Computer Prediction of Organic Crystal Structures Using Partial X-ray Diffraction Data

Angelo Gavezzotti* and Giuseppe Filippini*

Contribution from the Dipartimento di Chimica Strutturale e Stereochimica Inorganica and Centro CNR per lo Studio delle Relazioni tra Struttura e Reattività Chimica, Università di Milano, Milano, Italy

Received December 4, 1995[⊗]

Abstract: This paper describes a computational procedure for the determination of complete crystal structures when the cell dimensions and space group only are known from X-ray crystallography. Molecular structure and conformation are assumed, and cannot be refined. When diffraction intensity data are available, the procedure offers an alternative to standard methods for the solution of the phase problem. The procedure applies to a wide range of organic molecules thanks to the evolution of the force field and of the computer programs. While the full ab initio prediction of crystal structures is still, in our opinion, a faraway goal, an important and fruitful application of this kind of computer modeling is in the completion of partial X-ray determinations when single crystals of suitable quality are not available, a rather frequent occurrence. Examples of this application are given, and its success implies that the need for producing good quality single crystals of newly synthesized organic compounds is nowadays less stringent, especially when only a knowledge of the intermolecular organization pattern in the crystal is sought.

Introduction

There has recently been considerable interest in the computer prediction of the crystal structure of organic compounds. The most ambitious goal—the ab initio prediction from a knowledge of chemical constitution only—has been approached, with a limited amount of success.[1]

When the synthesis of the compound has been achieved, performing a single-crystal X-ray structure analysis is nowadays a routine matter. In fact, recent advances in the technology of single-crystal X-ray diffraction equipment now make it possible to collect a complete set of diffraction data in a matter of hours, and efficient methods are available for the solution of the phase problem. Unfortunately, advances in the understanding of crystal nucleation and growth have not been equally astonishing, and cases where no single crystals suitable for exhaustive X-ray work can be obtained are still frequent. Sometimes, the cell dimensions and space group can be acquired from imperfect or unstable specimens, or powder spectra can be recorded. Besides, even with good quality crystals, in a few cases standard direct methods for solving the phase problem fail. The frequency of such occurrences is difficult to assess, since no traces of such partly or totally unsuccessful work appear in the literature.

Sometimes, two or more polymorphic forms (crystals with the same constituting molecule, but different intermolecular, and possibly also intramolecular, structure) are obtained. In such cases, preliminary X-ray work is frequently carried out on all polymorphs, but the intensity data collection is completed for one of them only, usually the more suitable one from a crystallographic standpoint. Thus, for the less attractive polymorph, only the cell parameters and space group may be determined.

Computational modeling can be employed in such situations to determine the crystal structure; this is a much more modest accomplishment than full ab initio crystal structure prediction, but can still be a useful complement to extant X-ray diffraction techniques. The present method relies on a previously developed procedure[2] (PROMET) by which many polymorphic crystal structures can be generated starting from a known, rigid molecular structure; ameliorations and extensions have been introduced, and we have optimized an approximate but efficient force field[3,4] for general organic molecules. This had led to an updated and generalized version of the crystal modeling package. The results presented here prove that, when molecular structure and conformation can be reasonably guessed, the intermolecular structure of the crystal can be determined unequivocally; this is an important step in the assessment of the physical properties of the material.

Procedure

The procedure essentially involves[2] the generation of clusters of molecules under the action of the most common symmetry operators in organic crystallography, and the evaluation of their cohesive energies by empirical intermolecular potentials. These clusters can then be combined and expanded into full three-dimensional crystal structures; known cell parameters are enforced by appropriate choices of the cluster parameters (for example, a cluster over a screw axis with translation x will lead to a cell parameter of $2x$), while known space group symmetry can be enforced by choosing the appropriate combination of symmetry operators. Total lattice energies are then calculated by the same empirical potentials.

