Acta Crystallographica Section A
Foundations and Advances
ISSN 2053-2733

Received 16 February 2014
Accepted 20 August 2014

Prediction of molecular crystal structures by a crystallographic QM/MM model with full space-group symmetry

Philipp Mörschel and Martin U. Schmidt*

Institut für Anorganische und Analytische Chemie, Goethe-Universität, Max-von-Laue-Strasse 7, D-60438 Frankfurt am Main, Germany. Correspondence e-mail: m.schmidt@chemie.uni-frankfurt.de

A crystallographic quantum-mechanical/molecular-mechanical model (c-QM/MM model) with full space-group symmetry has been developed for molecular crystals. The lattice energy was calculated by quantum-mechanical methods for short-range interactions and force-field methods for long-range interactions. The quantum-mechanical calculations covered the interactions within the molecule and the interactions of a reference molecule with each of the surrounding 12–15 molecules. The interactions with all other molecules were treated by force-field methods. In each optimization step the energies in the QM and MM shells were calculated separately as single-point energies; after adding both energy contributions, the crystal structure (including the lattice parameters) was optimized accordingly. The space-group symmetry was maintained throughout. Crystal structures with more than one molecule per asymmetric unit, *e.g.* structures with $Z' = 2$, hydrates and solvates, have been optimized as well. Test calculations with different quantum-mechanical methods on nine small organic molecules revealed that the density functional theory methods with dispersion correction using the B97-D functional with 6-31G* basis set in combination with the DREIDING force field reproduced the experimental crystal structures with good accuracy. Subsequently the c-QM/MM method was applied to nine compounds from the CCDC blind tests resulting in good energy rankings and excellent geometric accuracies.

1. Introduction

Can crystal structures be predicted? 'No' said Angelo Gavezzotti in 1991 (Gavezotti, 1991). However, in the last 20 years computational methods have made considerable progress, and computer speed has increased considerably, too. In the blind tests for crystal structure prediction, which are regularly organized by the Cambridge Crystallographic Data Centre (CCDC), the breakthrough came in 2008 with the application of dispersion-corrected density functional theory (d-DFT) calculations (Neumann *et al.*, 2008*a*,*b*; Day *et al.*, 2009). These methods, as they are implemented *e.g.* in the program *GRACE* (Neumann & Perrin, 2005), have reached a degree of accuracy and reliability which was previously unknown. Today, this method is already commercially used by pharmaceutical companies to predict possible polymorphic forms of active pharmaceutical ingredients. For these ingredients, it is essential to know the polymorphic forms – including those forms that have not yet been observed experimentally, but which are energetically possible and may emerge unintendedly during development, production or storage. The polymorphic form determines not only the mechanical properties, the morphology and processability, but also the storage stability and the dissolution behaviour, and, hence, the bioavailability. All polymorphic forms must be known, as any uncontrolled change of the polymorphic form during production or storage is prohibited by law.

Polymorphism also plays a role in other fields. For example, in organic pigments the polymorphic form determines the optical properties (colour, hue) and the photostability. For explosives, the mechanical and thermal stability as well as the explosive power depend on the polymorphic form.

The five blind tests on crystal structure prediction carried out hitherto contained 21 molecules in total (Lommerse *et al.*, 2000; Motherwell *et al.*, 2002; Day *et al.*, 2005, 2009; Bardwell *et al.*, 2011). The hitherto most successful program, *GRACE*, was able to calculate 12 out of 21 molecules as lowest-energy structures, either by prediction in the fourth and fifth blind tests (Day *et al.*, 2009; Bardwell *et al.*, 2011) or in a post-test investigation (Asmadi *et al.*, 2009). In all cases the geometrical

deviations between the experimental and calculated structures were very small. *GRACE* uses the plane-wave code *VASP* (Kresse & Furthmüller, 1996) for single-point DFT calculations. The calculations were done with the PBE functional. *GRACE* added an empirical dispersion correction (Neumann & Perrin, 2005) and performed the lattice-energy optimization with free lattice parameters.

d-DFT calculations with other codes, *e.g. CASTEP* (Clark *et al.*, 2005), or *VASP* itself can be expected to yield similar results. However the calculation times of all d-DFT calculations on crystals are still very long. Other methods, especially force-field methods, are much faster, but show a limited geometrical accuracy and a less reliable energy ranking of the predicted crystal structures. The method described in this article fills the gap between full d-DFT and force-field methods. It is faster than full d-DFT, but more accurate than force-field methods. In a full d-DFT approach, all interactions are calculated by quantum-mechanical methods. Calculation time can be reduced by a QM/MM (quantum-mechanical/molecular-mechanical) shell approach: only the interactions in the inner shell (QM) are calculated by high-level quantum mechanics, whereas the other interactions are calculated by force fields (MM) or by low-level quantum mechanics [oniom model (Dapprich *et al.*, 1999)]. QM/MM methods are widely used for large systems, *e.g.* biological systems. They have also been applied *e.g.* for the calculation on an excited molecule within a cluster of non-excited molecules (Karfunkel & Gdanitz, 1992), or for the calculation of anisotropic displacement parameters of a given molecule surrounded by other molecules (Thorn *et al.*, 2012). Most QM/MM calculations were done using a non-periodic (cluster) model consisting of a correspondingly high number of molecules.

The prediction of crystal structures requires a periodic model. The lattice parameters must be optimized simultaneously with the atomic coordinates. The full crystal symmetry has to be respected, *i.e.* all atoms in the crystal are symmetry copies of the atoms in the asymmetric unit. The symmetry has to be maintained during the optimization; *i.e.* all atoms must move in accordance with the crystal symmetry. Therefore we developed a crystallographic QM/MM shell approach with full space-group symmetry for the calculation of molecular crystals. We will denote the method as c-QM/MM. Our approach is similar to the independently developed fragment QM/MM approach described by Beran (Wen & Beran, 2011; Nanda & Beran, 2012). However, the two methods differ in two major aspects. The c-QM/MM method uses full crystallographic space-group symmetry in each step of the calculations, whereas the fragment QM/MM method performs all calculations in *P*1 with a correspondingly high number of symmetrically independent molecules. Furthermore, our QM calculations are performed in a point-charge field in order to include three-body interactions.

The c-QM/MM method is coded in a program called *QuantumCRYSCA* (which stands for 'Quantum-mechanical CRYstal Structure CAlculations').

In this report we describe the method and its application for the prediction of molecular crystal structures.

2. Method

2.1. General

The crystal structure is described by the lattice parameters, the space-group symmetry and the content of the asymmetric unit. In the easiest (and very frequent) case, the asymmetric unit contains one molecule. Our procedure allows more than one molecule per asymmetric unit (*e.g.* $Z' = 2$, hydrates, solvates *etc.*). The following description is given for the case of one molecule per asymmetric unit, which is called 'reference molecule'; other cases are treated accordingly.

2.2. Concept of QM and MM shells

The lattice energy of a molecular crystal is calculated as interaction between the reference molecule and all other molecules in the crystal.

