# Ab Initio Crystal Structure Predictions for Flexible Hydrogen-Bonded Molecules

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

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

*Received November 8, 1999*

**Abstract:** Crystal structure predictions for glycol and glycerol are reported. A series of increasingly accurate energy calculations is applied, and the final predictions are based solely on ab initio derived energies. A recently developed transferable ab initio potential is used for intermolecular interactions, augmented with ab initio derived conformational energies. The experimental structure of glycol was predicted with a low energy, 1.1 kJ/mol above the global minimum. For glycerol the experimental structure corresponded to the global minimum. This latter result provides a proposal for the positions of the hydrogen atoms in the crystal structure of glycerol. A three-dimensional hydrogen-bonded network is formed which consists only of intermolecular hydrogen bonds. Together with previous work, the ab initio intermolecular potential has now been applied to predict the crystal structures of six different compounds. The energy difference between the observed crystal structure and the global energy minimum varied from 0 to 2 kJ/mol. Standard force fields fail to consistently produce such low values. This demonstrates the importance of highly accurate force fields in crystal structure prediction.

## Introduction

Can crystal structures be predicted solely from their molecular constitution? Since John Maddox in his citation-classic stated that "one of the continuing scandals in the physical sciences is that it remains in general impossible to predict the structure of even the simplest crystalline solids from a knowledge of their chemical composition",[1] research into this area has intensified, driven by both academic and industrial interest. The way molecules are packed in a crystal influences the properties of many practical materials, such as pigments, pharmaceuticals, explosives, and nonlinear optical materials. Therefore, the ability to predict the crystal packing for an unknown compound would be desirable. This packing need not be unique, as is shown by the existence of polymorphic forms for many organic crystals (see, e.g., ref 2 and references therein). Polymorphism is of particularly high interest to the pharmaceutical industry, where new crystalline forms can cause problems with the handling, shelf life, bio-availability, and patent protection of a product.

Nowadays a multitude of computational strategies exist to generate possible crystal structures. Nearly all these approaches search for structures with a low energy in some empirical force field (e.g., refs 3−5). More recently, approaches based on database statistics have been proposed (e.g., ref 6). We will not describe the details of all methods, as recent reviews are available.[7,8] At present such methods can produce a list of possible crystal structures which most probably contains the experimental structure(s), at least for fairly rigid molecules that crystallize with one independent molecule in the unit cell. In more complex cases, the structure generation can still easily fail to locate the experimental structure.[9] In energy-based predictions it is assumed that the experimental structure should correspond to the structure with the lowest energy, although it is realized that due to force field inaccuracies as well as more fundamental objections, one can only expect the experimental structure not to have an unacceptably high energy.[4] Indeed, in the approach of calculating lattice energies, thermal and kinetic effects are neglected, which is rather questionable. Such considerations may lead one to the conclusion that crystal structures are not predictable at all.[10]

Nevertheless, many crystal structure prediction studies have reported satisfactory results. Recently, the Cambridge Crystallographic Data Centre organized a blind test on crystal structure prediction.[11] Eleven participants were allowed to propose at most three structures for a few compounds. On one hand, this test showed that reliable ab initio structure prediction is still a distant goal, since no single participant predicted all structures correctly. On the other hand, four out of the five structures were predicted successfully by at least one participant, and in total seven successful predictions were done. These were exclusively based on lattice-energy calculations. So, despite all theoretical limitations, such calculations have proven to be a reasonable approach for crystal structure prediction.

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

(1) Maddox, J. *Nature* **1988**, *335*, 201.
(2) Gavezzotti, A.; Filippini, G. *J. Am. Chem. Soc.* **1995**, *117*, 12299−12305.
(3) Gavezzotti, A. *J. Am. Chem. Soc.* **1991**, *113*, 4622−4629.
(4) Karfunkel, H. R.; Gdanitz, R. J. *J. Comput. Chem.* **1992**, *13*, 1171−1183.
(5) van Eijck, B. P.; Mooij, W. T. M.; Kroon, J. *Acta Crystallogr.* **1995**, *B51*, 99−103.
(6) Hofmann, D. W. M.; Lengauer, T. *Acta Crystallogr.* **1997**, *A53*, 225−235.
(7) 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.
(8) Gdanitz, R. *J. Curr. Opin. Solid State Mater. Sci.* **1998**, *3*, 414−418.
(9) van Eijck, B. P.; Kroon, J. *Acta Crystallogr. B*, in press.
(10) Gavezzotti, A. *Acc. Chem. Res.* **1994**, *27*, 309−314.
(11) 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. *Acta Crystallogr. B* submitted.

