ORIGINAL PAPER

Application of the PM6 method to modeling the solid state

James J. P. Stewart

Abstract The applicability of the recently developed PM6 method for modeling various properties of a wide range of organic and inorganic crystalline solids has been investigated. Although the geometries of most systems examined were reproduced with good accuracy, severe errors were found in the predicted structures of a small number of solids. The origin of these errors was investigated, and a strategy for improving the method proposed.

Keywords PM6 · Solid state · Periodic boundary conditions · Minerals · Unit cell · Hydrogen bonding

Introduction

The semiempirical method PM6 [1] was designed primarily for the investigation of molecular species of biochemical interest. That is, the objective of parameter optimization was to reproduce the properties of molecules. When other semiempirical methods, e.g., MNDO [2, 3], AM1 [4], and PM3 [5, 6], were developed, initial reports indicated that they were significantly more accurate than earlier methods. But later, when each new method was used for modeling species that were significantly different from those used in the training set, average errors typically increased quite significantly. This unfortunate result was a natural consequence of the way in which semiempirical method development was done: if, during a survey, a systematic fault was

J. J. P. Stewart (✉)
Stewart Computational Chemistry,
15210 Paddington Circle,
Colorado Springs, CO 80921, USA
e-mail: MrMOPAC@OpenMOPAC.net

identified, then the training set would be modified in such a way as to correct the fault. The close relationship between the survey set and the training set meant that, by its nature, properties of species in the survey set were reproduced with a higher accuracy than those of species not in the survey set.

During the development of PM6, efforts were made to minimize the potential for this increase in error. Among these were the construction and use of very large survey and training sets. In contrast to previous methods in which the training set was a subset of the survey set, during the development of PM6 the training set was a superset of the survey set.

No solids were used in either the training set or the survey set while PM6 was being developed because inclusion of even one solid in the parameter optimization would have made the whole process extremely slow, which in turn would have precluded optimization of the parameters in any reasonable time. Because of this, solids were excluded from the parameterization, and therefore they form an ideal, clearly defined set of systems for testing the applicability of PM6 to species that were not used in the development of the method.

Theory

There are several problems associated with solid-state calculations that do not exist when discrete molecules are modeled, all of which are related to the fact that there are an infinite number of interacting atoms. The most obvious consequence of this is that the electric potential experienced by each of these atoms is the result of the contributions of an infinitely large number of electrostatic terms arising from the partial charges of all the other atoms. Another implication is that the number of one-electron wavefunctions contributing to the density matrix during the solution of the self-consistent field (SCF) equations is also infinite. Various techniques

have been developed for solving these problems. Thus, in all solid-state methods, the assumption is made that the wave-functions exhibit a perfect periodicity; this assumption is formalized in the Born–von Kármán [7] periodic boundary conditions.

The electrostatic contribution or Madelung integral can be solved analytically using the Ewald sum [8]. In this procedure, an elegant mixture of real and reciprocal space contributions is used in the evaluation of the potential. To assist in the solution of the SCF equations, the near infinite number of occupied wavefunctions contributing to the density matrix is replaced by an integration over the Brillouin zone. In turn, this integration is approximated by a Simpson's rule technique involving a weighted sampling of points within the zone.

Several complete procedures have been developed for modeling solids using semiempirical methods. One of these, the MOSOL program [9], used sampling of the Brillouin Zone but, because it used complex mathematics, it was impractical for application to anything more complicated than simple binary solids. If the unit cell used is sufficiently large, then, instead of sampling the Brillouin Zone using a regular mesh of points, only one point need be used and, if the point chosen is at the origin of k-space, i.e. the Γ point, then complex mathematics can be avoided entirely. This is the basis for the large unit cell [10] or cluster [11] approximation. More recently, Gale [12] has addressed the problem of solving the Ewald sum when neglect of diatomic differential overlap (NDDO) methods are used, and developed a technique that would allow the Madelung integral to be solved more rapidly. In turn, this has allowed the structures and energies of some crystalline oxides, such as corundum and some of the polymorphs of silica, and of ice, to be modeled.

During the development of a procedure to allow PM6 to be applied to solids, various deficiencies and limitations were found in earlier procedures. Some of these, and the resulting modifications that had to be made in order to allow PM6 to be used for modeling solids, will now be described.

NDDO error

The NDDO methods pioneered by Dewar and Thiel use the Dewar-Sabelli [13–15]-Klopman [16] (DSK) approximation, Eq. 1, which is equivalent to the Ohno approximation [17], for the two-electron two center integral γ_{AB} involving atoms A and B separated by a distance R_{AB}.

$$\gamma_{AB} = \frac{1}{\sqrt{R_{AB}^2 + \frac{1}{4}\left(\frac{1}{G_A} + \frac{1}{G_B}\right)^2}} \tag{1}$$

The DSK approximation has the correct behavior at the extremes. That is, it converges to the exact point-charge expression as the interatomic distance becomes very large, and also converges to the exact two-electron, one-center term, G_A, when the interatomic separation becomes zero. Additionally, it has good behavior at chemical bonding distances. Over the past 40 years, the DSK approximation has proven very successful in NDDO models when applied to both discrete species and to polymers.

Surprisingly, in its unmodified form, the DSK approximation gives rise to an infinite error when applied correctly to any non-elemental crystalline solid. This error arises from the fact that the one-center two-electron integrals differ from element to element. The origin of the error can be understood by considering the potential at an atom A in a simple binary solid, AB, arising from all atoms on the surface of a spherical shell of radius R; in such a solid, if the charge on atom A is Q, then the charge on atoms of type B would be $-Q$. When R becomes large enough, the fraction of atoms of type A and B at that distance will be the same, and the resulting electric potential, V, at atom A could then be represented by Eq. 2.

$$V \propto 4\pi R^2 \left(\frac{Q}{\sqrt{R^2 + \frac{1}{4}\left(\frac{1}{G_A} + \frac{1}{G_A}\right)^2}} - \frac{Q}{\sqrt{R^2 + \frac{1}{4}\left(\frac{1}{G_A} + \frac{1}{G_B}\right)^2}} \right) \tag{2}$$

A Taylor series expansion of this function shows that V is proportional to the reciprocal of the distance. For all values of R greater than about 10 Å this potential is clearly very small. In solids, however, the potential at an atom is the result of the summed electric fields of all such shells, out to infinity. For convenience, this sum can be replaced by an integral, Eq. 3.

$$V \propto \int_{R=10}^{\infty} \frac{1}{R} \tag{3}$$

The value of this integral is infinity, which means that, if the DSK approximation is used and the integration is done correctly, the potential experienced by an atom of type A arising from the electrostatic contributions of all other atoms would then be either plus infinity or minus infinity, depending on the sign of its partial charge. This is an obviously unphysical result.

This catastrophe can readily be avoided by modifying the DSK approximation to ensure that it converges to the point charge expression for large values of R. The simplest modification would be to ensure that, as the interatomic separation increased, a smooth transition is made from the DSK equation to the exact point charge equation. Several trial functions were examined and, from these, a Gaussian

function was selected as having the best characteristics; this function is shown in Eq. 4. Below 5 Å, the unmodified DSK equation would be used; at larger distances, Eq. 4 would be used. This function is well-behaved in that it is single-valued and has finite first and second derivatives.

$$\gamma_{AB} = \frac{1}{R}\left(1 - e^{-0.05(R-5)^2}\right) + e^{-0.05(R-5)^2}\left(R^2 + \frac{1}{4}\left(\frac{1}{G_A} + \frac{1}{G_B}\right)^2\right)^{-1/2}$$

(4)

Electrostatic interaction

Evaluating the electrostatic potential at an atom in a crystal involves summing interactions from all surrounding atoms, and since there are, for all practical purposes, an infinite number of these, the direct sum must be replaced by a tractable alternative. The simplest and most efficient method of evaluating the electrostatic potential arising from an infinite lattice of point charges is the Ewald sum [8]. In this summation, the contribution to the potential is divided into two terms, a real-space and a reciprocal- or Fourier-space term. When an appropriate error function is used, the Ewald sum is both accurate and readily evaluated, and is the method of choice when the model used represents the electrostatic potential as the sum of contributions from classical point charges. Gale successfully applied a modified version of the Ewald sum [12] in evaluating the electrostatic potential used in MNDO, AM1, and PM3 solid state calculations.

In all NDDO [18, 19] methods, including PM6, the electrostatic contribution to the potential at an atom arising from the charges on distant atoms can be represented by the classical point charge equation, at small distances by the DSK, and at intermediate distances by the modified DSK approximation. Gale [12] noted that a modification must be made to the potential in order for the Ewald summation to be used in an NDDO method. This change requires the point-charge contribution to the potential of each atom that arises from all nearby atoms to be replaced by the exact NDDO contribution. Derivatives of the energy with respect to geometry require all potential functions to be continuous, but if corrections of the type just described were made, the resulting function would obviously be discontinuous, and further corrections would be needed. So, although the Ewald sum is aesthetically attractive, its practical implementation would necessarily involve aesthetically unattractive corrections.

An alternative to the Ewald sum would be to modify the way in which the electrostatic sum is evaluated. In this approach, use is made of the fact that an integer number of interacting unit cells are used in a solid state calculation. If the DSK equation, either unmodified or modified as in

Eq. 4, is used, then the potential at any given atom arising from the direct summation of the NDDO electrostatic terms from all the other atoms would contain artifacts reflecting the asymmetric environment. In other words, the presence of boundary effects introduces spurious terms into the potential. If these terms were not eliminated, they would have a perturbative effect on the optimized geometry that would severely compromise the validity of the results. A method for removing these spurious effects was developed that involves modifying the distance term in the DSK approximation.

The potential experienced by each atom in a solid that arises from the partial charges on other atoms falls off rapidly with increasing distance. This is a natural result of the fact that the net charge arising from all atoms in a spherical shell must rapidly converge to zero as the radius increases. An implication of this is that, for large radii, the precise value of the radius used in evaluating the potential is unimportant. Conversely, when the radius is small, and there are relatively few atoms, the potential arising from the associated partial charges is large. In that case, the value of the interatomic separation used is of great importance. This behavior can be used as the basis for modifying the electrostatic sum. At large distances, because the electrostatic effect of the distant atoms is small, the value of the interatomic distance used in calculating the potential can be different from the actual value, and, in fact, can be set to any arbitrary large fixed value. That is, all potentials arising from distant atoms can be treated as if their partial charges were moved in to the surface of a sphere of fixed radius. A result of this is that the gradient or force arising from a charge that was initially outside the sphere would be exactly zero: any potential motion of the central atom in response to the presence of a charge on the surface of the sphere would be accompanied by a simultaneous motion of that charge. A consequence of this is that the gradient of the potential arising from a charge on the surface of the sphere is precisely zero. This modification of the effective interatomic distance (EID) used in evaluating the electrostatic potential completely eliminates all directional effects, in particular all artifacts arising from the use of a finite number of interacting unit cells.

If no further modifications to the EID were made, then there would be a discontinuity in the gradient arising from the presence of the sphere. The gradient arising from a partial charge just inside the sphere would be finite, but if that charge were to move just outside the sphere, its gradient would now become zero, and there would be a discontinuity. The presence of such discontinuities would then preclude the gradients being used in subsequent operations such as geometry optimization and calculation of vibrational frequencies. To avoid them, the EID must be further modified to ensure that the gradient arising from an

atom near the surface of the sphere drops smoothly to zero as the atom approaches the surface of the sphere. This is most simply accomplished by reducing the EID of an atom as it approaches the surface of the sphere.

To summarize: the value of the EID is set to a constant for all atoms separated by a large distance, is set less than the actual distance for intermediate distances, and is equal to the actual distance when the interatomic separation is small. A simple function that satisfies these criteria can be defined using three domains, as shown in Fig. 1.

For atoms that are at a distance less than some predefined value, $2/3C$, the exact DSK approximation is used. Between $2/3C$ and $4/3C$, the EID to be used in Eq. 4 would be reduced as shown in Eq. 5.

$$R'_{AB} = 2R_{AB} - C/3 - \frac{3R_{AB}^2}{4C}, \qquad (5)$$

and at distances greater than $4/3C$ the value of the EID would be a constant C. The effect of these changes when applied to an example set of charges is illustrated in Fig. 2, with the original charges shown in black, and the locations of the charges that would be used in evaluating the electrostatic potential shown in green.

Provided enough unit cells are used to ensure that all atoms within the sphere of radius $4/3C$ are present, the effect of this modification is to remove any directional influence, specifically surface effects, arising from the presence or absence of distant atoms. As with the unmodified DSK equation, the potential arising from an atom at any distance is single-valued and its first derivative is finite. An integer number of unit cells is always used in the evaluation of the electrostatic potential; therefore, the net charge on the surface of the sphere of radius $4/3C$ precisely counterbalances the sum of all the charges within the sphere, regardless of how many unit cells are outside the sphere. This is a natural and necessary consequence of the requirement that unit cells in a solid must have a zero charge.