The QM shell contains the reference molecule and all surrounding molecules. In organic crystals the first coordination sphere typically consists of 12–15 molecules (Schmidt, 1991). It is advantageous to include all molecules which have interatomic distances with the reference molecule shorter than a cutoff of about 6 Å. For the comparison of lattice energies of different polymorphs, the number of molecules in the QM shell should be identical for all structures.

In this approach the QM shell is not a sphere. In a sphere-type shell covalent bonds cross the border between QM and MM shells, which generates complications in the QM calculations. These difficulties are avoided if the chosen shell consists of complete molecules only.

The MM shell contains all other molecules. For the van der Waals interactions, a cutoff of typically 20 Å is used. Coulomb interactions are calculated within a range of typically $5 \times 5 \times 5$ unit cells (*e.g.* the unit cell of the reference molecule and two unit cells in every direction).

In each step the energies in the QM and MM shells are calculated separately as single-point energies; after adding both energy contributions, the crystal structure is changed accordingly.

2.3. Interactions in the QM shell

Ideally, the quantum-mechanical calculations should be performed on the complete QM shell. However, a simultaneous calculation of a medium-sized reference molecule with its 12 to 15 neighbouring molecules on a sufficient theoretical level (*e.g.* B97-D/6-31G*) results in a too demanding calculation effort. Thus the interaction between the reference molecule and the surrounding molecules was treated as a sum of 12–15 individual interactions between the reference molecule and each of the surrounding molecules ('dimer calculations').

The dimers were calculated in a point-charge field. The point charges were situated on the atomic positions of all other molecules in the QM shell and all molecules in the MM shell. These atomic charges were derived by quantum-mechanical calculations on an isolated molecule, preferably using the electrostatic potential (ESP) approach (Chirlian &

Francl, 1987). The charges were kept fixed during the optimizations. Through the point charges, the main component of three-molecule interactions (*e.g.* for OH···OH···OH chains in alcohols) was taken into account.

Different functionals and basis sets were tested (see §3.1.1). To account for the basis-set superposition error, each of the molecules of the dimer was calculated with the orbitals of both molecules (and the same point-charge field as for the dimer). The resulting energies were subtracted from the dimer energy, so that only the interaction *between* the two molecules in the dimer remained.

These interactions were summed for all 12 to 15 dimers. Additionally the energy of one monomer (without point-charge field) was added to account for the intramolecular energy of the reference molecule.

2.4. Interactions with the MM shell

The interactions of the reference molecule with the molecules in the MM shell were calculated by force-field methods. The intermolecular energy consists of a van der Waals potential and a Coulomb term:

$$E = \frac{1}{2}\sum_i \sum_j [-A_{ij} r_{ij}^{-6} + B_{ij} \exp(-C_{ij} r_{ij})] + \frac{1}{4\pi\varepsilon\varepsilon_0} \times \frac{q_i q_j}{r_{ij}},$$

where i denotes atoms of the reference molecule, j atoms of all molecules in the MM shell, r_{ij} interatomic distance between atoms i and j, A_{ij}, B_{ij}, C_{ij} empirical van der Waals parameters for the interaction between atoms i and j, q_i, q_j atomic charges; ε is the relative dielectric constant, $\varepsilon = 1$.

For A, B and C we used the DREIDING-X6 parametrization (Mayo *et al.*, 1990), but another force field may be used as well. Atomic charges q were calculated with the ESP approach (see above).

2.5. Molecular geometry

The starting molecular geometry can be derived *e.g.* from gas-phase calculations on the MP2/6-31G* level. Intramolecular degrees of freedom can be optimized together with the other packing variables (see the following paragraph).

2.6. Packing variables

The crystal structure is described by the molecular geometry (which is given in internal coordinates as the z-matrix) and the following packing variables:

(a) Lattice parameters a, b, c, α, β, γ;

(b) Position of the centre of the molecule, given in fractional coordinates (m_x, m_y, m_z);

(c) Spatial orientation of the molecule, given as rotation angles φ_x, φ_y, φ_z, which describe the rotation around the x, y and z axes from a reference orientation to the actual orientation;

(d) Internal degrees of freedom: bond lengths, bond angles and torsion angles (as given in the z-matrix), as far as necessary.

2.7. Optimization

In a local optimization, the energy of the crystal structure was minimized by varying the packing variables. Different optimization algorithms were implemented; all of them used the analytical gradient of the energy function, which was calculated in every optimization step. The step direction was determined by the steepest descent (Curry, 1944) or the conjugate gradient (Hestenes & Stiefel, 1952) method. The step length for each optimization step was determined by a line search algorithm according to Brent (1973). Once the optimization had converged to a stationary point, the neighbourhood of the minimum was searched to ensure that a local minimum had been found.

A local optimization may start from a given crystal structure. Alternatively a crystal structure prediction could be performed, where a large number of local optimizations starting from random structures were carried out. The combination of random structure generation and subsequent local optimization was then continued, until the lowest-energy minimum had been found several times. If so, one can assume that the search space had been scanned sufficiently completely and the global minimum had been found.

2.8. Calculation details

The program *QuantumCRYSCA* was developed in ISO-C and written for use in a Linux environment. For the QM calculations, an interface with the *ab initio* program *Gaussian09* (Frisch *et al.*, 2009) was developed. *QuantumCRYSCA* writes the input files for the monomer and dimer calculations, starts the *ab initio* calculations and reads the output files. The MM calculations were done directly within *QuantumCRYSCA*.

All calculations were carried out using 64-bit multiprocessor machines. Most of the calculations were performed on the Fuchs cluster at the Center for Scientific Computing (CSC) of the Goethe University Frankfurt.

3. Application

3.1. Evaluation of QM methods

3.1.1. Methods and basis sets. The crystallographic QM/MM shell model allows the use of different quantum-mechanical methods. We tested standard Hartree–Fock theory with and without second-order Møller–Plesset perturbation theory (HF/6-31G* and MP2/6-31G*), as well as density functional theory, using different functionals (B97-D, B3LYP, BP86) with the 6-31G* basis set. Additionally several calculations were performed with a correlation-consistent polarized valence-only triple-zeta (cc-pVTZ) basis set.

3.1.2. Selected compounds. A reliable method for crystal structure prediction fulfils two criteria:

(i) Geometric criterion: the predicted crystal structures should be similar to the experimental ones.

(ii) Energetic criterion: the energy ranking should be correct. The experimental structures should correspond to the lowest-energy minima. For polymorphic systems, the calcu-

Table 1
Ethane: experimental and optimized crystal structures, using different quantum-mechanical methods for the QM shell.

For the MM shell the DREIDING-X6 force field was used for all compounds. The space group is $P2_1/n$, $Z = 2$.

	a (Å)	b (Å)	c (Å)	β (°)	$V \times Z^{-1}$ (Å^3)
Experimental†	4.23	5.62	5.85	90.4	69.5
Optimized					
B97-D/6-31G*	4.22	5.44	5.59	90.4	64.2
B97-D/6-31+G*	4.24	5.52	5.71	90.4	66.8
B97-D/cc-pVTZ	3.59	5.72	5.97	90.4	61.3
B3LYP/6-31G*	4.25	5.62	5.68	90.4	67.8
B3LYP/cc-pVTZ	3.91	5.09	6.15	90.4	61.2
BP86/6-31G*	3.91	5.89	6.15	90.4	70.8
BP86/cc-pVTZ	4.30	5.94	6.22	90.4	79.4
HF/6-31G*	4.29	5.97	6.22	90.4	79.6
MP2/6-31G*	4.32	5.88	6.13	90.4	77.9
MP2/aug-cc-pVTZ	4.25	5.57	5.75	90.4	67.9

† Mark & Pohland (1925).