10.1021/ja993945t CCC: $19.00 
Published on Web 03/24/2000

To increase the accuracy of the calculated lattice energies, most attention has been given to the electrostatic interactions. Impressive results in crystal structure predictions for rigid N−H hydrogen-bonded molecules have been obtained by using an elaborate ab initio derived electrostatic model (e.g., ref 12−14). Problems arose in cases where the molecules were not completely rigid.[11,15] Although the electrostatic part of such models is theoretically sound, the potential as a whole is not, since the ab initio electrostatics are combined with an empirical dispersion-repulsion potential. Recently we have gone beyond such a mixed ab initio/empirical approach and derived a complete intermolecular potential from high-level ab initio calculations on methanol dimers and trimers.[16] This potential was seen to be transferable from methanol to other alcohols, alkanes, and ethers, as well as from the gas phase to the solid phase. Crystal structure predictions based on this theoretically well-founded model were in all cases superior to predictions based on standard empirical force fields.[17]

In this work we report crystal structure predictions for flexible, hydrogen-bonded molecules, viz., glycol and glycerol. Molecular flexibility requires a highly accurate intramolecular model to complement the ab initio intermolecular potential.[16] It seems unlikely that empirical force fields can achieve the desired accuracy for conformational energies of molecules such as glycol and glycerol. Therefore, we resort in the end to ab initio calculations to obtain accurate intramolecular energetics.

During the process of structure generation many thousands of possible crystal structures are investigated. This number is so large that it is computationally too demanding to use our elaborate ab initio potential throughout the procedure. Even more so, it is out of the question to perform ab initio calculations of the intramolecular energies for all these structures. Therefore, we perform the procedure in different stages. In each consecutive stage we increase the accuracy of the calculated energies at the expense of computational effort.

In the initial stages of structure prediction a huge number of structures are considered. Therefore it is essential to use a simple model, such as a standard force field. Apart from being simple, this force field should be accurate enough to identify the experimental polymorph(s) as reasonably favorable. The best few hundred structures are taken over to the second stage where the ab initio intermolecular potential is employed. Still, the number of structures is too large to allow accurate ab initio calculations of the conformational energies. Therefore, we employ molecular mechanics for all intramolecular interactions, and optimize the crystal structures in the combined ab initio + MM potential (AI+MM). Finally, for the most favorable structures we replace the molecular mechanics conformational energies by ab initio derived values. So, in the end the predictions are based solely on nonempirical energies.

## Computational Methods

**Stage 1: Standard Force Fields.** Crystal structures were generated with the grid search algorithm of the UPACK program,[18] using the standard program settings. Crystal structure generation starts from a molecular geometry, and for flexible molecules all possible conformations have to be considered. For glycol this involves two different conformations, for glycerol six. This excludes the conformations of the hydroxyl groups, which are treated in a special manner within the UPACK program: in the initial stages united OH "atoms" are used, only in a second stage explicit hydrogen atoms are introduced.[5,18] In addition to the molecular geometry, some space groups of interest have to be chosen. In this work the eight most abundant space groups for crystal structures with one independent molecule[17,19] were investigated. The united-atom force field UNITAT[18] was used in the initial stages of the search procedure, later on UNITAT, OPLS,[20] and the carbohydrate version of CHARMM[21,22] due to Reiling, Schlenkrich, and Brickmann[23] (CHARMM-RSB) were employed.

**Stage 2: Ab Initio Intermolecular Potential + Molecular Mechanics Intramolecular Potential.** The most favorable structures to come from the search procedure were subsequently minimized within our recently developed ab initio intermolecular potential for alkanes, alcohols, and ethers,[16] augmented with an intramolecular force field. The intermolecular potential involves atomic multipole moments (AMMs), atomic dipole polarizabilities, a damped atom−atom $r^{-6}$ 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 fitting to the electrostatic potential of a SCF/DZ($2d^O$) wave function, as described previously.[16] For intramolecular interactions the MM3(96) force field[24,25] as implemented in TINKER[26] was used, including its own van der Waals parameters and bond dipoles. Our choice for MM3(96) was based on the belief that this force field currently gives the best description of intramolecular energy and structure for small organic molecules. The crystal structure optimizations were performed with a previously developed program,[17] using a cutoff radius of 20 Å. Ab initio calculations were performed using GAUSSIAN94;[27] the AMMs were derived by using fitting routines that were implemented in MOLDEN.[28]