The electrostatic potential is, of course, dependent on the value of C. With increasing values of C, the potential converges rapidly to a constant, but also as C increases the number of unit cells that need to be used increases rapidly. The value of C was set to 30 Å, this being the best compromise between computational effort and numerical stability.

Unlike the Ewald summation, this modified DSK approximation can be used directly in evaluating the electrostatic potential. The new approximation is relatively simple in that the use of error functions and reciprocal space terms are avoided.

Solids with unpaired electrons

Many solids, particularly those containing transition metals, have unpaired electrons. Of the two standard methods available for modeling such systems when only molecules are involved, unrestricted Hartree Fock (UHF) and restricted Hartree Fock followed by configuration interaction (RHF-CI), only UHF is suitable for modeling solids. The use of RHF-CI methods is precluded because of the very large active space involved. For example, consider the garnet

Fig. 1 Truncation approximation for Madelung integral. C=30 Å
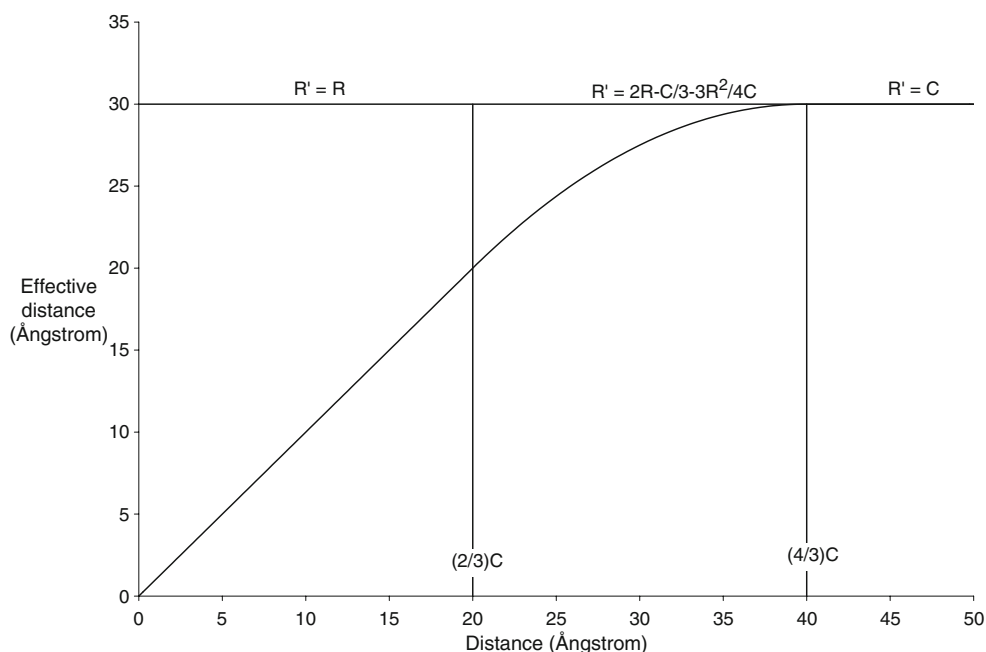
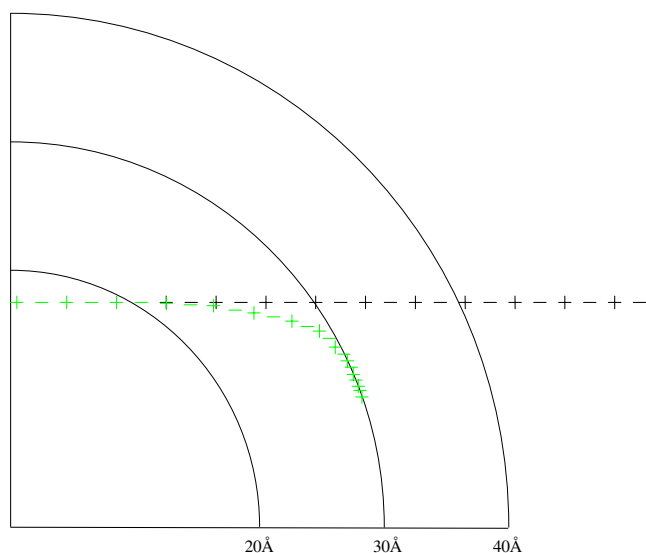**Fig. 2** Effect of truncation on apparent position of charges. For a set of charges (*black*), the position used in evaluating the electrostatic interactions is shown in *green*

uvarovite, calcium chromium silicate, $Ca_3Cr^{III}_2Si_3O_{12}$. Each chromium ion in this mineral has three unpaired d electrons. The unit cell contains eight formula units, so, if the RHF-CI procedure was used, at a minimum the active space would need to include all 80 molecular orbitals of predominantly d character. Even if the reasonable assumption was made that all the unpaired electrons had the same spin, the number of microstates involved would still be very large, $80!/(32!48!) \approx 2.2 \times 10^{22}$, and evaluation of the gradients for the resulting non-degenerate state would be prohibitively slow.

For solids in which the ions with unpaired electrons are well separated, that is, where the ions are electronically isolated from each other, the assumption can be made that the spin-state of one ion will not interact significantly with the spin-state of any adjacent ion. In addition, if the atom in question is a transition metal ion, then the spin state can usually be inferred from its environment. In the case of uvarovite, each chromium ion is in an almost octahedral environment (the exact symmetry is S_6), being surrounded by six oxygen atoms, so the three d electrons would be in a t_{2g} manifold, and would therefore be unpaired. If the Hund's rule assumption is made that the spin state is a maximum, then each chromium atom would be in a local $^4A_{2g}$ state, and any Jahn-Teller tendency to geometric distortion to a lower symmetry would be avoided. This assumption can be formalized in the calculation when a UHF method is used by defining the difference between the number of electrons of α and β spin to correspond to the maximum possible spin state of the entire system. For uvarovite, the unit cell would then be defined as having a

spin of $M_s = 24$, and therefore would have 48 more α than β electrons.

Applications

Organic compounds

Data sets were constructed for each organic compound, with, in each case, the starting geometry being the X-ray structure: i.e., the observed geometry. In contrast to molecular calculations, where internal coordinates are normally used, in the work reported here Cartesian coordinates were used exclusively. An attempt was made initially to use internal coordinates, but the numerical instabilities associated with the geometric gradients at the interfaces of the unit cells rendered their use impractical; no such difficulties were encountered when Cartesian coordinates were used. Each cluster consisted of between 100 and 200 atoms, and geometries were converged until the gradient norm had dropped below 5 kcal mol^{-1} Å^{-1}, this corresponding to an uncertainty in the optimized geometry of about 0.001 Å. All unit cell parameters were optimized, as were the coordinates of all atoms within the unit cell. Unless indicated otherwise, symmetry was not used to accelerate the optimization. All calculations were done using MOPAC2007 [20] on a 3.6 GHz Pentium computer, and each geometry optimization took between 20 min and 1 day, with most taking about 1 h. No problems were encountered in any of the optimizations.

With the possible exception of polymers, crystalline organic compounds consist of discrete molecules held together by relatively weak forces. As PM6 has been shown [1] to reproduce bond lengths and angles of simple organic compounds with useful accuracy, in this work attention was focused on the prediction of the structures of entire molecules and on the forces and energies arising from intermolecular interactions. A useful measure of accuracy of prediction of molecular structure is the root-mean-square (RMS) difference between the calculated and reference geometries of a single molecule or ion in a crystal. This quantity differs from the geometric quantities reported earlier [1] in that it measures the accuracy of prediction of the overall structure of a molecule, not just the accuracy of prediction of individual bond lengths and angles. It is possible for only relatively small distortions to exist in individual angles and, at the same time, for the overall structure to be severely in error. The RMS error is therefore complementary to the errors in individual bond lengths and angles. In order to probe the suitability of PM6 for modeling organic solids, compounds were selected that illustrate a wide range of common intermolecular interactions, the most

Table 1 Calculated and X-ray structural parameters for organic compounds

	RMS error	PM6						X-ray						Density (g/cm^3)^d
		a	b	c	α	β	γ	a	b	c	α	β	γ	
(Z Z)-N N'-Dimethylurea (NIJHUJ)	0.11	9.69	4.36	11.38	90.9	92.3	96.4	10.35	4.57	11.40	90.0	90.0	102.7	1.23 (1.11)
1, 4 Dioxocine 6-carboxylic acid chloride (CUWWIA)	0.29	14.67	12.30	3.90	89.6	91.7	89.2	15.15	12.49	3.81	90.0	87.7	90.0	1.63 (1.59)
1, 3, 5-Triaminobenzene, 1, 3, 5-trinitrobenzene (NIBZAM)	0.05	14.69	6.84	13.82	77.5	90.3	89.9	15.08	6.98	14.06	76.5	90.0	90.0	1.65 (1.55)
1, 3-bis((Pyrid-2-ylamino)carbonyl)adamantane α-ketoglutaric acid (RIZWUS)	0.14	9.97	13.66	10.72	92.0	113.1	67.2	10.53	13.73	10.73	91.8	114.7	67.8	1.42 (1.35)
1,4-diazabicyclo[2.2.2]octane azelaic acid (UNEGEZ)	0.50	30.36	8.33	5.93	91.1	89.1	91.3	28.15	8.60	6.98	90.0	93.0	90.0	1.33 (1.18)
1-diazonia-4-azabicyclo[2.2.2]octane glutarate (UNEFIC)	0.29	10.56	11.67	10.39	89.4	83.1	91.4	10.28	12.46	10.68	90.0	66.1	90.0	1.42 (1.30)
2-(2-(3-Carboxypyridyl)-4-isopropyl-4-methyl-5-oxo-imidazole (JAZCOC01)	0.10	16.07	10.74	7.201	88.7	86.5	85.5	16.31	10.78	7.17	85.6	90.0	90.0	1.40 (1.38)
2, 4, 4, 6, 6-Pentachloro-2-(piperidyl)cyclotriphosphazene (POTKEO)	0.23	8.12	22.50	17.20	89.5	90.1	90.1	8.32	22.01	17.26	90.0	90.0	90.0	1.68 (1.67)
2, 4, 6-Tribromoaniline (BRANIL)	0.09	4.20	13.59	12.99	90.0	90.0	90.0	4.26	14.62	13.44	90.0	90.0	90.0	2.96 (2.62)
2, 4, 6-Trinitro-N-methyl-N-nitroaniline (Tetryl) (MTNANL)	0.30	11.37	7.27	15.96	90.4	72.3	90.4	10.61	7.37	14.13	84.9	90.0	90.0	1.52 (1.73)
2-Pyridone (PYRIDO04)	0.03	5.77	5.93	12.75	90.2	90.4	90.1	5.60	5.79	13.56	90.0	90.0	90.0	1.45 (1.43)
4, 5-bis(Dimethylamino)-1,8-dihydroxynaphthalene (RISBIE)	0.06	10.33	13.75	8.91	107.7	90.3	90.1	10.65	13.77	9.32	107.0	90.0	90.0	1.36 (1.25)
4-Aminobenzoic acid 4-nitroaniline (RILJEB)	0.05	30.90	8.66	4.86	95.0	89.5	91.2	31.25	8.64	4.87	93.7	90.0	90.0	1.41 (1.39)
4-Fluoro-2-(phosphonomethyl)benzenesulfonic acid monohydrate (KIXQIR)	0.17	8.49	9.71	7.60	105.8	108.6	99.4	8.49	9.25	7.92	104.8	110.0	97.4	1.74 (1.75)
4-Hydroxybenzoic acid isonicotinamide (VAKTOR)		5.48	22.29	9.37	90.3	90.0	91.9	6.07	20.67	9.40	90.0	90.0	95.3	1.51 (1.47)
5H-Dibenz(b,f)azepine-5-carboxamide saccharin solvate (UNEZAO)	0.17	12.23	11.46	7.16	69.6	84.6	82.7	12.68	10.45	7.51	75.4	85.7	83.6	1.50 (1.46)
5-Methyl-2-((2-nitrophenyl)amino)-3-thiophenecarbonitrile (QAXMEH)	0.23	18.50	15.96	4.24	88.7	89.6	88.9	18.69	16.40	3.95	86.2	90.0	90.0	1.38 (1.43)
5-Methyl-2-((2-nitrophenyl)amino)-3-thiophenecarbonitrile (QAXMEH01)	0.11	16.42	8.32	8.57	90.0	89.6	89.7	16.41	8.54	8.50	88.2	90.0	90.0	1.47 (1.47)
5-Methyl-2-((2-nitrophenyl)amino)-3-thiophenecarbonitrile (QAXMEH02)	0.11	7.48	11.90	7.91	104.2	117.0	78.6	7.49	11.91	7.79	104.5	116.4	77.80	1.42 (1.44)
5-Methyl-2-((2-nitrophenyl)amino)-3-thiophenecarbonitrile (QAXMEH03)	0.17	8.02	11.36	13.66	90.6	90.4	106.5	7.98	11.68	13.32	90.0	90.0	104.8	1.44 (1.44)
5-Methyl-2-((2-nitrophenyl)amino)-3-thiophenecarbonitrile (QAXMEH04)	0.20	10.96	12.05	4.37	89.6	90.5	73.0	11.25	12.32	4.59	90.2	91.8	71.2	1.56 (1.43)
5-Methyl-2-((2-nitrophenyl)amino)-3-thiophenecarbonitrile (QAXMEH05)	0.12	13.29	22.57	7.91	90.6	90.1	90.5	13.18	22.80	8.02	90.0	90.0	90.0	1.45 (1.43)
5-Methyl-2-((2-nitrophenyl)amino)-3-thiophenecarbonitrile (QAXMEH12)	0.17	7.93	12.16	12.25	90.8	90.9	102.8	8.23	12.31	11.82	90.0	90.0	102.5	1.50 (1.47)
6-(Pyran-4-one-2-yl)-1,4-dioxocine (CUWWEW)	0.19	7.69	10.91	10.96	90.7	74.4	90.6	7.91	10.80	11.58	90.0	70.4	90.0	1.53 (1.46)
9-(2-Hydroxyethyl)adenine (FABFUJ)	0.11	7.18	8.34	13.96	90.2	90.0	89.4	7.34	8.27	13.57	90.0	90.0	90.0	1.42 (1.45)
Acetophenone (ACETPH)	0.04	8.71	8.45	9.75	90.9	58.5	90.6	8.56	8.68	10.26	90.0	59.0	90.0	1.31 (1.22)
Acetylacetone (LIWPIQ01)	0.04	7.85	4.11	15.48	92.5	90.1	89.8	8.46	4.15	16.03	90.0	90.0	90.0	1.33 (1.18)
Acetylcholine chloride (ACHOLC01)	0.17	15.29	5.94	9.62	90.0	89.3	90.2	15.32	6.30	9.89	90.0	90.0	90.0	1.38 (1.26)
Adenosine-3''-phosphate dihydrate (ADPOSD)	0.35	5.91	11.62	10.55	88.6	90.0	90.2	6.34	11.90	9.94	87.8	90.0	90.0	1.76 (1.70)
Adenosine-5''-diphosphate dihydrate (HMADPH)	0.29	18.50	6.76	8.89	91.7	89.3	93.0	18.44	6.89	9.20	90.0	87.5	90.0	1.75 (1.66)
α Glycine	0.10	5.09	5.50	10.77	90.1	90.0	69.6	5.10	5.46	11.97	90.0	90.0	68.3	1.77 (1.61)
α Resorcinol	0.05	5.60	9.59	10.38	89.7	89.9	90.1	5.60	9.53	10.53	90.0	90.0	90.0	1.31 (1.30)
Anthracene (ANTCEN14)	0.02	10.36	7.32	6.84	89.9	109.4	89.8	11.17	6.02	8.55	90.0	124.6	90.0	1.21 (1.25)
Aqua-tris(2-((dimethylamino)methyl)phenyl)borane (QEVYEV)	0.16	19.81	8.41	14.99	89.9	75.1	89.6	20.35	8.57	15.23	90.0	72.3	90.0	1.19 (1.13)