Table 2
Propane: experimental and optimized crystal structures, using different quantum-mechanical methods for the QM shell.

The space group is $P2_1/n$, $Z = 4$. In all optimizations the angle β changed by less than 0.01°.

	a (Å)	b (Å)	c (Å)	β (°)	$V \times Z^{-1}$ (Å^3)
Experimental†	4.15	12.61	6.98	91.28	91.3
Optimized					
B97-D/6-31G*	4.18	12.55	6.80	91.28	89.2
B97-D/6-31+G*	4.17	12.59	6.83	91.28	89.6
B3LYP/6-31G*	4.23	12.57	6.85	91.28	91.0
BP86/6-31G*	4.23	12.52	6.96	91.28	92.1
BP86/cc-pVTZ	4.15	12.61	6.98	91.28	91.3
HF/6-31G*	4.23	12.56	6.94	91.28	92.2
MP2/6-31G*	4.21	12.54	6.97	91.28	92.0
MP2/aug-cc-pVTZ	4.14	12.62	6.97	91.28	91.0

† Boese *et al.* (1999*a*,*b*).

Table 3
Congressane: experimental and optimized crystal structures, using different quantum-mechanical methods for the QM shell.

The space group is $Pa\bar{3}$, $Z = 4$.

	a (Å)	$V \times Z^{-1}$ (Å^3)
Experimental†	10.11	258.3
Optimized		
B97-D/6-31G*	9.62	222.6
B97-D/6-31+G*	9.65	224.7
B3LYP/6-31G*	9.67	226.1
BP86/6-31G*	9.65	224.7
HF/6-31G*	9.67	226.1
MP2/6-31G*	10.11	258.3
MP2/aug-cc-pVTZ	9.77	233.1

† Karle & Karle (1965).

lated energy ranking should reproduce the experimental order of stabilities of the individual polymorphs.

The geometrical accuracy of our crystallographic QM/MM shell model was tested on nine compounds [(1) to (9), see Fig. 1]. Ethane (1) and propane (2) are simple, non-polar molecules. Congressane (3) and hexamethylenetetramine (4) are molecules that crystallize in the cubic space group $Pa\bar{3}$ and were chosen to prove the ability of *QuantumCRYCSA* to calculate systems with high symmetries. The other compounds, benzene (5), methanol (6), ethanol (7), acetic acid (8) and oxalic acid (9), are polymorphic and were used to investigate whether the different QM methods were able to correctly reproduce the relative energies of the polymorphs. Methanol was also used as an example of a structure with $Z' = 3$.

3.1.3. Calculation details. The molecular geometry of all compounds was calculated in the gas phase from scratch on the MP2/6-31G* level using *Gaussian09* (Frisch *et al.*, 2009). Subsequently atomic charges were derived from the MP2/6-31G* calculations using the ESP approach with the CHELPG formalism (Chirlian & Francl, 1987).

In all crystal structures the experimentally determined molecules were replaced by molecules with *ab initio* optimized structures from the MP2/6-31G* calculations. For this replacement the optimized molecules must have the (approximately) correct positions and orientations in the unit cells. This was done by minimizing the difference between old and new atomic positions using a home-made least-squares-fitting algorithm implemented in *QuantumCRYSCA*.

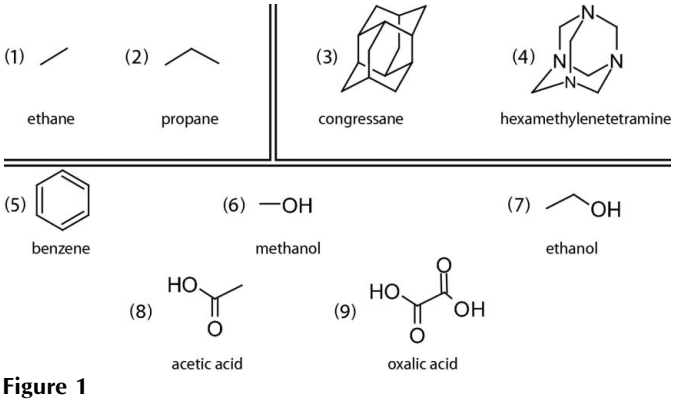

Figure 1
Compounds used as a testing set for the crystallographic QM/MM model.

Crystal structures with molecules on special positions were calculated with complete molecules in one of the corresponding subgroups. For example, the high-pressure modification of benzene in $P\,2_1/b\,2_1/c\,2_1/a$ (*Pbca*) was calculated in the subgroup $P2_12_12_1$ with the centre of the benzene molecule fixed to (0, 0, 0).

The number of molecules in the QM sphere was manually set, according to the crystal structures, and ranged from 12 to 15. In all cases, all molecules that had interatomic distances of less than 3.9 Å with the reference molecule were taken into account. For the polymorphic systems (5)–(9), the number of molecules in the QM sphere was identical for both polymorphs.

Interactions with the MM sphere were calculated with the DREIDING parametrization, using the 6-exp potential suggested by Mayo *et al.* (1990).

During the optimizations, the lattice parameters, the molecular positions (if not fixed) and the spatial orientations were optimized. The molecular geometry was kept rigid. The optimization was done with the conjugate-gradient optimizer

Table 4
Hexamethylenetetramine: experimental and optimized crystal structures, using different quantum-mechanical methods for the QM shell.

The space group is $I\bar{4}3m$, $Z = 2$.

	a (Å)	$V \times Z^{-1}$ (Å^3)	T (K)
Experimental†	7.02	173.0	298
	6.93	166.4	100
	6.91	165.0	34
Optimized			
B97-D/6-31G*	6.84	160.0	
B97-D/6-31+G*	6.93	166.4	
B97-D/cc-pVTZ	6.85	160.7	
B3LYP/6-31G*	6.88	162.8	
BP86/6-31G*	6.90	164.3	
HF/6-31G*	6.93	166.4	
MP2/6-31G*	6.92	165.7	
MP2/aug-cc-pVTZ	6.91	165.0	

† Hexamethylenetetramine at three different tempeatures (Becka & Cruickshank, 1963).

with line search until an energy shift of 0.0001 kJ mol^{-1} or a structural shift of 0.0001 Å was achieved.

3.1.4. Results. The crystallographic data of the experimental and optimized structures (1) to (9) are given in Tables 1–9. The overall geometric accuracy is quite high, although the results of the different quantum-mechanical methods partially differ from each other.

The DFT approaches, especially B97-D (Antony & Grimme, 2006), work well. The lattice parameters depend on the method, but the molecular positions and orientations remained close to the experimental values for all calculations (see Figs. 2 and 3). The $P\bar{1}$ polymorph of methanol, having three independent molecules per asymmetric unit, could be optimized without problems, too.

