VIP Crystal-Structure Prediction

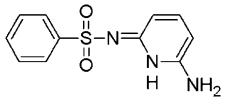
DOI: 10.1002/anie.201007488

# Molecule VI, a Benchmark Crystal-Structure-Prediction Sulfonimide: Are Its Polymorphs Predictable?**

*H. C. Stephen Chan, John Kendrick, and Frank J. J. Leusen**

Many organic compounds are known to crystallize in more than one distinct crystal structure, a phenomenon known as polymorphism,[1] which can cause problems but also offers exploitation opportunities because of the variation in physical properties among polymorphs. The ability to predict the crystal structures of a compound would make a significant contribution to crystal engineering. In 1988, John Maddox commented that the general failure in crystal-structure prediction (CSP) remained as "one of the continuing scandals in physical sciences".[2] Since then, steady progress has been made. To assess the technological advances in CSP, a series of blind tests was organized by the Cambridge Crystallographic Data Centre. In the 2007 blind test a new approach correctly predicted, for the first time, all four target structures.[3] This result was achieved using a DFT(d) method which combines density functional theory simulations and an empirical correction for dispersive forces.[4] Recently, we used the same methodology to re-evaluate the lattice energies and energy rankings of the experimental structures and all submitted predictions in the 1999, 2001, and 2004 blind tests, with very encouraging results.[5]

None of the 2001 blind-test participants had predicted the then only known experimental structure (form I, a $Z'=1$ structure) of molecule VI (6-amino-2-phenylsulfonylimino-1,2-dihydropyridine, see Scheme 1).[6] Two additional polymorphs, forms II[7] (a $Z'=2$ structure) and III[8] (a $Z'=1$ structure), were discovered after the 2001 blind test. Neither of these new structures had been reported by the blind-test participants. It was concluded that CSP had failed because the observed polymorphs were kinetically favored and the methods used in CSP are designed to locate thermodynamically favored forms.[7–9] These findings have led to comments on the state of CSP[9–11] and it was suggested that "structure prediction, which would be most valuable for



