

# A strategy for predicting the crystal structures of flexible molecules: the polymorphism of phenobarbital

G. M. Day,<sup>\*a</sup> W. D. Sam Motherwell<sup>b</sup> and W. Jones<sup>a</sup>

Receipt/Acceptance Data [DO NOT ALTER/DELETE THIS TEXT]

5 Publication data [DO NOT ALTER/DELETE THIS TEXT]

DOI: 10.1039/b000000x [DO NOT ALTER/DELETE THIS TEXT]

A computational exploration of the low energy crystal structures of the pharmaceutical molecule phenobarbital is presented as a test of an approach for the crystal structure prediction of flexible molecules. Traditional transferable force field methods of modelling flexible molecules are  
10 unreliable for the level of accuracy required in crystal structure prediction and we outline a strategy for improving the evaluation of relative energies of large sets of crystal structures. The approach involves treating the molecule as a set of linked rigid units, whose conformational energy is expressed as a function of the relative orientations of the rigid groups. The conformational energy is calculated by electronic structure methods and intermolecular interactions using an  
15 atomic multipole description of electrostatics. A key consideration in the design of our approach is the scalability to more typical pharmaceutical molecules of higher molecular weight with many more atoms and degrees of flexibility.

## Introduction

The prediction of crystalline structure starting from nothing  
20 more than the connectivity of atoms within a molecule has been a great challenge for the modelling community [1-4] and the pursuit of such a computational procedure has been a driving force for the development of increasingly accurate methods of modelling molecular crystals. The most common  
25 approach to crystal structure prediction (CSP) has been to search the multidimensional lattice energy surface for all local minima that correspond to possible crystal packings of the molecule, then to rank these possibilities by their lattice [5] or free energy [6-8]. While the growth conditions may influence  
30 which crystal is formed, the observed crystal structure should be amongst the thermodynamically most stable structures found in such a lattice energy search.

A lattice energy search often finds a myriad of distinct crystal structures within a reasonable energy range and the  
35 energy differences between structures are typically on the order of 0.1-1 kJ/mol [9, 10]. Thus, ranking the thermodynamic stability of structures is a stringent test of the model potentials used to calculate the energies of these crystal structures. Developments in the modelling of intermolecular  
40 interactions have led to improved reliability of crystal structure prediction for small rigid molecules and, in most cases, the observed crystal structures of such molecules are found amongst a small set (5-10) of the lowest energy predictions [9]. Such results can help us to understand the  
45 crystallisation behaviour of a molecule and are a useful

starting point for investigations of other factors that determine which structure is actually obtained experimentally (eg. nucleation and growth).

One potentially important application of crystal structure  
50 prediction is in the anticipation and understanding of polymorphism in the pharmaceutical industry. Polymorphism is the ability of a molecule to adopt more than one crystal structure and the sudden appearance of an unexpected polymorph can have critical consequences [11]. In order to  
55 make an impact on the pharmaceutical industry, computational methods of CSP that have been successful in dealing with simple rigid molecules must now be extended to the more complex and challenging systems typically encountered in pharmaceutical materials science. One source  
60 of complexity is the conformational flexibility of a molecule; small changes in molecular conformation can have large effects on calculated energies [12] and there are difficulties in treating inter- and intra-molecular interactions in an accurate

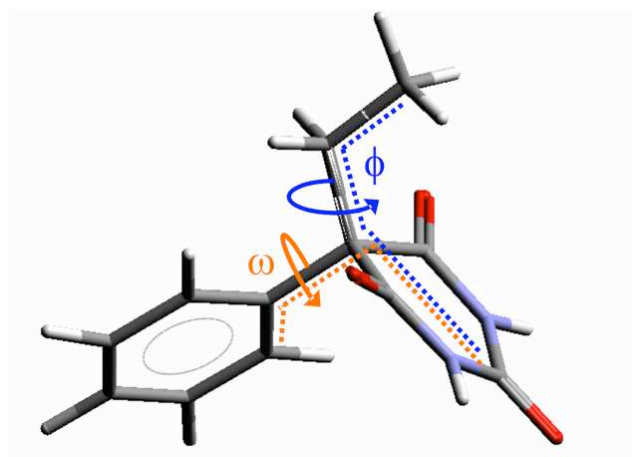
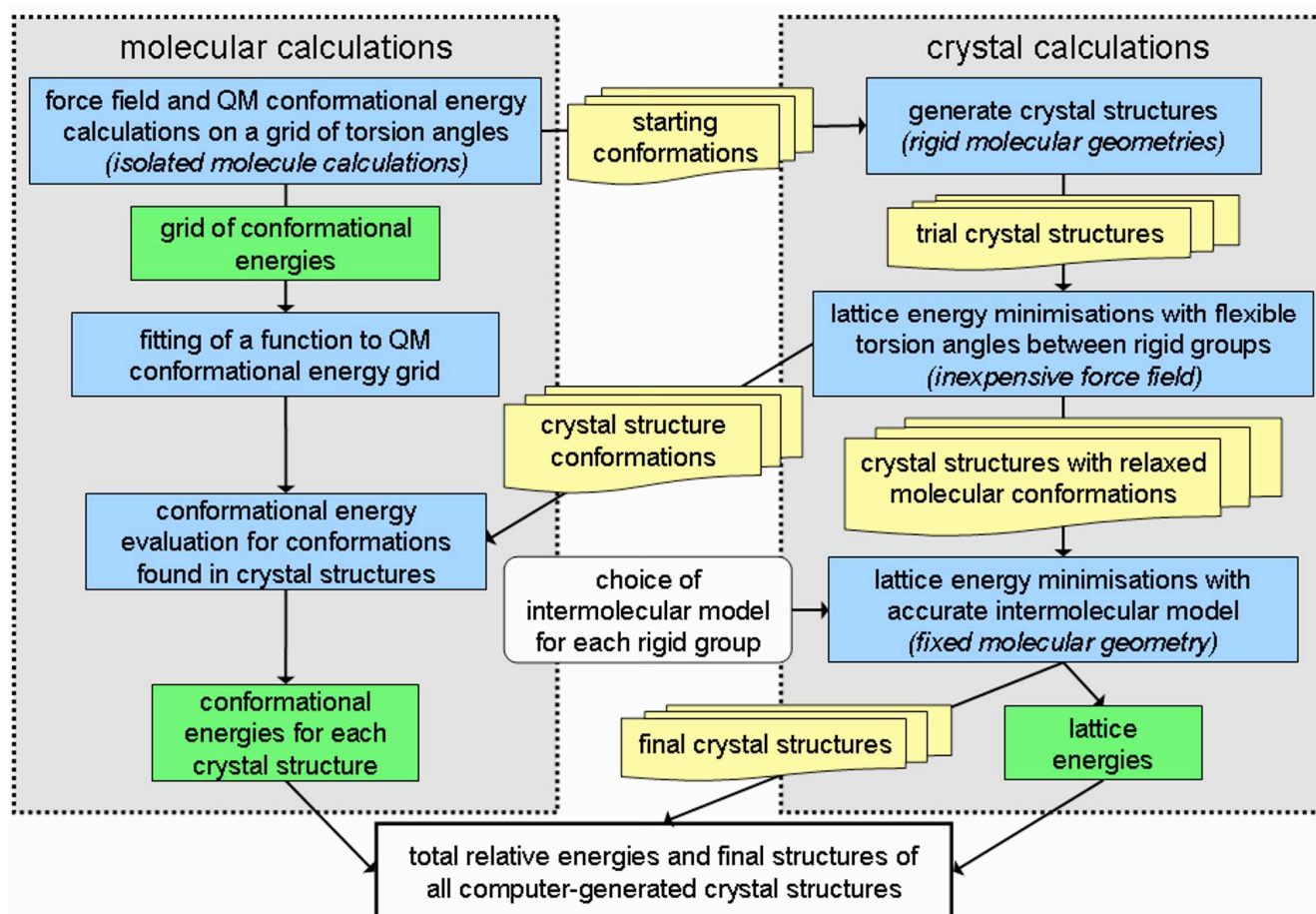


Figure 1. The molecular structure of phenobarbital, with flexible torsion angle definitions.

<sup>a</sup> Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, United Kingdom. Fax: +44 01223 336362; Tel: +44 01223 336390; E-mail: gmd27@cam.ac.uk

<sup>b</sup> Cambridge Crystallographic Data Center, Union Road, Cambridge, United Kingdom.

