Transferable ab Initio Intermolecular Potentials. 2. Validation and Application to Crystal Structure Prediction

Wijnand T. M. Mooij,*,† Bouke P. van Eijck,‡ and Jan Kroon§

Department of Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

Received: May 19, 1999; In Final Form: September 7, 1999

We have tested the performance of a previously developed ab initio potential in the simulation of crystalline phases. First, the model was validated by performing energy minimizations for the experimental crystal structures of several small organic molecules such as hydrocarbons, ethers, alcohols, and carbohydrates. Generally, the experimental structures were maintained very well. Calculated packing energies were in good agreement with experimental heats of sublimation. For flexible molecules, change of the molecular charge distribution with conformation was seen to be important. Secondly, crystal structure predictions were performed for methanol, ethanol, 1,4-dioxane, and propane. For methanol the experimental structure corresponded to the most favorable structure with one independent molecule, although a few structures with two independent molecules had a marginally lower energy. In the case of ethanol the experimental structure, which contains two independent molecules, was among the best ones, but in this case three structures with one independent molecule were slightly more favorable. For dioxane the high- and the low-temperature phase were predicted with low energy, but in the wrong order. The experimental structure of propane was predicted correctly: it corresponded to the most favorable structure in our ab initio potential. In all cases the predictions using the ab initio potentials were superior to predictions based on standard force fields.

1. Introduction

In many areas of solid-state chemistry it would be desirable to be able to predict crystal structures from molecular constitution alone. This has been an active field of research over the past decade, and considerable progress has been made. For fairly rigid molecules, it is nowadays possible to generate a large amount of plausible crystal structures which includes the experimentally observed structure(s). A number of different approaches exist, which have been reviewed recently.[1,2] Nevertheless, it is still very hard to reliably pin-point the correct solution from the overwhelming number of possibilities,[2,3] although some encouraging results have been obtained (see refs 1−4 and references therein). A major problem is that the calculated energy differences between hypothetical crystal structures can be extremely small. In spite of this fact, most methods rely on standard force fields for their energy calculations. More accurate potentials are an essential first step towards more reliable structure predictions.

In a companion article[5] we have described the development of a potential based on ab initio calculations on methanol dimers and trimers. This potential was seen to be accurate to within a few tenths of a kcal/mol for methanol dimers and related systems containing methane, water, and dimethyl ether. For crystal structure prediction relative energies are more important than absolute energies. The former will be more accurate due to compensation of errors. Because our potential was completely parameterized on ab initio data for dimers and trimers, its transferability to the condensed phase had to be assessed. Also the transferability to other molecules had to be further validated.

With the derivation of this potential we have focused on the energetic differences between hypothetical crystals. In the end, thermodynamic and kinetic effects cannot be neglected: phase transitions prove that temperature can play a decisive role, and crystallization of different polymorphs from different solvents shows the importance of crystallization conditions on kinetics. However, the static approach will be a reasonable first approximation and an appropriate starting point for any more sophisticated study.

In this article we report the performance of our ab initio potential in solid state simulations by energy minimization for experimental crystal structures of small organic molecules. The test set contains alkanes, ethers, alcohols, and carbohydrates. Subsequently, crystal structure predictions for methanol, ethanol, 1,4-dioxane, and propane are reported.

2. Methods

2.1. Program. Crystal energy minimizations were performed in a new crystal energy minimizer, which was written on the basis of modules from the UPACK[6,7] and the TINKER[8] program suites. The program will be made available through the TINKER distribution. Basically it combines the crystal capabilities of the UPACK minimization algorithm with the polarizable multipole electrostatics of the TINKER package. Due to the complexity of that formulation no Ewald summation was implemented, and so the program uses a direct summation for nonbonded interactions, based on a molecular cutoff. A pair list is used, which is only updated when a parameter changes more than a certain value. Minimizations were performed by a variable metric (BFGS) algorithm[9] under the constraints of space-group symmetry.

The electrostatic energy of polar crystals is treated using van Eijck's long-range correction.[10] Inclusion of this correction term

* To whom correspondence should be adressed.
† E-mail: w.t.m.mooij@chem.uu.nl.
‡ E-mail: b.p.vaneijck@chem.uu.nl.
§ E-mail: j.kroon@chem.uu.nl.

10.1021/jp991640v CCC: $18.00
Published on Web 12/02/1999

yields results that are in accordance with the (more common) approach of Ewald summation with tin foil boundary conditions. In this work we include a similar long-range correction for the polarization energy. To calculate the polarization energy one needs to know the electrostatic field (**E**) at the position of each atom. One would normally calculate this field by summing over all molecules within the cutoff sphere ($\mathbf{E}_{\text{cutoff}}$). For a polar crystal, the field due to the charge distribution outside the cutoff sphere is not negligible, and has to be taken into account as well. This contribution to the electrostatic field is implicit in the derivation of the long-range correction for the electrostatic energy of van Eijck et al.:[10] their eq 6 can be interpreted as the energy of a dipole in the field of a charge density $\boldsymbol{p}\cdot\boldsymbol{n}/V$ on the cutoff sphere, where **p** is the cell dipole, V is the cell volume, and **n** is a unit vector perpendicular to the cutoff sphere. The field created by this charge density is easily calculated by integration, resulting in a total electric field at a certain atom:

$$\mathbf{E} = \mathbf{E}_{\text{cutoff}} + \frac{4\pi\mathbf{p}}{3V}$$

The van der Waals energy was calculated using the same pair list as for the electrostatic energy, augmented with a continuum tail correction[11] for the dispersion energy.