Compound		a	b	c	α	β	γ	a	b	c	α	β	γ	ρ (calc/exp)
Aspartic acid (LASPRT)	0.20	5.13	6.38	7.70	90.0	106.2	90.0	5.14	6.98	7.62	90.0	99.8	90.0	1.83 (1.64)
Barium oxalate oxalic acid dihydrate (BAHOXH11)	0.01	4.75	10.68	13.09	68.9	90.1	90.1	5.41	12.46	14.45	63.8	90.0	90.0	3.77 (2.67)
Benzene (BENZEN)	0.15	9.27	6.79	7.15	90.1	90.2	89.8	9.55	6.92	7.44	90.0	90.0	90.0	1.15 (1.06)
Berberine sulfate (CISREB)	0.08	19.91	7.43	27.51	97.6	117.3	90.7	20.37	7.44	27.43	97.7	116.2	85.5	1.61 (1.56)
β Glycine	0.04	5.34	6.09	5.14	90.0	65.7	89.8	5.38	6.27	5.08	90.0	66.8	90.0	1.64 (1.59)
β Resorcinol		8.81	4.99	12.34	89.9	88.9	90.0	7.81	5.43	12.62	90.0	90.0	90.0	1.35 (1.37)
bis(1,4-Diazoniabicyclo(2.2.2)octane) bis(1-aza-4-azoniabicyclo(2.2.2)octane) tetrakis(tribromide) dibromide[26] (DAHGUO)		15.39	20.60	12.81	103.7	89.9	90.0	16.17	21.37	13.23	101.0	90.0	90.0	2.65 (2.33)
bis(3,5-Dinitro-4-methylbenzoic acid) 1,2-bis(4-pyridyl)ethane (LAPTUS)	0.18	12.01	16.11	7.27	75.4	89.7	94.1	12.77	15.77	7.27	73.1	91.1	93.4	1.56 (1.51)
bis(Pyridinium) oxalate oxalic acid (DEFCUM)		4.12	11.39	7.77	102.1	99.8	91.0	4.00	11.32	8.44	102.0	97.3	88.6	1.60 (1.52)
bis(Urea) oxalic acid (UROXAL01)		12.34	7.00	5.13	84.2	89.8	90.0	12.37	6.88	5.05	83.6	90.0	90.0	1.58 (1.64)
Bromo-tris(2-dimethylaminoethyl)amine-manganese(ii) (DAEAMN)	0.27	11.85	11.85	11.83	90.2	90.3	89.7	12.22	12.22	12.22	90.0	90.0	90.0	1.78 (1.62)
Calcium acetate chloride pentahydrate (CALCLA)	0.04	13.26	11.20	6.66	119.1	89.7	88.6	13.72	11.51	6.82	116.7	90.0	90.0	1.73 (1.55)
Camphor (UGAHUF)	0.06	8.45	26.77	7.32	90.8	91.1	89.9	8.93	27.04	7.38	90.0	90.0	90.0	1.22 (1.14)
Cholesteryl acetate (CHOLAD04)	0.12	16.98	9.21	16.28	87.9	71.7	88.4	17.62	9.22	16.52	90.0	72.8	90.0	1.18 (1.11)
cis Copper Diglycinate (CUGLYM02)	0.19	5.33	12.60	10.54	89.9	90.1	89.9	5.19	13.55	10.69	90.0	90.0	90.0	2.16 (2.03)
Citric acid (CITRAC10)	0.31	11.92	5.38	12.18	89.5	115.2	89.6	11.47	5.62	12.81	90.2	111.5	90.0	1.81 (1.66)
Coronene (CORONE)	0.03	16.02	4.79	9.88	90.1	68.8	90.0	16.12	4.70	10.10	90.0	69.1	90.0	1.41 (1.40)
Cyclohexane (CYCHEX)	0.02	8.12	6.41	11.23	90.0	107.1	90.0	8.20	6.44	11.23	90.0	108.8	90.0	1.00 (1.00)
Cyclotrimethylene-trinitramine (RDX)	0.22	11.24	11.48	14.32	89.8	88.4	89.9	10.71	11.57	13.18	90.0	90.0	90.0	1.60 (1.81)
Cystine (LCYSTI10)	0.42	5.34	54.25	5.37	88.4	59.9	89.2	5.42	56.28	5.42	90.0	60.0	90.0	1.78 (1.67)
Cytosine (CYTOSM03)	0.04	10.21	7.09	7.79	87.2	90.4	88.4	9.82	7.52	7.73	79.5	90.0	90.0	1.52 (1.53)
Dinicotinic acid (DINICA10)	0.18	6.71	11.09	9.68	89.7	111.9	90.1	6.59	11.15	9.70	90.0	107.8	90.0	1.66 (1.64)
Dipyridinium bis(hydrogen-oxalate) oxalic acid (DUVLUB)		27.71	8.30	7.35	90.0	83.8	90.0	26.84	8.91	7.45	90.0	87.1	90.0	1.69 (1.60)
Disodium adenosine-triphosphate trihydrate (ADENTP)	0.76	6.24	20.58	31.37	90.0	91.4	88.3	7.07	20.88	30.45	90.0	90.0	90.0	2.00 (1.79)
DL-2-Amino-4-phosphonobutyric acid monohydrate (CAXDIO01)	0.11	8.11	4.83	19.89	87.6	102.3	90.6	8.40	4.95	20.60	90.0	100.4	90.0	1.76 (1.58)
γ Glycine	0.07	5.52	6.93	8.94	114.4	50.9	90.7	5.48	7.04	8.92	113.2	52.1	90.0	1.65 (1.59)
Glucose (GLUCSA)	0.19	10.27	4.20	15.12	90.0	90.0	90.0	10.36	4.97	14.84	90.0	90.0	90.0	1.84 (1.57)
Guanine	0.06	3.62	16.84	9.88	88.2	85.9	96.2	3.55	16.34	9.69	90.0	90.0	90.0	1.68 (1.79)
8-Azaguanine (AZGUAN01)	0.06	3.67	11.81	16.38	89.6	102.7	89.7	3.56	11.44	16.47	90.0	85.9	90.0	1.46 (1.51)
Hydroquinone bis(aniline) (HIBFUT)		5.03	8.21	18.42	76.8	90.0	90.0	5.39	8.22	18.74	90.0	95.1	90.0	1.33 (1.22)
Isonicotinamide (EHOWIH)	0.08	9.98	5.80	10.11	90.0	79.3	89.9	10.03	5.73	10.17	77.5	90.0	90.0	1.41 (1.40)
Isonicotinamide 3-hydroxybenzoic acid (LUNMEM)		20.70	5.16	22.19	89.9	96.2	90.0	20.87	5.14	22.42	90.0	81.8	90.0	1.47 (1.45)
Keggin pentakis(Tetraethylammonium) bis(meso-tetraphenyl-porphyrinato-zinc) tetraconta-oxo-silicon-dodeca-molybdenum bromide(PIJFUJ)		15.14	32.01	14.98	90.1	90.3	90.1	15.22	32.28	15.22	90.0	97.7	90.0	1.88 (1.83)
l-Alanine	0.06	5.93	11.67	5.86	90.0	89.9	89.9	5.79	12.26	5.93	90.0	90.0	90.0	1.46 (1.41)
Leucine (LEUCIN02)	0.13	13.85	5.24	9.44	89.2	82.4	90.2	14.52	5.30	9.56	90.0	85.8	90.0	1.28 (1.19)
Lithium acetate (Li(CH3COO) 2(H2O))	0.02	11.23	6.38	7.15	98.3	94.4	88.2	10.88	6.62	6.82	90.0	90.0	90.0	1.34 (1.38)
l-Lysine monohydrochloride dihydrate (LYSCLH)	0.11	12.35	5.76	14.38	81.6	90.4	89.8	13.32	5.88	14.98	81.2	90.0	90.0	1.44 (1.25)
Malonic acid bis(isonicotinamide) (ULAWEJ)		14.65	14.68	12.25	65.8	81.8	68.8	15.68	14.86	11.94	67.3	85.6	65.7	1.55 (1.49)
m-Aminophenol (MAMPOL02)	0.02	8.24	6.12	11.33	90.0	89.5	89.9	8.28	6.10	11.23	90.0	90.0	90.0	1.27 (1.28)
m-Aminopyridine (AMIPYR)	0.03	6.30	5.68	15.23	90.1	89.5	111.4	6.19	5.71	15.30	90.0	90.0	110.5	1.23 (1.24)
m-Cresol N,N,N′,N′-tetraisopropyloxamide (DUGRIG)		6.62	26.72	11.88	93.4	89.8	89.8	6.86	27.19	12.08	96.1	90.2	89.8	1.15 (1.07)
m-Diaminobenzene, 3,5-dinitro-1-cyanobenzene (REDDEJ)		13.07	10.39	9.85	84.0	89.5	90.4	13.26	10.45	9.88	86.7	90.0	90.0	1.51 (1.46)
Methionine (LMETON02)	0.11	5.18	16.40	9.31	85.3	93.8	83.3	5.20	14.83	9.49	81.0	90.0	90.0	1.27 (1.37)