† Electronic Supplementary Information (ESI) available: CIFs of all low energy computer-generated crystal structures of phenobarbital and of the most likely crystal structures of phenobarbital form V See <http://dx.doi.org/10.1039/b000000x/>



**Figure 2.** Flowchart illustrating the methodology applied to crystal structure prediction of phenobarbital. Calculations are represented as blue boxes, transfer of structures as yellow boxes and calculated energies in green boxes.

and balanced manner. The inadequacy of current methods used to assess the relative stabilities of putative crystal structures of flexible molecules is clear from the results of the three blind tests of crystal structure prediction, in which there has been only one successful prediction of a crystal structure of a flexible molecule [1-3]. This occasional success of current force field methods is not satisfactory and the reliability of such calculations must be improved. Here, we investigate the known and hypothetical crystal structures of phenobarbital (5-ethyl-5-phenylbarbituric acid, Figure 1), a moderately flexible anticonvulsant drug molecule. An earlier computational study of the similar molecule primidone [13] demonstrated the limitations of standard force fields for crystal structure prediction. Even though the two known polymorphs could be located by Monte Carlo sampling of the energy surface, with the study limited to an incomplete sampling of the two space groups of the known polymorphs and using the molecular conformations from the X-ray determined crystal structures, the true crystal structures were ranked higher in energy than several other computer-generated crystal structures.

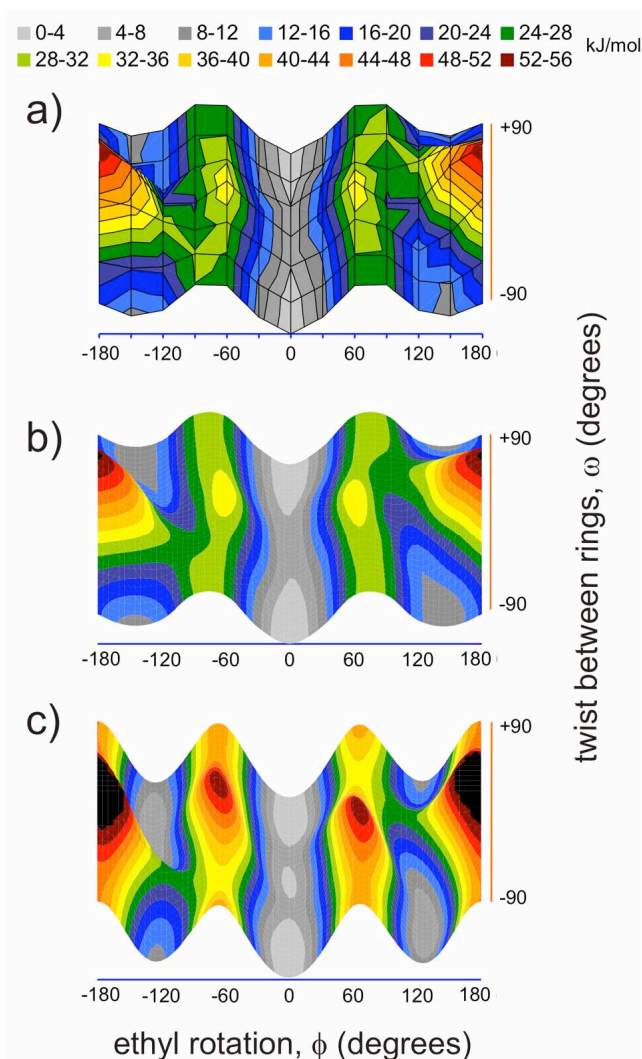
The polymorphism of phenobarbital has been well-studied and is of practical importance because of the polymorphic dependence of its stability [14], dissolution [15] and tableting [16] properties. At least five pure anhydrous polymorphs exist (although as many as twelve anhydrous forms have been

reported [17]), as well as a monohydrate. The crystal structure of form III, in the space group  $P2_1/c$  with one molecule in the asymmetric unit ( $Z'=1$ ), was first determined by Williams in 1974 [18], who also reported the structure of the monohydrate and unit cell dimensions and space groups of forms I ( $P2_1/c$ ,  $Z'=3$ ), II ( $P\bar{1}$ ,  $Z'=3$ ) and V ( $P2_1/c$ ,  $Z'=2$ ) [19]. Form IV was too unstable to characterise, so no structural information has been reported. The complete structures of the two  $Z'=3$  polymorphs (I and II) have been recently determined by Platteau *et al* [20] from powder X-ray diffraction data. To our knowledge, no other crystal structures of phenobarbital have been reported.

Our main motivation for studying this molecule was to develop a strategy for modelling the relative energies of the real and hypothetical crystal structures of flexible molecules. The presence of multiple independent molecules in the asymmetric unit – salts, solvates, cocrystals and  $Z'>1$  crystals – represents another challenge for crystal structure prediction because of the increased dimensionality of the lattice energy surface. Therefore, we also present results of  $Z'=2$  and  $Z'=3$  crystal structure searches for phenobarbital to highlight the problems encountered in searching the phase space for such systems.

Phenobarbital is a relatively simple molecule and many molecules of interest will have much greater conformational freedom. Therefore, in approaching the CSP for this molecule,

we have kept in mind that the scaling of computing costs with molecular complexity is an important consideration in developing a methodology that will be of general use in pharmaceutical materials applications, as well as other areas of crystal engineering.



**Figure 3.** Conformational energy surfaces of phenobarbital, calculated a) on a 30° grid with DFT, b) interpolated from the DFT grid and c) calculated with the CVFF force field.

## Methods

The overall approach to predicting the crystal structures of phenobarbital is summarised in Figure 2. We first use a minimally expensive computational model for generating trial crystal structures, although this has to be good enough to generate the set of important crystal structures, with the structures and energies then refined with higher level calculations. Molecular conformations were fixed after the initial search and energy minimisations, after which conformational and intermolecular energies were calculated separately using higher levels of theory.

### Molecular energy as a function of conformation

The conformational energy of the isolated phenobarbital molecule was evaluated using both atomistic (force field) and density functional theory (DFT) calculations. DFT calculations were performed using the DMol3 program [21], with the PW91 functional and double numerical polarised (DNP) basis set. Force field calculations were performed on an evenly spaced grid for the two torsion angles defined in Figure 1, at 10° intervals in the region  $\phi \in [0^\circ, +180^\circ]$ ,  $\omega \in [-90^\circ, +90^\circ]$ ; the  $\phi \in [-180^\circ, 0^\circ]$ ,  $\omega \in [-90^\circ, +90^\circ]$  regions were generated by symmetry. At each point, a harmonic restraint (force constant = 4000 kJmol<sup>-1</sup>radian<sup>-1</sup>) was placed on the two torsion angles and the remaining degrees of freedom (all other torsions, angles and bond lengths) were allowed to fully relax. The DFT energy scan was performed over the same range for the two angles, but these more expensive calculations were made on a coarser 30° grid, with torsion angles constrained at each point. This rough surface was then interpolated using a third-order piecewise polynomial, giving a smooth surface on which conformational energies could be evaluated at any ( $\phi, \omega$ ) pair.

Many force fields are available for organic molecules, differing in the functional form of the model for bonded and non-bonded interactions, as well as the approach taken to parameterise these functions. We did not intend to use the force field energies in our final evaluation of the computer-generated crystal structures. Instead, our main criterion for the choice of force field was that the molecular conformations in the crystal structures should be as reliable as possible, *ie.* that the conformations are as close as possible to what would be found using a higher quality model potential. Therefore, conformational energy scans were performed using several force fields and we looked for an overall qualitative agreement with the higher quality DFT calculated surface; in particular, the force field should produce as closely as possible the energy minima at the correct geometries. Since molecular conformations in crystals need not be minima for the isolated molecule, we also looked for a good representation of the shape of the basins around the minima, where we expect to find molecular conformations in the crystal structures. After testing several force fields that have been parametrised for small organic molecules, we chose the CVFF-950 force field [22], as implemented in the Accelrys Cerius2 modelling software [23]. CVFF was used in the crystal structure searches and initial energy minimisation of the computer-generated crystal structures.

In the final energy minimisations, the crystal structures were re-minimised using the W99 *exp-6* model potential [24-26] for intermolecular repulsion-dispersion interactions. Molecular geometries were fixed during this second energy minimisation, at the final conformation from the CVFF lattice energy minimisations, and only molecular positions, orientations and lattice parameters were optimised. For the models of intermolecular electrostatic interactions, atomic charges and multipoles were fitted using the program MOLDEN [27] to the molecular electrostatic potential of the lowest energy conformation, calculated at the B3LYP/6-31G\*\* level of theory using Gaussian03 [28]. Atomic multipoles up to quadrupole were defined with respect to a

local axis frame within the rigid unit to which that atom belongs – these rigid units are defined below.

