of molecules in the known crystal, diagrams of the 5-formyluracil crystal structures, a table comparing the crystal structures and energies calculated with the different methods, the nine crystal structures in .res format. This information is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: +39 0649913819. E-mail: irrera@caspur.it.

Present Addresses

[§]Dipartimento di Chimica, "Sapienza" Università di Roma, P.le Aldo Moro 5, 00185 Rome, Italy.

■ ACKNOWLEDGMENT

This work has been supported by the Newton International Fellowship scheme (SI), the Ramsay Memorial Fellowships Trust (MH) and EPSRC via Control & Prediction of the Organic Solid State, www.cposs.org.uk (EP/F03573X). We thank Dr Alston Misquitta for assistance with determining the distributed polarizabilities of 5-formyluracil. Computing Support was provided by HECToR High Performance Computing resources via membership of the EPSRC-funded Materials Chemistry Consortium (EP/D504872).

■ REFERENCES

- (1) Bernstein, J. Cryst. Growth Des. 2011, 11, 632-650.
- (2) Bernstein, J. *Polymorphism in Molecular Crystals*; Clarendon Press: Oxford, United Kingdom, 2002; pp 240–296.
- (3) Price, C. P.; Grzesiak, A. L.; Matzger, A. J. J. Am. Chem. Soc. 2005, 127, 5512–5517.
- (4) Trask, A. V.; Shan, N.; Motherwell, W. D. S.; Jones, W.; Feng, S. H.; Tan, R. B. H.; Carpenter, K. J. Chem. Commun. 2005, 880–882.
- (5) Bauer, J.; Spanton, S.; Henry, R.; Quick, J.; Dziki, W.; Porter, W.; Morris, J. *Pharm. Res.* **2001**, *18*, 859–866.
 - (6) Price, S. L. Acc. Chem. Res. 2009, 42, 117-126.
 - (7) Lehmann, C. W. Angew. Chem., Int. Ed. 2011, 50, 5616-5617.
- (8) Svard, M.; Rasmuson, A. C. Ind. Eng. Chem. Res. 2009, 48, 2899–2912.
- (9) Day, G. M.; Motherwell, W. D. S.; Jones, W. Cryst. Growth Des. **2005**, *5*, 1023–1033.
- (10) Asmadi, A.; Neumann, M. A.; Kendrick, J.; Girard, P.; Perrin, M. A.; Leusen, F. J. J. *J. Phys. Chem. B* **2009**, *113*, 16303–16313.
- (11) Civalleri, B.; Zicovich-Wilson, C. M.; Valenzano, L.; Ugliengo, P. CrystEngComm 2008, 10, 405–410.
- (12) Sorescu, D. C.; Rice, B. M. J. Phys. Chem. C 2010, 114, 6734–6748.
- (13) Hongo, K.; Watson, M. A.; Sanchez-Carrera, R. S.; Iitaka, Y.; Aspuru-Guzik, A. J. Phys. Chem. Lett. **2010**, 1, 1789–1794.
- (14) Maschio, L.; Usvyat, D.; Civalleri, B. CrystEngComm 2010, 12, 2429–2435.
- (15) Beran, G. J. O.; Nanda, K. J. Phys. Chem. Letts **2010**, 1, 3480–3487.
- (16) Guin, M.; Patwari, G. N.; Karthikeyan, S.; Kim, K. S. Phys. Chem. Chem. Phys. 2009, 11, 11207–11212.
- (17) Maity, S.; Patwari, G. N.; Karthikeyan, S.; Kim, K. S. Phys. Chem. Chem. Phys. **2010**, 12, 6150–6156.
- (18) Gray, A. E.; Day, G. M.; Leslie, M.; Price, S. L. Mol. Phys. 2004, 102, 1067–1083.
- (19) Karamertzanis, P. G.; Raiteri, P.; Parrinello, M.; Leslie, M.; Price, S. L. J. Phys. Chem. B 2008, 112, 4298–4308.
- (20) Day, G. M.; Cooper, T. G.; Cruz-Cabeza, A. J.; Hejczyk, K. E.; Ammon, H. L.; Boerrigter, S. X. M.; Tan, J.; Della Valle, R. G.; Venuti, E.; Jose, J.; Gadre, S. R.; Desiraju, G. R.; Thakur, T. S.; van Eijck, B. P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Neumann, M.; Leusen, F. J. J.; Kendrick, J.; Price, S. L.; Misquitta, A. J.; Karamertzanis, P. G.; Welch, G. W. A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; van de Streek, J.; Wolf, A.; Schweizer, B. *Acta Crystallogr., Sect. B* **2009**, *65*, 107–125.
- (21) Neumann, M. A.; Perrin, M. A. J. Phys. Chem. B 2005, 109, 15531–15541.
 - (22) Price, S. L. Int. Rev. Phys. Chem. 2008, 27, 541-568.
- (23) Bjelland, S.; Eide, L.; Time, R. W.; Stote, R.; Eftedal, I.; Volden, G.; Seeberg, E. *Biochemistry* **1995**, 34, 14758–14764.
- (24) Bjelland, S.; Anensen, H.; Knaevelsrud, I.; Seeberg, E. *Mutat. Res.* **2001**, 486, 147–154.
- (25) Portalone, G.; Colapietro, M. Acta Crystallogr., Sect. C 2007, 63, O650-O654.
- (26) Holden, J. R.; Du, Z. Y.; Ammon, H. L. J. Comput. Chem. 1993, 14, 422–437.