Table 1 (continued)

	RMS error	PM6						X-ray						Density (g/cm³)[d]
		a	b	c	α	β	γ	a	b	c	α	β	γ	
Methyl (+)-(1α,2β,8α,9α,10β)-2-chloro-4-aza-3-oxatetracyclo(8.4.0.0^{2,9}.0^{4,8}) tetradecane-9-carboxylate oxalic acid monohydrate (HUZKOC)	0.15	6.19	18.35	8.11	96.5	101.6	81.7	6.61	17.89	8.07	102.3	99.8	83.2	1.47 (1.43)
m-Hydroxybenzoic acid (BIDLOP)	0.03	23.80	4.97	5.04	90.0	74.3	90.0	23.89	4.94	5.49	90.0	74.3	90.0	1.60 (1.47)
tetrakis(2-Carboxypyridinium) octacyano-molybdenum(iv) (PYCMOA) (4[$C_5H_6NO_2$] [$Mo^{IV}(CN)_8$]$^{4+}$)	0.09	17.96	10.46	9.70	93.3	91.4	89.2	18.51	10.15	9.17	89.8	90.0	90.0	1.46 (1.54)
m-Toluidine (FANDOO)	0.08	9.13	5.79	23.12	91.7	80.9	89.4	8.76	5.81	24.87	90.0	79.9	90.0	1.18 (1.14)
Natroxalate ($Na_2(C_2O_4)$)	0.58	4.23	9.29	5.38	90.1	90.2	86.5	3.45	10.38	5.24	90.0	90.0	90.0	2.11 (2.38)
N, N'-(6 6''-dimethylbiphenyl-2 2''-diyl) bis(methylamine) (ENIWIH)	0.07	10.67	11.77	10.15	90.1	90.1	90.0	11.09	12.02	10.57	90.0	90.0	90.0	1.25 (1.13)
N, N-2 6-Tetramethyl-4-nitroaniline (FOCVOI)	0.12	7.85	16.17	7.67	90.7	95.6	88.9	7.64	18.08	7.64	90.0	90.0	90.0	1.33 (1.22)
N, N-Dimethylaniline (DMAFBZ01)	0.08	7.09	13.20	8.71	82.4	101.2	95.4	6.78	12.33	8.26	80.5	74.0	92.0	1.29 (1.57)
Nicotinamide adenine dinucleotide tetrahydrate (CEVYEH11)	0.25	8.89	8.20	11.33	83.2	69.7	103.2	8.84	8.59	11.19	89.4	70.4	103.9	1.66 (1.58)
N-Methylurea (MEUREA)	0.09	6.85	6.64	8.08	90.1	91.2	90.3	6.92	6.98	8.48	90.0	90.0	90.0	1.34 (1.20)
N, N-Dimethylbenzamide (ODOTOQ)	0.08	6.51	16.19	7.49	90.0	90.6	90.8	6.63	16.30	7.65	90.0	90.0	90.0	1.26 (1.20)
N, N-Dimethylurea (WIFKEB)	0.09	9.27	8.49	6.06	90.4	89.6	109.3	9.27	8.60	6.04	90.0	90.0	108.8	1.30 (1.29)
o-Aminophenol (AMPHOM02)	0.05	7.50	8.10	19.70	90.0	90.5	90.0	7.85	7.25	19.75	90.0	90.0	90.0	1.21 (1.29)
o-Aminopyridine (AMPYRD)	0.04	7.41	5.65	11.62	90.0	85.9	89.9	7.59	5.67	11.71	90.0	84.5	90.0	1.29 (1.25)
o-Diaminobenzene (BAGFIY)	0.04	7.72	7.57	9.96	89.9	82.5	90.0	7.72	7.54	10.32	90.0	80.0	90.0	1.25 (1.21)
o-Dimethoxybenzene (TUKGEL)	0.06	12.94	9.35	5.58	90.1	89.8	89.9	13.35	9.92	5.53	89.4	90.0	90.0	1.36 (1.25)
Oxalic acid (OXALAC06)	0.03	6.91	6.08	8.16	89.9	89.9	90.0	6.56	6.09	7.85	90.0	90.0	90.0	1.74 (1.91)
Oxalic acid dihydrate (OXACDH26)	0.04	4.92	11.45	3.44	89.7	90.0	106.9	6.09	11.93	3.47	89.7	90.0	106.9	2.26 (1.73)
Oxytocin 1XY10001 [23]	0.44	22.06	8.89	26.73	90.6	99.1	88.4	23.04	9.04	27.27	90.0	102.2	90.1	1.43 (1.33)
p-Aminophenol (AMPHOL01)	0.03	7.95	12.90	4.78	90.1	90.1	90.1	8.18	12.95	5.26	90.0	90.0	90.0	1.48 (1.30)
p-Aminopyridine (AMPYRE)	0.04	5.48	11.92	7.19	89.6	90.3	91.0	5.57	12.12	7.32	90.0	90.0	90.0	1.33 (1.26)
p-Chloroaniline (CLANIC05)	0.03	8.37	7.30	8.98	90.0	89.7	90.2	8.59	7.24	9.19	90.0	90.0	90.0	1.54 (1.48)
p-Diaminobenzene (PENDAM)	0.03	8.29	23.08	5.94	89.9	90.0	93.5	8.37	22.95	5.97	90.0	90.0	93.6	1.27 (1.26)
Phenanthrene (PHENAN08)	0.04	9.16	5.73	8.56	90.2	80.9	89.9	9.44	6.14	8.44	90.0	82.0	90.0	1.34 (1.22)
Phenolphthalein (NIMDAO)	0.09	11.24	14.58	19.01	90.0	89.8	89.5	11.39	14.82	19.27	90.0	90.0	90.0	1.34 (1.28)
Phloroglucinol (PHGLOL)	0.03	12.71	9.35	4.87	89.8	89.6	89.9	12.56	9.37	4.83	90.0	90.0	90.0	1.45 (1.47)
p-Hydroxybenzoic acid (JOZZIH)	0.04	6.21	10.05	18.46	89.8	85.7	90.1	6.34	10.46	18.51	90.0	86.8	90.0	1.60 (1.50)
Picric acid (PICRAC)	0.15	9.63	19.49	9.38	88.9	90.4	89.8	9.70	19.13	9.25	90.0	90.0	90.0	1.73 (1.77)
Potassium hydrogen acetate	0.01	3.91	6.77	25.00	90.3	89.8	82.1	4.01	7.19	23.88	90.0	90.0	81.7	1.60 (1.54)
1-Dimethylamino-8-dimethylammonionaphthalene saccharin dihydrate (AJOHUC)	0.09	8.81	9.54	24.50	90.2	99.6	90.1	9.25	9.17	24.93	90.0	95.8	90.0	1.42 (1.37)
Salicylaldoxime (SALOXM)	0.07	12.91	5.45	10.08	90.3	65.4	89.8	13.60	5.08	10.41	90.0	67.1	90.0	1.41 (1.38)
Sodium acetate $Na(CH_3COO)\cdot3(H_2O)$ (NAACET01)		11.65	11.39	15.40	88.9	78.8	91.7	10.40	10.47	12.35	90.0	68.3	90.0	0.90 (1.45)
Sodium hydrogen acetate		16.33	16.33	16.33	89.9	89.9	90.0	15.92	15.92	15.92	90.0	90.0	90.0	1.30 (1.40)
Sucrose (SUCROS01)	0.24	10.36	8.59	7.50	90.0	103.0	90.0	10.82	8.68	7.72	90.0	103.0	90.0	1.75 (1.61)
Tetramethylammonium dihydrogen phosphate monohydrate (FIJHEL)	0.02	8.01	8.07	12.32	90.1	90.0	98.1	8.52	8.44	12.90	90.0	90.0	99.1	1.59 (1.37)
Tetramethylurea (TIDBIR)	0.16	8.69	6.00	11.73	92.1	93.1	91.6	9.97	6.26	10.63	90.0	93.0	90.0	1.27 (1.17)

Compound	RMS	a	b	c	α	β	γ	a	b	c	α	β	γ	Density
Thymine (THYMIN01)	0.05	6.79	6.96	11.71	78.1	89.8	89.6	6.85	6.78	12.89	75.1	90.0	90.0	1.55 (1.45)
trans Copper digicinate (IWUBEH)	0.32	9.53	27.27	5.32	89.0	92.3	88.5	9.44	27.93	5.06	90.0	90.0	90.0	1.85 (1.91)
Trinitrotoluene (ZZZMUC01)	0.11	5.43	22.63	14.72	90.1	91.4	90.1	6.08	20.02	14.99	90.0	90.0	90.0	1.67 (1.66)
tris(Acetylacetonato) titanium(iv) perchlorate (TIACPC)	0.25	8.97	11.73	9.78	91.0	83.5	92.3	8.68	11.74	9.89	88.5	84.3	92.0	1.45 (1.47)
Tryptophan (TRYPTC)	0.13	15.26	5.17	6.90	89.3	78.3	88.9	14.67	5.30	7.45	90.0	81.2	90.0	1.50 (1.40)
Tyrosine (LTYROS10)		6.85	20.39	5.90	89.9	90.0	90.0	6.91	21.12	5.83	90.0	90.0	90.0	1.46 (1.41)
Urea (UREAXX13)	0.03	4.67	5.58	5.58	90.0	90.0	90.0	4.69	5.57	5.57	90.0	90.0	90.0	1.37 (1.37)
Urea nitrate (UREANT02)		7.81	8.20	9.70	89.9	52.8	90.1	7.50	8.20	9.54	90.0	55.8	90.0	1.65 (1.69)
Weddellite (Ca(C$_2$O$_4$)·2(H$_2$O))		7.23	11.72	11.72	89.7	89.9	89.9	7.36	12.37	12.37	90.0	90.0	90.0	2.20 (1.94)
Whewellite (Ca(C$_2$O$_4$)(H$_2$O) (CALOXM03)		6.04	9.77	14.57	89.9	90.5	107.7	6.29	10.12	14.58	90.0	90.0	109.5	2.37 (2.22)
tris(2,2,6,6-tetramethylheptane-3,5-dionato)-yttrium(iii) (HAHTOZ01)	0.28	9.57	8.89	18.95	89.6	86.4	90.2	10.63	9.98	17.87	90.0	90.0	90.0	1.32 (1.12)

[d] X-ray densities in parenthesis
Values of a, b, c in Ångstroms; α, β, γ in degrees. Reference crystal structures from the Cambridge Structural Database [21]. *RMS* Root-mean-square

important of these being, in order of the energies involved: ionic, hydrogen bonding, and π-stacking.

Densities

Another useful measure of the accuracy of prediction of organic and inorganic solids is the density. In most cases when the density is accurately reproduced the internal structure of the unit cell is also accurately predicted. This is not an infallible rule, in that it is possible for the density to be predicted with good accuracy and, at the same time, the unit cell structure to be significantly distorted. This rare occurrence can usually be detected by distortions of the unit cell parameters. No cases were found where the unit cell parameters were predicted with good accuracy and, at the same time, significant errors existed in the internal structure of the unit cell. A comparison of PM6 and X-ray unit cell parameters for 124 organic solids is presented in Table 1. In this table, the unit cell used was often different from that reported in the literature, particularly so in hexagonal crystals, that is, crystals in which the interface angles are 90°, 90°, and 60°. Unit cells were chosen that would maximize the size of sphere that could be contained in a given cluster; to this end, most hexagonal unit cells were replaced by equivalent orthorhombic unit cells. Predicted densities were reproduced with good accuracy, the average unsigned error in density being 6.9%, with the bulk of this error arising from errors in the calculated intermolecular distances. Although most systems optimized with only small changes in the geometry, in three instances quantitative changes occurred.

In the solid state, individual molecules of oxalic acid in oxalic acid dihydrate [Cambridge Structural Database [21] (CSD) entry OXACDH26] exist as the neutral species. PM6 incorrectly predicts them to be fully ionized, as oxalate, $[C_2O_4]^{2-}$, plus two hydronium ions, $[H_3O]^+$. This change was accompanied by a very large increase in density, from 1.73 to 2.23 g/cc, a direct consequence of going from a neutral to an ionic species.

A related system is barium oxalate oxalic acid dihydrate (CSD entry BAHOXH11) in which there exist polymeric chains of oxalate groups connected by bridging protons. As with oxalic acid, in the optimized PM6 geometry the proton is abstracted by the water molecule to give oxalate groups and hydronium, resulting in an increase in density of almost 40%.