### crystal structure searches

The computational search for low energy crystal structures was performed using the simulated annealing algorithm [29-32] implemented in the Cerius2 software suite [23]. The nine most commonly observed space groups ( $P2_1/c$ ,  $P\bar{1}$ ,  $P2_12_12_1$ ,  $P2_1$ ,  $C2/c$ ,  $Pbca$ ,  $Pnma$ ,  $Pna2_1$ , and  $Pbcn$ ) were searched with one molecule in the asymmetric unit, while the two most common space groups with  $Z' > 1$ ,  $P2_1/c$  and  $P\bar{1}$ , were also searched with both  $Z' = 2$  and  $Z' = 3$ .

Rigid molecular geometries were used in the simulated annealing sampling, to generate loosely packed starting crystal structures. Searches were performed with five molecular conformations, with  $\phi$  and  $\omega$  chosen from the stationary points (local minima and saddle points) on the CVFF conformational energy surface (see Results section). The simulated annealing sampling was repeated 3 times in each space group for each of the five starting conformations. Three rigid units were defined within the molecule (the  $C_2H_5$  ethyl group,  $C_6H_5$  phenyl group and the  $C_3O_3N_2H_2$  pyrimidine ring) and the internal geometries of these groups were kept fixed, with bond lengths, angles and torsions taken from the global minimum DFT optimised molecular geometry. Following the rigid molecule sampling of phase space, the trial crystal structures (291897 with  $Z'=1$ , 134314 with  $Z'=2$  and 196524 with  $Z'=3$ ) were energy minimised using the CVFF force field, fixing the geometries of the three rigid groups, and allowing the relative orientations of these groups to adjust to the crystal environment. The purpose of fixing the three rigid parts of the molecule was to ensure that the geometries of these groups were identical in all of the final crystal structures; this improved the transferability between conformations of atomic charges and higher multipoles in the final lattice energy calculations.

All structures within 25 kJ/mol of the global minimum crystal structure from the CVFF minimisations were then reminimised with the W99 model potential and fixed

molecular geometries, using the crystal structure modelling program DMAREL [33, 34]. We evaluated the intramolecular contribution to the total crystal energies from the interpolant of the DFT calculated conformational energy surface (Figure 3b).

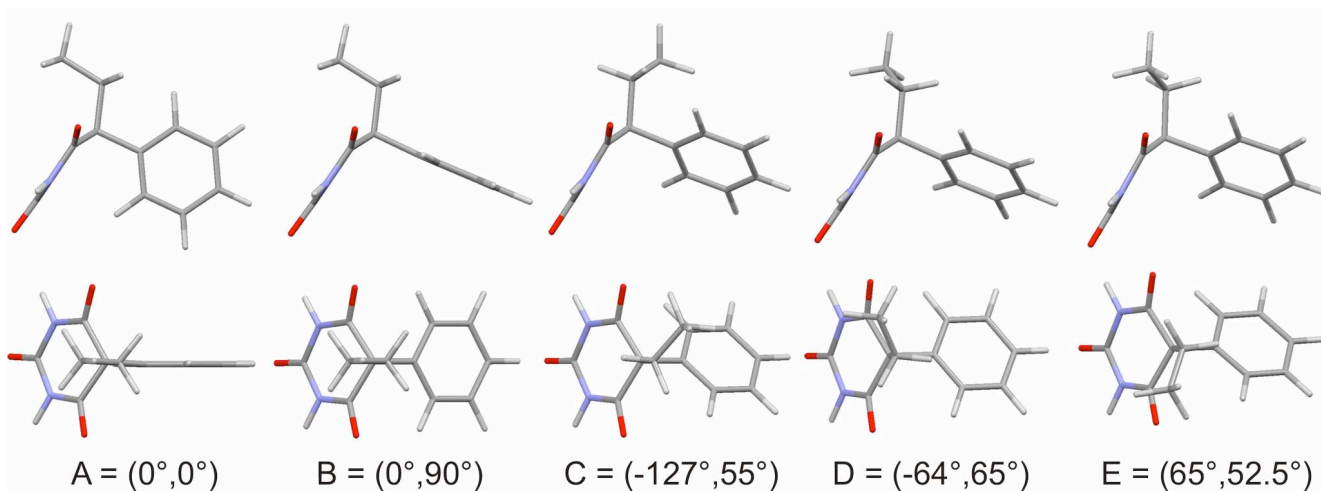
## Results and Discussion

### Conformational energy surface

The phenobarbital conformational energy surfaces, calculated with both the CVFF force field and by DFT, are qualitatively very similar (Figure 3). There is one long “valley” centred at  $\phi = 0^\circ$ , corresponding to conformations with the ethyl group aligned with the symmetry plane of the pyrimidine ring. The molecule has minimum energy at  $(\omega, \phi) = (0^\circ, 90^\circ)$ , although the phenyl group is relatively free to rotate with little energetic cost. The other two minima are symmetry related; molecules at  $(\omega, \phi)$  and  $(-\omega, -\phi)$  are mirror images. The minimum centred at  $(-127^\circ, 55^\circ)$  on the CVFF surface ( $(-150^\circ, 75^\circ)$  on the DFT surface) has the ethyl group pointing away from the pyrimidine ring, but at an angle with the phenyl ring to avoid unfavourable steric interactions. This minimum is 4 kJ/mol less stable than the  $(0^\circ, 90^\circ)$  conformation on the CVFF energy surface, or 10 kJ/mol from the DFT calculations.

### $Z'=1$ results

Structure searches were performed using five different initial molecular conformations (Figure 4), which were chosen to sample the entire relevant conformational energy surface during the generation of trial crystal structures. Three of these conformations were chosen as minima on the CVFF energy surface – two in the long valley with  $\phi = 0^\circ$ , and one at the higher energy local minimum; we labelled these as A  $(0^\circ, 0^\circ)$ , B  $(0^\circ, 90^\circ)$  and C  $(-127^\circ, 55^\circ)$ . Note that, for the purposes of crystal structure prediction with  $Z'=1$ , crystal structures with conformations of  $(\omega, \phi)$  and  $(-\omega, -\phi)$  are considered the same. Either both enantiomers are present, with the other generated by symmetry or, for chiral space groups, the absolute



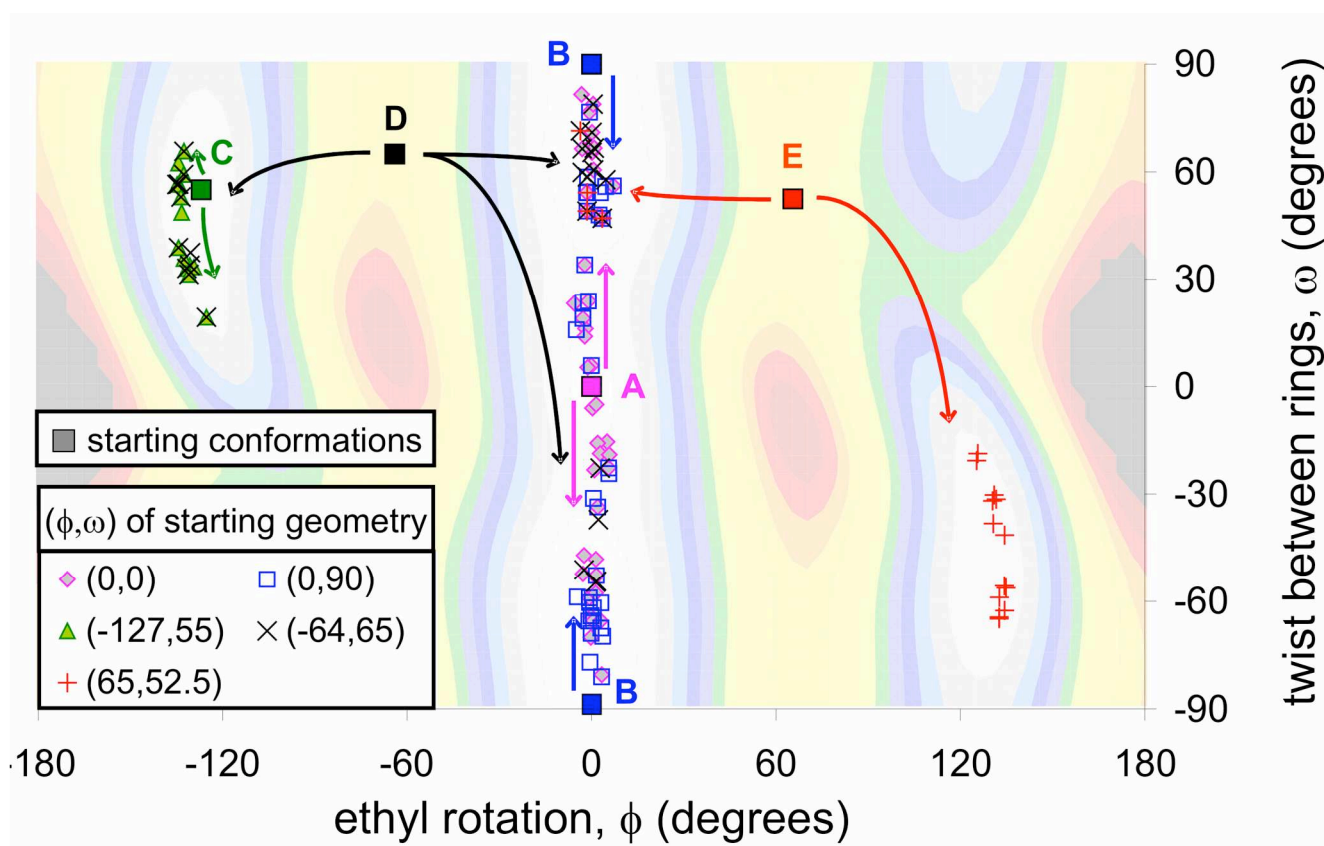
**Figure 4.** Starting molecular geometries A-E used in the crystal structure generation, indicating  $(\phi, \omega)$  values. Views are taken perpendicular to (top) and in the plane of (bottom) the phenyl ring in the  $(0^\circ, 0^\circ)$  conformation.