Just as some standard force fields require the calculation of electrostatic potential derived charges, our intermolecular potential requires the calculation of atomic multipole moments. The AMMs are defined in local axes systems. In this way the moments maintain their orientation relative to the local atomic environment, to follow changes in the molecular conformation. However, it has been shown that AMMs derived in one conformation do not achieve an accurate description of intermolecular interactions when rotated to another conformation: the molecular charge distribution and the corresponding atomic multipole moments are conformation dependent.[29] For glycerol this effect can be very large: rotating the AMMs derived from one conformation to a completely different conformation can cause errors of 10 kJ/mol in the intermolecular electrostatic part of the lattice energy.

(12) Price, S. L.; Wibley, K. S. *J. Phys. Chem.* **1997**, *A101*, 2198−2206.
(13) Coombes, D. S.; Nagi, G. K.; Price, S. L. *Chem. Phys. Lett.* **1997**, *265*, 532−537.
(14) Potter, B. S.; Palmer, R. A.; Withnall, R.; Chowdry, B. Z.; Price, S. L. J. *Mol. Struct.* **1999**, *485*/*486*, 349−361.
(15) Aakeröy, C. B.; Nieuwenhuyzen, M.; Price, S. L. *J. Am. Chem. Soc.* **1998**, *120*, 8986−8993.
(16) 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**, *A103*, 9872−9882.
(17) Mooij, W. T. M.; van Eijck, B. P.; Kroon, J. *J. Phys. Chem.* **1999**, *A103*, 9883−9890.
(18) van Eijck, B. P.; Kroon, J. *J. Comput. Chem.* **1999**, *20*, 799−812.
(19) Cole, J. C. Ph.D. Thesis, University of Bristol, Bristol, 1995.
(20) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(21) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(22) Nilsson, L.; Karplus, M. *J. Comput. Chem.* **1986**, *7*, 591−616.
(23) Reiling, S.; Schlenkrich, M.; Brickmann, J. *J. Comput. Chem.* **1996**, *17*, 450−468.
(24) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551−8567.
(25) Lii, J.-H.; Allinger, N. L. *J. Phys. Org. Chem.* **1994**, *7*, 591−609.
(26) 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/.
(27) 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, PA, 1995.
(28) Schaftenaar, G.; Noordik, J. H. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 123−134. MOLDEN is available from: http://www.cmbi.kun.nl/~schaft/molden/molden.html.
(29) Koch, U.; Stone, A. J. *J. Chem. Soc., Faraday Trans.* **1996**, *92*, 1701−1708.

**Table 1.** Experimental Structures of Glycol and Glycerol as Found in Crystal Structure Generations Using Various Models[a]

| glycol ($P2_12_12_1$) | $a$, Å | $b$, Å | $c$, Å | $\Delta X$ | $\Delta\theta$ | $\tau_1$ | $\tau_2$ | $\tau_3$ | ranking | $\Delta E$ |
|---|---|---|---|---|---|---|---|---|---|---|
| exptl | 5.01 | 6.92 | 9.27 | | | −64 | 81 | 139 | | |
| UNITAT | 4.67 | 7.31 | 9.65 | 0.53 | 25.0 | −72 | 70 | 155 | 71 | 3.0 |
| OPLS | 4.75 | 6.81 | 9.09 | 0.09 | 6.4 | −56 | 59 | 140 | 261 | 15.0 |
| CHARMM-RSB | 4.98 | 6.82 | 9.18 | 0.02 | 2.2 | −63 | 71 | 132 | 46 | 5.1 |
| AI+MM | 4.76 | 6.99 | 9.76 | 0.22 | 3.6 | −67 | 90 | 152 | 11 | 2.9 |
| ab initio | | | | | | | | | 8 | 1.1 |

| glycerol ($P2_12_12_1$) | $a$, Å | $b$, Å | $c$, Å | $\Delta X$ | $\Delta\theta$ | $\tau_4$ | $\tau_5$ | $\tau_6$ | $\tau_7$ | $\tau_8$ | ranking | $\Delta E$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| exptl | 7.00 | 9.96 | 6.29 | | | 67 | −60 | | | | | |
| UNITAT | 6.84 | 9.95 | 6.47 | 0.10 | 6.4 | 70 | −71 | −177 | 66 | 92 | 12 | 1.9 |
| OPLS | 6.87 | 9.82 | 6.06 | 0.05 | 1.4 | 66 | −60 | 167 | 62 | 81 | 2807 | 35.4 |
| CHARMM-RSM | 6.93 | 9.88 | 6.31 | 0.08 | 5.6 | 62 | −60 | 173 | 61 | 80 | 1 | 0 |
| AI+MM | 7.01 | 9.86 | 6.21 | 0.13 | 1.6 | 58 | −69 | 160 | 65 | 75 | 1 | 0 |
| ab initio | | | | | | | | | | | 1 | 0 |