2.2. Intermolecular Force Field. The intermolecular force field was described in the companion article.[5] This potential involves atomic multipole moments (AMMs), atomic dipole polarizabilities, a damped r^{-6} dispersion contribution, and an exponential repulsion term, which is anisotropic for oxygen. The AMMs include charges and dipoles on hydrogens and charges, dipoles, and quadrupoles on carbon and oxygen. They were derived by ESP fitting to a SCF/DZ($2d^{O}$) wavefunction, as described in the previous article (AMMs-I).[5] For methanol dimer this was seen to be an acceptable level for the calculation of AMMs. To further validate this approach, we also derived AMMs at a higher level (MP2/IOM) for methanol, ethanol, and 1,4-dioxane (AMMs-II). Ab initio calculations were performed using GAMESS-UK[12] or GAUSSIAN94;[13] the AMMs were derived using fitting routines that were implemented in MOLDEN.[14]

2.3. Intramolecular Force Field. Because our program does not allow for rigid body minimization we had to add an intramolecular force field, even for essentially rigid molecules. Generally, mixing force fields is not a recommended procedure. In any case, torsional parameters cannot be separated from van der Waals and electrostatic parameters that were used in their parameterization. Therefore, we decided to use completely separated intra- and intermolecular potentials. This has also been used in the MM2x force field developed by Halgren.[15] For intramolecular interactions the MM3 force field[16] as implemented in TINKER[8] was used, including its own van der Waals parameters and bond dipoles, after adding the MM3(96) directional hydrogen-bonding term.[17]

2.4. Conformational Dependency of Multipoles. The charge distribution of a molecule is only approximately transferable from one conformation to another.[18−20] For many of the molecules considered in this work conformational dependency of multipoles is not a problem, because they are rather rigid. So, we can safely calculate the multipole moments only once. For a flexible molecule, one can assume transferability in the following way: multipoles are fixed in a local axes system, defined in terms of nearby (bonded) atoms. In this way dipoles and quadrupoles stay fixed with respect to those atoms when the conformation of the molecule is changed. However, for ethanol it has been shown that this approach does not lead to a fully correct description of intermolecular interactions.[19] A solution to this problem is to calculate the multipole moments for different conformations separately. This was done in the energy minimizations for the experimental crystal structure of ethanol, which contains both the *trans* and the *gauche* conformation.

For crystal structure prediction we developed a more general approach. Optimization of a crystal structure was interrupted when a torsional parameter had changed more than 5°. Subsequently, an ab initio calculation was performed to obtain new AMMs, after which the optimization was resumed. This iterative procedure was also employed for optimization of the hexapyranoses. When we did not use this procedure, but used multipoles calculated for the experimental geometry instead, some structures showed considerable change in both energy and structure; β-D-glucose even changed towards a structure with a different hydrogen-bond scheme. We think that this breakdown of the concept of rotation of AMMs can be explained in the following way. For example, on a hydroxyl oxygen certain contributions to the multipole moments are induced by other nearby polar groups, while others describe the lone pairs on the oxygen. Upon rotation of the hydroxyl group the first contributions stay fixed with respect to those inducing groups, while the latter stay fixed with respect to the directly bonded atoms. In the present approach we rotate the complete multipole moments, which leads to an (artificial) deformation of the electrostatic energy surface. Recalculation of AMMs solves this problem, but at a considerable computational cost.

It is possible that (more) transferable multipole distributions exist. One way to obtain these would be to use a multipole distribution obtained by averaging over multiple conformations.[20] One can hope that contributions from intramolecular polarization are more or less averaged out, resulting in more transferable multipoles. Better would be to fit multipole moments (in their local axes system) to the ESP of multiple conformations, which by definition will result in the most transferable multipole distribution. Such an approach has previously been used for point charges.[21] Still, transferability to any conformation that was not present in the fit would be uncertain.

3. Energy Minimization for Experimental Crystal Structures

Experimental crystal structures were chosen to separately probe the description of different types of interactions. Therefore the set contained hydrocarbons to test the C,H part of the potential, ethers for the C−H···O interactions, and alcohols for the hydrogen bonds. Six hexapyranoses were added to assess the validity of our approach of separated intra- and intermolecular force fields for somewhat larger molecules.

Crystal structures were taken from the CSD[22] and will be generally cited by their refcodes in order to avoid an excessive amount of references. For the calculation of the AMMs the experimental geometry was used, with C−H and O−H bond lengths normalized to standard bond lengths.[23] For hexane no hydrogen atoms were given and a SCF/3-21G optimized geometry was used. The energies were minimized under the constraint of space-group symmetry, until the root-mean-square gradient was below 0.0001 kcal mol^{-1} Å^{-1}. The cutoff radius was taken to be 20 Å, which was sufficient for convergence of the electrostatic energy to within a few hundredths of a kcal/mol. The final geometries and lattice energies were compared

TABLE 1: Optimized Crystal Structures and Their Packing Energy[a]

	a	*b*	*c*	α	*β*	*γ*	Δ*X*	Δ*θ*	*ρ*	*E*
					Hexane (HEXANE) $P\bar{1}$					
exptl	4.17	4.70	8.57	96.6	87.2	105.0			0.89	12.8
min	4.03	4.46	8.64	97.6	86.8	102.3	0	6.5	0.95	12.6
					Heptane (HEPTAN01) $P\bar{1}$					
exptl	4.15	19.97	4.69	91.3	74.3	85.1			0.89	14.5
min	4.02	20.02	4.44	90.9	77.9	84.3	0.01	4.0	0.96	14.5
					Octane (OCTANE10) $P\bar{1}$					
exptl	4.22	4.79	11.02	94.7	84.3	105.8			0.89	17.2
min	4.01	4.45	11.11	96.1	83.1	102.0	0	7.0	0.99	16.9

	a	*b*	*c*	*β*	Δ*X*	Δ*θ*	*ρ*	*E*
				Cyclohexane (CYCHEX) *C*2/*c*				
exptl	11.23	6.44	8.20	108.8			1.00	11.8
min	11.26	6.28	7.92	108.7	0	1.0	1.05	12.1
				Tetrahydrofurane (BUNJAV) *C*2/*c*				
exptl	6.08	8.91	7.74	106.1			1.19	10.9[b]
min	6.07	9.04	7.52	107.2	0.01	0.6	1.22	12.9
				Dioxane-I (CUKCIU10) High-T phase (stable 272.9-285 K) $P2_1/c$				
exptl	4.58	9.18	5.82	99.6			1.21	13.5[b]
min(AMMS-I)	4.18	9.29	5.73	98.3	0	5.1	1.33	16.6
min(AMMs-II)	4.17	9.38	5.66	98.6	0	5.3	1.34	17.1
				Dioxane-II (CUKCIU02) Low-T phase (stable < 272.9 K) $P2_1/c$				
exptl	5.72	6.46	6.13	99.9			1.31	14.0[b]
min(AMMs-I)	5.70	6.13	6.37	104.5	0	5.8	1.36	16.2
min(AMMs-II)	5.64	6.12	6.50	106.0	0	4.6	1.36	16.5
				Diethylether (DETYLE) $P2_12_12_1$				
exptl	11.81	8.07	10.85	90			0.95	9.4[b]
min	11.98	7.81	10.48	90	0.04/0.08	3.9/1.5	1.00	11.4
				Ethanol (ETANOL) *Pc*				
exptl	5.38	6.88	8.26	102.2			1.02	12.5[b]
min(AMMs-I)	5.34	6.58	8.39	99.6	0.02/0.03	2.9/12.5	1.05	14.2
min(AMMs-II)	5.35	6.66	8.37	101.3	0.01/0.04	1.0/8.6	1. 05	13.3
				1,6-Hexanediol (FECCOF) $P2_1/c$				
exptl	8.03	5.10	18.30	111.1			1.12	31.0
min	7.83	4.93	18.27	112.9	0.04	4.0	1.21	32.3