265 configuration of the crystal is unimportant for our purposes; the energy of the crystal with  $(\omega, \phi)$  is identical to that with  $(-\omega, -\phi)$ . We also performed crystal structure searches with initial molecular conformations at saddle points on the CVFF energy surface, *i.e.* on the ridge separating the two minima: 270 one at  $(-64^\circ, 65^\circ)$ , labelled D, and another at  $(65^\circ, 52.5^\circ)$ , labelled E. To establish how efficiently packing space was sampled with each of these starting molecular models, we monitored the distribution of final conformations in the lowest energy minimised crystal structures; those within 20 kJ/mol in 275 total energy of the global minimum crystal structure from each starting conformation (Figure 5).

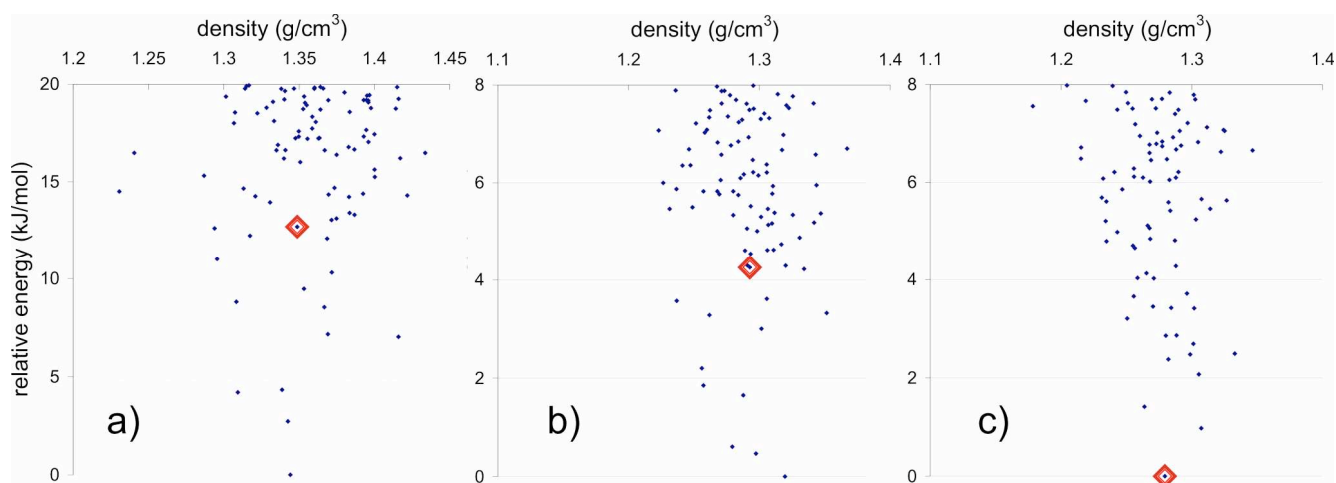
Both starting geometries in the main valley (A and B) resulted exclusively in energy minimised crystal structures with the molecular geometries within this same valley,  $\phi \sim 0^\circ$ , 280 with  $\omega$  values distributed over the entire range (red diamonds and blue squares in Figure 5). The calculations show that the relative orientation of the pyrimidine and phenyl rings can move freely during energy minimisation of the crystal structures, allowing  $\omega$  to find its optimum value from either of 285 the initial molecular conformations. Rotations about the single bonds are more restricted in the second minimum and structures generated from molecular model C remain localised near the starting conformation (black diamonds in Figure 5). Those searches that were started from saddle point molecular

290 geometries resulted in energy minimised crystal structures in each of the conformational energy basins. In the 20 kJ/mol range of total (lattice + conformational) energies, no crystal structures contain molecular conformations higher than 4 kJ/mol above the nearest conformational energy minimum (inside the first contour on the conformational energy surface, Figure 5). Both saddle point starting molecular models (D & E) gave a similar sampling of crystal structures near the higher energy local minimum conformation (maroon x's and turquoise crosses, respectively, in Figure 5), as well as leading 295 to crystal structures with molecular conformations along the  $\phi \sim 0^\circ$  valley. Searches using starting point D sampled this low energy region of conformation space more effectively than those using E, the former leading to final structures covering the entire range of  $\omega$ , while starting point E only led to 300 structures with  $\omega$  between  $45^\circ$  and  $75^\circ$ .

The searches started from saddle point molecular geometries have thus been shown to sample all of the low energy regions of conformation space, so could, in theory, provide a complete sampling of all low energy crystal structures. The strategy of starting from saddle point molecular geometries could become useful for more complex molecules, with more degrees of freedom, where there may be 310 too many individual basins on the conformational energy surface to search each individually.



**Figure 5.** Distribution of molecular geometries in the low energy crystal structures (within 20 kJ/mol in total energy of the global minimum crystal structure). Only conformations from the search in space group  $P2_1/c$  are shown, to avoid overcrowding of the figure. Large filled squares are the geometries used to start the search. Each small open symbol and cross represents an energy minimised crystal structure, while the symbol indicate the starting geometry from which each final crystal structure was generated. The plot is overlaid on the CVFF conformational energy surface, with contour lines at 4 kJ/mol intervals, as in Figure 3.



**Figure 6.** Relative energy vs density plot of the lowest energy predicted structures of phenobarbital ( $Z'=1$  only), a) using the CVFF force field, b) after re-minimisation with the W99 + atomic charges model potential and DFT conformational energies and c) after re-minimisation with the final energy model: W99 + multipoles model potential and DFT conformational energies. Each point is a distinct lattice energy minimum and the known  $Z' = 1$  polymorph (form III, CSD refcode PHBARB) is highlighted by a red diamond. Note the different energy and density scales between a, b and c.

315 We felt that the searches gave a good coverage of conformational and packing space, within the crystal symmetries considered ( $Z'=1$ , common space groups), so we merged the sets of structures resulting from each starting conformation and removed any duplicates using the COMPACT algorithm [35], with a 15 molecule coordination sphere and 15% tolerance on inter-atomic distances. We also used the COMPACT algorithm to compare the list of computer generated crystal structures to the one known  $Z'=1$  crystal structure (Form III,  $P2_1/c$ ). A very close reproduction of the X-ray determined structure was found amongst the predictions – the same energy minimum that is obtained by energy minimising the X-ray structure with the partly-rigid molecular model, atomic charges and CVFF force field. It is noteworthy that form III was located in the searches, indicating that, although the observed molecular conformation ( $\phi = +0.4^\circ$ ,  $\omega = -52.3^\circ$ ) is *not* close to one of the starting molecular models used in the search, the choice of starting models and molecular flexibility allowed in the first stage of energy minimisations provided a sufficient sampling of conformational space.

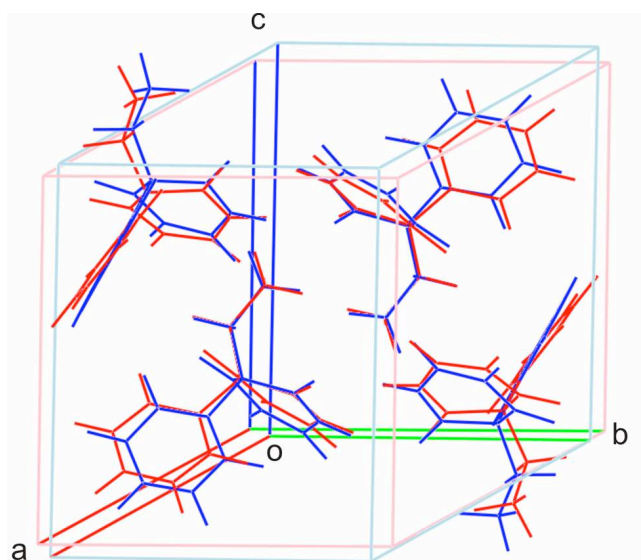
335 However, even though the form III crystal structure was located in the search, there were 14 computer-generated crystal structures with lower calculated energies, and form III was 12.7 kJ/mol above the global minimum energy crystal structure (Figure 6a). While the observed crystal structure need not necessarily correspond to the global minimum energy structure from CSP calculations, it is unlikely that 14 lower energy polymorphs exist and that form III is really so high in energy; in CSP of rigid molecules, the observed crystal structure is almost always found within 5 kJ/mol of the global minimum [9]. Thus, the CVFF force field does not seem to be adequate for crystal structure prediction of this molecule.