Generally, DFT methods are problematic for the calculation of molecular crystals, since the dispersion interaction, which is usually the most important one in organic crystals, is not correctly described by DFT; correspondingly, periodic DFT calculations frequently cause a significant expansion and distortion of the lattice, with large changes in the lattice parameters. To correct for this effect, one can use an empirical dispersion correction or a functional including a dispersion term, such as B97-D. Astonishingly, such an expansion and distortion of the lattice was not observed (except for the *Pbca* polymorph of oxalic acid) in *QuantumCRYSCA* calculations with B3LYP or BP86 functionals, which do not have a dispersion correction. Apparently the expansion and distortion of the lattice is successfully prevented by the attractive r^{-6} part of the van der Waals force field between the reference molecule and the next-neighbouring molecules at distances of about 4 to 20 Å.

The use of the larger cc-pVTZ basis set in the DFT calculations apparently leads to an overestimation of the intermolecular interactions, resulting in a shortening of the lattice parameters.

If the calculations start from slightly different parameters, *e.g.* with lattice parameters changed by 10%, the optimization leads to the same minimum.

The energetical reliability was tested on the energy ranking for the polymorphs of compounds (5)–(9). Since temperature and pressure were neglected in all calculations, the experimental low-temperature structures should appear as the most stable ones (*i.e.* with lowest energies), whereas high-pressure polymorphs should be calculated as less stable. For oxalic acid two polymorphs are available at room temperature. The *Pbca* polymorph is experimentally more stable than the $P2_1/c$ polymorph by 1.17 kJ mol^{-1}.

The calculated energy rankings of compounds (5)–(9) strongly depend on the quantum-mechanical method, the basis set and the dispersion correction (if any) (see Tables 5–9). For methanol (6) and acetic acid (8) all methods correctly reproduce the experimental order of stability. For oxalic acid (9) all methods fail. For ethanol (7) only HF/6-31G* and for benzene (5) only MP2/6-31G* give the correct order of stability. For some compounds we additionally tested B97-D/6-31+G* and MP2/aug-cc-pVTZ, but the results were not convincing. All quantum-mechanical methods have their limitations, and obviously further work is necessary to find an improved combination of quantum-mechanical method, basis set, dispersion correction and force field.

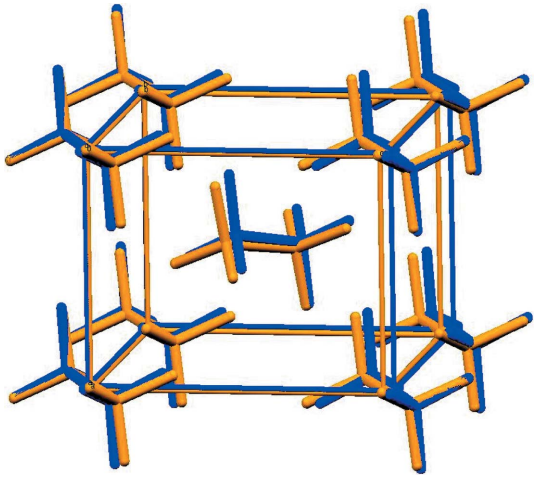

Figure 2
Superposition of the experimental crystal structure of ethane (in yellow) on the optimized structure (in blue) calculated by d-DFT with the B97-D functional for the QM shell. For the MM shell the DREIDING-X6 force field was used for all compounds.

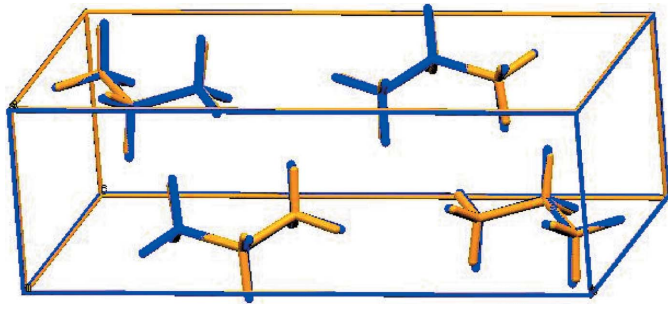

Figure 3
Superposition of the experimental crystal structure of propane (in yellow) on the optimized structure (in blue) calculated by d-DFT with the B97-D functional for the QM shell.

Table 5
Benzene: experimental and optimized crystal structures of the high-pressure (HP) and low-temperature (LT) polymorphs, using different quantum-mechanical methods for the QM shell.

	Space group	Z	$E_{\rm Latt}$ (kJ mol^{-1})	ΔE (kJ mol^{-1})	a (Å)	b (Å)	c (Å)	β (°)	$V \times Z^{-1}$ (Å^3)	
Experimental†	*Pbca*	4			7.39	9.42	6.81	90.0	118.5	LT
	$P2_1/c$	2			5.42	5.38	7.53	110.0	103.2	HP
Optimized										
B97-D/6-31G*	*Pbca*	4	−32.41	1.71	7.34	9.11	6.70	90.0	112.0	
B97-D/6-31G*	$P2_1/c$	2	−34.12	0.0	5.42	5.44	7.53	110.03	104.3	
B3LYP/6-31G*	*Pbca*	4	−26.63	1.25	7.39	9.42	6.81	90.0	118.5	
B3LYP/6-31G*	$P2_1/c$	2	−27.88	0.0	5.42	5.38	7.53	110.0	103.2	
BP86/6-31G*	*Pbca*	4	−26.08	0.83	7.39	9.24	6.68	90.0	114.0	
BP86/6-31G*	$P2_1/c$	2	−26.91	0.0	5.46	5.43	7.53	110.0	104.9	
HF/6-31G*	*Pbca*	4	−26.54	0.98	7.25	9.20	6.78	90.0	113.1	
HF/6-31G*	$P2_1/c$	2	−27.52	0.0	5.42	5.50	7.53	110.0	105.5	
MP2/6-31G*	*Pbca*	4	−26.47	0.0	7.30	9.15	6.81	90.0	113.7	
MP2/6-31G*	$P2_1/c$	2	−26.06	0.41	5.39	5.32	7.53	110.0	101.4	

† Cox (1958).

Table 6
Methanol: experimental and optimized crystal structures of the high-pressure (HP) and low-temperature (LT) polymorphs, using different quantum-mechanical methods for the QM shell.