[a] Cell parameters are given in Ångstroms and degrees. Δ*X* is the net translation of the center of mass (Ångstroms), calculated from the difference in fractional coordinates in order to exclude contributions from deformation of the cell. Δ*θ* is the net rotation of the molecule (degrees), defined by three non-hydrogen atoms in the molecule. This may contain deformation of the molecule, which makes this number more uncertain with increased flexibility. Two values for Δ*X* and Δ*θ* are given for structures containing two independent molecules. *ρ* is the density (g cm^{-3}). *E* is the packing energy (kcal/mol), defined as $E_{\text{optimized molecule}} - E_{\text{crystal}}$. Experimental data are calculated from heats of sublimation in ref 24 by applying a 2*RT* correction. [b] ΔH_s° is taken to be $\Delta H_{\text{fusion}} + \Delta H_{\text{vaporization}}^\circ$.

with the experimental values. The results are given in Table 1. The packing energy is calculated as $-(E_{\text{crystal}} - E_{\text{optimized molecule}})$. This definition includes molecular deformation energy, which is in all cases rather small (maximum ∼0.5 kcal/mol for 1,6-hexanediol).

Generally, the experimental structures are very well maintained, only the density is increased upon minimization. This is not unexpected, because the energy minima correspond to structures at 0 K, while experimental structures were measured at a certain temperature. In general, one cannot expect a better agreement than a few tenths of an Ångstrom and several degrees, due to the neglect of thermal motions.[25] Results for hydrocarbons, ethers and alcohols are equally satisfactory, which supports the transferability of the potential to molecules other than methanol. Strangely, one of the larger deviations occurs for ethanol, in the orientation of the second molecule. The largest structural changes occur for dioxane-I. This is, however, a high-temperature phase, measured only 6° below the melting point. Temperature effects are expected to play a considerable role under these circumstances.

Comparing lattice energies to sublimation enthalpies is full of uncertainties.[26] Sublimation enthalpies are measured at a certain temperature $\Delta H_s(T)$, and calculated packing energies correspond to 0 K. In theory reduction to 0 K can be achieved by

$$\Delta H_s(T) = \Delta H_s(0\ K) + \int_0^T \Delta C_p(T')\mathrm{d}T'$$

The ΔC_p integral cannot be easily evaluated. Although C_p values are tabulated up to very low temperatures for the condensed phases of many of the compounds considered in this study, these are useless without accurate C_p values of the gas. Therefore, one usually assumes that the intramolecular vibrations are not influenced by the crystal packing (which is especially questionable for flexible molecules). If one then approximates the remaining contributions to the specific heat by 6*R* for the crystal and 4*R* for the gas, the ΔC_p integral is reduced to $-2RT$. Another problem is that the sublimation enthalpy at 0 K (ΔH_s-(0 *K*)) also contains the zero-point contribution to the lattice energy, which is missing from calculated packing energies. For naphthalene this has been calculated to be ∼0.5 kcal.[27]

Considering all these uncertainties in the comparison, deviations of a few kcal/mol cannot be considered significant. Then, all values are in agreement with the experimental heats of sublimation. The largest deviation occurs for 1,4-dioxane. We

TABLE 2: Optimized Crystal Structures for Six Hexapyranoses[a]

	a	b	c	ΔX	$\Delta\theta$	ρ	$\Delta\tau_{OH}$	$\Delta\tau_{CO}$
α-D-galactose (ADGALA01)								
exptl	5.90	7.84	15.68			1.65		
min	5.88	7.66	15.77	0.09	2.3	1.68	7.7	6.6
α-D-glucose (GLUCSA01)								
exptl	10.37	14.85	4.97			1.56		
min	10.20	15.01	4.85	0.07	3.4	1.61	4.4	2.8
α-D-talose (ADTALO01)								
exptl	8.10	12.13	7.66			1.59		
min	8.25	11.78	7.62	0.06	4.4	1.62	13.4	1.5
β-D-allose (COKBIN)								
exptl	4.92	11.93	12.81			1.59		
min	4.83	11.74	12.85	0.12	2.2	1.64	12.0	2.6
β-D-galactose (BDGLOS01)								
exptl	12.66	7.77	7.70			1.58		
min	11.49	8.22	7.61	0.09	5.4	1.66	17.6	1.6
β-D-glucose (GLUCSE01)								
exptl	6.60	9.01	12.72			1.58		
min	6.62	8.63	12.92	0.05	5.7	1.62	20.6	8.4

[a] All structures in space group $P2_12_12_1$. ΔX, $\Delta\theta$, and ρ as in Table 1. $\Delta\tau_{OH}$ is the average deviation of the torsion involving the hydroxyl hydrogens, $\Delta\tau_{CO}$ is the deviation of the exocyclic O−C−C−O torsion (both in degrees).

think that the experimental value is most probably underestimated, being only slightly larger than for the hydrocarbons of equal size. Our calculations indicate that there is a substantial electrostatic contribution to the energy, resulting in a higher packing energy than for the alkanes, where this contribution is practically zero. The AMMs-I overestimate the packing energy somewhat compared to the AMMs-II results. However, the differences are small and so are the differences in the optimized structures. This confirms that the level at which the AMMs-I are obtained is a reasonable choice.