[a] AI+MM: ab initio intermolecular force field + molecular mechanics (MM3(96)) for intramolecular interactions. Ab initio: ab initio intermolecular force field + MP2/6-311+G(2d,2p) conformational energies; geometries as for AI+MM. $\Delta X$ is the net translation of the center of mass (Å), calculated via fractional coordinates. $\Delta\theta$ is the net rigid-body rotation of the molecule (degrees). For glycol this is the average of the rotations for the two O−C−C planes; for glycerol it is the rotation of the C−C−C plane. $\Delta E$ (kJ/mol) is the energy difference with the lowest-energy structure that was found. A ranking of $r$ means that the structure was the $r$th lowest in energy. $\tau_1$, O−C−C−O; $\tau_2$, $\tau_3$, H−O−C−C; $\tau_4$, O1−C1−C2−O2; $\tau_5$, O2−C2−C3−O3; $\tau_6$, H1−O1−C1−C2; $\tau_7$, H2−O2−C2−C3; $\tau_8$, H3−O3−C3−C2 (atom numbering according to Figure 3, torsional angles in degrees).

We developed a procedure to incorporate this conformational dependency by calculating the atomic multipole moments more or less on-the-fly during a crystal structure optimization:[17] upon any significant change in the conformation of the molecule (5° in a torsional angle), optimization of the crystal structure is interrupted and an ab initio calculation is performed to obtain new AMMs, after which the crystal structure optimization is continued.

**Stage 3: Ab Initio Intermolecular Potential + ab Initio Conformational Energies.** Finally, for a limited number of favorable structures in the AI+MM potential, we can afford to replace the MM3 intramolecular force field by accurate ab initio calculations of the intramolecular energies. Ideally, we would have to optimize the crystal structure in combination with an ab initio treatment of the conformational energies. Since we saw no possibility to achieve this, we performed ab initio calculations on the isolated molecules in the conformations as obtained at stage 2.

Ab initio intramolecular energies are already obtained during the calculation of the AMMs. However, these energy values are not reliable, as the SCF/DZ(2d$^O$) level is aimed at optimal intermolecular electrostatic interactions[16] rather than accurate conformational energies. Moreover, the geometries at stage 2 contain bond lengths and bond angles close to the optimal values for MM3. The optimal values for the ab initio calculation are somewhat different, and the energies that are associated with these differences will be large. This geometric offset hampers the direct use of the molecular geometries obtained at stage 2 in ab initio calculations of intramolecular energies. To remedy this, these geometries were first optimized at the HF/6-31G(d,p) level. The H−O−C−C and O−C−C−O torsional angles were constrained during this geometry optimization in order to maintain the essential features of the geometry within the crystal packing. Allowing for full optimization would produce predominantly intramolecular hydrogen bonds, and the corresponding energies would bear no resemblance to the energies of the conformations present in the crystal structures.

We calculated MP2/6-311+G(2d,2p) energies at the geometries obtained in this way. It has been shown that at least this level is needed to obtain an accuracy in the order of 1 kJ/mol for conformational energies of glycol. It was seen to produce results similar to computationally much more costly MP2/6-311++G(3df,3dp) calculations.[30]

## Results

The final result of a crystal structure generation is a list of possible crystal structures ordered according to their energy, and each different energy model will result in a different list. If energy at 0 K would be all that determines a crystal structure, and the model for the energy would be perfect, then the experimental structure would be the first on this list (neglecting differences in zero-point vibrational energy). In general, we will find it at a certain place on the list (ranking), and we define the energy difference with the most favorable structure as $\Delta E$. Since all the generated structures are energy minima, the closest thing to the experimental structure one can find is the energy-minimized experimental structure. In a good force field, this should be close to the true experimental structure, otherwise crystal structure prediction cannot succeed at all.