Sodium acetate trihydrate is an ionic solid consisting of sodium ions surrounded by an acetate group and five water molecules, four of which form bridges between pairs of sodium ions. PM6 completely fails to predict the observed structure: the distance between the sodium ions increases considerably, effectively destroying any tendency of the water molecules to form bridges.

When these three solids were removed from consideration, the average unsigned error (AUE) in density decreased to 6.1%. The most common intermolecular interaction is hydrogen bonding, which PM6 predicts to be too short by about 0.1 Å, with the result that the average signed error in calculated densities of organic compounds is too high by 3.9%. When a systematic correction to the density was made, the AUE in density decreased still further to 4.8%.

Heats of formation

Table 2 presents a comparison of experimental and calculated heats of formation of organic solids. The largest difference occurs with 2, 4, 6-tribromoaniline, which is predicted to be too stable by 44.5 kcal/mol. The optimized structure revealed an unrealistically short intermolecular Br–N distance of 2.14 Å, indicating that the Br–N core-core repulsion was severely underestimated. Examination of the values of the PM6 parameters for the Br–N core-core interaction revealed that the Voityuk interaction would be negligible at chemical bonding distances. This error in PM6 can be attributed to the absence of appropriate reference data in the training set, a fault that could be readily corrected in future work.

Heats of sublimation

The heat of sublimation is a measure of the intermolecular interaction energy. In some cases sublimation is accompanied by large geometric and electronic changes. For example, the simple amino acids exist as the Zwitterion in the crystal phase, but in the gas phase they are unionized. Representative values for calculated and observed heats of sublimation are presented in Table 3. Determining the accuracy of PM6 for the prediction of heats of sublimation is made difficult by the acknowledged unreliability of many experimental measurements. Thus in the reference compendium of sublimation enthalpies [22], the authors indicate that the reported value for aspartic acid, 22.9±1.0 kcal/mol, was likely unreliable.

Biomolecules

The primary objective in developing PM6 was to more accurately model systems of biochemical interest. The applicability of PM6 to the study of crystals of biochemical importance was therefore of interest.

Oligopeptides

The X-ray structures of many small polypeptides have been determined and are readily available in the Protein Data

Table 2 Comparison of calculated and experimental heats of formation of organic compounds (kcal/mol)

	PM6	Reference[a]	Difference
(Z Z)-N, N′-Dimethylurea (NIJHUJ)	−62.0	−76.3	14.3
2, 4, 6-Tribromoaniline (BRANIL)	−30.7	13.8	−44.5
α Glycine	−122.9	−126.1	3.2
α Resorcinol	−79.3	−88.0	8.7
Anthracene	26.4	30.0	−3.6
β Glycine	−121.8	∼−126.1	∼4.3
Camphor	−66.3	−76.3	10.0
Citric acid	−348.7	−369.0	20.3
Cyclotrimethylene-trinitramine (RDX)	−5.9	18.9	−24.8
Cystine	−237.9	−246.8	8.9
γ Glycine	−120.8	∼−126.1	∼5.3
L-Alanine	−132.0	−134.1	2.1
Leucine	−142.7	−152.3	9.6
m-Aminophenol (MAMPOL02)	−37.6	−47.9	10.3
m-Aminopyridine (AMIPYR)	23.2	14.4	8.8
m-hydroxybenzoic acid	−123.7	−142.0	18.4
o-Aminophenol (AMPHOM02)	−38.2	−48.1	9.9
o-Aminopyridine (AMPYRD)	21.4	9.4	12.0
o-Diaminobenzene (BAGFIY)	12.9	9.3	3.6
Oxalic acid	−175.5	−198.4	22.9
p-Aminophenol (AMPHOL01)	−40.7	−46.4	5.7
p-Aminopyridine (AMPYRE)	19.0	10.0	9.0
p-Chloroaniline (CLANIC05)	2.5	−8.0	10.5
p-Diaminobenzene (PENDAM)	5.1	10.1	−5.1
Phenanthrene	24.1	26.2	−2.1
p-Hydroxybenzoic acid	−125.9	−145.0	19.2
Picric acid	−43.4	−52.1	8.7
Salicylaldoxime	−27.0	−43.9	16.9
Sucrose	−535.8	−532.0	−3.8
Trinitrotoluene	−5.7	−15.1	9.4
Tyrosine	−150.7	−163.7	13.0
Urea	−65.7	−79.6	13.9

[a] Reference values taken from the CRC Handbook [32]

Bank [23]. One representative entry in this collection, 1XY1, is the nonapeptide deamino-oxytocin, Cys-Tyr-Ile-Glu-Asp-Cys-Pro-Leu-Gly. The structure of this hormone had been refined to a resolution of slightly better than 1.1 Å [24]. Each deamino-oxytocin molecule contains a disulfide bridge between atoms S1 and S6, and two strong hydrogen bonds between N2 and O5, and N5 and O2. The unit cell contains four deamino-oxytocin molecules related by a pseudo C$_2$ operation, that is, there are two inequivalent polypeptide molecules and 26 water molecules. The coordinates of the hydrogen atoms in the peptide were given, but not those of the hydrogen atoms on the water

Table 3 Comparison of calculated and experimental heats of sublimation

	Reference [22]	PM6	Difference
(Z Z)-N, N′-Dimethylurea (NIJHUJ)	22.1	15.9	−6.2
α Glycine	32.6	29.7	−2.9
Anthracene	23.8	33.1	9.3
Aspartic acid	22.9	37.8	14.9
Benzene	10.0	3.2	−6.8
Camphor	12.4	6.5	−5.9
Cyclohexane	9.0	2.1	−7.0
Guanine	44.5	25.2	−19.3
L-Alanine	31.7	32.6	0.9
Leucine	36.0	28.5	−7.5
m-Aminophenol (MAMPOL02)	23.6	11.8	−11.8
m-Aminopyridine (AMIPYR)	19.3	9.5	−9.8
Methionine	32.0	27.2	−4.8
m-Hydroxybenzoic acid	29.5	15.3	−14.3
N-Methylurea	22.6	17.0	−5.6
N, N-Dimethylbenzamide	21.4	10.0	−11.4
o-Aminophenol (AMPHOM02)	22.3	17.6	−4.7
o-Aminopyridine (AMPYRD)	18.3	9.2	−9.1
o-Diaminobenzene (BAGFIY)	20.4	9.1	−11.3
Oxalic acid	22.3	19.2	−3.1
p-Aminophenol (AMPHOL01)	24.2	19.3	−4.9
p-Aminopyridine (AMPYRE)	20.8	10.4	−10.4
p-Chloroaniline (CLANIC05)	21.7	8.7	−13.0
p-Diaminobenzene (PENDAM)	22.0	16.0	−6.1
Phenanthrene	22.0	30.2	8.2
p-Hydroxybenzoic acid	29.5	14.5	−15.1
Picric acid	25.1	14.6	−10.5
Trinitrotoluene	26.9	14.7	−12.2
Tyrosine	24.1	33.7	9.6
Urea	21.7	17.5	−4.2

molecules. That water of crystallization exists implies that peptide–water hydrogen bonds also exist.

Because the positions of the hydrogen atoms on the water molecules were not given in the X-ray structure, an estimate of the locations of the 52 hydrogen atoms had to be made before the geometry could be optimized. For this operation, the "ice rules" were used: each oxygen atom in a water molecule was involved in forming two hydrogen bonds and each hydrogen atom formed one hydrogen bond. Of necessity, some of these bonds involved atoms on the peptide.

Several candidate structures, each of which satisfied these conditions, were constructed, and the positions of all hydrogen atoms optimized while those of the other atoms were held constant. An incidental benefit of this operation was that any potential errors in the X-ray positions of some

hydrogen atoms, in particular the apparently faulty location reported for H147, were automatically corrected. After the positions of the hydrogen atoms were optimized, an unconstrained optimization on the unit cell was carried out. This involved the optimization of the Cartesian positions of all the atoms in the unit cell of formula $(C_{43}H_{65}N_{11}O_{12}S_2)_4 \cdot 26(H_2O)$, and the unit cell dimensions, i.e., the simultaneous optimization of 1,839 coordinates.

Each initial geometry optimized to give a different final structure. That is, the optimized geometry was very sensitive to the choice of initial locations of the hydrogen atoms assigned to the water molecules. As a result, it was not possible to unambiguously define an optimized PM6 structure; however, all the fully optimized structures were within a few kcal/mol of each other, so one structure was chosen arbitrarily and used in the following analysis.

The optimized PM6 unit cell dimensions are shown in Table 1. The optimized PM6 structure of the entire unit cell had an RMS error of 0.61 Å, and an RMS error of 0.44 Å for a single molecule of deamino-oxytocin. Deviations from the pseudo-C_2 symmetry were small, and were very sensitive to the initial choice of hydrogen bonds; it is likely that the time-average would be exactly pseudo-C_2. All the weak intra-chain bonds were preserved: PM6 predicted the disulfide bridges, S1–S6 to be 2.06 Å and 2.05 Å compared with the X-ray values of 2.08 Å and 1.95 Å, and the hydrogen bonds between N2 and O5 to be 1.98 Å versus the X-ray value of 1.93 Å, and N5 and O2 1.98 Å, compared to the X-ray, 1.90 Å.

Acetylcholine One of the simplest of the important biochemicals is the neurotransmitter acetylcholine, $[CH_3COOCH_2CH_2N(CH_3)_3]^+$ (ACh). In one form, this ion exists in the solid state as the chloride (CSD entry ACHOLC01). PM6 reproduced the structure of ACh in this salt with good accuracy, the RMS error for a single ACh being only 0.21 Å. Part of this error can be attributed to the C–H and N–H bond-lengths from the X-ray structure being about 0.2 Å too small; when only the heavy atoms are used in the comparison, the RMS error decreased to 0.13 Å. An estimate of the effect of the crystal-packing forces can be obtained by comparing the PM6 predicted structure of the gas-phase ion with that found in the crystal; when this was done the RMS difference increased to 0.54 Å. This is unequivocal evidence that inclusion of crystal-packing forces is essential in order to reproduce the observed structure.

Adenosine diphosphate Adenosine diphosphate (ADP) is an important intermediate in energy transfer in biochemistry. ADP contains an pyrophosphate group, P_2O_7, which, in the tris(hydroxymethyl)-methylammonium dihydrate salt (CSD entry HMADPH), is described as being doubly

ionized, with the counterions being the adenine and the quaternary ammonium ion. Hydrogen atoms were added to the structure given in the CSD and, in a preliminary calculation, their positions were optimized using PM6, the rest of the geometry being fixed at the X-ray structure. One of the hydrogen atoms in the resulting geometry was located between an oxygen of the terminal PO_4 group on one ADP ion and a nitrogen atom on an adenine in an adjacent ADP ion. Because these oxygen and nitrogen atoms were only 2.57 Å apart, it is likely that a bridging hydrogen bond exists between them in the crystal. The geometry of the entire system was then optimized. This resulted in an insignificant increase in the N–O distance to 2.58 Å, and gave an RMS error for a single ADP ion of 0.41 Å. This suggests that a better description of the pyrophosphate moiety would be that it is singly ionized and that it is part of a two molecule ADP–adenine system connected by a bridging hydrogen. As with acetylcholine, the structure of ADP depends strongly on its crystal environment: when the geometry of a single ATP ion was optimized using PM6, the structure distorted dramatically and the resulting RMS error relative to the X-ray structure increased to 2.08 Å.

Adenosine triphosphate The nucleotide adenosine-triphosphate (ATP) is a source of energy for many biochemical reactions and, as such, its structure and properties are of considerable interest. Only one simple compound of ATP was found in the CSD, disodium adenosine-triphosphate trihydrate, ADENTP. Its structure was very badly reproduced when PM6 was used, with the RMS error of the ATP molecule being 1.18 Å. Examination of the unit cell revealed that the sodium ions had formed spurious weak bonds to nearby hydrogen atoms, and that this was partially responsible for the distortion of the ATP ion. This was confirmed when potassium was used in place of sodium and the RMS error dropped to 0.70 Å. Because of this and other results, there is convincing evidence that PM6 sodium has severe problems when solid state systems are being modeled, and it is highly probable that part of the error in prediction of the structure of ADENTP can be attributed to faults in sodium parameters.

Nicotinamide adenine dinucleotide In contrast to ATP, the structure of nicotinamide adenine dinucleotide (NAD) tetrahydrate (CSD entry CEVYEH11) was reproduced with good accuracy, the RMS error being only 0.33 Å. Like ATP, NAD contains a polyphosphate group.

Hydrogen bonding

Because of the importance of hydrogen bonding in biochemistry, a range of types of hydrogen bond were examined. Most of the important hydrogen bonds in biochemistry involve a proton positioned between either two oxygen atoms, two nitrogen atoms, or an oxygen and a nitrogen atom, the more exotic bonds, such as those involving halogen ions, while interesting, being of secondary importance.