A problem is that the energy ordering of dioxane-I and -II is in contradiction with the thermodynamic data, which report a value 0.56 kcal/mol for the enthalpy of the phase change II → I at 272.9 K.[24] Because one of the larger structural changes is observed upon optimizing dioxane-I, the calculated energy difference does not really correspond to the experimental dioxane-I structure anymore. Indeed, if we optimize the two structures with their cell parameters fixed to the experimental values, dioxane-II is either 0.16 kcal/mol more favorable than dioxane-I (AMMs-I) or equally favorable (AMMs-II). Nevertheless, without generating any hypothetical crystal structures, we predict at least one structure to have a better packing energy than the experimentally observed low-temperature phase of dioxane-II.

For the six hexapyranoses, the potential performs rather similarly (Table 2). Only for β-D-galactose there occurs a rather large change in a cell axis. Changes for the hydroxyl torsions ($\Delta\tau_{OH}$) should not be given too much significance, since hydrogen atoms cannot be determined accurately by X-ray diffraction. The α-D-glucose structure was determined by neutron diffraction, and it is encouraging to see that $\Delta\tau_{OH}$ is lowest for this structure. In general, the combination of MM3 with the ab initio potential is seen to give a satisfactory reproduction of experimental crystal structures, even for these more complex molecules.

4. Crystal Structure Prediction

4.1. General Considerations. In the first stages of crystal structure prediction thousands of crude structures are generated, and it would be computationally too expensive to apply the ab initio potential in this part of the process. It is more efficient to use a standard force field in the search for possible structures. In this work we generated structures either using the method of Gdanitz and Karfunkel as implemented in MSI's Polymorph Predictor,[28,29,1] or using our own program UPACK.[6,7] We employed the standard force fields DREIDING[30] and OPLS,[31] respectively; in the first stages of crystal structure generation in UPACK the united-atom force field UNITAT[7] was also used. Only structures with a reasonably low energy were subsequently studied with the elaborate ab initio potential. For comparison, these structures were also minimized in all the mentioned standard force fields.

TABLE 3: Space-Group Statistics for Organic Crystals Consisting of One Type of Molecule, with Either One or Two Independent Molecules: Taken from Cole[33]

$Z''=1$		$Z''=2$	
SPGR	%	SPGR	%
$P2_1/c$	43.9	$P2_1/c$	31.5
$P\bar{1}$	16.4	$P\bar{1}$	28.7
$P2_12_12_1$	14.7	$P2_1$	14.6
$P2_1$	7.1	$P2_12_12_1$	7.9
$Pbca$	5.5	$P1$	3.7
$C2/c$	4.0	$Pca2_1$	2.8
$Pna2_1$	2.0	$Pna2_1$	2.4
Cc	1.1	$Pbca$	2.4
$Pca2_1$	0.9	$C2/c$	1.5
$C2$	0.6	Cc	1.0
$P1$	0.4	Pc	0.8
$Pbcn$	0.4	$C2$	0.8
Pc	0.4	$P2/c$	0.5

After each minimization, structures that were initially different may have converged towards the same minimum. Such duplicate entries were removed by applying a clustering algorithm based on lists of distances between specific atom types. This method was described previously,[32] but the program has been extensively improved to speed up the clustering process.

The normal procedure in most structure predictions is to generate structures in the most abundant space groups in the CSD. However, the overall statistics include cocrystals, and structures with occupied special positions. This is not consistent with the structures that are generated in the predictions. Therefore, one should use space-group statistics for crystals consisting of one type of molecule, with the appropriate number of crystallographically independent molecules (Z'' [4]). Cole[33] derived such statistics by considering only entries in which the number of discrete bonded units in the asymmetric unit equals the relative multiplicity Z' given in the CSD (Table 3). Indeed, a cocrystal or a crystal with two occupied special positions may have $Z'=1$, but these cases will be excluded from the statistics because there is more than one molecule in the asymmetric unit ($Z''>1$). The main difference between these statistics and the overall space-group statistics is the complete absence of space groups possessing mirror planes (like *Pnma* and $P2_1/m$).[33,34]

4.2. Methanol. Methanol has a least three crystalline phases. We can expect only to find the low-temperature α-phase ($P2_12_12_1$),[35,36] because the high-temperature β-phase is disordered due to a rapid interchange of two puckered forms of the infinite hydrogen-bonded chains.[37,36] In the α-phase the hydrogen-bonded chains are puckered: the carbon atoms are no longer in the plane defined by the oxygen atoms. The structures of the intermediate γ -phase and an additional, metastable, phase[35] are unknown.

The search for hypothetical crystal structures was performed with MSI's Polymorph Predictor,[28,29,1] using the standard

TABLE 4: The Experimental α-Phase of Methanol, as Found in Structure Predictions with Various Force Fields[a]

					search-1		search-2	
	a	*b*	*c*	*E*	Δ*E*	rank	Δ*E*	rank
exptl	4.87	4.64	8.87	11.1				
AMMs-I	5.02	4.68	8.79	11.7	0	1	0.06	4
AMMs-II	5.07	4.66	8.71	11.6	0	1	0.06	3
DREIDING	4.90	5.42	8.42	12.8	0	1	0	1
OPLS	4.72	4.95	8.33	13.1	0	1	0	1
UNITAT	4.48	5.11	8.75	8.5	0.03	5	0.03	11

[a] Cell parameters are given in Ångstroms and degrees. Δ*E* (kcal/mol) is the energy difference with the lowest energy structure that was found. A rank of *r* means that the structure was the *r*th lowest in energy. *E* is the packing energy (kcal/mol), defined as $E_{\text{optimized molecule}} - E_{\text{crystal}}$. Experimental value is $\Delta H_{\text{fusion}} + \Delta H^{\circ}_{\text{vaporization}} + 2RT$.[24] Search-1 was performed with one independent molecule and search-2 with two independent molecules.

program settings. The DREIDING-2.21 force field[30] as implemented in Cerius[2] was used as standard force field. Charges were derived by fitting to the electrostatic potential at the HF/6-31G** level of theory in a HF/6-31G** optimized geometry, using the programs GAMESS-UK[12] and MOLDEN.[14]

In a first run (Search-1) structures with one independent molecule were generated, in the eight most abundant space groups (Table 3). The results for all space groups were clustered together. The remaining 95 structures were subsequently optimized within our potential (cutoff 20 Å) with two approaches for the AMMs. In the first, the AMMs-II were used, calculated in the structure from microwave spectroscopy.[38] In the second approach, the AMMs-I were used with the procedure of iteratively recalculating the AMMs.