	Space group	Z	$E_{\rm Latt}$ (kJ mol^{-1})	ΔE (kJ mol^{-1})	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)	$V \times Z^{-1}$ (Å^3)	
Experimental†	$P\bar{1}$	6			7.67	4.41	7.20	88.1	102.9	93.9	39.4	HP
‡	$P2_12_12_1$	4			4.87	4.64	8.87	90.0	90.0	90.0	50.1	LT
Optimized												
B97-D/6-31G*	$P\bar{1}$	6	−14.69	1.56	7.67	4.41	7.24	88.1	102.9	93.9	39.6	
B97-D/6-31G*	$P2_12_12_1$	4	−16.25	0.0	4.87	4.64	8.86	90.0	90.0	90.0	50.1	
B97-D/6-31+G*	$P\bar{1}$	6	−14.20	3.78	7.67	4.44	7.23	88.1	102.9	93.9	39.9	
B97-D/6-31+G*	$P2_12_12_1$	4	−17.98	0.0	4.89	4.62	8.79	90.0	90.0	90.0	49.6	
B97-D/cc-pVTZ	$P\bar{1}$	6	−11.81	1.62	7.68	4.42	7.26	88.1	102.9	93.9	39.9	
B97-D/cc-pVTZ	$P2_12_12_1$	4	−13.43	0.0	4.87	4.64	8.85	90.0	90.0	90.0	50.0	
B3LYP/cc-pVTZ	$P\bar{1}$	6	−9.68	1.78	7.67	4.41	7.20	88.1	102.9	93.9	39.4	
B3LYP/cc-pVTZ	$P2_12_12_1$	4	−11.46	0.0	4.87	4.64	8.85	90.0	90.0	90.0	50.0	
BP86/cc-pVTZ	$P\bar{1}$	6	−8.16	2.52	7.68	4.46	7.47	88.1	102.9	93.9	41.4	
BP86/cc-pVTZ	$P2_12_12_1$	4	−10.68	0.0	4.87	4.64	8.85	90.0	90.0	90.0	50.0	
HF/6-31G*	$P\bar{1}$	6	−13.74	3.61	7.67	4.41	7.25	88.1	102.9	93.9	39.7	
HF/6-31G*	$P2_12_12_1$	4	−17.35	0.0	4.88	4.63	8.74	90.0	90.0	90.0	49.4	
MP2/6-31G*	$P\bar{1}$	6	−9.35	6.98	7.67	4.41	7.19	88.1	102.9	93.9	39.4	
MP2/6-31G*	$P2_12_12_1$	4	−16.33	0.0	4.89	4.63	8.73	90.0	90.0	90.0	49.4	
MP2/aug-cc-pVTZ	$P\bar{1}$	6	−7.20	2.44	7.69	4.43	7.20	88.3	102.9	94.1	39.7	
MP2/aug-cc-pVTZ	$P2_12_12_1$	4	−9.66	0.0	4.86	4.64	8.79	90.0	90.0	90.0	49.6	

† Allan *et al.* (1998). ‡ Kirchner *et al.* (2008).

Table 7
Ethanol: experimental and optimized crystal structures of the high-pressure (HP) and low-temperature (LT) polymorphs, using different quantum-mechanical methods for the QM shell.

	Space group	Z	$E_{\rm Latt}$ (kJ mol^{-1})	ΔE (kJ mol^{-1})	a (Å)	b (Å)	c (Å)	β (°)	$V \times Z^{-1}$ (Å^3)	
Experimental†	*Pc*	2			5.38	6.88	8.26	102.2	74.7	LT
	$P2_1/c$	4			7.60	4.77	7.27	114.8	59.8	HP
Optimized										
B97-D/6-31G*	*Pc*	2	−21.55	3.64	5.47	6.96	8.40	100.6	78.6	
B97-D/6-31G*	$P2_1/c$	4	−25.19	0.0	7.83	5.60	7.85	114.8	78.1	
B97-D/6-31+G*	*Pc*	2	−13.12	3.85	5.40	6.95	8.22	102.2	75.4	
B97-D/6-31+G*	$P2_1/c$	4	−16.97	0.0	7.70	5.13	7.61	114.8	68.2	
B97-D/cc-pVTZ	*Pc*	2	−13.66	5.72	5.38	6.88	8.26	102.2	74.7	
B97-D/cc-pVTZ	$P2_1/c$	4	−19.38	0.0	7.60	5.00	7.84	114.8	67.6	
B3LYP/6-31G*	*Pc*	2	−9.52	6.55	5.38	6.88	8.26	102.2	74.7	
B3LYP/6-31G*	$P2_1/c$	4	−16.07	0.0	7.60	5.13	7.26	114.8	64.2	
HF/6-31G*	*Pc*	2	−22.78	0.0	5.37	7.21	8.26	102.2	78.1	
HF/6-31G*	$P2_1/c$	4	−20.90	1.88	7.67	5.63	7.83	114.8	76.7	
MP2/6-31G*	*Pc*	2	−11.75	3.63	5.38	7.01	8.26	102.2	76.1	
MP2/6-31G*	$P2_1/c$	4	−15.38	0.0	7.60	4.99	7.83	114.8	67.4	
MP2/aug-cc-pVTZ	*Pc*	2	−14.33	3.52	5.40	6.95	8.22	102.2	75.4	
MP2/aug-cc-pVTZ	$P2_1/c$	4	−17.85	0.0	7.65	5.01	7.57	114.8	65.8	

† Allan & Clark (1999).

Table 8
Acetic acid: experimental and optimized crystal structures of the high-pressure (HP) and low-temperature (LT) polymorphs, using different quantum-mechanical methods for the QM shell.

	Space group	Z	E_{Latt} (kJ mol^{-1})	ΔE (kJ mol^{-1})	a (Å)	b (Å)	c (Å)	β (°)	$V \times Z^{-1}$ (Å^3)	
Experimental†	$P2_1/n$	4			3.94	13.02	5.64	93.2	72.2	HP
	$Pna2_1$	4			13.31	4.09	5.77	90.0	78.5	LT
Optimized										
B97-D/6-31G*	$P2_1/n$	4	−26.07	2.98	3.95	13.13	5.55	93.2	71.8	
B97-D/6-31G*	$Pna2_1$	4	−29.05	0.0	13.31	4.09	5.74	90.0	78.1	
B97-D/6-31+G*	$P2_1/n$	4	−23.40	3.61	3.94	13.12	5.64	93.2	72.9	
B97-D/6-31+G*	$Pna2_1$	4	−27.01	0.0	13.31	4.09	5.75	90.0	78.3	
B3LYP/cc-pVTZ	$P2_1/n$	4	−21.64	3.68	3.94	13.03	5.67	93.18	72.7	
B3LYP/cc-pVTZ	$Pna2_1$	4	−25.32	0.0	13.31	4.09	5.75	90.0	78.3	
HF/6-31G*	$P2_1/n$	4	−22.74	7.89	3.94	13.03	5.64	93.2	72.3	
HF/6-31G*	$Pna2_1$	4	−30.63	0.0	13.30	4.09	6.90	90.0	93.8	
MP2/6-31G*	$P2_1/n$	4	−21.43	5.73	3.94	13.03	5.64	93.2	72.3	
MP2/6-31G*	$Pna2_1$	4	−27.16	0.0	13.31	3.99	6.92	90.0	91.9	
MP2/aug-cc-pVTZ	$P2_1/n$	4	−19.83	3.74	3.97	13.09	5.60	93.2	72.6	
MP2/aug-cc-pVTZ	$Pna2_1$	4	−23.57	0.0	13.31	4.09	5.72	90.0	77.8	

† Allan & Clark (1999).

Table 9
Oxalic acid: experimental and optimized crystal structures, using different quantum-mechanical methods for the QM shell.