Individual types of hydrogen bonds

O–H–O A common example of hydrogen bonds is provided by simple organic compounds that contain hydroxyl groups. Examples of such systems are sucrose, aspartic acid, citric acid, glucose, oxalic acid, and the hydroxybenzoic acids. In all such simple hydrogen bonded systems, the geometry predicted by PM6 was similar to that observed. Of more interest are those hydrogen bonds that occur in systems in which relatively large charges are involved.

The tendency for a proton to form a bridging structure between two oxygen atoms in organic acids was investigated. Such a situation occurs in the polymeric solid potassium hydrogen acetate, where a single proton is shared between two acetate groups, the whole assembly, two acetates plus the proton, having a formal charge of −1. In catena-((μ_4-acetato)-(μ_2-acetic acid)-potassium) (CSD entry KHACET02), this structure has the geometry O–O′: 2.49 Å, O–H: 1.09 Å and O′–H 1.40 Å, the three atoms forming a straight line. For this solid, PM6 predicts the following: O–O′: 2.52 Å, O–H: 1.22 Å and O′–H 1.30 Å, with the O–H–O angle being 178°.

Solid 4-fluoro-2-(phosphonomethyl)benzenesulfonic acid monohydrate (CSD entry KIXQIR) exists as the Zwitterion. In this system, the sulfonic acid group donates a proton to the phosphono group, giving –$[SO_3]^-$ and its counterion –$[P(OH)_3]$. One of the hydrogen atoms of the phosphono group then forms a strong hydrogen bond with the nearby water molecule. PM6 incorrectly predicts this bridging hydrogen to be nearer to the water than to the phosphono group. Where the X-ray structure has the PO···H distance of 1.05 Å, PM6 predicts 1.48 Å; the corresponding distances for the H···OH2 are 1.37 Å, observed, and 1.10 Å, calculated.

Another interesting hydrogen bond exists in crystalline acetylacetone (CSD entry LIWPIQ01). In this system the X-ray structure shows that the hydrogen atom involved in hydrogen bonding is disordered over two equivalent positions within a single molecule, with the result that the observed bond lengths and angles are symmetric about the central C–H unit. As expected, when the geometry is optimized using PM6, the extra symmetry is destroyed. This is a consequence of the requirement that, in a quantum chemical calculation, each atom must be in a defined position. However, although the calculated structure was of lower symmetry, when the optimized geometries of the two

halves of the molecule were averaged, the result was in good agreement with the X-ray structure: where PM6 predicted the C–O bond to be 1.25 Å and 1.35 Å, the X-ray structure gave 1.28 Å, and where PM6 gave the C_2–C_3 distance as 1.36 Å and 1.44 Å, the observed value is 1.40 Å. The structure of the hydrogen bond was also well reproduced, with the calculated O–O distance being 2.62 Å compared with the observed 2.54 Å, albeit the calculated O–H bond length, 1.06 Å was much greater than the reported value of 0.92 Å.

N–H–N In 1-dimethylamino-8-dimethylammonionaphthalene saccharin dihydrate (CSD entry AJOHUC), saccharine donates a proton to the "proton sponge" 1,8-bis(dimethylamino)-naphthalene to form an ionic crystal. The reported structure has that proton asymmetrically positioned between the two nitrogen atoms: N–N′: 2.56, N–H: 1.35, N′–H: 1.26 Å. The optimized PM6 structure gives N–N′: 2.68, N–H: 1.71, N′–H: 1.11. PM6 thus both underestimates the bridging power of the proton and exaggerates the asymmetry of the bond.

A closely related species, 4,5-bis(dimethylamino)-1,8-dihydroxynaphthalene, exists as the Zwitterion in the solid (CSD entry RISBIE). In this system, the geometry of the N–H–N′ structure is symmetric, N–H: 1.27 Å, N–N′: 2.57 Å, and the O–H–O structure is unsymmetric, O–H: 1.00 Å, O–O′: 2.45 Å. PM6 predicts both the N–H–N′ (N–H: 1.12 Å, N–N′: 2.68 Å) and the O–H–O′ structures (O–H: 1.09 Å, O–O′: 2.49 Å) to be unsymmetric. In the gas phase, 4,5-bis(dimethylamino)-1,8-dihydroxynaphthalene would most likely exist as the neutral species; a B3LYP/6–31G (d) calculation predicts the energy of the Zwitterionic form to be 0.6 kcal/mol above that of the neutral form; however, PM6 incorrectly predicts that the Zwitterion should be 15.8 kcal/mol more stable than the neutral form.

In 2005, an even stronger proton sponge, 1,8-bis (hexamethyltriaminophosphazenyl)naphthalene, HMPN, was reported [25]. The effect of steric crowding in HMPN arising from the $–N{=}P(N(Me_2))_3$ groups distorts the naphthalene skeleton so that the reported C_1–C_9–C_{10}–C_5 torsion angle, Fig. 3, is 173.9°. The fully optimized PM6 crystal structure predicted this angle to be 170.3°. For the gas-phase structure, B3LYP/6–31G* predicted the torsion angle to be 172.8° [25] while PM6 gave an angle of 165.8°, indicating that the PM6 model was producing either a less rigid naphthalene structure or a greater steric repulsion energy.

An additional example of the importance of crystal packing forces is provided by the accuracy of prediction of the structure of HMPN in the gas and solid state phases. Excluding hydrogen atoms, the RMS difference between the B3LYP gas-phase structure and the X-ray structure was 0.286 Å, while for the PM6 crystal structure the equivalent difference was 0.267 Å, and for the PM6 gas-phase structure the RMS difference was 0.677 Å. That is, the B3LYP gas-phase structure was a significantly better fit to the observed crystal structure than that given by PM6, but when crystal forces were included in the PM6 calculation, the PM6 gave a slightly better fit than the B3LYP result.

The energetics involved in the sublimation process for HMPN can be modeled in three stages. In the solid phase, PM6 gives a ΔH_f of −47.8 kcal/mol for HMPN. Using the optimized geometry for the crystal form, PM6 gives a ΔH_f of −17.9 kcal/mol for HMPN in the gas phase. When the geometry is allowed to relax, the ΔH_f of the optimized gas-phase geometry decreased to −27.9 kcal/mol. From this it follows that crystal packing forces distort the geometry of HMPN so that its energy increases by 10 kcal/mol. This increase is, however, more than offset by the intermolecular stabilization energy of 29.9 kcal/mol resulting in a net sublimation energy of 19.9 kcal/mol.

Because HMPN has a very high proton affinity, it is interesting to speculate about the minimum energy structure of gas-phase 4,5-bis(hexamethyltriaminophosphazenyl)-1,8-dihydroxynaphthalene. As with 4,5-bis(dimethylamino)-1,8-dihydroxynaphthalene, PM6 predicts this system to exist as the Zwitterion in the gas phase, yet, like 4,5-bis(dimethylamino)-1,8-dihydroxynaphthalene, a BLYP/6–31G* calculation suggests that the neutral form is the more stable form, but only by 0.3 kcal/mol. That is, the possibility exists that 4,5-bis(hexamethyltriaminophosphazenyl)-1,8-dihydroxynaphthalene might form a stable gas-phase Zwitterion, but the probability of this being the case is low.

An unusual complex resulting from the reaction of elemental bromine and 1, 4-diazabicyclo[2.2.2]octane (DABCO), $(DABCO)_4{\cdot}Br_{14}$ has been reported [26] to contain an almost linear structure of three DABCO groups and seven bromine atoms (Fig. 4) with the bromine atoms
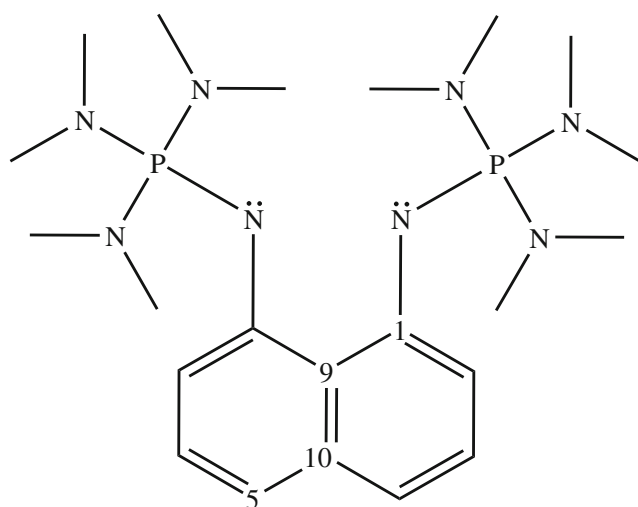

Fig. 3 1,8-bis(hexamethyltriaminophosphazenyl)naphthalene

split into three fragments of Br_3^-, Br_3^-, and Br^-, resulting in a formal charge of −3. From the X-ray structure (CSD entry DAHGUO), the authors concluded that partial protonation had occurred, that the complex of three DABCO units included four protons, and that the remaining isolated DABCO moiety was doubly protonated.

When the structure suggested by the authors was optimized using PM6, only very small changes occurred in the geometry, as shown in Table 4; all the essential features proposed by the original authors were reproduced, and are therefore confirmed. These included the conjectures that, in the set of three DABCO units, the central DABCO was doubly ionized, that the terminal DABCOs were singly ionized, and that the isolated DABCO was doubly charged. The two tribromide ions had charges of −0.98 and −1.00, with the central atom in each group carrying only an insignificant charge, in accordance with the postulated existence of a tribromide anion, and the isolated bromide ion had a partial charge of −0.78, again in accordance with the proposed structure.

N–H–O The simple amino acids exist as Zwitterions in the solid state. In this form, an –NH_3^+ group on one ion electrostatically interacts strongly with a –COO^- ion on a neighboring ion. In all cases examined, the structure of the resulting salt was reproduced with good accuracy. This included the structures of three polymorphs of glycine, α, β, and γ, a set of systems in which any differences in heats of formation could arise only from the different intermolecular interactions. PM6 predicted all three polymorphs to have heats of formation within 2 kcal/mol of each other, albeit not in the order observed experimentally. Of more interest are those cases where the energy of interaction is smaller and, consequently, the geometry of the N–H–O system would be more sensitive to the environment. A good example of such a system is methyl (+ −)-(1α,2β,8α,9α,10β)-2-chloro-4-aza-3-oxatetracyclo($8.4.0.0^{2,9}.0^{4,8}$) tetradecane-9-carboxylate oxalic acid monohydrate (CSD entry HUZKOC), in which a molecule of oxalic acid hydrogen bonds to a neutral nitrogen on the large organic fragment. In this system, PM6 predicts the N–O distance to be slightly too large, 2.61 Å compared to the reported 2.57 Å.

The existence of a uniquely short heteroatom separation in a hydrogen-bonded compound, 2-(2-(3-carboxypyridyl))-4-isopropyl-4-methyl-5-oxo-imidazole, Fig. 5, was reported in 1989 [27]. In this system (CSD entry JAZCOC01), the

pyridine and imidazole rings are almost coplanar, a condition essential to the formation of the strong hydrogen bond. The optimized gas-phase geometry predicted by AM1 and PM3, and now PM6, all have large twist angles, 47, 37, and 40°, respectively. In the solid state, however, PM6 predicts the rings to be coplanar, and the N–O distance to be 2.50 Å, which is very close to the reported 2.47 Å. In the observed crystal structure the hydrogen atom is asymmetrically positioned, with the O-H and N-H distances being 1.16 Å and 1.32 Å, and the O–H–N angle being 170°. PM6 also predicts the hydrogen to be asymmetrically positioned, but in the opposite sense: the PM6 O–H and N–H distances being 1.46 and 1.12 Å, and the O–H–N angle to be 150°.

The tendency of PM6 to incorrectly favor the Zwitterionic form of a hydroxy–amine over the neutral was investigated by modeling the three isomers of hydroxyaniline. All three isomers exist in the solid state as the neutral form; however, on optimizing the geometry using PM6, the ortho and para forms spontaneously transitioned to the Zwitterion. Meta-hydroxyaniline optimized to the neutral form, but when the calculation was repeated, starting with the Zwitterionic form, the structure optimized to the Zwitterionic form, and in that form was 1.1 kcal/mol lower in energy than the neutral form. As expected, the density increased from 1.27 to 1.34 on going from the neutral to the Zwitterion. PM6 thus has a definite and demonstrable error in its exaggerated tendency to form Zwitterions.