It should be noted that the occurrence of only small deviations between the energy-minimized and the true experimental structure is a necessary but not sufficient condition for a good force field. It merely proves that the part of the potential energy surface that is sampled by the experimental structure is adequately described. In structure prediction, hypothetical structures may sample other, less well-modeled parts of the potential energy surface. This would result in erroneous energies for these alternative crystal packings. For flexible molecules it is obvious that minimization of the experimental structure does not test the ability of the force field to describe the relative energies of the various conformers of the molecule.

**Glycol.** The standard search procedure in the UNITAT force field delivered ∼500 structures. The geometry of the experimental structure is poorly reproduced in this force field: the molecule is reoriented by 25°. The UNITAT force field appears to be unsuited for structure predictions on glycol, and ranking and $\Delta E$ of such a highly distorted structure (Table 1) are of very limited value. Such problems were also encountered for propane and methanol.[17] It has been observed before that crystal structures of small, symmetric molecules are often poorly modeled.[31]

For the process of structure generation this caused a problem. The deformation in the UNITAT force field is so large that subsequent minimization within the AI+MM potential did not recover the experimental structure. So, if structure generation was performed through the UNITAT optimized structures, the experimental structure was missed. Therefore, we decided to use another force field. From the point of view of true ab initio structure prediction, this is not fair, since this ad hoc choice was made with knowledge of the experimental structure. This

(30) Reiling, S.; Brickmann, J.; Schlenkrich, M.; Bopp, P. A. *J. Comput. Chem.* **1996**, *17*, 133−147.

(31) Nobeli, I.; Price, S. L. *J. Phys. Chem.* **1999**, *A103*, 6448−6457.

illustrates a pitfall that is still present in structure prediction. No matter how accurate the final calculations shall become, predictions can still fail when the search procedure does not yield a structure close enough to the experimental structure.

We employed the OPLS force field,[20] which gave satisfactory results for carbohydrates.[18] In OPLS the experimental geometry is reproduced much better than in UNITAT. However, in this force field the ranking and relative energy were surprisingly poor. Therefore, all resulting structures (∼700) were taken over to the AI+MM stage. If we would have continued in stage 2 only for structures within, say, 10 kJ/mol of the global minimum in OPLS, the experimental structure would have been missed again. This shows another pitfall present in current structure prediction. Predictions can fail if the model that is used in the initial screening is not able to identify the experimental structure as reasonably favorable.

The main problem in OPLS seems to be in the intramolecular part of the force field. All low-energy structures have a trans O−C−C−O conformation. Within this force field a large 1-fold cosine term (39.8 kJ/mol) is applied to this torsion, which destabilizes the (experimental) gauche conformation by about 30 kJ/mol. A speculative explanation for this flaw is that this torsional parameter compensates for a too favorable description of intramolecular hydrogen bonding which, in the gas phase, is present in the gauche conformations, but absent in trans conformations. In a crystal environment both trans and gauche conformations form preferably intermolecular hydrogen bonds.

In search of a simple force field capable of giving both a reasonable ranking and structure for glycol, we additionally tried CHARMM-RSB.[23] This force field is based on CHARMM22,[21,22] for which essential torsional parameters were reparametrized based on MP2/6-311+G(2d,2p) calculations for glycol and glycerol. Intermolecular energies are still fully based on the empirical CHARMM parametrization. This force field gives much more satisfactory results than UNITAT and OPLS and excellently reproduces the experimental structure of glycol (Table 1). With hindsight, it would have been ideal for use in the first stages of the structure generations.

Within the AI+MM potential 33 structures fall within 5 kJ/mol, and for these structures ab initio conformational energies were calculated. Results for the standard force fields and the ab initio models are summarized in Table 1. Cell parameters for the experimental structure[32] are given, together with rankings and $\Delta E$'s for the experimental structure as found in the structure generations with the various models. (These structures are identical to the energy-minimized experimental structures, as they should be.) The results show that improving the accuracy of the potential improves the ranking and reduces the relative energy. We note that in the AI+MM potential the cell shows a 5% deviation in both the *a*- and the *c*-axis, which is large compared to deviations previously encountered for other molecules.[17] As can be seen in Figure 1, the changes in cell axes are at least partially related to the small changes in the hydroxyl torsional angles. From this observation it can, however, not be concluded whether these deformations are due to inaccuracies in the intra- or the intermolecular potential. With a value of 2.9 kJ/mol, $\Delta E$ is considerably larger than what we encountered for methanol and ethanol.[17] This can mainly be attributed to inaccuracies in the intramolecular potential: improving the description of the intramolecular energy from the MM3(96) level to the ab initio level reduces $\Delta E$ to 1.1 kJ/mol. This is well within the error limits of the energy, as the intramolecular energies alone are probably not yet accurate to within 1 kJ/mol.[30]