350 The crystal structures obtained with the CVFF force field were then re-minimised with the W99 model potential used for the repulsion-dispersion contribution to intermolecular interactions. The molecular geometry was kept fixed from the CVFF lattice energy minimisations, but we re-evaluated the

conformational energy from the values of  $\phi$  and  $\omega$  using the interpolated DFT calculated energy surface. We first paired the W99 potential with the simplest electrostatic model: atomic charges on each atomic site. The relative energies of the computer-generated crystal structures were changed significantly from those calculated with the CVFF force field (Figure 6b), the main effect being to condense the set of crystal structures into a smaller energy range, which is likely due to the flatter DFT conformational energy surface compared to that of the force field (Figure 3). Thus, the calculated difference in energy between form III and the global minimum from the search is now decreased to 4.25 kJ/mol, with 12 crystal structures calculated to be more stable than the known structure – a more reasonable relative energy, but little improvement in the absolute ranking of the known structure. The position of the observed crystal structure amongst the predictions is now in accord with typical crystal structure prediction results for rigid molecules using the same simple intermolecular model potential [10].

370 For smaller rigid molecules, it has been demonstrated that an important improvement in the reliability of crystal structure prediction is achieved through the use of a more elaborate representation of electrostatic interactions [9, 36]. We therefore further refined our model potential, using a distributed multipole description of electrostatics, as has been advocated by Price and coworkers for accurate molecular crystal structure modelling [37, 38]. Such elaborate electrostatic models have been shown to improve the modelling of rigid molecules, but their combination with traditional force fields for flexible molecules has been problematic when the atomic multipole electrostatic model is used to model both inter- and intramolecular interactions [39]. In this work, by calculating conformational energies from the DFT surface, the multipole-based electrostatic model only needs to accurately model intermolecular interactions.

390 Studies on rigid molecules have shown that the form of the electrostatic model is most important for polar hydrogen bonding molecules, while the effect for non-polar molecules, such as hydrocarbons, is very small [9]. Learning from these



**Figure 7.** Overlay of the observed form III (blue) and global minimum predicted crystal structure (W99 + multipoles model, red).

observations for rigid molecule systems, we treated the three rigid units within phenobarbital with different levels of theory: the electrostatic model was refitted to the molecular electrostatic potential, with an atomic multipole model up to quadrupoles for all atoms in the pyrimidine ring, while the simpler atomic charge model was used for the two hydrocarbon moieties ( $C_6H_5$  and  $C_2H_5$ ).<sup>‡</sup> Reminiscing the entire set of crystal structures with the W99 model potential and this electrostatic model has an important effect on the energy ranking of the crystal structures. In particular, form III is now ranked as the global minimum crystal structure, approximately 1 kJ/mol more stable than any of the other computer generated crystal structures (Figure 6c).

This improvement in the ranking of the experimentally observed crystal structure gives us confidence that this final approach to calculating relative total energies – the

combination of DFT conformational energies with the W99 + atomic multipoles intermolecular energies – is promising for the crystal structure prediction of flexible molecules. The structural deviation between the predicted (ie. energy minimised) and observed crystal structures is another measure of the quality of the energy model – for form III, the RMS error in the lengths of the lattice vectors (a,b,c) is 4.8% (Table 1, Figure 7). This is not as good as the typical errors for crystal structures of small rigid molecules, but reflects the extra source of error in the molecular geometry. There is a noteworthy discrepancy between the calculated and observed orientation of the phenyl group relative to the pyrimidine ring: the predicted conformation has  $\phi = 0.6^\circ$ ,  $\omega = -65.5^\circ$ , compared to the observed values of  $\phi = 0.4^\circ$  and  $\omega = -52.3^\circ$ .

### The challenge of $Z' > 1$ polymorphism

Phenobarbital is polymorphic, with several polymorphs known to have multiple independent molecules in the unit cell ( $Z' > 1$ ). The known structures of forms I and II both have three independent molecules in the asymmetric unit ( $Z'=3$ ), so could not have been present in the initial crystal structure searches, which only considered  $Z'=1$ . Similarly, from the unit cell dimensions and space group, we know that form V has two independent molecules ( $Z'=2$ ). We therefore attempted predictions of  $Z'=2$  and  $Z'=3$  crystal structures, with three aims:

- to test whether the  $Z'=3$  polymorphs (forms I and II) would have been predictable;
- to see if low energy  $Z'=2$  crystal structures could be found, with unit cell dimensions close to those reported by Williams [19] from powder XRD of form V and
- to see if there are other low energy  $Z'>1$  polymorphs comparable in energy with the  $Z'=1$  predictions, as possibilities for form IV and further polymorphism.

The prediction of crystal structures with more than one independent molecule is very important. Cocrystals, solvates

**Table 1.** Relative energies, unit cell parameters and molecular conformations of the observed and energy minimised (using the W99 + multipoles model) polymorphs of phenobarbital.

Crystal form	symmetry		unit cell parameters						molecular conformation		density / $gcm^{-3}$	energy (relative energy) / $kJmol^{-1}$
			a / Å	b / Å	c / Å	$\alpha / ^\circ$	$\beta / ^\circ$	$\gamma / ^\circ$	$\phi / ^\circ$	$\omega / ^\circ$		
I	$P2_1/c, Z'=3$	obs	10.691	47.125	6.800	90	94.18	90	-2.9 -0.9 -2.3	+49.0 -39.5 -33.1	1.354	-
		min.	10.978 (+2.68%)	47.718 (+1.26%)	6.833 (+0.49%)	90	93.30	90	-0.8 -1.7 +2.4	+41.9 -54.5 -39.1	1.295 (-4.36%)	-93.22 (+3.39)
II	$P1, Z'=3$	obs	10.731	23.511	6.783	90.97	94.48	88.15	+0.1 -13.8 +7.3	+44.2 -42.6 -41.2	1.357	-
		min	11.049 (+2.96%)	23.691 (+0.77%)	6.850 (+0.99%)	90.94	93.20	88.01	+0.3 -3.0 -1.5	+47.0 -53.6 -49.5	1.293 (-4.72%)	-93.20 (+3.41)
III	$P2_1/c, Z'=1$	obs	9.534	11.855	10.794	90	111.56	90	+0.4	-52.3	1.360	-
		min	10.112 (+6.06%)	12.126 (+2.29%)	10.241 (-5.12%)	90	109.04	90	+0.6	-65.5	1.300 (-4.41%)	-96.61 (0.0)
V	$P2_1/c, Z'=2$	obs	12.66	6.75	27.69	90	106.9	90			1.384	-

<sup>a</sup> The lattice energy is the intermolecular energy, calculated using the W99 + atomic multipoles model, plus the DFT molecular energy relative to the lowest energy conformation..

445 and salts all fall into this category, as well as single-  
 component crystals with  $Z' > 1$ ; over 10% of single-component  
 crystal structures in the Cambridge Crystallographic Database  
 have  $Z' > 1$ . Despite their importance, only a few studies have  
 reported crystal structure prediction for structures with 2  
 450 independent molecules [2, 40-43] and, to our knowledge, only  
 one  $Z'=3$  crystal structure has so far been predicted [44].

To test whether our method for modelling the energies of  
 phenobarbital crystals extends to the known  $Z'=3$  polymorphs,  
 we energy minimised the two crystal structures using the same  
 455 methodology as outlined above for the  $Z'=1$  CSP.<sup>§</sup> The  
 structures of forms I and II are reproduced very well, with  
 RMS errors in cell lengths of 1.73% and 1.86% respectively  
 (Table 1) and the calculated energies are 3.39 and 3.41 kJ/mol  
 higher than form III. These energy differences are well within  
 460 the reasonable range for polymorphism<sup>\*\*</sup> and the small energy  
 difference between forms I and II is not surprising considering  
 their structural similarities [20] – the  $P2_1/c$  form I is nearly a  
 cell-doubled version of form II ( $P\bar{1}$ ).