- (27) Stone, A. J. J. Chem. Theory Comput. 2005, 1, 1128-1132.
- (28) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03; Gaussian Inc.: Wallingford, CT, 2004.
- (29) Barnett, S. A.; Hulme, A. T.; Issa, N.; Lewis, T. C.; Price, L. S.; Tocher, D. A.; Price, S. L. New J. Chem. **2008**, 32, 1761–1775.
- (30) Misquitta, A. J.; Stone, A. J. J. Chem. Theory Comput. 2008, 4, 7–18.
- (31) Misquitta, A. J.; Stone, A. J.; Price, S. L. J. Chem. Theory Comput. **2008**, *4*, 19–32.
- (32) Welch, G. W. A.; Karamertzanis, P. G.; Misquitta, A. J.; Stone, A. J.; Price, S. L. *J. Chem. Theory Comput.* **2008**, *4*, 522–532.
- (33) Misquitta, A. J.; Stone, A. J. *CamCASP*; University of Cambridge: Cambridge, U. K., 2007. http://www-stone.ch.cam.ac.uk/programs. html#CamCASP (accessed Dec, 2009).
- (34) Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. Phys. Chem. Chem. Phys. **2010**, 12, 8478–8490.
- (35) Cooper, T. G.; Hejczyk, K. E.; Jones, W.; Day, G. M. J. Chem. Theory Comput. 2008, 4, 1795–1805.
- (36) Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. J. Chem. Phys. 2002, 117, 43–45.
 - (37) Grimme, S. J. Comput. Chem. 2006, 27, 1787-1799.
- (38) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I.; Dal Corso, A.; de Gironcoli, S.; Fabris, S.; Fratesi, G.; Gebauer, R.; Gerstmann, U.; Gougoussis, C.; Kokalj, A.; Lazzeri, M.; Martin-Samos, L.; Marzari, N.; Mauri, F.; Mazzarello, R.; Paolini, S.; Pasquarello, A.; Paulatto, L.; Sbraccia, C.; Scandolo, S.; Sclauzero, G.; Seitsonen, A. P.; Smogunov, A.; Umari, P.; Wentzcovitch, R. M. J. Phys.: Condens. Mat. 2009, 21, 395502.
- (39) Zhao, Y.; Truhlar, D. G. J. Chem. Theory Comput. 2007, 3, 289-300.
- (40) Byrd, E. F. C.; Scuseria, G. E.; Chabalowski, C. F. J. Phys. Chem. B 2004, 108, 13100–13106.
- (41) Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A. *J. Appl. Crystallogr.* **2008**, *41*, 466–470.
- (42) Karamertzanis, P. G.; Day, G. M.; Welch, G. W. A.; Kendrick, J.; Leusen, F. J. J.; Neumann, M. A.; Price, S. L. *J. Chem. Phys.* **2008**, 128, 244708–244717.
- (43) Day, G. M.; Cooper, T. G. CrystEngComm 2010, 12, 2443–2453.

■ NOTE ADDED AFTER ASAP PUBLICATION

This article was published ASAP on July 27, 2011 with an incorrect affiliation for the third author. The correct version was published on August 2, 2011.