The software has been updated[5] with respect to the original formulation,[2] largely increasing the efficiency of the energy hypersurface search. The computer programs can be used on any modern workstation, but a reasonable knowledge of geometrical crystallography and of space group symmetry is required; at some stages, human intervention cannot be dispensed with. At present, the procedure applies only to rigid molecules containing C, H, N, O, S, and Cl atoms, in the

[⊗] Abstract published in *Advance ACS Abstracts,* July 1, 1996.

(1) (a) Holden, J. R.; Du, Z.; Ammon, H. L. *J. Comput. Chem.* **1993**, *14*, 422. (b) Karfunkel, H. R.; Gdanitz, J. R. *J. Comput. Chem.* **1992**, *13*, 1171. (c) van Eijck, B. P.; Mooij, W. T. M.; Kroon, J. *Acta Crystallogr.* **1995**, *B51*, 99. (d) Williams, D. E. *Acta Crystallogr.* **1996**, *A52*, *326.*

(2) Gavezzotti, A. *J. Am. Chem. Soc.* **1991**, *113*, 4622. The procedure relies in its final stage on the PCK83 software (ref 13).

(3) Filippini, G.; Gavezzotti, A. *Acta Crystallogr.* **1993**, *B49*, 868.

(4) Gavezzotti, A.; Filippini, G. *J. Phys. Chem.* **1994**, *98*, 4831.

(5) Gavezzotti, A. PROMET(5): A Program for the Generation of Possible Crystal Structures from the Molecular Structure of Organic Compounds, Mark 5 version, University of Milano, Italy, 1995 (available from the author upon request).

S0002-7863(95)04063-7 CCC: $12.00

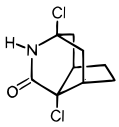
space groups $P1$, $P\bar{1}$, $P2_1$, $P2_1/c$, Pc, Cc, $P2_12_12_1$, $Pna2_1$, $Pca2_1$, $Pbca$, and $C2/c$, which account for over 90% of organic crystal structures.[6] Computing times run from minutes for the most favorable cases to several hours for complicated cases and large (above 40 atoms) molecules.

A complete molecular model must be supplied as input; when the molecule has internal (point group) symmetry, in some cases the procedure can be applied to the generation of crystal structures with half a molecule in the asymmetric unit, provided that proper care is taken of certain geometrical requirements in the generation of the space group symmetry.[5] A nonsymmetrical molecular dimer can also be constructed, and then packed in a crystal structure with two molecules in the asymmetric unit. This last procedure however introduces a further degree of freedom and a large uncertainty in the basic building block, and its performance has not yet been tested thoroughly.

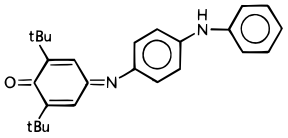
Nothing prevents the use of the procedure for attempting a complete ab initio crystal structure prediction. The implications of such an enterprise have been discussed.[7,8]

Validation Tests

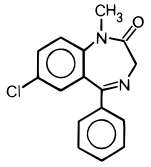
Four fully determined structures, 1,7-dichloro-9-azatricyclo-[4,3,1,$0^{3,7}$]decan-8-one,[9] **1**, 3,5-dimethoxybenzoic acid,[10] **2**, 2,6-di-*tert*-butyl-1,4-benzoquinone 4-(4-anilinophenyl)imine,[11] **3**, and diazepam (7-chloro-1,3-dihydro-1-methyl-5-phenyl-2*H*-1,4-benzodiazepin-2-one),[12] **4**, were chosen to test the procedure



1



3



4

and force field; the successful reproduction of the known crystal structures implies the solution of the phase problem in X-ray crystallography. The first two compounds were selected as carrying an obvious bias toward intermolecular recognition in their hydrogen-bonding ability, the other two as intermediate (**4**) and extreme (**3**) cases of molecular complexity. Admittedly, many other similar molecules could have served the same purpose equally well. Molecular structures were fixed as resulting from the X-ray determinations, and the crystal structure search was restricted to the known unit cell and space group (except for **3**). Since the crystal structures generated by PROMET are eventually refined using a lattice energy optimizer,[13] for the comparisons also the X-ray structures were relaxed in the same way under the action of the potential field (this is the case for all entries labeled "opt" in the tables).