***Scheme 1.*** The molecular structure of molecule VI from the 2001 blind test.

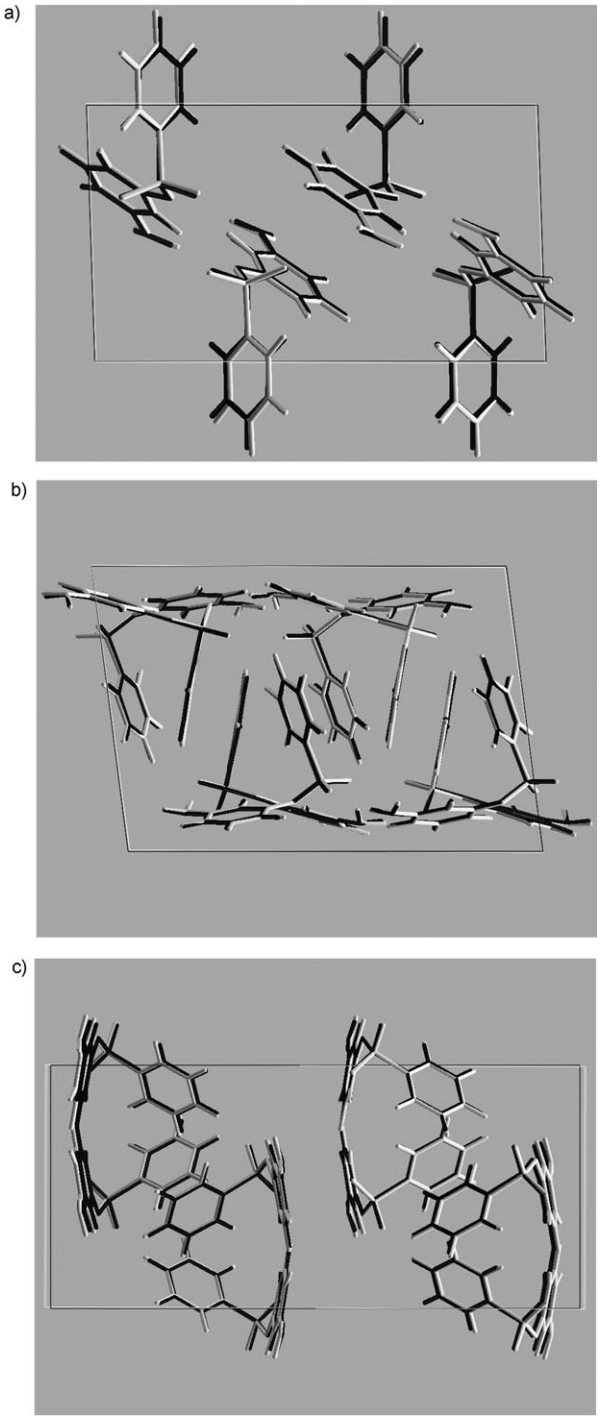



***Figure 1.*** Superpositions of the experimental (black) and the predicted (white) structures of a) form I, b) form II, and c) form III. All images are viewed along the *b* axis. See also Supporting Information.

[*] H. C. S. Chan, Dr. J. Kendrick, Dr. F. J. J. Leusen
School of Life Sciences, University of Bradford
Bradford, BD7 1DP (UK)
Fax: (+44) 1274-236155
E-mail: f.j.j.leusen@bradford.ac.uk

[**] We thank Avant-garde Materials Simulation for providing a courtesy license to the GRACE software package and the School of Life Sciences at the University of Bradford for funding this project. Molecule VI is 6-amino-2-phenylsulfonylimino-1,2-dihydropyridine.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201007488.

***Table 1:*** The ten most stable predicted structures compared to the three experimental polymorphs.

| CSP ranking | | $\Delta E_{DFT(d)}$[a] [kJ mol$^{-1}$] | Density [g cm$^{-3}$] | $Z'$[b] | Space group | Unit cell parameters[c] | | | | RMSD[d] [Å] | HB[e] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DFT(d) rank | TMFF rank | | | | | *a* [Å] | *b* [Å] | *c* [Å] | $\beta$ [°] | | |
| form I[f] | | | 1.464 | 1 | $P2_1/c$ | 8.48 | 8.96 | 14.89 | 91.86 | | 1D |
| 1 | 1 | 0.000 | 1.452 | 1 | $P2_1/c$ | 8.54 | 9.00 | 14.84 | 92.10 | 0.113 | 1D |
| form III[g] | | | 1.452 | 1 | *Pbca* | 10.62 | 9.32 | 23.06 | 90.00 | | 1D |
| 2 | 20 | 0.887 | 1.428 | 1 | *Pbca* | 10.59 | 9.32 | 23.50 | 90.00 | 0.093 | 1D |
| form II[h] | | | 1.463 | 2 | $P2_1/c$ | 12.11 | 10.79 | 17.46 | 97.32 | | 2D |
| 3 | 140 | 0.911 | 1.471 | 2 | $P2_1/c$ | 12.11 | 10.78 | 17.41 | 97.96 | 0.091 | 2D |
| 4 | 18 | 1.144 | 1.417 | 2 | $P2_1/c$ | 23.84 | 10.61 | 9.30 | 83.50 | | 1D |
| 5 | 21 | 1.357 | 1.438 | 2 | *Pbca* | 10.60 | 9.31 | 46.72 | 90.00 | | 1D |
| 6 | 2 | 1.362 | 1.445 | 1 | $P2_1/c$ | 11.70 | 10.61 | 9.24 | 87.45 | | 1D |
| 7 | 5 | 1.524 | 1.425 | 2 | *Pbca* | 10.62 | 9.26 | 47.26 | 90.00 | | 1D |
| 8 | 16 | 1.679 | 1.408 | 1 | *Pbca* | 10.62 | 9.28 | 23.85 | 90.00 | | 1D |
| 9 | 17 | 1.726 | 1.428 | 2 | *Pbca* | 10.60 | 9.33 | 46.88 | 90.00 | | 1D |
| 10 | 111 | 1.820 | 1.436 | 2 | $P2_1/c$ | 14.17 | 10.98 | 17.79 | 56.45 | | 2D |

[a] Relative DFT(d) lattice energy. [b] Number of independent molecules in the asymmetric unit. [c] Angles $\alpha$ and $\gamma$ are 90° because of symmetry constraints. [d] The root-mean-square deviations in atomic positions calculated by the crystal-packing similarity tool in Mercury CSD 2.2,[13] with a 16 molecule comparison for $Z'=1$ structures and a 30 molecule comparison for $Z'=2$ structures. [e] Dimensionality of the hydrogen-bonding network; 1D = one dimensional; 2D = two dimensional. [f] CSD reference UJIRIO.[7] [g] CSD reference UJIRIO05.[8] [h] CSD reference UJIRIO02.[7]

process chemistry, has still a way to go".[11] However, our previous work indicated that the DFT(d) method is capable of representing forms I and II of molecule VI correctly in terms of lattice energy, as well as in terms of geometry, although a full DFT(d) CSP study had not been performed.[5] To evaluate the relative stability of the new form III and to consider the validity of the previous comments on the predictability of the polymorphs of molecule VI,[7–9] we have conducted a full CSP of molecule VI using the GRACE software.[12]