The energy range of the structures is ∼5 kcal/mol. Of course, all structures with a low energy have infinite hydrogen-bonded chains. The variations are in the puckering and in the stacking of the chains. Some structures differ only in the direction of the hydrogen-bonded chains: parallel in the one and antiparallel in the other. The lowest energy structures all contain the same puckered hydrogen-bonded chain that is present in the α-phase. The results for the AMMs-I and AMMs-II calculations are very similar, both in terms of energies and structures.

The experimental low-temperature phase is ranked first both in the AMMs-I and -II calculation (Table 4). For comparison, the cell parameters and rankings in the DREIDING, OPLS, and UNITAT force fields are also given. In all these force fields the experimental structure is ranked first as well, but the structure is reproduced less satisfactorily. It should also be noted that in those force fields low-energy structures occur which have a flat hydrogen-bond geometry. Such a flat geometry occurs in the β-phase only as the dynamical average of two puckered structures. The prediction of such geometries as favorable lattice energy minima is most likely caused by an inadequate description of the directionality of hydrogen bonding in those force fields.

A more extended search (search-2) for possible crystal structures was performed using two independent molecules in the twelve most abundant space groups (Table 3). This resulted in ∼ 900 structures. The number of possible structures increases drastically upon raising the number of independent molecules, which is illustrated in Figure 1. Ideally, all structures from the previous run should be present in this run as well. This is certainly not the case, but from the plot one can see that at least the lowest-energy structures from search-1 were found again. The completeness of the search can probably be improved considerably by changing some parameters of the search and minimization protocol, which will of course not be without computational costs. We did not investigate this any further, but just note that our set of $Z'' = 2$ structures is not exhaustive. It worthwhile to notice that among the $Z'' = 1$ structures that were found in search-2 some unusual space groups have been encountered ($I\bar{4}$, $P4_1$, $P4_12_12$, $P4_32_12$, *Aba*2, *Iba*2, *Fdd*2). These would have been hard to find by systematically going down the list of most abundant space groups for one independent molecule. The occurrence of those space groups is in accordance with the observation that monoalcohols are much more likely to crystallize in high-symmetry space groups.[39]

For 500 of those structures the energy was minimized in the ab initio potential, using either the AMMs-I or -II. The results are very similar to those of search-1. Some $Z'' = 2$ structures have energies slightly lower than the lowest energy $Z'' = 1$ structure. Upon visual inspection it was seen that most of the low-energy structures with two independent molecules are more variations in packing of the same puckered hydrogen-bonded chain. With two independent molecules one can, for example, alternate the stacking of layers of hydrogen-bonded chains from two different $Z'' = 1$ structures. Obviously, this results in energies that are very similar to both these $Z'' = 1$ structures. An example is structure C in Figure 2, which is a alternation of A and B stackings. Structure D, the global minimum in the ab initio potential, is yet another type of variation that is possible with two independent molecules. Here the direction of the hydrogen-bonded chains alternates: in structure A neighboring hydrogen-bonded chains are antiparallel, whereas in D they are antiparallel in pairs. It is somewhat hard to imagine that such alternations would grow without disorder.

4.3. Ethanol. Ethanol crystallizes with two independent molecules in space group *Pc*.[40] The molecules differ in conformation: the hydroxyl hydrogen is either *trans* or *gauche*. Hypothetical crystal structures for ethanol were generated using MSI's Polymorph Predictor,[28,29,1] using the standard program settings. Charges were calculated in the *trans* conformation, at the same level of theory as for methanol. The conformational flexibility of the hydroxyl group was treated with an approach developed by Verwer.[41] In this approach a modified version of the DREIDING force field is used, in which the van der Waals parameters for O and H_O (hydroxyl hydrogen) are changed in such a way that the hydroxyl hydrogen lies completely within the oxygen atom. Also the torsional potential on the hydroxyl hydrogen atom is removed, while its charge is left unchanged. The idea of the method is that in this way the hydroxyl hydrogen can adjust freely to the packing requirements. The advantage over the UPACK united atom approach[6,7] is that one does not have to bother where to place the hydrogen atom after the first minimizations. The disadvantage is that it is more uncertain whether all possible packing arrangements are found. The structures were subsequently refined in the normal DREIDING-2.21 force field.

The predictions were started with ethanol in the trans geometry as building block. The search was performed using two independent molecules, in the twelve most abundant space groups (Table 3). The Polymorph Predictor run resulted in ∼1200 different structures. Although we started out from two ethanol molecules in the *trans* geometry, the search produced structures containing all combinations of *trans* and *gauche* conformations. We did not investigate whether starting from different conformations would produce different results. Probably, the starting conformation does bias the sampling of the

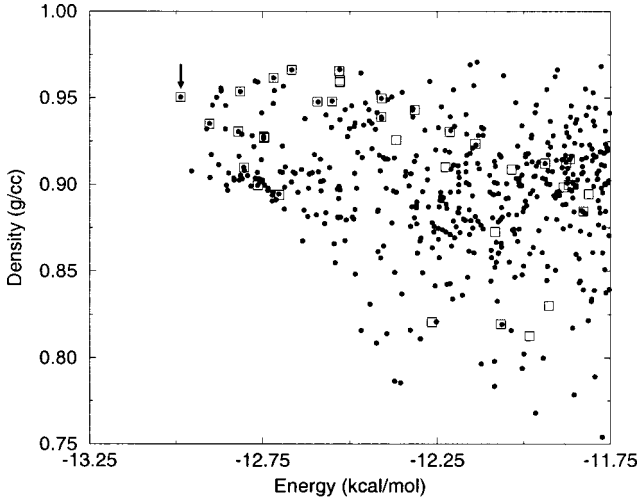

Figure 1. Energy versus density plot of the results of Polymorph Predictor search for methanol, using either one (□) or two (●) independent molecules. The energy is the total energy in the DREIDING force field, the arrow indicates the experimentally observed low-temperature structure.

search, but for now we note that Verwer's modified OH approach was successful, as it located the experimental crystal structure.

For computational reasons, we only used the 500 best structures within the DREIDING force field as starting structures within the ab initio force field (cutoff 20 Å), using the procedure of iteratively recalculating the AMMs. Results are given in Table 5. The predictions using the ab initio model are much better than those using any of the standard force fields, both in energy and in rankings. It is worth noting that the first three structures in the ab initio model contain only one independent molecule. For monoalcohols, it has been argued that a packing problem exists that favors multiple independent molecules or high-symmetry space groups.[39] Apparently, such a packing problem is not a determining factor for the crystallization of ethanol with two independent molecules.