Compounds **1** and **2** most likely[4,14] form hydrogen-bonded dimers in the crystal; building such dimers and packing them in the known space group, we reproduced the full crystal structures (Table 1) in a few minutes to a few hours of cpu time (Silicon Graphics Indy). For compound **3**, the search started by forming a chain of N−H···N hydrogen bonds, in keeping with the general principle that whenever a molecule carries both hydrogen bond donor and acceptor sites, such a bond is almost invariably formed in the crystal.[15] In space group $P2_1/c$, the search was biased with the known cell parameters.

Table 1. Results of Computational Experiments on Crystal Structure Solution

structure		Z	cell parameters a (Å)	b (Å)	c (Å)	β (deg)	packing energy (kcal/mol)
			Compound **1**				
$P2_1/n$	expt[a]	4	6.241	12.296	12.758	95.58	−24.3
	calc	4	6.09	12.38	12.47	94	−25.0
			Compound **2**				
$P2_1/c$	expt[b]	4	10.879	4.924	16.388	104.42	−30.8
	calc	4	11.08	4.71	16.24	103	−31.0
			Compound **3**				
$P2_1/n$	expt[c]	4	6.082	12.546	30.045	94.56	−42.9
$P2_1/c$	opt[d]	4	6.01	12.30	29.57	96.1	−45.8
	calc	4	6.01	12.32	29.53	96	−45.9
$P2_1$	calc	2	6.10	12.14	17.44	61	−44.2
$P2_12_12_1$	calc	4	6.00	12.43	31.54		−42.1
			Compound **4**				
$P2_1/c$	expt[e]	4	7.976	13.354	12.928	90.01	−31.1
	opt[b]	4	8.02	12.67	12.97	89	−32.1
	calc	4	8.02	12.64	12.97	89	−32.1

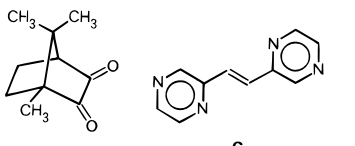
[a] Reference 9. [b] Reference 10. [c] Reference 11. [d] After structure relaxation under the potentials (see the text). [e] Reference 12.

Results for **3** and **4** (Table 1) show that the X-ray structures were correctly reproduced, as was further checked by comparing powder spectra[16] for the X-ray and calculated crystal structures (Figure 1). For **3**, the lattice energies of several computational polymorphs are, as expected,[17] quite similar.

As already suggested by previous experience,[8,17] and confirmed by many other successful tests not reported here in detail, the performance of the procedure for the generation of the full crystal structure from known cell dimensions and space group is very satisfactory.

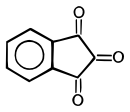
Structure Determination for Polymorphs

Literature cases were collected where the crystal structure of one polymorph was completely characterized, and mention was made of the existence of another polymorph whose crystals were not suitable for complete X-ray work, so that only the cell parameters and space group had been determined. Such was the case for 1-camphorquinone,[18] **5**, (*E*)-2,2′,5,5′-tetraazastilbene,[19] **6**, triketoindan,[20] **7**, 1,8-dinitronaphthalene,[21,22] **8**, and

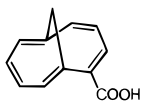


5

6



7



10

3-chlorocinnamic acid,[23,24] **9**. The computer generation of the complete crystal structure for the undetermined polymorphs