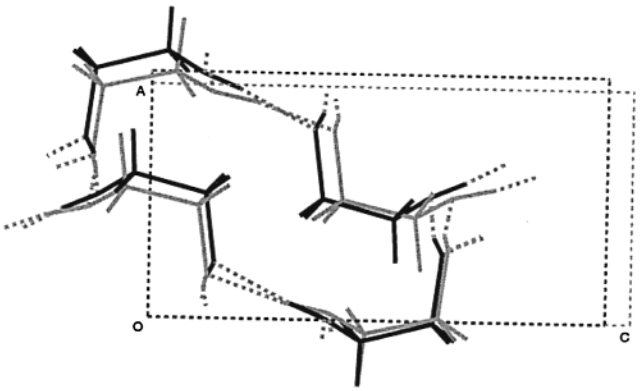


**Figure 1.** Experimental structure of glycol (black) superimposed with the energy-minimized experimental structure, which ranked 11th in the AI+MM potential (gray).

Relative energies at the MM3(96) level differ from the ab initio values up to 5 kJ/mol. A general observation is that upon switching from the MM3 to the ab initio level *tTt*-like conformations become ∼2 kJ/mol higher in energy relative to the experimental *g′Gt* conformation (nomenclature as in ref 30). The lowest-energy structures in the AI+MM potential are *tTt*, so this explains the reduction in $\Delta E$. On the other hand *g′Tg* conformations become ∼2 kJ/mol lower in energy relative to *g′Gt*, and as a result some of those structures become slightly more favorable than the experimental structure. This is why the ranking is hardly improved. Considering the magnitude of these energy shifts, it is unlikely that any structure outside the present energy window of 5 kJ/mol in the AI+MM potential will reach a lower energy than the present minimum. So, if we would extend the calculation of ab initio conformational energies to include more structures, this would not change $\Delta E$, although the ranking could deteriorate somewhat.

**Glycerol.** Due to the conformational flexibility of glycerol excessive numbers of hypothetical crystal structures were generated. In the end, ∼6200 structures were produced by the UPACK search within the UNITAT force field. This list is certainly not exhaustive: at different stages of the UPACK procedure only a limited number of structures is passed on to the next stage. All structures within 8 kJ/mol of the global minimum (∼800) in UNITAT were taken over to stage 2. Finally the 10 structures within 5 kJ/mol of the global minimum in the AI+MM potential were taken over to stage 3. Additionally, the UNITAT structures were also studied in OPLS and CHARMM-RSB.

The experimental crystal structure for glycerol[33] does not contain hydrogen atoms. Therefore, the list of proposed crystal structures was searched for structures that were consistent with the cell axes and heavy atom positions. Only one possible structure was common in all the different models. CHARMM-RSB identified an additional possibility which was only 2.3 kJ/mol higher in energy. Within both ab initio models this structure was 8.8 kJ/mol higher in energy compared to the first possibility. In addition, some ab initio calculations (MP2/6-311+G(2d,2p)) on the two conformations indicated that CHARMM-RSB overestimates the relative stability of the conformation in the second possible structure by approximately 2 kJ/mol. Therefore, we propose the first option to be the experimental structure. Fractional coordinates for this structure within the AI+MM

(32) Boese, R.; Weiss, H.-C. *Acta Crystallogr.* **1998**, *C54*, 9800024.

(33) van Koningsveld, H. *Recl. Trav. Chim. Pays-Bays* **1968**, *87*, 243−254.