The methodology we used for our  $Z'=2$  and  $Z'=3$   
 465 predictions was essentially the same as for the  $Z'=1$  case,  
 except that all possibilities for combinations of the molecular  
 conformations had to be considered. The five starting models  
 used in the  $Z'=1$  search seemed to cover conformational space,  
 so for  $Z'=2$  we started searches with all possible combinations  
 470 of two of these conformations – the full list of 21  
 combinations is given as supplementary information.<sup>††</sup> The  
 computational cost of starting  $Z'=3$  searches with all 65  
 possible combinations of three of the conformations was too  
 great. Therefore, we limited ourselves to the 13 combinations  
 475 of 3 conformational minima (A, B & C) and the 10  
 combinations of the saddle point conformations (D & E),  
 repeating some combinations twice to give a total of 31  
 separate  $Z'=3$  runs of the Monte Carlo simulated annealing  
 search. Because of the unreliability of the CVFF energies, we

480 considered all structures up to a high relative energy (25  
 kJ/mol above the global minimum) for refinement with the  
 more reliable energy model, reminimising all crystal  
 structures in this range using the W99 + multipoles model  
 potential.

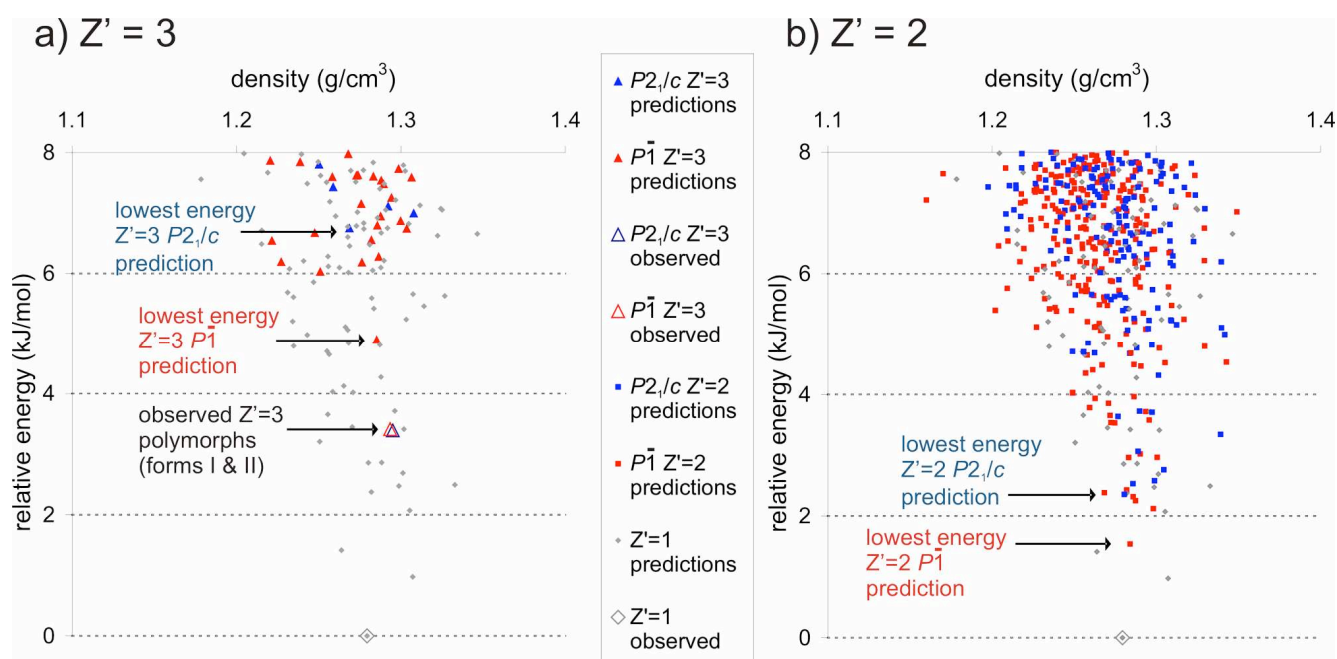
### $Z'=3$

The sampling with  $Z'=3$  produced very few crystal structures  
 comparable in energy to the lowest energy  $Z'=1$  predictions.  
 The lowest energy  $P\bar{1}$  and  $P2_1/c$  structures generated by the  
 490 search were 4.9 and 6.8 kJ/mol above the global minimum,  
 respectively (Figure 8a). These are significantly higher in  
 energy than the energy minimised forms I and II. Neither  
 $Z'=3$  polymorph was generated by the search method,  
 showing that the sampling of conformational and packing  
 495 space was incomplete.

The molecular conformations in forms I and II are all quite  
 close to  $(0^\circ, 45^\circ)$ , which is far from any of the starting  
 conformations in the searches. To test if the searches might  
 have been successful with starting molecular conformations  
 500 closer to what is observed, an additional 3 simulated  
 annealing repeats were performed in each space group,  
 starting with the observed torsion angles:  $(0.1^\circ, 44.2^\circ)$ ;  $(-13.8^\circ, -42.6^\circ)$ ;  $(7.3^\circ, -41.2^\circ)$  for the three molecules in  $P\bar{1}$   
 and  $(-2.9^\circ, 49.0^\circ)$ ;  $(-0.9^\circ, -39.5^\circ)$ ;  $(-2.3^\circ, -33.1^\circ)$  in  $P2_1/c$ .  
 505 These additional searches failed to produce any lower energy  
 structures than the initial set of searches, so failed to locate  
 the known polymorphs, highlighting the difficulty of sampling  
 all of packing space with three independent molecules.

### $Z'=2$

In contrast, the  $Z'=2$  searches did produce many crystal  
 structures with comparable energies to the  $Z'=1$  predictions  
 (Figure 8). In fact, many of the  $Z'=1$  structures, including the  
 global minimum form III, were generated during the  $Z'=2$



**Figure 8.** Calculated (W99 + multipoles + DFT) relative energies and densities of a) the predicted and observed  $Z'=3$  and b) predicted  $Z'=2$  crystal structures of phenobarbital.

515 searches; the extra symmetry elements were located after energy minimisation using the program PLATON.[45] These observation gave us confidence that the searches with two independent molecules provided a more complete sampling of packing space than we had for  $Z'=3$ .

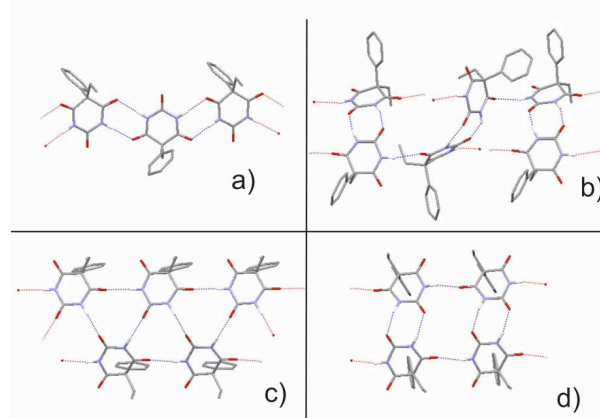
520 The list of  $Z'=2$   $P2_1/c$  predicted structures was searched for any obvious candidates for form V, which is expected to be comparable in energy to forms I, II and III, *i.e.* within about 5-6 kJ/mol of the global minimum. To compare the unit cells of the predicted crystal structures to those reported by  
525 Williams [19], we calculated the reduced unit cells of all predicted structures and looked for structures with similar dimensions and unique axis **a** (the reduced cell of Williams' reported unit cell is  $a = 6.75 \text{ \AA}$ ,  $b = 12.66 \text{ \AA}$ ,  $c = 26.89 \text{ \AA}$ ,  $\alpha = 99.9^\circ$ ). The 6 lowest energy structures with root-mean-squared  
530 % deviations in the lengths of **a**, **b** and **c** of less than 7% are presented in Table 2.

None of these  $Z'=2$  structures have the same hydrogen bonding as seen in forms I, II and III, where the primary interactions form ribbons of hydrogen bonded dimers (Figure  
535 9a). The structures at 4.77, 8.12 and 8.15 kJ/mol above the global minimum are formed of  $R_2^2(8)$  dimers; these are linked into a 3-dimensional network (Figure 9b) in 4.77 kJ/mol structure, and into 1-dimensional tapes (Figure 9d) in the other two. The three others (at 5.16, 6.70 and 7.71 kJ/mol  
540 above the global minimum) all contain ribbons of trimers (Figure 9c). This variety of structures with very similar unit cells shows the wide choice of hydrogen bonding available for this molecule. In fact, there are known crystal structures of other 5,5-disubstituted barbituric acids with each of the motifs  
545 observed in the phenobarbital predictions: 5-(1-cyclohepten-2-yl)-5-ethylbarbituric acid, 5-ethyl-5-(1-methylbutenyl)-barbituric acid and polymorph IV of 5,5-diethylbarbituric acid contain linked dimers as in Figure 9b; 5-ethyl-5-(3,3-dimethyl-n-butyl)barbituric acid hydrogen bonds as in Figure  
550 9c; polymorph I of unsubstituted barbituric acid, 5-ethylbarbituric acid, form I of 5,5-diethylbarbituric acid and 5-methyl-5-phenylbarbituric acid form hydrogen bonds as in Figure 9d.<sup>‡‡</sup>

**Table 2.** Calculated energies and reduced unit cells of  $Z'=2$   $P2_1/c$  predicted crystal structures with similar unit cells to those reported for form V.