(6) Brock, C. P.; Dunitz, J. D. *Chem. Mater.* **1994**, *6*, 1118.
(7) Gavezzotti, A. *Acc. Chem. Res.* **1994**, *27*, 309.
(8) Gavezzotti, A. *Acta Crystallogr.* **1996**, *B52*, 201.
(9) Barlow, M. G.; Pritchard, R. G.; Sibous, L.; Tipping, A. E. *Acta Crystallogr.* **1994**, *C50*, 550.
(10) Lynch, D. E.; Smith, G.; Byriel, K. A.; Kennard, C. H. L. *Acta Crystallogr.* **1994**, *C50*, 1259.
(11) Hiller, W. Personal communication.
(12) Camerman, A.; Camerman, N. *J. Am. Chem. Soc.* **1972**, *94*, 268.
(13) Williams, D. E. PCK83; QCPE Program 548; Quantum Chemistry Program Exchange; Chemistry Department, Indiana University, Bloomington, IN, 1983.
(14) (a) Leiserowitz, L. *Acta Crystallogr.* **1976**, *B32*, 775. (b) Leiserowitz, L.; Hagler, A. T. *Proc. R. Soc. London* **1983**, *A388*, 133.
(15) Etter, M. C. *Acc. Chem. Res.* **1990**, *23*, 120.
(16) Young, R. A.; Sakthivel, A.; Moss, T. S.; Paiva-Santos, C. O. DBWS, Programs for Rietveld Analysis of X-ray and Neutron Powder Diffraction Patterns, 1994.
(17) Filippini, G.; Gavezzotti, A. *J. Am. Chem. Soc.* **1995**, *117*, 12299.
(18) Bright, W. M.; Cannon, J. F.; Langs, D. A.; Silverton, J. V. *Cryst. Struct. Commun.* **1980**, *9*, 251.
(19) Vansant, J.; Smets, G.; Declerq, J. P.; Germain, G.; Van Meersche, M. *J. Org. Chem.* **1980**, *45*, 1557.
(20) Bolton, W. *Acta Crystallogr.* **1965**, *18*, 5.
(21) Ciechanowicz-Rutkowska, M. *J. Solid State Chem.* **1977**, *22*, 185.
(22) (a) Kozhin, V. M. *Zh. Strukt. Khim.* **1961**, *2*, 46. (b) Kozhin, V. M. *Zh. Strukt. Khim.* **1964**, *5*, 324.
(23) Kanao, S.; Kashino, S.; Haisa, M. *Acta Crystallogr.* **1990**, *C46*, 2436.
(24) Schmidt, G. *J. Chem. Soc.* **1964**, 2015.