	Space group	Z	E_{Latt} (kJ mol^{-1})	ΔE (kJ mol^{-1})	a (Å)	b (Å)	c (Å)	β (°)	$V \times Z^{-1}$ (Å^3)
Experimental†	$Pbca$	4			6.56	6.09	7.85	90.0	78.4
	$P2_1/c$	4			5.33	6.02	5.44	115.8	78.6
Optimized									
B97-D/6-31G*	$Pbca$	4	−51.54	3.09	6.50	6.09	8.16	90.0	80.8
B97-D/6-31G*	$P2_1/c$	4	−54.63	0.0	5.85	6.02	4.97	115.8	78.8
B97-D/6-31+G*	$Pbca$	4	−44.48	3.58	6.69	6.40	8.14	90.0	87.1
B97-D/6-31+G*	$P2_1/c$	4	−48.06	0.00	5.88	5.56	5.08	115.8	74.8
B97-D/cc-pVTZ	$Pbca$	4	−45.54	5.9	6.55	6.11	7.92	90.0	79.2
B97-D/cc-pVTZ	$P2_1/c$	4	−51.44	0.0	5.87	6.02	5.44	115.8	86.5
B3LYP/6-31G*	$Pbca$	4	−53.52	5.76	6.47	6.09	8.42	90.0	82.9
B3LYP/6-31G*	$P2_1/c$	4	−47.76	0.0	5.86	5.78	4.81	115.8	73.3
BP86/6-31G*	$Pbca$	4	−41.87	5.89	6.98	8.33	6.39	90.0	92.9
BP86/6-31G*	$P2_1/c$	4	−47.76	0.0	5.86	5.78	4.81	115.8	73.3
BP86/cc-pVTZ	$Pbca$	4	−41.03	4.85	6.67	7.46	7.26	90.0	90.3
BP86/cc-pVTZ	$P2_1/c$	4	−45.88	0.0	5.79	5.81	4.79	115.8	72.5
HF/6-31G*	$Pbca$	4	−44.30	4.76	7.00	8.25	6.43	90.0	92.8
HF/6-31G*	$P2_1/c$	4	−49.06	0.0	5.94	5.79	4.76	115.8	73.7
MP2/6-31G*	$Pbca$	4	−41.62	6.51	6.48	7.15	7.40	90.0	85.7
MP2/6-31G*	$P2_1/c$	4	−48.13	0.0	5.87	5.55	5.07	115.8	74.4

† Derissen & Smith (1974).

In a comparison between the high-pressure and the low-temperature polymorphs of compounds (5)–(8) the high-pressure polymorphs were predicted to have a more dense crystal structure by all applied methods. As a compromise between the accuracy and the calculation effort we chose the B97-D functional with 6-31G* basis set for all further calculations. This method has also proven to give a good geometrical accuracy throughout.

3.2. Blind test molecules

3.2.1. On the CCDC blind tests. In the blind tests for crystal structure prediction organized by the CCDC, the participants had to predict the crystal structures of small molecular compounds from scratch, using only a sketch of the molecule. Hitherto ten to 20 different research groups have participated in each of the five performed blind tests. A broad variety of methods were applied, including standard force fields, statistical potentials, force fields with electrostatic multipoles, tailor-made force fields fitted to *ab initio* calculations, and different combinations of *ab initio* and force-field methods.

Several groups used a two-step approach. The global search for possible crystal structures, which typically requires 10^5 to 10^7 local optimizations in different space groups, with varying lattice parameters and molecular packings, was carried out with a fast method (force field). In the second step the ten to 100 most promising structures were optimized by quantum-mechanical methods.

Every group was allowed to submit three 'proposed' crystal structures per compound. Subsequently the experimental crystal structures were laid open. The list of these proposed crystal structures is available from the CCDC. This list of three structures each, derived by a broad variety of methods, serves as a useful set of starting structures for the evaluation and comparison of new methods for crystal structure prediction (Asmadi *et al.*, 2009).

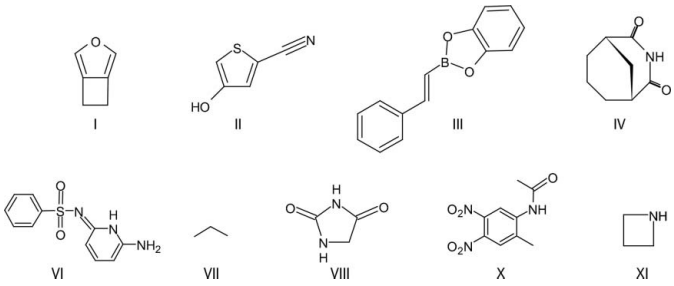

Figure 4
Compounds of the CCDC blind test used to evaluate the *QuantumCRYSCA* approach. The numbering (I to XI) was taken from the blind tests.

We used this list of structures to evaluate the accuracy and reliability of our new approach. Owing to limitations on computing time we restricted the calculations to a selection of the compounds (Fig. 4). This selection included compounds with two polymorphs (compound I), molecules with unusual elements such as sulfur or boron (compounds II, III, VI), structures with $Z' = 2$ (compound XI) and a compound that is known to have a huge number of possible crystal structures within a narrow energy window (compound VII).

3.2.2. Procedure. In the reference list of proposed crystal structures, the molecular geometries strongly depend on the applied calculation method. All molecular geometries were calculated from scratch in the gas phase on the MP2/6-31G* level using *Gaussian*. Atomic charges were derived using the ESP approach on the same level, applying the CHELPG formalism.

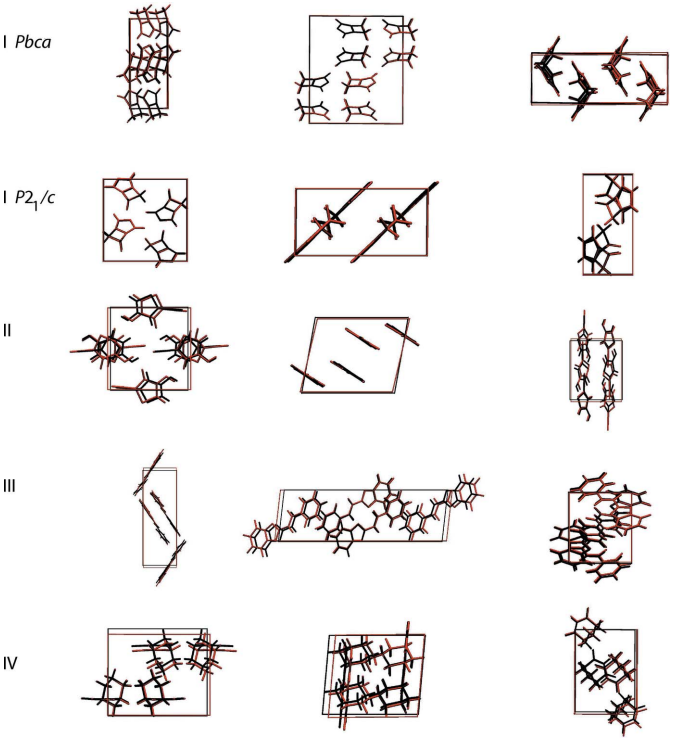

Figure 5
Overlay of experimental (black) and optimized (red) structures of the blind test molecules (I–IV). First column: view direction [100], *b* axis right, *c* axis down; second column: view direction [010], *c* axis right, *a* axis down; third column: view direction [001], *a* axis right, *b* axis down.