4.4. Dioxane. Structures for 1,4-dioxane with one independent molecule in the asymmetric unit were generated using the UPACK program.[6,7] The search was performed with the UNITAT force field, but afterwards the energy was minimized in OPLS. DREIDING was also employed, for which charges were calculated at the same level of theory as for methanol. The search was performed in the twelve most abundant space groups, and delivered 455 structures.

Subsequently, the structures were optimized in the ab initio potential, using both AMMs-I and -II calculated at the MM3 optimal geometry. Because the molecule is fairly rigid there was no need for recalculation of AMMs. Since the molecule has no dipole moment, we used a cutoff radius of 15 Å to speed up computation.

Results are given in Table 6. All force fields have problems to maintain the experimental structure of dioxane-I, for which changes up to 1.3 Å in cell axes occur. It must be remembered that dioxane-I is only stable just below the melting point, so thermal effects are expected to be very significant. Although the optimized structure of dioxane-I differs quite substantially among the force fields, it is in all cases the energetically most favorable one. By this mere consistency, one might be led to believe that this is not just due to model errors, but that it really has the best packing energy at 0 K. The only way to explain this from an energetic point of view is to assume that above a certain temperature dioxane-II is favored over such an optimized dioxane-I structure by vibrational entropy, and that above 273 K the true dioxane-I structure is again favored over dioxane-II. For the moment, this is purely speculation, and force field errors are a more likely cause of this wrong energy ordering.

We note that dioxane-I has the lowest electrostatic energy of all hypothetical structures, whereas dioxane-II has lower (but not the lowest) dispersion energy. This suggests that when errors in the repulsion term are the cause of artificially favorable electrostatic energy for dioxane-I, repairing these flaws would increase the relative favorability of dioxane-II. Anyhow, the predictions based on the ab initio potential are superior to those based on the standard force fields. The ab initio potential predicts the smallest energy difference between the two phases, and the lowest ranking for dioxane-II.

4.5. Propane. The crystal structure of propane was predicted for a workshop on crystal structure prediction that was organized by the Cambridge Crystallographic Data Centre. This workshop involved a blind test on three compounds, and participants were also challenged to predict the crystal structure of propane. To us, this also formed a blind test, because the crystal structure of propane was not yet published when we submitted our results to the referee of the meeting. Complete results of this workshop will be published elsewhere.[42]

Propane crystallizes below 83 K, and the structure was recently solved at 30 K.[43] The crystal structure contains one independent molecule in space group $P2_1/n$. Structures for propane with one independent molecule in the asymmetric unit were generated using the UPACK program.[6,7] All computational details of the search and the subsequent minimizations were identical to those for dioxane.

Results are given in Table 7. The search delivered ~600 structures, and the energetic differences between the best hypothetical crystal structures are extremely small. Still, the ab initio model correctly predicts the experimental structure as the energetically most favorable one. This result is further support for the applicability of our potential to saturated hydrocarbons. Moreover, because the electrostatic energy contribution in propane is very small, this is also a validation of our dispersion-energy model. DREIDING and OPLS do not predict the experimental structure to have the best packing energy, although the ΔE's are very small. Optimization of the experimental structure in the UNITAT force field (where propane consists of only three united C "atoms") leads to a large rotation of the molecule (23°), resulting in a different space group. Remarkably, this is also the structure with the lowest energy in that force field; nevertheless, the experimental structure is so deformed that this can hardly be considered a correct prediction.

5. Discussion and Conclusions

We have shown that is it possible to derive from ab initio data an accurate potential that is transferable to the crystalline phase. Crystal structure relaxations show that the potential can be successfully applied to crystals of molecules such as alkanes, ethers, and alcohols. Crystal structure generations indicate that the potential is accurate enough to correctly predict the crystal structures of methanol and ethanol within an energy range of 0.1 kcal/mol.

For dioxane the results are somewhat less satisfactory: the energy ordering of phase I and II is not correct. A more detailed investigation, including some new ab initio interaction energy calculations on the one hand, and lattice dynamics calculations and molecular dynamics simulations on the other hand, would be needed to pin-point its exact cause. Still, the energy difference between the low temperature phase and the energetically most