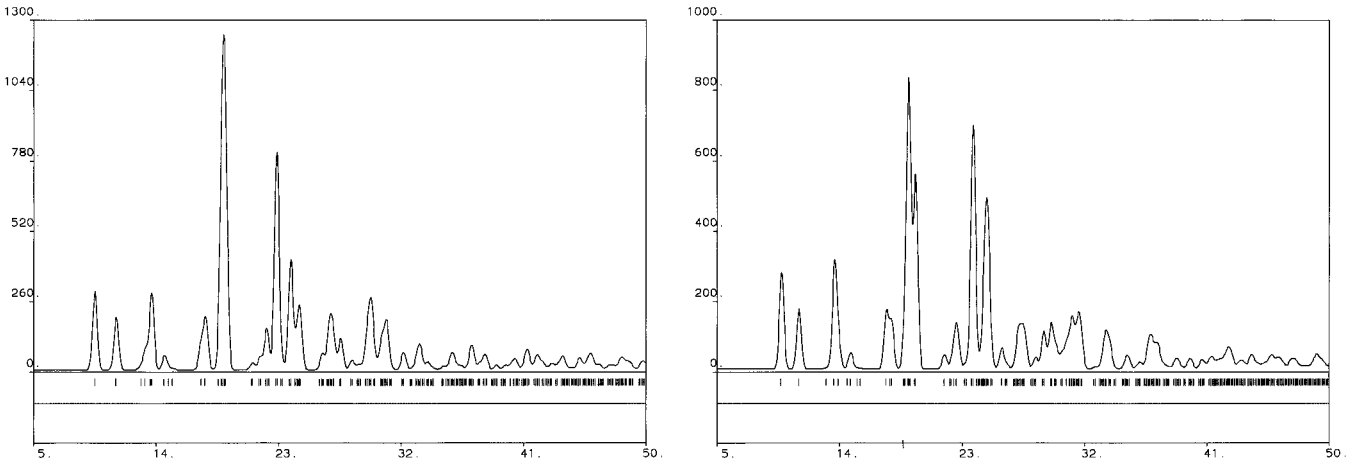


Figure 1. Powder spectra (ref 16) of **4** calculated from (a) the crystal structure as determined by single-crystal X-ray analysis (ref 12) and (b) the best crystal structure obtained from the PROMET search.

Table 2. Predicted Crystal Structures for Polymorphs with Known Cell Parameters and Space Group

			cell parameters						
structure		Z	a (Å)	b (Å)	c (Å)	α (deg)	β (deg)	γ (deg)	packing energy (kcal/mol)
					1-Camphorquinone, **5**				
$I2$	expt[a,b]	4	12.081	6.731	23.43		96.25		−18.6
	opt[c]	4	11.82	6.62	22.51		96		−19.8
$P2_12_12_1$	expt[b,d]	4	6.7	11.3	12.7				
	calc	4	6.61	11.03	12.00				−19.5
$P\bar{1}$	calc	2	6.65	12.22	13.11	57	69	118	−19.8
$P2_1/c$	calc	4	10.73	6.52	12.44		89		−19.3
					(*E*)-2,2′,5,5′-Tetraazastilbene, **6**				
$P2_1/a$	expt[a,e]	4	9.190	5.894	9.080		114.77		−27.8
	opt[c]	4	8.87	5.91	9.04		118		−29.9
$P1$, $P\bar{1}$	expt[d,e]	–	11.52	5.26	3.85	101.2	92.4	95.1	
$P\bar{1}$	calc	1	11.39	5.19	3.57	81[f]	93	89[f]	−30.8
					Triketoindan, **7**				
$I4_1cd$	expt[a,g]	8	7.058	7.058	28.77				−17.5
	opt[c]	8	7.03	7.03	28.32				−20.4
$Pbca$	expt[d,g]	8	15.524	14.160	6.380				
	calc	8	15.59	14.12	6.39				−20.6
					1,8-Dinitronaphthalene, **8**				
$P2_12_12_1$	expt[a,h]	4	11.375	14.974	5.388				−28.9
	opt[c]	4	11.31	14.73	5.27				−29.6
	calc	4	11.42	14.76	5.19				−29.4
$C2/c$	expt[d,i]	4	17.49	8.05	14.91		115.25		
$I2/a$	calc	4	17.14	7.86	14.70		113		−28.6
$P2_1/c$	calc	4	7.21	8.26	14.88		88		−29.5
$Pbca$	calc	8	9.29	13.73	14.26				−28.8
$P\bar{1}$	calc	2	7.68	7.80	9.20	104	102	117	−29.7
					3-Chlorocinnamic Acid, **9**				
$P\bar{1}$	expt[a,j]	2	8.618	13.627	3.909	106.77	96.26	75.71	−27.4
	opt[c]	2	8.49	13.30	3.78	105	97	76	−28.5
$P2_1/c$	expt[d,k]	4	14.1	4.93	12.5		94.0		
	calc	4	14.31	4.90	11.90		96		−26.7
	calc	4	14.68	3.82	14.08		94		−28.9

[a] Full X-ray structure. [b] Reference 18. [c] After structure relaxation under the potentials (see the text). [d] X-ray cell and space group only. [e] Reference 19. [f] Match α and γ of the experimental structure upon changing the direction of the *b* cell axis. [g] Reference 20. [h] Reference 21. [i] Reference 22. [j] β phase, ref 23. [k] γ phase, ref 24.

amounts to real crystal structure prediction, assisted by partial X-ray analysis data. The molecular conformation was assumed to be the same as in the fully determined polymorph.