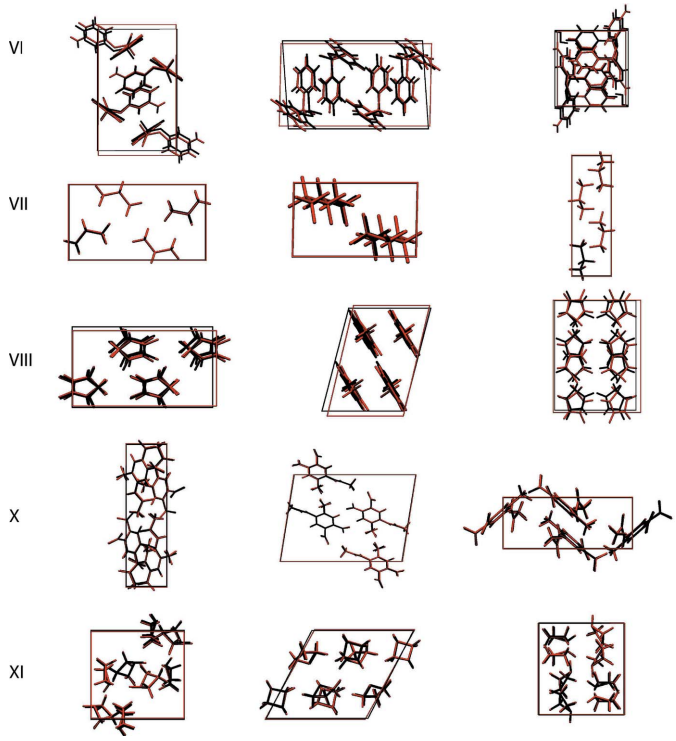

Figure 6
Overlay of experimental (black) and optimized (red) structures of the blind test molecules (VI-VIII, X, XI). View direction and axes as in Fig. 5.

The number of molecules in the QM shell was chosen separately for each compound, so that in all proposed structures of a given compound all molecules of the first coordination shell were included (typically 12 to 15).

Quantum-mechanical interactions were calculated by d-DFT with the B97-D functional using the program *Gaussian*. The force-field interactions were calculated using the DREIDING-X6 parametrization. Structures with molecules on special positions were calculated with a complete molecule in the corresponding subgroups.

3.2.3. Results and discussion. The optimized and the experimental structures are given in the supporting information.[1] All crystal structures were reproduced with a very good accuracy (see Figs. 5 and 6). The lattice parameters are almost identical to the experimental values. The atomic coordinates match very well, too. This even holds for the structures containing sulfur and boron atoms. In most cases, the difference between the experimental and the structures optimized by *QuantumCRYSCA* is considerably smaller than that for the structures optimized by other methods. Obviously the intermolecular interactions are described very well by the applied combination of B97-D and DREIDING methods.

The crystal structures of four of the nine compounds were found as lowest-energy structures, *i.e.* with energy rank 1. The

[1] Supporting information for this paper is available from the IUCr electronic archives (Reference: KX5031).

Table 10
Energy ranking for the treated blind test molecules derived by the c-QM/MM approach.

ΔE denotes the energy difference between the optimized and the predicted lowest-energy structure.

Molecule		Rank	ΔE (kJ mol^{-1})
3-Oxabicyclo[3.2.0]hepta-1,4-diene†	I	1/14	0.0/4.12
2-Cyano-4-hydroxythiophene	II	1	0.0
(*E*)-2-(2-Phenylethenyl)-1,3,2-benzodioxaborole	III	1	0.0
3-Azabicyclo[3.3.1]nonane-2,4-dione	IV	3	2.74
6-Amino-2-(phenylsulfonylimino)-1,2-dihydropyridine	VI	1	0.0
Propane	VII	4	0.12
Hydantoin	VIII	3	8.39
2-Acetamido-4,5-dinitrotoluene	X	7	6.86
Azetidine	XI	2	2.31

† Polymorphs in the order form 1 (*Pbca*)/form 2 ($P2_1/c$).

other structures were obtained with the energy ranks 2, 3, 4 or 7 (see Table 10).

Compound I is known to have two polymorphic forms. At room temperature, I is a liquid. During the low-temperature crystallization experiments, the *Pbca* form was obtained only once, whereas the $P2_1/c$ form was obtained several times (Boese, 2000); from this observation the $P2_1/c$ form is considered as being the thermodynamically stable polymorph, whereas the *Pbca* form is considered as kinetically controlled. *QuantumCRYSCA* predicted the *Pbca* form as energy rank 1 and the $P2_1/c$ form as energy rank 14 with an energy difference of 4.1 kJ mol^{-1}.

However, not only *QuantumCRYSCA*, but also most other methods including *GRACE* predicted the *Pbca* structure to be the more stable form. This raises the question of whether the $P2_1/c$ polymorph is indeed the thermodynamically stable one, or if its appearance might be kinetically controlled, too.

For propane (compound VII), force-field calculations carried out during the first blind test revealed several hundred structures in an energy window of 6 kJ mol^{-1} (Lommerse *et al.*, 2000). In the blind test the experimental structure was found by three groups, with the energy ranks of 1, 5 (ΔE = 0.23 kJ mol^{-1}) or 15 (ΔE = 0.64 kJ mol^{-1}). *QuantumCRYSCA* yields the structure on energy rank 4 with an energy difference of only 0.12 kJ mol^{-1}.

In total *QuantumCRYSCA* ranked seven out of ten structures with energy rank 1 (four times), 2 (one time) or 3 (two times). Since we did not perform a full search in all space groups, there might be a few other crystal structures that were not predicted by other groups, but have a low energy when optimized with *QuantumCRYSCA*. Nevertheless, under the conditions of the blind tests, where three submissions per compound are allowed, *QuantumCRYSCA* would have had a success rate of about 7 out of 10. This is not as good as *GRACE*, which found all these structures at ranks 1 (eight times) and 2 (two times) in post-test calculations, but the success rate is higher than for most other participating methods.

The geometrical accuracy of the new c-QM/MM approach is as high as with periodic d-DFT calculations. This is not surprising, since both approaches use d-DFT for the most important, close interactions. Temperature effects are not included in the d-DFT calculations. The force field is parametrized for crystal structures at room temperatures. However, the attractive part of the van der Waals potential changes only slightly with temperature, and the repulsive part of the potential is without importance in our calculations. Hence, the c-QM/MM calculations represent structures close to 0 K. Correspondingly, the structure of propane, which was measured at 30 K, is predicted with higher accuracy than the structures which were determined at room temperature; and the structures of ethane, hexamethylenetetramine and ethanol ($P\,c$) determined at intermediate temperatures are in between (see Table S1 in the supporting information). In the case of hexamethylenetetramine, where the structure could be determined at three different temperatures, the prediction corresponds to the low-temperature structures.