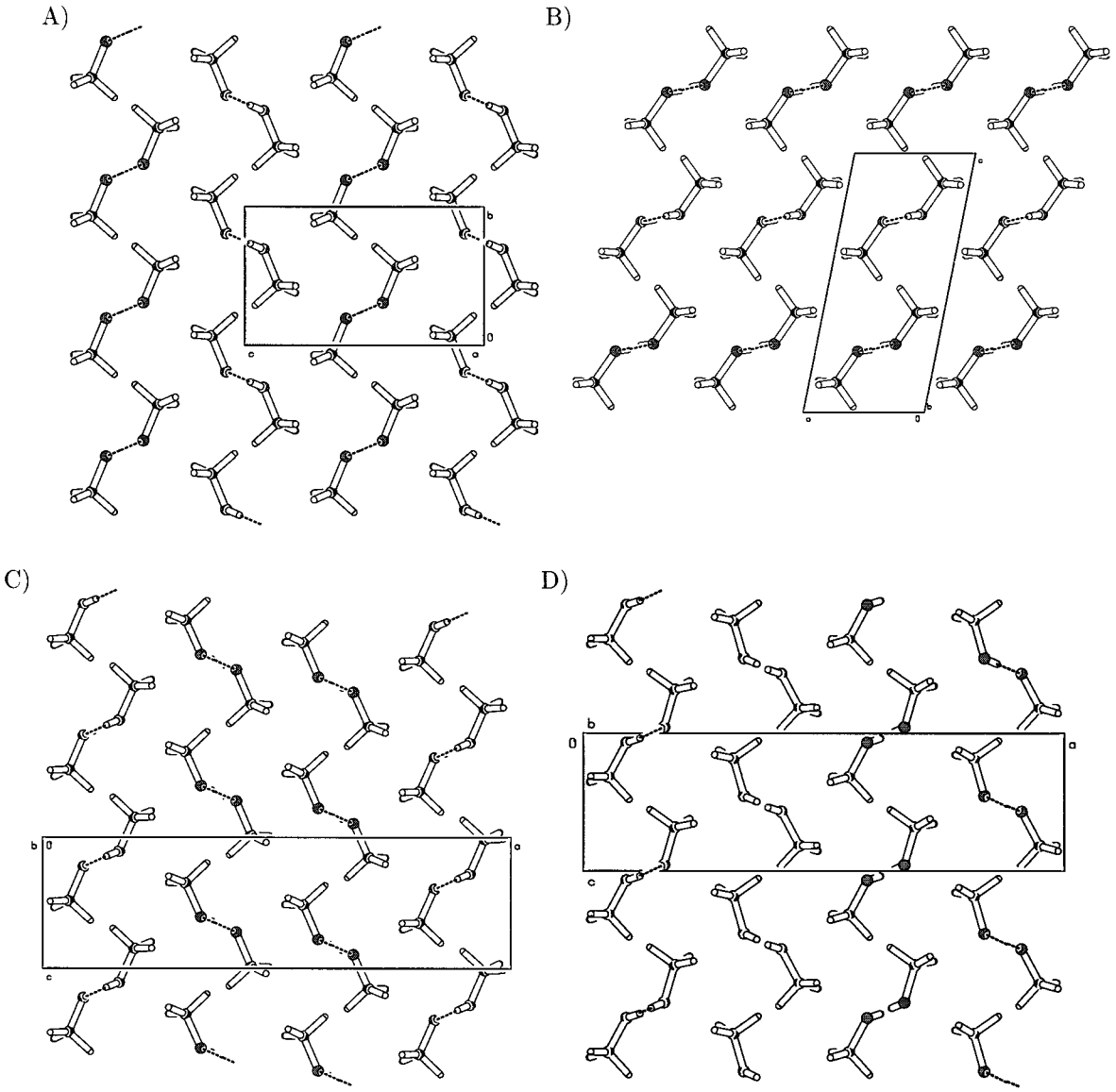

Figure 2. Methanol structures with one (A,B) or two (C,D) independent molecules. A is the low-temperature α phase, B is the third $Z'' = 1$ structure, with an alternative stacking of the hydrogen-bonded chains, D is the minimum energy $Z'' = 2$ structure. Hydrogen-bonded chains run perpendicular to the paper, with all OH vectors either pointing *up* (open O atoms) or *down* (black O atoms). In A the hydrogen-bonded chains run *down, up, down, up* (from left to right), while in D they run *up, up, down, down.*

TABLE 5: The Experimental Structure of Ethanol, as Found in Structure Predictions with Various Force Fields[a]

	a	*b*	*c*	*β*	*E*	Δ*E*	rank
exptl	5.38	6.88	8.26	102.2	12.5		
AMMs-I	5.34	6.58	8.39	99.6	14.2	0.09	4
DREIDING	5.46	7.05	8.74	101.9	14.3	0.61	201
OPLS	5.27	6.67	8.23	103.6	15.1	0.43	62
UNITAT	5.16	6.98	8.44	101.7	10.7	0.21	72

[a] Symbols defined as in Table 4.

TABLE 6: Experimental Structures of Dioxane, as Found in Structure Predictions with Various Force Fields[a]

	a	*b*	*c*	*β*	*E*	Δ*E*	rank
			Dioxane-I				
exptl	4.58	9.18	5.82	99.6	13.5		
AMMs-I	4.18	9.29	5.73	98.3	16.6	0	1
AMMs-II	4.17	9.38	5.66	98.6	17.1	0	1
DREIDING	4.33	8.12	6.88	93.8	14.7	0	1
OPLS	4.13	7.82	6.68	94.0	15.7	0	1
UNITAT	4.09	8.59	6.79	92.0	13.5	0	1
			Dioxane-II				
exptl	5.72	6.46	6.13	99.9	14.0		
AMMs-I	5.70	6.13	6.37	104.5	16.2	0.59	14
AMMs-II	5.64	6.12	6.50	106.0	16.5	0.45	10
DREIDING	5.95	6.56	6.54	100.2	13.0	1.71	40
OPLS	5.64	6.23	6.42	102.2	14.1	1.64	34
UNITAT	5.76	6.66	6.57	104.6	12.1	1.39	145

[a] Symbols defined as in Table 4.

favorable structure is not very dramatic (~0.5 kcal/mol) and superior to the results of the standard force fields. For propane our model predicts the experimentally observed structure as the energetically most favorable one, unlike any of the standard force fields. This result is further support for the accuracy of the hydrocarbon part of our potential.

Some preliminary work on hypothetical structures of the six hexapyranoses indicates that for those molecules still structures can be proposed that are significantly more favorable than the experimental one. One of the reasons for this may be the use of only intermolecular polarization, which becomes problematic in these larger molecules. In the present model intramolecular polarization is treated a priori, within the calculation of the AMMs. This leads to an incorrect description of nonadditivity,

TABLE 7: Experimental Structure of Propane, as Found in Structure Predictions with Various Force Fields[a]

	a	*b*	*c*	*β*	*E*	Δ*E*	rank
exptl	4.15	12.61	6.98	91.3	7.1		
AMMs-I	4.07	12.58	6.78	91.8	6.6	0	1
AMMs-II	4.06	12.57	6.78	91.7	6.7	0	1
DREIDING	4.39	12.83	6.99	91.6	6.7	0.11	12
OPLS	4.14	12.41	6.60	91.2	7.6	0.03	5
UNITAT	4.10	12.40	7.00	90[b]	6.6	0	1

[a] Symbols defined as in Table 4, apart from the experimental *E* value, which is calculated from ΔH_s in ref 24 by applying a $2RT$ correction. [b] Space group changed to *Pnma* , $Z' = 1/2$.

which is most serious in the case of intramolecular hydrogen bonds. The solution would be to incorporate intramolecular polarization as well in our model. This, however, raises many unresolved issues concerning electrostatic and polarization interactions between neighboring atoms.

As we expected,[32] a very accurate force field does not suddenly isolate one structure from all other ones. Improving the accuracy of potential energy functions does not reduce the number of low-energy structures drastically, but it allows us to limit the energy range of structures that have to be considered as true candidates for experimental observation. Of such a limited number of structures one could then investigate in more detail their thermodynamic and kinetic properties, which will be needed to arrive at genuine structure prediction.