Results are shown in Table 2. Success was judged by coincidence of cell parameters within the limits set by previous experience using the PROMET procedure; a deviation up to 10% is considered acceptable, also in view of the unavoidable cell shrinkage introduced by the final optimization (a well-known spurious effect due to the neglect of molecular libration in the force field). Truly, however, a wrong structure with the right cell parameters is an unlikely but not impossible occurrence.

The calculated $P2_12_12_1$ structure of **5** is shown in Figure 2. Its packing energy (−19.5 kcal/mol) is virtually indistinguishable from that of the $I2$ polymorph (−19.8 kcal/mol). Calculated crystal structures in centrosymmetric space groups have quite similar packing energies, but the compound may have been isolated in optically pure form, so that these space groups may not have been accessible.

Calculations on compound **6** were performed assuming a centrosymmetric molecule and using space group $P1$, so that the resulting space group is in fact $P1$, $Z = 1$; the experimental work could not discriminate between $P1$ and $P\bar{1}$, nor could the

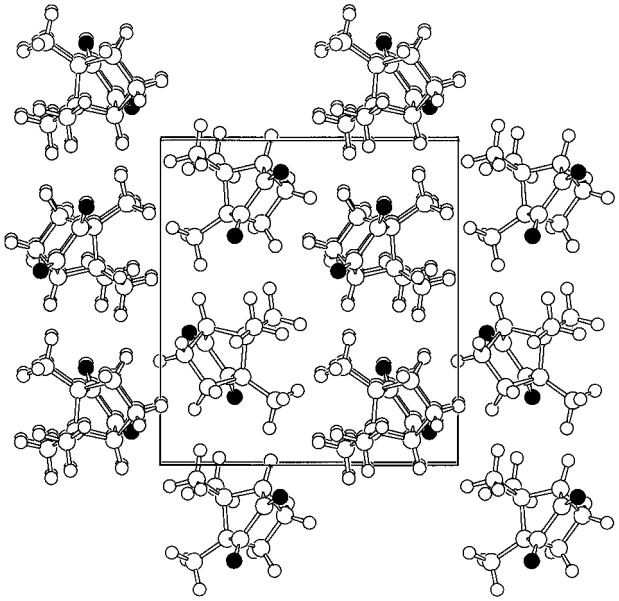


Figure 2. Packing diagram (ref 28) for the predicted $P2_12_12_1$ crystal structure of **5**. Oxygen atoms are shaded.

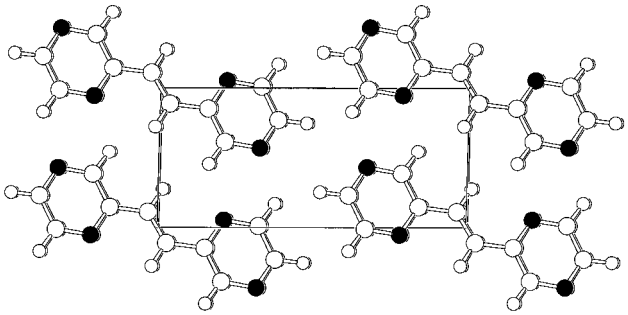


Figure 3. Packing diagram (ref 28) for the predicted $P\bar{1}$ crystal structure of **6**. Nitrogen atoms are shaded.

calculations, given the assumption of molecular centrosymmetry. A packing diagram for the calculated structure is given in Figure 3. The packing energy of the $P\bar{1}$ polymorph is more cohesive than that of the fully characterized one by 0.9 kcal/mol.

The *Pbca* crystal structure of **7** (Figure 4) was determined, in a few minutes of cpu, by forming a highly cohesive, head-to-tail centrosymmetric molecular dimer (an obvious possibility for this highly dipolar molecule) and then adding two perpendicular screw axes. The packing energies of the two polymorphs are, again, virtually indistinguishable.