The computational procedure used in this study and its limitations are described in the Supporting Information. In summary, a tailor-made force field (TMFF) was constructed specifically for molecule VI using data sets generated by the DFT(d) method. This TMFF was used for crystal-structure generation with one ($Z'=1$) and two ($Z'=2$) flexible, independent molecule(s) in the asymmetric unit. Low-energy TMFF structures were re-ranked by the DFT(d) method, resulting in 118 predicted structures with $Z'=1$ and 173 structures with $Z'=2$. The crystal-packing similarity tool in Mercury CSD 2.2[13] was used to calculate geometric deviations between experimental and predicted structures. All DFT(d) optimized crystal structures are provided in the Supporting Information.

Table 1 shows the CSP results for the ten lowest-energy DFT(d) structures. The rank 1, 3, and 2 predicted structures correspond to the experimental forms I, II, and III, respectively. The predicted order of polymorph stability is consistent with differential scanning calorimetry results.[8] Superpositions of the experimental and the predicted structures are shown in Figure 1 and the Supporting Information. Note that the TMFF is not accurate enough for a successful CSP of this molecule, but it is capable of generating all relevant structures to be considered with the more accurate DFT(d) method.

In form II, neighboring molecules adopt a dimer structure which is embedded in a two-dimensional hydrogen-bonding network, consistent with synthon A discussed elsewhere.[7,8] Forms I and III exhibit a one-dimensional hydrogen-bonding network, which corresponds to synthon B.[7,8] Figure 2 shows a plot of lattice energies versus densities of the predicted structures, mapping the dimensionalities of the hydrogen-bonding patterns across the energy spectrum. One-dimensional hydrogen-bonding patterns are prevalent among the low-energy structures, whilst two-dimensional networks occur infrequently. Zero- and three-dimensional motifs do not appear at all among the low-energy structures.

As a general class of compounds sulfonamides are well documented as being polymorphic.[14] Molecule VI has already been found to exist as three polymorphs and it is possible that more polymorphs may be discovered under specific crystallization conditions. The most likely candidates are the other low-energy crystal structures predicted in this study.

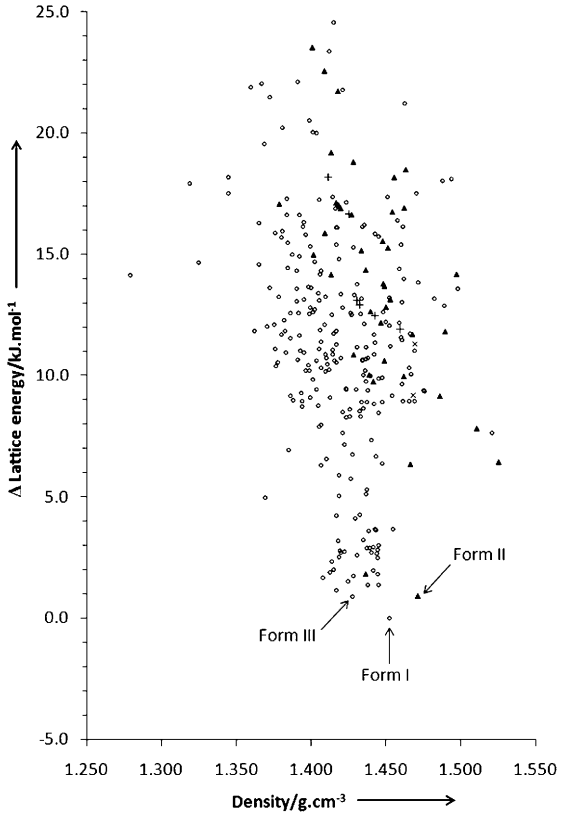



***Figure 2.*** Plot of relative lattice energies versus densities of the predicted structures, showing the distribution of hydrogen-bond dimensionalities (× 0D, ○ 1D, ▲ 2D, + 3D hydrogen bonding; see also Supporting Information).