Acknowledgment. This work was supported by the Council for Chemical Sciences of the Netherlands Organization for Scientific Research (CW-NWO), in the framework of the PPM/CMS-crystallization project. We thank Mr. S. Polling for his work on the UPACK predictions for dioxane and Dr. P. Verwer for providing us with his modified force field for hydroxyl groups. We are grateful to Professor R. Boese for helpful discussions on the crystal structures of methanol.

References and Notes

(1) Verwer, P.; Leusen, F. J. J. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 1998; Vol. 12, Chapter 7.
(2) Gdanitz, R. J. *Curr. Opin. Solid State Mater. Sci.* **1998**, *3*, 414.
(3) Gavezzotti, A. *Acc. Chem. Res.* **1994**, *27*, 309.
(4) van Eijck, B. P.; Spek, A. L.; Mooij, W. T. M.; Kroon, J. *Acta Crystallogr.* **1998**, *B54*, 291.
(5) Mooij, W. T. M.; van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Eijck, B. P. *J. Phys. Chem.* **1999**, *103*, 9872.
(6) van Eijck, B. P.; Mooij, W. T. M.; Kroon, J. *Acta Crystallogr.* **1995**, *B51*, 99.
(7) van Eijck, B. P.; Kroon, J. *J. Comput. Chem.* **1999**, *20*, 799.
(8) Ponder, J. W. *TINKER: Software Tools for Molecular Design*, Version 3.6. Washington University of Medicine, St. Louis, 1998. Available from: http://dasher.wustl.edu/tinker/.
(9) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes, the Art of Scientific Computing*; Cambridge University Press: Cambridge, 1989; p 308.
(10) van Eijck, B. P.; Kroon, J. *J. Phys. Chem.* **1997**, *B101*, 1096.
(11) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, 1987; p 65.
(12) GAMESS-UK is a package of ab initio programs written by M. F. Guest, J. H. van Lenthe, J. Kendrick, K. Schoffel, P. Sherwood, and R.J. Harrison, with contributions from R.D. Amos, R.J. Buenker, M. Dupuis, N.C. Handy, I.H. Hillier, P.J. Knowles, V. Bonacic-Koutecky, W. von Niessen, V.R. Saunders, and A.J. Stone. The package is derived from the original GAMESS code due to M. Dupuis, D. Spangler and J. Wendoloski, NRCC Software Catalog, Vol. 1, Program No. QG01 (GAMESS), 1980.
(13) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Johnson, P. M. W. G. B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian94*, Revision B.2; Gaussian, Inc.: Pittsburgh, Pennsylvania, 1995.
(14) Schaftenaar, G.; Noordik, J. H. *J. Comput.-Aided Mol. Design.* **1999**. In press. MOLDEN is available from: http://www.caos.kun.nl/~ schaft/molden/molden.html.
(15) Halgren, T. A. *J. Am. Chem. Soc.* **1992**, *114*, 7827.
(16) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.
(17) Lii, J.-H.; Allinger, N. L. *J. Phys. Org. Chem.* **1994**, *7*, 591.
(18) Price, S. L.; Stone, A. J. *J. Chem. Soc., Faraday Trans.* **1992**, *88*, 1755.
(19) Koch, U.; Stone, A. J. *J. Chem. Soc., Faraday Trans.* **1996**, *92*, 1701.
(20) Dudek, M. J.; Ponder, J. W. *J. Comput. Chem.* **1995**, *16*, 791.
(21) Reynolds, C. A.; Essex, J. W.; Richards, W. G. *J. Am. Chem. Soc.* **1992**, *114*, 9075.
(22) Allen, F. H.; Kennard, O. *Chem. Des. Automation News* **1993**, *8*, 31.
(23) Allen, F. H.; Kennard, O.; Watson, D. G.; Brammer, L.; Orpen, A. G.; Taylor, R. J. *J. Chem. Soc., Perkin Trans II* **1987**, S1.
(24) Mallard, W. G.; Linstrom, P. J., Eds. *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*; National Institute of Standards and Technology, Gaithersburg, MD, 20899, November 1998. See: http://webbook.nist.gov/.
(25) Pertsin, A. J.; Kitaigorodsky, A. I. *The Atom-Atom Potential Method*; Springer: Berlin, 1986; p 81.
(26) Gavezzotti, A.; Filippini, G. In *Theoretical Aspects and Computer Modeling of the Molecular Solid State*; Gavezzotti, A., Ed.; John Wiley & Sons: New York, 1997; Chapter 3.
(27) Filippini, G.; Gramaccioli, C. M.; Simonetta, M.; Suffritti, G. B. *Chem. Phys.* **1975**, *8*, 136.
(28) Gdanitz, R. J. *Chem. Phys. Lett.* **1992**, *190*, 391.
(29) Karfunkel, H. R.; Gdanitz, R. J. *J. Comput. Chem.* **1992**, *13*, 1171.
(30) Mayo, S. L.; Olafson, B. D.; Goddard, W. A., III. *J. Phys. Chem.* **1990**, *94*, 8897.
(31) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
(32) Mooij, W. T. M.; van Eijck, B. P.; Price, S. L.; Verwer, P.; Kroon, J. *J. Comput. Chem.* **1998**, *19*, 459.
(33) Cole, J. C. PhD thesis, University of Bristol, Bristol, England, 1995.
(34) Brock, C. P.; Dunitz, J. D. *Chem. Mater.* **1994**, *6*, 1118.
(35) Torrie, B. H.; Weng, S. X.; Powell, B. M. *Mol. Phys.* **1989**, *67*, 575.
(36) Nussbaumer, M. PhD thesis, Universität Essen, Essen, Germany, 1996.
(37) Tauer, K. J.; Lipscomb, W. N. *Acta Crystallogr.* **1952**, *5*, 606.
(38) Gerry, M. C. L.; Lees, R. M.; Winnewisser, G. *J. Mol. Spectrosc.* **1976**, *61*, 231.
(39) Brock, C. P.; Duncan, L. L. *Chem. Mater.* **1994**, *6*, 1307.
(40) Jönsson, P.-G. *Acta Crystallogr.* **1976**, *B32*, 232.
(41) Verwer, P. To be published.
(42) Lommerse, J.; Motherwell, S.; all participants of the workshop. In preparation.
(43) Boese, R.; Weiss, H.-C.; Bläser, D. *Angew. Chem. Int. Ed.* **1999**, *38*, 988.