**Table 2.** Fractional Coordinates of the Glycerol Structure[a]

| | experimental | | | predicted (AI+MM) | | |
|---|---|---|---|---|---|---|
| atom | *x* | *y* | *z* | *x* | *y* | *z* |
| C1 | 0.0200 | −0.3068 | 0.2760 | 0.0188 | −0.3042 | 0.2871 |
| H4 | | | | −0.1305 | −0.2952 | 0.2300 |
| H5 | | | | 0.0482 | −0.2202 | 0.4018 |
| O1 | 0.1241 | −0.3042 | 0.0791 | 0.1469 | −0.2979 | 0.1098 |
| H1 | | | | 0.1003 | −0.2305 | 0.0109 |
| C2 | 0.0483 | −0.4428 | 0.3772 | 0.0495 | −0.4421 | 0.3967 |
| H6 | | | | 0.0172 | −0.5243 | 0.2793 |
| O2 | 0.2460 | −0.4526 | 0.4354 | 0.2449 | −0.4454 | 0.4555 |
| H2 | | | | 0.2719 | −0.5351 | 0.5096 |
| C3 | −0.0777 | −0.4610 | 0.5702 | −0.0754 | −0.4599 | 0.5986 |
| H7 | | | | −0.2272 | −0.4445 | 0.5538 |
| H8 | | | | −0.0399 | −0.3844 | 0.7258 |
| O3 | −0.0479 | −0.5861 | 0.6669 | −0.0562 | −0.5942 | 0.6783 |
| H3 | | | | 0.0654 | −0.5996 | 0.7474 |

[a] Atom labeling: see Figure 3. Unit-cell parameters are given in Table 1.

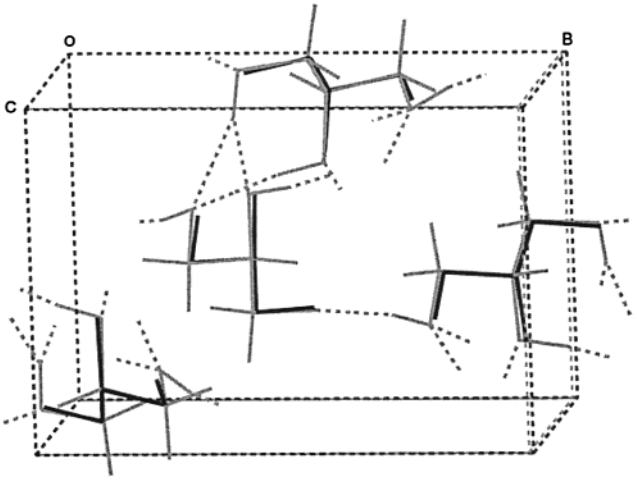


**Figure 2.** Experimental structure of glycerol (black) superimposed with the structure as predicted by the AI+MM potential (gray).

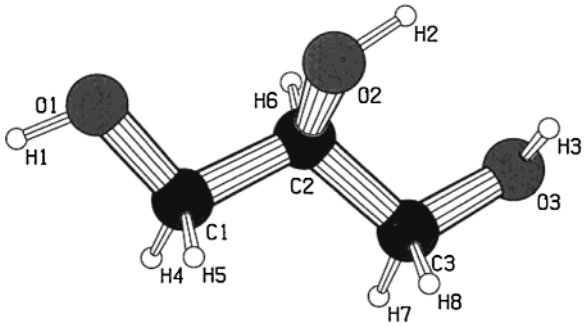


**Figure 3.** Molecular conformation of glycerol in the predicted crystal structure.

potential are given in Table 2. A superposition of the predicted and the experimental structure is given in Figure 2. Figure 3 shows the molecular conformation of glycerol in the predicted crystal structure.

In the structure a three-dimensional hydrogen-bonded network is present, involving only intermolecular hydrogen bonds. Details of the hydrogen-bond geometry are given in Table 3. CHARMM-RSB produces consistently too short O··O distances, whereas these are too long in the AI+MM potential. A small difference is that the hydrogen bond for O3··O2 is somewhat bifurcated to O1 in the ab initio potential, whereas in CHARMM-RSB it is fully directed to O2.

**Table 3.** Hydrogen Bonds in the Glycerol Structure[a]

| | | exptl | AI+MM | | CHARMM-RSB | |
|---|---|---|---|---|---|---|
| donor | acceptor | O··O | O··O | O−H··O | O··O | O−H··O |
| O1 | O3 | 2.72 | 2.76 | 170 | 2.66 | 173 |
| O2 | O1 | 2.74 | 2.81 | 173 | 2.65 | 165 |
| O3 | O2 | 2.73 | 2.81 | 156 | 2.65 | 168 |
| O3 | O1 | 3.21 | 3.09 | 128 | 3.08 | 110 |

[a] Distances in angstroms, angles in degrees. Atom labeling: see Figure 3.