Crystallization is a kinetic process, which explains the phenomenon of polymorphism. On the other hand, nature strives towards structures with the lowest possible energy, which explains why it is feasible to predict crystal structures by only considering lattice energy. A thermodynamically favorable crystal structure can only be obtained experimentally if crystallization conditions can be chosen such that a kinetic pathway exists to that structure. CSP failures have often been attributed to the kinetic nature of crystallization; molecule VI is a good example.[7–9] The most important finding of this work is that the three experimental structures of molecule VI are predicted correctly, both in terms of stability and in terms of geometry. Therefore, the failure of previous CSP studies to predict correctly the polymorphs of molecule VI was not due to the kinetic nature of the crystallization process, but due to the inaccuracy of the force fields used.[6–8] Even the TMFF developed specifically for molecule VI as part of this study is not capable of correctly ranking the three known polymorphs according to their relative stability.

The results presented herein, taken together with other results obtained with the DFT(d) approach,[3,5,15–17] suggest that a purely thermodynamic approach can predict the likely structures resulting from a crystallization experiment, at least for small molecules and provided that a sufficiently accurate method of calculating the lattice energy is used. However, despite the excellent prediction results achieved with the DFT(d) method, it is not yet possible to predict the outcome of a specific crystallization experiment. Such a task would have to take into account the influence of solvent, concentration, and temperature amongst other things, which is far beyond current computational capabilities.

Received: November 29, 2010
Published online: February 24, 2011

**Keywords:** crystal engineering · crystal-structure prediction · density functional calculations · molecular mechanics · polymorphism

[1] J. Bernstein, *Polymorphism in Molecular Crystals*, Clarendon, Oxford, **2002**.
[2] J. Maddox, *Nature* **1988**, *335*, 201.
[3] M. A. Neumann, F. J. J. Leusen, J. Kendrick, *Angew. Chem.* **2008**, *120*, 2461 – 2464; *Angew. Chem. Int. Ed.* **2008**, *47*, 2427 – 2430.
[4] M. A. Neumann, M. A. Perrin, *J. Phys. Chem. B* **2005**, *109*, 15531 – 15541.
[5] A. Asmadi, M. A. Neumann, J. Kendrick, P. Girard, M. A. Perrin, F. J. J. Leusen, *J. Phys. Chem. B* **2009**, *113*, 16303 – 16313.
[6] W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Dzyabchenko, P. Erk, A. Gavezzotti, D. W. M. Hofmann, F. J. J. Leusen, J. P. M. Lommerse, W. T. M. Mooij, S. L. Price, H. Scheraga, B. Schweizer, M. U. Schmidt, B. P. van Eijck, P. Verwer, D. E. Williams, *Acta Crystallogr. Sect. B* **2002**, *58*, 647 – 661.
[7] R. K. R. Jetti, R. Boese, J. A. R. P. Sarma, L. S. Reddy, P. Vishweshwar, G. R. Desiraju, *Angew. Chem.* **2003**, *115*, 2008 – 2012; *Angew. Chem. Int. Ed.* **2003**, *42*, 1963 – 1967.
[8] S. Roy, A. J. Matzger, *Angew. Chem.* **2009**, *121*, 8657 – 8660; *Angew. Chem. Int. Ed.* **2009**, *48*, 8505 – 8508.
[9] M. T. Kirchner, L. S. Reddy, G. R. Desiraju, R. K. R. Jetti, R. Boese, *Cryst. Growth Des.* **2004**, *4*, 701 – 709.
[10] C. H. Arnaud, *Chem. Eng. News* **2009**, *87*(40), 37.
[11] T. Laird, *Org. Process Res. Dev.* **2010**, *14*, 1.
[12] GRACE software from Avant-garde Materials Simulation, www.avmatsim.eu.
[13] J. A. Chisholm, W. D. S. Motherwell, *J. Appl. Crystallogr.* **2005**, *38*, 228 – 231.
[14] S. S. Yang, J. K. Guillory, *J. Pharm. Sci.* **1972**, *61*, 26 – 40.
[15] J. Kendrick, M. D. Gourlay, M. A. Neumann, F. J. J. Leusen, *CrystEngComm* **2009**, *11*, 2391 – 2399.
[16] A. Asmadi, J. Kendrick, F. J. J. Leusen, *Phys. Chem. Chem. Phys.* **2010**, *12*, 8571 – 8579.
[17] A. Asmadi, J. Kendrick, F. J. J. Leusen, *Chem. Eur. J.* **2010**, *16*, 12701 – 12709.