Crystallization kinetics is the big unknown for all crystal structure predictions. The crystallization rate is heavily affected by a wide variety of factors such as the number and direction of screw dislocations, the solvent, the concentration of impurities and their molecular structures. For example, an impurity concentration of a few p.p.b. is sufficient to cover the whole surface of all crystals in a batch with crystals of a size of 1 µm. It is hardly possible to include all these imponderabilities in the crystal structure prediction. Furthermore, the single-crystal X-ray analysis is not always carried out on the thermodynamically stable phase, but on the polymorph that crystallizes first, forms the largest crystals, exhibits the nicest morphology, or is not affected by difficult twinning or severe diffuse scattering. To prove that the obtained crystal structure corresponds to the thermodynamically stable phase, one should always perform a polymorph screening, and record an X-ray powder pattern and compare it with the pattern simulated from the single-crystal data. However, in most cases, none of these investigations are done and it remains questionable if the published structure corresponds to the thermodynamically stable one.

The geometrical accuracy of the new c-QM/MM approach is as high as with periodic d-DFT calculations. The deviations between experimental and calculated structures are generally smaller than the effects of thermal expansion, when the single-crystal structure determination is carried out at low temperature instead of at room temperature.

4. Conclusion

The crystallographic QM/MM model is a useful approach for the optimization and prediction of molecular crystal structures. The intermolecular interactions are well described by the combination of d-DFT methods with the B97-D functional for close interactions and the DREIDING force field for long interactions. Apparently neither the division of the QM calculations into a sum of two-molecule calculations (and the corresponding incomplete treatment of three-body interac-

tions), nor the basis-set superposition error or the treatment of long-range interactions by force field instead of QM methods are too serious a drawback. Hitherto we tested only a small number of DFT functionals, basis sets and dispersion corrections. We hope that with better d-DFT methods some of the deviating structures, especially those where the *GRACE* code performs better, give improved results. However, even d-DFT methods have a limited accuracy and it is challenging to achieve an accuracy of even 2 kJ mol^{-1} for intermolecular interactions.

The calculation times on these small molecules are considerably higher than with force-field methods, but smaller than with plane-wave DFT codes, such as *GRACE* or *VASP*.

The crystallographic QM/MM method provides very accurate crystal structures with a good energy ranking. Hence, this method may be a good alternative to periodic d-DFT methods for crystal structure prediction.

The authors thank Dirk Rehn for initializing the project, Konstantin Mörschel for implementing the optimization procedures, Manuela Thurn, Markus Dietz, Robert Binder and Christian Czech (all University of Frankfurt) for evaluation and application of *QuantumCRYSCA*, the Centre for Scientific Computing of the University of Frankfurt for calculation time and the referees for helpful comments.

References

Allan, D. R. & Clark, S. J. (1999). *Phys. Rev. B*, **60**, 6328–6334.
Allan, D. R., Clark, S. J., Brugmans, M. J. P., Ackland, G. J. & Vos, W. L. (1998). *Phys. Rev. B*, **58**, R11809.
Antony, J. & Grimme, S. (2006). *Phys. Chem. Chem. Phys.* **8**, 5287–5293.
Asmadi, A., Neumann, M. A., Kendrick, J., Girard, P., Perrin, M. A. & Leusen, F. J. (2009). *J. Phys. Chem. B*, **113**, 16303–16313.
Bardwell, D. A. *et al.* (2011). *Acta Cryst.* B**67**, 535–551.
Becka, L. N. & Cruickshank, D. W. J. (1963). *Proc. R. Soc. London Ser. A*, **273**, 435–454.
Boese, R. (2000). Presentation on the workshop to the first CCDC blind test, Cambridge.
Boese, R., Weiss, H.-C. & Bläser, D. (1999*a*). *Angew. Chem.* **111**, 1042–1045.
Boese, R., Weiss, H.-C. & Bläser, D. (1999*b*). *Angew. Chem. Int. Ed.* **38**, 988–992.
Brent, R. P. (1973). *Algorithms for Minimization without Derivatives*, 1st ed. Englewood Cliffs: Prentice-Hall.
Chirlian, L. E. & Francl, M. M. (1987). *J. Comput. Chem.* **8**, 894–905.
Clark, S. J., Segall, M. D., Pickard, C. J., Hasnip, P. J., Probert, M. I. J., Refson, K. & Payne, M. C. (2005). *Z. Kristallogr.* **220**, 567–570.
Cox, E. G. (1958). *J. Phys. Chem. B*, **113**, 16303–16313.
Curry, H. B. (1944). *Q. Appl. Math.* **2**, 258–261.
Dapprich, S., Komáromi, I., Byun, S., Morokuma, K. & Frisch, M. J. (1999). *J. Mol. Struct. (Theochem)*, **462**, 1–21.
Day, G. M. *et al.* (2009). *Acta Cryst.* B**65**, 107–125.
Day, G. M. *et al.* (2005). *Acta Cryst.* B**61**, 511–527.
Derissen, J. L. & Smith, P. H. (1974). *Acta Cryst.* B**30**, 2240–2242.
Frisch, M. J. *et al.* (2009). *Gaussian09*. Revision A.2. Gaussian Inc., Wallingford, CT, USA.
Gavezotti, A. (1991). *J. Am. Chem. Soc.* **113**, 4622–4629.
Hestenes, M. R. & Stiefel, E. (1952). *J. Res. NIST*, **49**, 409–436.
Karfunkel, H. R. & Gdanitz, R. J. (1992). *J. Comput. Chem.* **13**, 1171–1183.
Karle, I. L. & Karle, J. (1965). *J. Am. Chem. Soc.* **87**, 918–920.
Kirchner, M. T., Das, D. & Boese, R. (2008). *Cryst. Growth Des.* **8**, 763–765.
Kresse, G. & Furthmüller, J. (1996). *J. Comput. Mater. Sci.* **6**, 15–50.
Lommerse, J. P. M., Motherwell, W. D. S., Ammon, H. L., Dunitz, J. D., Gavezzotti, A., Hofmann, D. W. M., Leusen, F. J. J., Mooij, W. T. M., Price, S. L., Schweizer, B., Schmidt, M. U., van Eijck, B. P., Verwer, P. & Williams, D. E. (2000). *Acta Cryst.* B**56**, 697–714.
Mark, H. & Pohland, E. (1925). *Z. Kristallogr.* **62**, 103–112.
Mayo, S. L., Olafson, B. D. & Goddard, W. A. III (1990). *J. Phys. Chem.* **94**, 8897–8909.
Motherwell, W. D. S. *et al.* (2002). *Acta Cryst.* B**58**, 647–661.
Nanda, K. D. & Beran, G. J. (2012). *J. Chem. Phys.* **137**, 174106.
Neumann, M. A., Leusen, F. J. J. & Kendrick, J. (2008*a*). *Angew. Chem.* **120**, 2461–2464.
Neumann, M. A., Leusen, F. J. J. & Kendrick, J. (2008*b*). *Angew. Chem. Int. Ed.* **47**, 2427–2430.
Neumann, M. A. & Perrin, M. A. (2005). *J. Phys. Chem. B*, **109**, 15531–15541.
Schmidt, M. U. (1991). *Geometrische Formen von Molekülen im Festkörper*. Thesis, Institut für Anorganische Chemie der RWTH Aachen, Germany.
Thorn, A., Dittrich, B. & Sheldrick, G. M. (2012). *Acta Cryst.* A**68**, 448–451.
Wen, S. & Beran, G. J. O. (2011). *J. Chem. Theory Comput.* **7**, 3733–3742.