DABCO forms several very unusual hydrogen-bonded systems, among which are 1,4-diazabicyclo[2.2.2]octane azelaic acid (CSD entry UNEGEZ), where each DABCO forms two different types of bridging N–H–O bonds to adjacent azelaic acid molecules. PM6 predicts one of these to have the structure N–O: 2.53(2.55), N–H: 1.29(1.21), and O–H: 1.27(1.40), (X-ray structures in parentheses) and the other to be N–O: 2.75(2.61), N–H: 1.75(1.50), and O–H: 1.06(1.11). The X-ray structure shows the existence of a polymeric chain of alternating azelaic acid and DABCO units, from which it follows that the different geometries of the two N–H–O structures can only be attributed to crystal packing forces. A closely related compound is 1-diazonia-4-azabicyclo[2.2.2]octane glutarate (CSD entry UNEFIC), in which a similar structure exists but now with the bridging hydrogen being nearer to the nitrogen than to the oxygen. In UNECIF, the N–H distance predicted by PM6 was 1.10 (0.95) Å and the O–H 1.65 (1.76) Å.
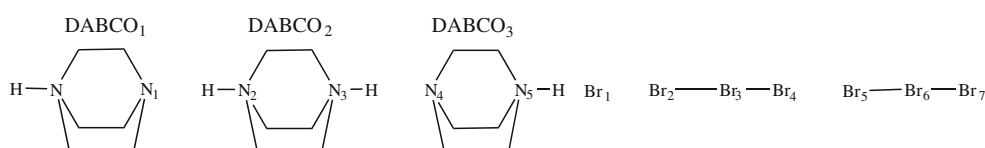

Fig. 4 Detail of bis(1,4-Diazoniabicyclo(2.2.2)octane) bis(1-aza-4-azoniabicyclo(2.2.2)octane) tetrakis(tribromide) dibromide

Table 4 Interatomic distances and charges in bis(1,4-Diazoniabicyclo(2.2.2)octane) bis(1-aza-4-azoniabicyclo(2.2.2)octane) tetrakis(tribromide) dibromide

Interatomic distances			Charges	
	X-ray	PM6		PM6
N_1-N_2	2.69	2.70	$DABCO_1$	+1.01
N_2-H		1.16	$DABCO_2$	+1.68
N_3-H		1.16	$DABCO_3$	+1.01
N_3-N_4	2.66	2.70	$DABCO_4$	+1.81
N_5-Br_1	3.17	2.97	Br_1	−0.78
Br_1-Br_2	3.60	3.00	Br_2	−0.54
Br_2-Br_3	2.51	2.48	Br_3	−0.05
Br_3-Br_4	2.59	2.50	Br_4	−0.39
Br_4-Br_5	3.25	2.98	Br_5	−0.45
Br_5-Br_6	2.45	2.47	Br_6	−0.09
Br_6-Br_7	2.69	2.53	Br_7	−0.47

$\pi-\pi$ stacking

$\pi-\pi$ stacking occurs in the polycyclic aromatic hydrocarbons where it is the result of Van der Waals (VDW) interactions between the rings. In general, VDW interactions are weaker than hydrogen bonding interactions, and, historically, have been the hardest to model using semiempirical methods. Thus, when the default single determinant wavefunction is used, VDW terms are completely absent, and, in order to mimic the effects of VDW attraction, the normal procedure is for modifications to be made to the core–core interaction. An estimate of how accurately PM6 can reproduce the VDW interaction is provided by anthracene (CSD entry ANTCEN14), benzene (CSD entry BENZEN), and coronene (CSD entry CORONE). In anthracene, the molecules are stacked in a staggered arrangement. While PM6 reproduces the density with good accuracy, the optimized structure predicts the parallel sheets of anthracene molecules to be separated by 5.37 Å rather than the observed 2.83 Å, and pairs of anthracene molecules forming "T" structures rather than the observed
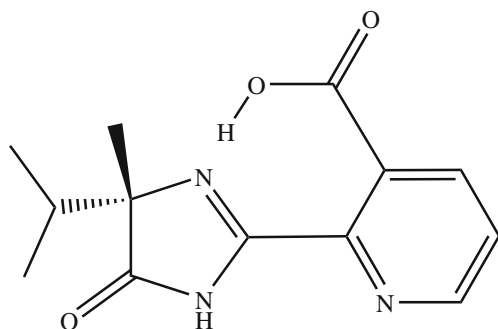
Fig. 5 2-(2-(3-Carboxypyridyl))-4-isopropyl-4-methyl-5-oxo-imidazole

"V" configuration. Conversely, both benzene and coronene crystallize with a perfect herringbone packing, and this structure is reproduced with very good accuracy by PM6, the calculated and observed inter-plane distance being essentially identical.

Very weak interactions

In small saturated hydrocarbons the strongest intermolecular interaction energy arises from instantaneous correlation or VDW forces. A consequence of this is that such compounds are, in general, highly volatile and most have very low melting points. Another characteristic of compounds of this type is that intermolecular separations are typically very large: in cyclohexane the smallest intermolecular separation is about 2.6 Å. When the structure of cyclohexane was optimized from the experimental structure using PM6, the final and X-ray structures agreed almost exactly. However, the heat of formation of each cyclohexane decreased from an initial +65.8 kcal/mol for the starting X-ray structure to the final −29.6 kcal/mol for the optimized PM6 structure, reflecting the relaxation of the X-ray C–H bond lengths, which are normally too short. However, the PM6 ΔH_f for isolated cyclohexane is predicted to be −27.5 kcal/mol, from which it follows that the energy of interaction amounted to only 2.1 kcal/mol, much less than the reported 9.0 kcal/mol. The implication is that, although PM6 was able to reproduce the observed crystal structure, the magnitude of the VDW interaction was grossly underestimated.

In solid methionine (CSD entry LMETON02) the molecules form double layers with the hydrophobic end, –CH_2–S–CH_3, on the outside. The X-ray structure shows that these layers are separated by about 2.21 Å; PM6 predicts the inter-layer separation to be much larger, 2.83 Å. Further investigation of the properties of sulfur predicted by PM6, see below, indicated the lack of any sulfur–sulfur VDW attraction. This deficiency is likely responsible for the unrealistically large inter-layer separation.

Most acetylacetonato transition metal complexes, such as tris(acetylacetonato) titanium(iv) perchlorate (CSD entry TIACPC), adopt an almost octahedral coordination of the oxygen atoms around the central metal ion. An exception is the hexamethyl acetylacetone complex of yttrium(iii), tris(2,2,6,6-tetramethylheptane-3,5-dionato)-yttrium(iii) (CSD entry HAHTOZ01), where the oxygen atoms form a trigonal prism. This structure was reproduced by PM6. The deviation from the conventional octahedral structure cannot be attributed to crystal packing forces—when the geometry of the isolated complex ion was optimized using PM6, the trigonal prism structure was retained. However, when the geometry of the unsubstituted complex, tris(acetylacetonato) yttrium(iii), was optimized, the expected D_{3d} structure

resulted, suggesting that the likely driving force was steric crowding arising from the tertiary butyl groups.

Polymorphs

5-Methyl-2-((2-nitrophenyl)amino)-3-thiophenecarbonitrile is unique in the CSD in that there are seven distinct polymorphs [28]; because three of these are red, orange, and yellow, this chemical is commonly referred to as ROY. To test the suitability of PM6 for modeling polymorphs, the structure of each of the polymorphs of ROY was optimized using PM6, starting with the X-ray geometry. In every case, the optimized PM6 structure was qualitatively the same as the X-ray structure; that is, the crystal packing arrangement was preserved. Within each unit cell, the structures of the individual molecules were reproduced with good accuracy (Table 1). Although the calculated heats of formation of the X-ray structures spanned a range of 31 kcal/mol, the optimized PM6 structures all had similar heats of formation, spanning a range of 4.3 kcal/mol, as would be expected for polymorphs.

Co-crystals

Designing crystal structures for active pharmaceutical ingredients presents an important challenge to crystal engineers. One promising avenue of research involves designing co-crystals, so determining the suitability of PM6 as a tool for this work is of obvious interest. The structures of several co-crystals were optimized using PM6; all the resulting geometries were in good agreement with the structures found in the CSD. In all of the systems investigated the two components were held together by hydrogen bonds. Thus in the bis-urea–oxalic acid co-crystal

(Fig. 6, CSD entry UROXAL01), each oxalic acid forms two strong O···H···O hydrogen bonds to the keto groups of the neighboring urea molecules, this resulting in a tri-molecular unit. In turn, these units pack together in the crystal using weaker N···H···O hydrogen bonds. PM6 predicts that the co-crystal would be significantly more stable than its two separate precursors: the predicted ΔH_f of the co-crystal is −313.4 kcal/mol, while the sum of the ΔH_f of the precursors, Table 2, is −306.9 kcal/mol.

Another example of such a co-crystal is provided by isonicotinamide 3-hydroxybenzoic acid, Fig. 7 (CSD entry LUNMEM), in which each hydroxybenzoic acid forms hydrogen bonds to three adjacent isonicotinamide molecules in a complicated three-dimensional structure. As with the previous co-crystal, PM6 predicts LUNMEM to be more stable than its precursors but, in this case, only by 0.1 kcal/mol.

Metal-containing species

Many organic compounds that contain elements of Groups I or II are ionic salts. A good example is calcium oxalate, which forms three well characterized solids: whewellite, $(Ca(C_2O_4)\cdot(H_2O))$, weddellite, $(Ca(C_2O_4)\cdot2(H_2O))$, and caoxite, $(Ca(C_2O_4)\cdot3(H_2O))$, at least one of which the author involuntarily prepared in vivo, the separation of which from the surrounding organic material was accompanied by severe algia. The structures of all three minerals are reproduced with good accuracy. Anhydrous disodium oxalate exists as the mineral natroxalate. In contrast to the calcium oxalates, the structure predicted by PM6 was completely different to that observed: within each oxalate dianion, one carboxylate group was rotated by ~90° to give an approximately D_{2d} structure. More seriously, some

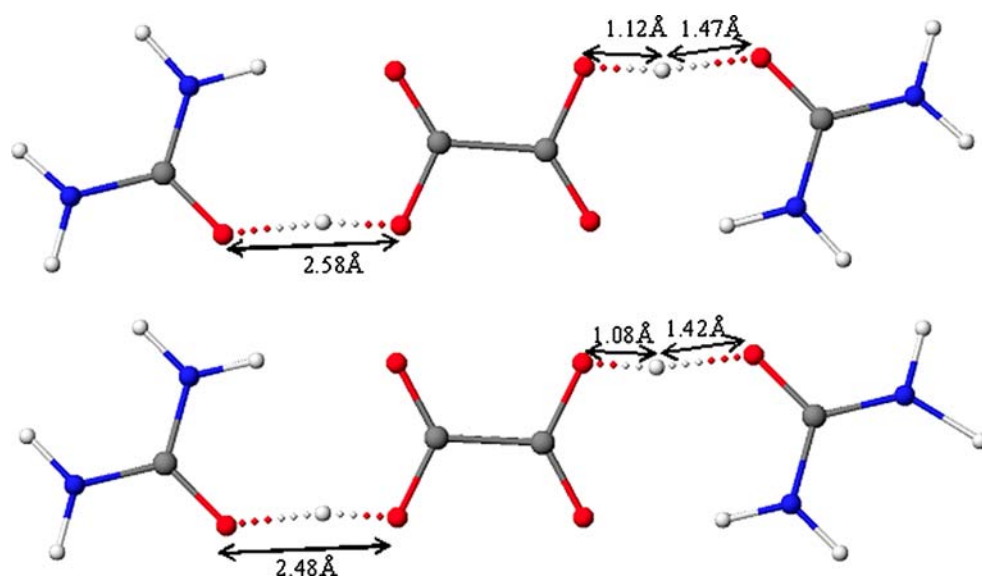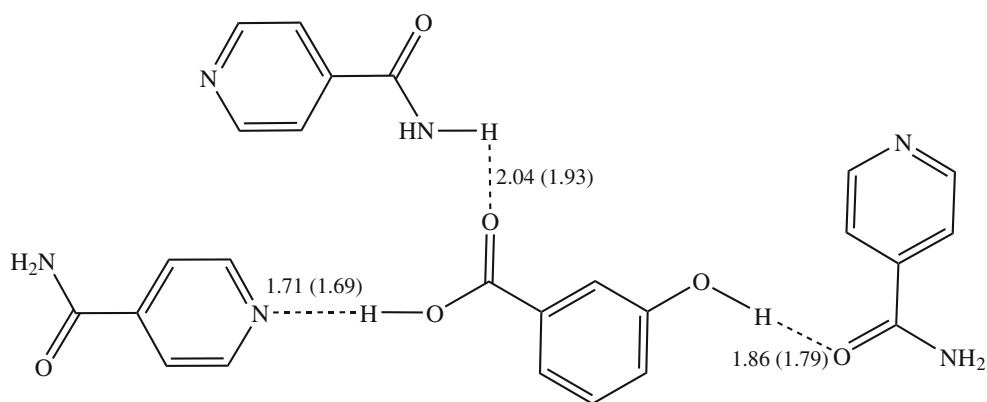Fig. 6 Crystal structure of the co-crystal of urea and oxalic acid

Fig. 7 Co-crystal of isonicotina-
mide and 3-hydroxybenzoic acid
showing PM6 hydrogen-bond
lengths (X-ray in parenthesis)

sodium–sodium distances became unrealistically short, 1.23 Å instead of the observed 3.30 Å. This specific error is attributable to the faulty PM6 values of the Na–Na core–core parameters.