$\Delta E / \text{kJmol}^{-1}$	$a / \text{ \AA}$	$b / \text{ \AA}$	$c / \text{ \AA}$	$\alpha / ^\circ$	RMS deviation (a,b,c) / %
+4.77	6.743 (-0.1%)	12.898 (+1.9%)	27.553 (+2.5%)	91.92	1.79
+5.16	7.028 (+4.1%)	13.285 (+4.9%)	25.448 (-5.4%)	98.34	4.84
+6.70	6.974 (+3.3%)	13.875 (+9.6%)	26.095 (-3.0%)	99.36	6.11
+7.71	7.009 (+3.8%)	12.561 (-0.8%)	28.120 (+4.6%)	90.63	3.47
+8.12	7.010 (+3.9%)	12.134 (-4.2%)	29.724 (+10.5%)	100.78	6.90
+8.15	6.877 (+1.9%)	12.433 (-1.8%)	28.638 (+6.5%)	90.03	4.04
Reduced cell of reported form V [19]					
-	6.75	12.66	26.89	99.9	-

<sup>‡‡</sup> CIF files of all of these predicted crystal structures are available as supplementary information.



**Figure 9.** Hydrogen bonding motifs in the observed (a) and predicted (b, c, d) crystal structures of phenobarbital. Dashed lines indicate hydrogen bonds. Hydrogen atoms not involved in hydrogen bonding are omitted from the diagrams.

### 555 Form IV and other possible polymorphs

There are many other possible structures within a reasonable energy range, so the abundant polymorphism of phenobarbital is unsurprising.<sup>§§</sup> The lowest energy crystal structure not corresponding to one of the observed phases is a  $Z'=1$   $C2/c$  structure, 0.98 kJ/mol higher in energy than form III, followed by  $Pbca$  ( $Z'=1$ ) and  $P\bar{1}$  ( $Z'=2$ ) structures at 1.41 and 1.54 kJ/mol (the first two of these show the same hydrogen bonding as in forms I, II and III, Figure 9a, while the structure at 1.54 kJ/mol has hydrogen bonding as in Figure 9b). The  
560 other energetically competitive computer-generated structures show the full range of hydrogen bonding patterns (Figure 9); these are all possibilities for the elusive form IV, which has been observed but not characterised [19], as well as any other (as yet) undiscovered forms. All those crystal structures  
570 within 5 kJ/mol of the global minimum – 72 distinct crystal structures – are supplied as supplementary information.

### Computational cost

All crystal structure searches here were performed on a modest 600 MHz SGI Octane. As a comparison of the relative  
575 costs of the  $Z'=1$ , 2 and 3 searches, the  $Z'=1$  searches in  $P\bar{1}$  and  $P2_1/c$  were complete after the 15 Monte Carlo simulated annealing repeats, averaging about 8 hours per repeat and taking a total of ~120 CPU hours. The  $Z'=2$  searches in these same space groups (21 Monte Carlo repeats, averaging 47  
580 hours each) required 990 CPU hours. The  $Z'=3$  searches (31 repeats, averaging 77 hours) took 2380 CPU hours and did not provide a complete sampling of the energy surface. The computational cost clearly increases steeply with increasing  $Z'$ .

### 585 Conclusions

We have presented an approach to modelling the total relative energies of the crystal structures of molecules with conformational flexibility, developed with the application of crystal structure prediction in mind. Typically, the relative  
590 energies of thousands of putative crystal structures must be

calculated in crystal structure prediction studies (in this study, over 620000 crystal structures were energy minimised), so the cost of the energy model must not be prohibitive.

Using the methodology presented here, which combines a force field search of the crystal packing energy surface with further refinements of the most promising low energy crystal structures, the  $Z'=1$  phenobarbital form III was located as the global minimum in total (lattice + conformational) energy. The stability ranking with the final energy model, which combined DFT calculated conformational energies and an atomic multipole based model for intermolecular interactions, was a clear improvement over the calculations using a traditional force field.

Other accurate methods of calculating the relative energies of molecular crystals, making even greater use of quantum mechanics calculations for conformational [46, 47] or total [48] energies, have been developed, but their greater computational demands restrict their use to a few candidate crystal structures in crystal structure prediction studies. Furthermore, the scaling of quantum mechanics calculations in proportion to the number of atoms in the unit cell places a limit on the complexity of system that can be studied with a reasonable computational cost.

By treating the flexible molecule as a set of linked rigid units, experience that has been gained from studies of many small molecule crystal structures can be transferred to larger molecules. Each rigid unit can be treated at the appropriate level of theory (eg. atomic charges for non-polar moieties, or atomic multipoles where polar interactions are important) and the success of such an approach is demonstrated here. The flexibility in the final energy model could be crucial to tackling more flexible molecules. Using atomic multipoles on the entire molecule can increase the computational expense of energy minimisations 10-fold [9], a feasible cost for small molecules, where energy minimisations are not time-consuming. However, with increasing molecular size and flexibility the cost of each lattice energy minimisation is increased and, with added conformational flexibility, the number of structures that must be considered (the number within a reasonable energy range) also seems to increase. Therefore, recognising where extra complexity in the energy model is required and where the simpler models are adequate will become vitally important as the size and flexibility of target molecules increases.

Furthermore, strategies for treating flexible molecules must be developed with other simulation methods in mind; lattice dynamical contributions to the relative free energies of crystal structures can be especially important for flexible molecules [7, 49], so static lattice energy minimisations might not be adequate for their reliable prediction. Approaches to modelling crystal energies should therefore be amenable to lattice dynamics or molecular dynamics simulations, where derivatives of the energy surface are required. The model used in this study has been developed with this requirement in mind: derivatives of both the intermolecular and conformational energy model are available.

As well as the influence of the model potential on ranking the crystal structures, the effect of starting molecular conformation on the sampling of crystal structures has also

been investigated. A set of stationary points on the conformational energy surface accesses the full range of important conformations in the energy minimised crystal structures. The approach seemed to give a good sampling of  $Z'=1$  crystal structures, but there were convergence problems in searches with more than one independent molecule. The results of the  $Z'=3$  searches for forms I and II highlight the need for other methods of generating all low energy crystal structures in these complex systems. Future studies of such complex systems, with multiple independent and conformationally flexible molecules, will require the use of algorithms that sample the energy surface more efficiently [50] or which are designed to take advantage of distributed computing resources [51-53].

The results of the calculations on phenobarbital show the great choice of low energy packing patterns available to this molecule and structures have been suggested for the as-yet uncharacterised forms IV, V and possible further polymorphism.

### Acknowledgements

We thank Dr Neil Feeder for valuable discussions and the Pfizer Institute for Pharmaceutical Materials Science for funding.