An extensive search of the crystal energy hypersurface was conducted for **8**. The X-ray $P2_12_12_1$ structure was identified among the calculated ones, a further validation of the procedure. Calculated cell dimensions of the *C*2/*c* polymorph match experimental ones only after transformation to *I*2/*a*. A restricted set of reflections having been collected,[22] a thorough analysis of extinctions may not have been possible; otherwise, the match between the *I*2/*a* and *C*2/*c* cells would be a surprising coincidence indeed. The packing energy (Table 2) of this last structure is less cohesive than that of the more stable polymorph by 1 kcal/mol or about 3% of the total lattice energy. As expected, the packing energy of other calculated crystal structures for this compound (Table 2) are quite close to those of the observed structure.

The cell parameters of the $P2_1/c$ polymorph of **9** are reproduced by the calculations, but its packing energy (Table 2) is less cohesive than that of the more stable polymorph by 1.8 kcal/mol or about 6% of the total lattice energy. The best calculated (but as yet not observed) $P2_1/c$ structure is of the β type[25] (4 Å axis), like the triclinic one, implying superposition of flat molecules at short distances. In the observed $P2_1/c$ γ phase (shortest axis 4.9 Å) there is no superposition of flat molecules, but a herringbone-type arrangement of the molecular planes (Figure 5).

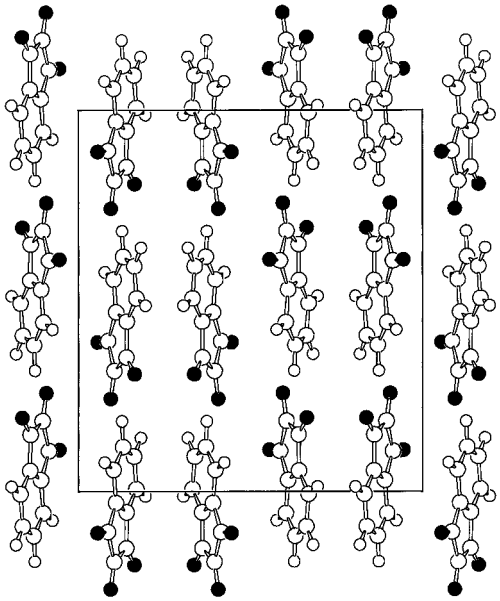


Figure 4. Packing diagram (ref 28) for the predicted *Pbca* crystal structure of **7**. Oxygen atoms are shaded.

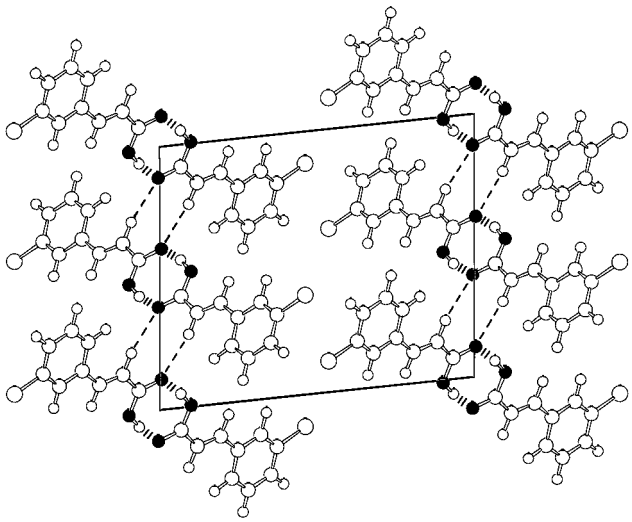


Figure 5. Packing diagram (ref 28) for the predicted $P2_1/c$ crystal structure of **9** corresponding to the observed γ phase. Oxygen atoms are shaded. O—H···O hydrogen bonds and weak C═C—H···O interactions are evidenced.