Results for the crystal structure predictions with the different models are summarized in Table 1. The experimental structure is predicted as the global minimum both within the AI+MM potential and in the fully ab initio model. It was also rather favorable within the UNITAT force field, with a very similar geometry. In the OPLS force field the experimental structure has a very poor ranking and very high relative energy. In this force field, all low-energy structures have both O−C−C−O conformations trans, which should be energetically unfavorable due to parallel C−OH dipoles (Hassel-Ottar effect).[34,35] Just as for glycol these results suggest that the high 1-fold torsion term for O−C−C−O (51.2 kJ/mol for triols) leads to an excessive destabilization of gauche conformations without an intramolecular hydrogen bond.

Using CHARMM-RSB, the experimental structure is predicted as the global minimum as well. This indicates that force fields that are much simpler than ours can be successful in structure prediction. Note that CHARMM-RSB was derived with the aid of high-level ab initio calculations on glycol and glycerol, so the conformational energies should be fairly accurate. Nevertheless, the intermolecular potential remains empirical and accurate results are not assured, as can be seen for glycol. For this molecule the force field should be equally well suited, but the experimental structure is found with a $\Delta E$ of 5.1 kJ/mol, compared to 1.1 kJ/mol at the ab initio level.

Within the AI+MM potential all low-energy conformations contain only gauche O−C−C−O conformations. Upon switching from MM3 to ab initio conformational energies, there is some reshuffling in the ordering of the predicted structures. However, in both cases the experimental structure remains the most favorable one, and the energy difference with second-best structure even increases from 1.4 to 2.4 kJ/mol. The MM3 conformational energy differences were shifted from the ab initio values by up to 4 kJ/mol. Some of these shifts involve conformations that differ only by one hydroxyl group. Considering the magnitude of these shifts, it seems unlikely that a structure outside the energy window of 5 kJ/mol in the AI+MM potential would get a lower energy than the experimental structure when its conformational energy would be calculated at the ab initio level.

## Conclusions

In this work we investigated energy-based crystal structure predictions for two flexible hydrogen-bonded molecules, viz., glycol and glycerol. To this end we employed a series of increasingly accurate models, ranging from standard force fields to an ab initio intermolecular potential, augmented with ab initio calculated conformational energies. Within the latter model the experimental structure of glycol was predicted with a reasonably low energy, 1.1 kJ/mol above the global minimum. For glycerol it was predicted as the global minimum. A practical application of such crystal structure calculations is the possibility to indicate

(34) Hassel, O.; Ottar, B. *Acta Chem. Scand.* **1947**, *1*, 929−942.
(35) Jeffrey, G. A. *Acta Crystallogr.* **1990**, *B46*, 89−103.

hydrogen positions, which can be difficult to determine from X-ray diffraction data. We propose a three-dimensional hydrogen-bonded network for the crystal structure of glycerol that consists only of intermolecular hydrogen bonds.

The energy of a crystal packing for a flexible molecule is a tradeoff between intermolecular and conformational energy. Therefore it is essential to use intra- and intermolecular potentials that are both as accurate as possible. In any case, the two parts should be in proper balance.[36] Hydrogen-bonded molecules are very demanding in this respect, since a delicate balance in the modeling of intramolecular and intermolecular hydrogen bonds is required. If a hydrogen bond would be somewhat too unfavorable in the intramolecular force field compared to the intermolecular force field, crystal structure predictions would be strongly influenced. In the present work this has not been rigorously put to the test, since both glycol and glycerol preferably form intermolecular hydrogen bonds in a crystal environment. This is not the case for molecules such as monosaccharides, which we are currently studying.

Together with our previous work,[17] the ab initio intermolecular potential has now been applied to predict the low-temperature crystal structures of six different compounds. The energy difference between the observed crystal structure and the global energy minimum amounted to zero for propane and glycerol, ~0.3 kJ/mol for methanol and ethanol, 1.1 kJ/mol for glycol, and 2.0 kJ/mol for dioxane. Standard force fields fail to consistently produce such low values. This demonstrates the importance of highly accurate models in crystal structure prediction. Of course, experiments prove that energy is not the only matter of importance, and temperature and kinetics can play a decisive role: polymorphic phase transitions occur, and different polymorphs can crystallize from different solvents. Ultimately, genuine structure prediction would involve the prediction of such phenomena as well, which is obviously impossible based on energy alone. However, predictions based on empirical force fields involve too much uncertainty in the calculated energies for their success or failure to be taken as an indication for the importance of kinetic and thermodynamic effects.

**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 Professor F. B. van Duijneveldt for useful discussions and critically reading the manuscript and Professor W. L. Jorgensen for providing us with the OPLS parameters for diols and triols.

JA993945T

(36) 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.