### Notes and references

- <sup>1</sup>  $Z'$  is the number of formula units in the asymmetric unit. For single component crystals, this is the number of independent molecules.
- <sup>1</sup> The electrostatic model was fitted to the global minimum energy molecular geometry from the DFT calculations – the conformational dependence of the multipoles is presented as supplementary information.
- <sup>1</sup> To be explicit: we first replaced the molecules in the PXRD determined structures by the partly-rigid molecular models, then energy minimised these using the CVFF force field, keeping the internal geometries of the three rigid units fixed but allowing molecular flexibility between these groups. Finally, the molecular geometries were fixed and the rigid molecule lattice energy minimisation was performed using the W99 model potential with atomic centred multipoles (pyrimidine ring) and charges (phenyl and ethyl groups); total energies were taken as a sum of the W99 intermolecular energy and the conformational energy taken from the DFT energy surface at  $(\phi, \omega)$  of the final molecular geometries.
- <sup>1</sup> A rule of thumb for the maximum energy difference between polymorphs is 10 kJ/mol.
- <sup>1</sup> For our searches with  $Z'>1$ , the mirror image versions of conformations C, D and E must sometimes be considered as distinct conformations from their original versions. For example, a  $Z'=2$  search with two  $(-127^\circ, +55^\circ)$  conformations is different from a search with one  $(-127^\circ, +55^\circ)$  and one  $(+127^\circ, -55^\circ)$ .
- <sup>1</sup> Cambridge Structural Database REFCODES of these crystal structures are: CHEBAR (5-(1-cyclohepten-2-yl)-5-ethylbarbituric acid); VINBAR (5-ethyl-5-(1-methylbutenyl)-barbituric acid); DETBAA3 (polymorph IV of 5,5-diethylbarbituric acid); EMBBAR20 (5-ethyl-5-(3,3-dimethyl-n-butyl)barbituric acid); BARBAC (polymorph I of unsubstituted barbituric acid); ETBARB (5-ethyl-barbituric acid); DETBAA11 (form I of 5,5-diethylbarbituric acid) and MPBRBL (5-methyl-5-phenylbarbituric acid).
- <sup>1</sup> Some of these energy minima might be separated by low energy barriers and would interconvert at temperatures above 0K. Dynamical simulations would be required to assess this and evaluate the number of distinct minima on the free energy surface.
1. W.D.S. Motherwell *et al.*, *Acta Cryst.*, 2002. **B58**, 647-661.
  2. G.M. Day *et al.*, *Acta Cryst.*, 2005. **B61**, 511-527.
  3. J.P.M. Lommerse *et al.*, *Acta Cryst.*, 2000. **B56**, 697 - 714.

4. G.M. Day, *Current Organic Chemistry*, 2006, *accepted for publication*.
5. T. Beyer, T. Lewis and S.L. Price, *CrystEngComm*, 2001, **3**, 178-212.
6. A.T. Anghel, G.M. Day and S.L. Price, *CrystEngComm*, 2002, **4**, 348-355.
- 715 7. B.P. van Eijck, *J. Comput. Chem.*, 2001, **22**, 816-826.
8. T.C. Lewis, D.A. Tocher, G.M. Day and S.L. Price, *CrystEngComm*, 2003, **5**, 3-9.
9. G.M. Day, W.D.S. Motherwell and W. Jones, *Crystal Growth & Design*, 2005, **5**, 1023-1033.
- 720 10. G.M. Day *et al.*, *Crystal Growth & Design*, 2004, **4**, 1327-1340.
11. S.R. Chemburkar *et al.*, *Organic Process Research & Development*, 2000, **4**, 413-417.
12. A.J. Cruz Cabeza, G.M. Day, W.D.S. Motherwell and W. Jones, *Crystal Growth & Design*, 2006, **6**, 1858-1866.
- 725 13. R.S. Payne, R.J. Roberts, R.C. Rowe and R. Docherty, *Int. J. Pharm.*, 1999, **177**, 231 - 245.
14. M. Otsuka, M. Onoe and Y. Matsuda, *Pharm. Res.*, 1993, **10**, 577-582.
- 730 15. J.A. Clements and D. Stanski, *Canadian Journal of Pharmaceutical Sciences*, 1971, **6**, 9.
16. M. Otsuka, M. Nakanishi and Y. Matsuda, *Drug Development and Industrial Pharmacy*, 1999, **25**, 205-215.
17. R.J. Mesley, R.L. Clements, B. Flaherty and K. Goodhead, *J. Pharm. Pharmacol.*, 1968, **20**, 329-340.
- 735 18. P.P. Williams, *Acta Cryst.*, 1974, **B30**, 12-17.
19. P.P. Williams, *Acta Cryst.*, 1973, **B29**, 1572-1579.
20. C. Platteau *et al.*, *Acta Cryst.*, 2005, **B61**, 80-88.
21. B. Delley, *J. Chem. Phys.*, 1990, **92**, 508-517.
- 740 22. P. Dauber-Osguthorpe *et al.*, *Proteins: Structure, Function and Genetics*, 1988, **4**, 31-47.
23. Cerius2, version 4.6, 2001, Accelrys Inc.: San Diego, USA.
24. D.E. Williams, *J. Mol. Struct.*, 1999, **486**, 321 - 347.
25. D.E. Williams, *J. Comput. Chem.*, 2001, **22**, 1154-1166.
- 745 26. D.E. Williams, *J. Comput. Chem.*, 2001, **22**, 1-20.
27. G. Schaftenaar and J.H. Noordik, *J. Comput.-Aided Mol. Design*, 2000, **14**, 123-134.
28. M.J. Frisch *et al.*, *GAUSSIAN 03*. 2004, Gaussian Inc.: Wallingford, CT.
- 750 29. H.R. Karfunkel and R.J. Gdanitz, *J. Comput. Chem.*, 1992, **13**, 1171.
30. R.J. Gdanitz, *Chem. Phys. Letters*, 1992, **190**, 391.
31. H.R. Karfunkel, F.J.J. Leusen and R.J. Gdanitz, *J. Comput.-Aided Mat. Design*, 1994, **1**, 177.
32. P. Verwer and F.J.J. Leusen, *Rev. Comput. Chem.*, 1998, **12**, 327-365.
- 755 33. D.J. Willock, S.L. Price, M. Leslie and C.R.A. Catlow, *J. Comput. Chem.*, 1995, **16**, 628 - 647.
34. DMAREL 3.1, S.L. Price, D.J. Willock, M. Leslie and G.M. Day, <http://www.ucl.ac.uk/~ucca17p/dmarelmanual/dmarel.html>, 2001.
- 760 35. J. Chisholm and S. Motherwell, *J. Applied Cryst.*, 2005, **38**, 228-231.
36. D.S. Coombes, S.L. Price, D.J. Willock and M. Leslie, *J. Phys. Chem.*, 1996, **100**, 7352 - 7360.
37. S.L. Price, *J. Chem. Soc. -Faraday Transactions*, 1996, **92**, 2997-3008.
- 765 38. S.L. Price, *Toward More Accurate Model Intermolecular Potentials for Organic Molecules*, in *Reviews in Computational Chemistry*, K.B. Lipkowitz and D.B. Boyd, Editors. 2000, Wiley-VCH, John Wiley and Sons, Inc.: New York. p. 225 - 289.
39. S.W. Brodersen, F.J.J. Leusen and G. Engel, *Phys. Chem. Chem. Phys.*, 2003, **5**, 4923-4931.
- 770 40. B.P. van Eijck and J. Kroon, *Acta Cryst.*, 2000, **B56**, 535-542.
41. F.J.J. Leusen, *Crystal Growth & Design*, 2003, **3**, 189-192.
42. B.P. van Eijck, *J. Comput. Chem.*, 2002, **23**, 456-462.
43. P.G. Karamertzanis and S.L. Price, *J. Phys. Chem. B*, 2005, **109**, 17134-17150.
- 775 44. I.D.H. Oswald *et al.*, *Crystal Growth & Design*, 2004, **5**, 1055-1071.
45. A.L. Spek, *PLATON*. 1980-2004, Utrecht University, The Netherlands: <http://www.cryst.chem.uu.nl/platon>.
46. P.G. Karamertzanis and S.L. Price, *J. Chemical Theory and Computation*, 2006, **2**, 1184-1199.
- 780 47. B.P. van Eijck, W.T.M. Mooij and J. Kroon, *J. Comput.Chem.*, 2001, **22**, 805-815.
48. M.A. Neumann and M.-A. Perrin, *J. Phys. Chem. B*, 2005, **109**, 15531-15541.
- 785 49. A. Gavezzotti and G. Filippini, *J. Amer. Chem. Soc.*, 1995, **117**, 12299-12305.
50. J. Pillardy *et al.*, *Proc. Nat. Acad. Sci. USA*, 2001, **98**, 12351-12356.
51. V.E. Bazterra, M.B. Ferraro and J.C. Facelli, *J. Chem. Phys.*, 2002, **116**, 5984-5991.
- 790 52. V.E. Bazterra, F. M.B. and J.C. Facelli, *J. Chem. Phys.*, 2002, **116**, 5992-5995.
53. P.G. Karamertzanis and C.C. Pantelides, *J. Comput.Chem.*, 2004, **26**, 304-324.

795 **Graphical and text abstract**

A computational exploration of the low energy crystal structures of the pharmaceutical molecule phenobarbital is presented as a test for an approach for the crystal structure prediction of flexible molecules. Traditional transferable force field methods of modelling flexible molecules are unreliable for the level of accuracy required in crystal structure prediction and we outline a  
800 strategy for improving the evaluation of relative energies of large sets of crystal structures. The approach involves treating the molecule as a set of linked rigid units, whose conformational energy is expressed as a function of the relative orientations of the rigid groups. The conformational energy is calculated by electronic structure methods and intermolecular interactions using an  
805 atomic multipole description of electrostatics. A key consideration in the design of our approach is the scalability to more typical pharmaceutical molecules of higher molecular weight with many more atoms and degrees of flexibility.