Table 2 shows that all differences in packing energy between polymorphs are within the usual range,[17] that is, a few percent of the total cohesive energy. Nevertheless, all structures reported in the table certainly correspond to minima in the potential energy hypersurface, having been reached many times over and from many different starting points throughout the search.

Discussion and Conclusion

In principle, polymorphs having less cohesive packing energies should be metastable ones. However, when the relative energies of polymorph structures are judged, the accuracy of the force field is crucial, given the smallness of the differences.

(25) Sarma, J. A. R. P.; Desiraju, G. R. *Acc. Chem. Res.* **1986**, *19*, 222 and references therein.

For example, we ran some test calculations for **9**, supplementing the 6-exp force field with R^{-1} terms using empirical charge parameters;[26] although minor changes appeared, the energy ordering was essentially unaffected. We may conclude that the γ phase is a metastable polymorph due to kinetically favored survival of T-shaped arrangements of molecules in the early stages of crystal formation; we stress that such conclusions are necessarily of a tentative nature, in view of the fact that energy differences between polymorphs are of the same order of magnitude as the uncertainties introduced by the force field and by the various generation and minimization algorithms embodied in the overall procedure. Not unexpectedly, previous experience[8] with the use of charge parameters produced a significant reshuffling of the energy ordering among polymorphs. We doubt that a general purpose empirical force field could ever be designed to the required level of accuracy for an absolute discrimination of the most stable polymorph.

The most important conclusion drawn from the present work is that, while the energy ordering of polymorphs depends on the force field, the location of minima in the potential energy hypersurface does not. When X-ray data point out the existence of a crystal energy minimum, as any observed crystal structure must be, then even a moderately accurate 6-exp force field is adequate for the determination of the full structure. The application of computer modeling we have described in this paper is therefore relatively free from the need of an extremely accurate force field, and in this respect, computer generation of crystal structures can turn into a useful complement to X-ray diffraction in the analysis of organic solids.

We summarize our results by stating that, allowance made for its present limitations, the procedure we have described may be helpful in the following cases: (1) It provides an alternative to direct methods for the solution of the phase problem. (2) When the cell parameters and space group only are available, it provides a method for obtaining the full crystal structure in the absence of single-crystal diffraction measurements. (3) When powder diffraction data are available, it will conceivably produce a good starting point for Rietveld treatment, by which even the molecular geometry could be refined, given the recent advances in this methodology.[27] For the above reasons, there is a substantial chance that in the near future the need for well-grown single crystals suitable for intensity measurements will be greatly reduced.

On the other hand, we believe that ab initio crystal structure prediction[1] is still a faraway goal. Packing energies of polymorphs are always quite similar; conformational differences may appear between polymorphs, while an accurate estimation of both intra- and intermolecular energy contributions seems still beyond reach, as it is very difficult to calibrate consistent inter- and intramolecular force fields, and also because the energy surface search is increasingly problematic. Possible polymolecular asymmetric units and formation of solvates add to the overall uncertainty. Finally, even if the correct energy ordering of polymorphic structures were available, their appearance would possibly depend on kinetic factors. In fact, in none of the cases presented in this paper could the observed crystal structures have been predicted without the support of X-ray diffraction data.

Supporting Information Available: Table S1, atomic coordinates for all the predicted structures in Tables 1 and 2 (8 pages). See any current masthead page for ordering and Internet access instructions.

JA9540637

(26) Charge separations: C−Cl, 0.2 electron; C−H, 0.1 electron; COOH charge parameters, 0.5, −0.4, −0.3, 0.2 electron, respectively.

(27) Masciocchi, N.; Moret, M.; Sironi, A.; Ardizzoia, G. A.; LaMonica, G. Submitted for publication.

(28) Keller, E. SCHAKAL92. A Program for the Graphic Representation of Molecular and Crystallographic Models, University of Freiburg, Germany, 1993.