Solids containing metal complex ions can be regarded as salts: that is, as molecular metal complexes, cations or anions, plus counterions. An example is bromo-tris(2-dimethylaminoethyl)amine-manganese(ii) (CSD entry DAEAMN), where the metal complex consists of the neutral tris(2-dimethylaminoethyl)amine that chelates a manganese dication, the whole complex behaving like a large dication. Two bromide counterions are present for each such complex ion in order to maintain electroneutrality. PM6 predicts the single N–Mn distance to be 2.09 Å versus 2.19 Å in the X-ray structure, and the three N–Mn to be 1.96 Å versus 2.27 Å in the X-ray.

Inorganic compounds

In general, most inorganic solids differ from crystalline organic compounds in that they do not involve discrete molecules. Instead, they exist as extended covalently or ionically bound infinite systems. A consequence of this is that identification of simple structural units in inorganic solids is often either difficult or impossible.

A more subtle consequence is that the band-structure of inorganic solids is usually more complicated than those of organic solids: for the same reciprocal distance, bands arising from inorganic solids generally have a much greater curvature than those for organic solids, this being a consequence of the strong bonds that extend throughout such solids. Conversely, in most organic solids there is an intermolecular gap that effectively confines molecular orbitals to individual molecules. This means that the band-structure of most organic compounds consists of relatively flat bands. In the cluster method, the Γ point represents the entire Brillouin zone so that a much larger cluster must be used when inorganic solids that are not composed of discrete molecules are modeled. In practice

this means that the cluster used has to be large enough to contain a sphere of radius 10–12 Å, in contrast to the 7–8 Å used in modeling organic solids.

A survey of the applicability of PM6 to a wide range of inorganic solids was carried out, using structural reference data obtained from the American Mineralogist Crystal Structure Database [29]. As with organic crystals, the starting geometry was the X-ray structure. But, in contrast to most organic solids, many inorganic solids had to be modified before a PM6 calculation could be started. Semiempirical calculations require that a definite structure be used. However, some minerals, such as forsterite, $(Fe^{II}, Mg)_2SiO_4$, are of variable composition, with one or more sites occupied by the two types of metal atoms at random. Before a calculation on such a system can be performed, all variable atoms have to be replaced by definite atoms. In the case of forsterite, all iron atoms were replaced by magnesium atoms. In a few solids there was still disorder in the lattice although the formula was stoichiometric. In spinel, for example, the formula is $MgAl_2O_4$ but, in the observed crystal structure, a small fraction of the aluminum sites are replaced by magnesium atoms, and vice versa. Of necessity, the model used in the PM6 calculation was idealized so that sites that were 80% magnesium were made 100% magnesium, and sites that were 90% aluminum were made 100% aluminum. Although this idealized structure does not occur naturally it could be considered a good approximation to the observed structure.

Some minerals that contain hydrogen atoms did not have the positions of these atoms reported. In those cases a preliminary calculation was carried out in which the hydrogen atoms were positioned in likely sites, and then the positions of those atoms optimized. During this operation, the positions all the other atoms were fixed at the experimental values.

The results can conveniently be partitioned according to the types of solids involved and sequenced in order of complexity. In most cases, the calculated structure was similar to the X-ray structure in that no bonds were made or

Table 5 Calculated and X-ray structural parameters for inorganic compounds. Values of a, b, c are in Ångstroms; α, β, γ in degrees

	PM6						X-ray						Density (g/cm³)[d]
	a	b	c	α	β	γ	a	b	c	α	β	γ	
Elements													
Diamond (C)	3.59	3.59	3.59	90.0	90.0	90.0	3.57	3.57	3.57	90.0	90.0	90.0	3.44 (3.52)
Glitter (C)[e]	2.60	2.60	5.96	90.1	91.0	90.0	2.60	2.60	5.90	90.0	90.0	90.0	2.98 (3.08)
Graphite (C)	2.47	2.478	7.17	89.6	90.0	60.0	2.46	2.46	6.70	90.0	90.0	60.0	2.12 (2.28)
Silicon (Si)	5.11	5.11	5.11	90.0	90.0	90.0	5.43	5.43	5.43	90.0	90.0	90.0	2.78 (2.33)
Sulfur (S)	11.77	14.01	26.81	90.4	90.2	89.8	10.43	12.83	24.36	90.0	90.0	90.0	1.54 (2.09)
Group I Halides													
Cesium bromide (CsBr)	4.35	4.35	4.35	90.0	90.0	90.0	4.30	4.30	4.30	90.0	90.0	90.0	4.29 (4.43)
Cesium chloride (CsCl)	4.17	4.18	4.18	90.3	89.8	89.7	4.12	4.12	4.12	90.0	90.0	90.0	3.84 (3.99)
Cesium fluoride (CsF)	6.02	6.02	6.02	90.0	90.0	90.0	6.01	6.01	6.01	90.0	90.0	90.0	4.62 (4.64)
Cesium iodide (CsI)	4.73	4.73	4.73	90.0	90.0	90.0	4.57	4.57	4.57	90.0	90.0	90.0	4.07 (4.51)
Potassium bromide (KBr)	6.79	6.75	6.75	87.6	90.0	90.0	6.61	6.61	6.61	90.0	90.0	90.0	2.55 (2.74)
Potassium chloride (KCl)	6.12	6.12	6.12	90.0	90.0	90.0	6.29	6.29	6.29	90.0	90.0	90.0	2.16 (1.99)
Potassium fluoride (KF)	5.03	5.03	5.03	90.0	90.0	90.0	5.38	5.38	5.38	90.0	90.0	90.0	3.02 (2.48)
Potassium iodide (KI)	7.17	7.17	7.17	90.1	90.0	90.0	7.07	7.07	7.07	90.0	90.0	90.0	2.99 (3.12)
Lithium bromide (LiBr)	5.49	5.48	5.48	90.0	90.0	90.0	5.50	5.50	5.50	90.0	90.0	90.0	3.50 (3.46)
Lithium chloride (LiCl)	5.13	5.13	5.13	90.0	90.0	90.0	5.14	5.14	5.14	90.0	90.0	90.0	2.09 (2.07)
Lithium fluoride (LiF)	4.03	4.04	4.05	89.9	89.9	90.0	4.03	4.03	4.03	90.0	90.0	90.0	2.62 (2.64)
Lithium iodide (LiI)	6.04	6.03	6.04	90.0	90.0	90.0	6.03	6.03	6.03	90.0	90.0	90.0	4.04 (4.06)
Sodium bromide (NaBr)	6.27	6.20	6.20	85.9	90.0	90.0	6.04	6.04	6.04	90.0	90.0	90.0	2.84 (3.10)
Sodium chloride (NaCl)	6.24	6.24	6.24	90.0	90.0	90.0	5.60	5.60	5.60	90.0	90.0	90.0	1.60 (2.21)
Sodium fluoride (NaF)	4.79	4.79	4.79	90.2	90.1	89.8	4.61	4.61	4.61	90.0	90.0	90.0	2.54 (2.85)
Sodium iodide (NaI)	7.02	7.02	7.02	90.0	90.0	90.0	6.47	6.47	6.47	90.0	90.0	90.0	2.87 (3.67)
Rubidium bromide (RbBr)	6.87	6.88	6.88	90.3	90.0	90.0	6.90	6.90	6.90	90.0	90.0	90.0	3.38 (3.35)
Rubidium chloride (RbCl)	6.70	6.68	6.68	89.1	90.0	90.0	6.63	6.63	6.63	90.0	90.0	90.0	2.69 (2.76)
Rubidium fluoride (RbF)	5.86	5.86	5.86	90.0	90.0	90.0	6.01	6.01	6.01	90.0	90.0	90.0	3.45 (3.20)
Rubidium iodide (RbI)	7.84	7.76	7.76	83.3	90.0	90.0	7.35	7.35	7.35	90.0	90.0	90.0	3.01 (3.55)
Halides													
Calcium iodide (CaI₂)	6.21	4.48	4.48	119.9	90.0	90.1	6.96	4.48	4.48	120.0	90.0	90.0	4.51 (4.04)
Carnallite (KMg(H₂O)₆Cl₃)	15.86	22.31	10.24	73.1	92.8	95.9	16.12	22.47	9.55	90.0	90.0	90.0	1.61 (1.60)
Chloromagnesite (MgCl₂)	19.40	3.75	3.75	119.9	89.1	90.9	17.59	3.60	3.60	107.9	90.0	90.0	2.00 (2.41)
Cryolite (Na₃AlF₆)	5.67	8.03	5.69	90.1	90.0	89.9	5.46	7.80	5.61	90.0	90.0	90.2	2.69 (2.92)
Fluorite (CaF₂)	5.49	5.45	5.48	90.0	90.0	90.0	5.46	5.43	5.46	90.0	90.0	90.0	3.17 (3.20)
Frankdicksonite (BaF₂)	6.22	6.22	6.23	89.9	90.0	90.0	6.20	6.20	6.20	90.0	90.0	90.0	4.83 (4.89)
Hydrophilite (CaCl₂)	4.28	6.42	6.38	89.9	90.0	90.0	4.20	6.43	6.24	90.0	90.0	90.0	2.10 (2.19)
Magnesium dibromide (MgBr₂)	5.83	4.02	4.02	119.1	91.6	88.3	6.26	3.81	3.81	120.0	90.0	90.0	3.71 (3.88)
Magnesium diiodide (MgI₂)	6.28	4.27	4.24	118.2	90.6	89.6	6.88	4.14	4.14	120.0	90.0	90.0	4.61 (4.52)
Na₃(TiCl₆)	10.23	7.35	7.00	90.0	87.1	89.8	9.81	7.09	6.68	90.0	89.7	90.0	2.08 (2.36)
Neighborite (NaMgF₃)	8.14	8.13	8.13	90.0	90.0	90.1	7.67	7.67	7.67	91.4	90.0	90.2	2.57 (3.07)
Sal ammoniac (NH₄Cl)	4.07	4.07	3.68	90.0	89.9	73.5	3.86	3.86	3.86	90.0	90.0	90.0	1.52 (1.54)
Sellaite (MgF₂)	3.07	4.62	4.62	90.0	90.0	90.0	3.05	4.63	4.63	90.0	90.0	90.0	3.16 (3.17)

Oxides													
α Crystobalite (SiO$_2$)	5.36	5.36	7.47	89.9	90.1	89.6	4.98	4.98	6.95	90.0	90.0	90.0	1.86 (2.32)
α Quartz (SiO$_2$)	5.19	5.18	5.74	90.0	90.0	120.0	4.91	4.91	5.41	90.0	90.0	120.0	2.24 (2.65)
Anatase (TiO$_2$)	10.29	3.92	3.92	90.1	90.2	89.6	9.52	3.78	3.78	90.0	90.0	90.0	3.36 (3.90)
Arkelite (ZrO$_2$)	6.09	4.91	5.50	90.0	89.9	90.1	5.13	5.13	5.13	90.0	90.0	90.0	4.97 (6.06)
Barium oxide (BaO)	5.64	5.64	5.64	89.8	90.0	90.0	5.63	5.63	5.63	90.0	90.0	90.0	5.69 (5.72)
β Quartz (SiO$_2$)	5.19	5.18	5.74	90.0	90.0	119.9	5.00	5.00	5.46	90.0	90.0	60.0	2.24 (2.54)
β Tridymite (SiO$_2$)	5.34	5.34	8.72	90.0	90.4	90.0	5.05	5.05	8.27	90.0	90.0	120.0	1.86 (2.18)
Brookite (TiO$_2$)	5.42	5.67	9.61	90.0	90.0	90.0	5.45	5.14	9.17	90.0	90.0	90.0	3.59 (4.13)
Cassiterite (SnO$_2$)	4.94	4.96	3.49	93.9	93.9	93.8	5.14	4.74	3.19	94.3	94.3	90.0	5.86 (7.00)
Chabazite (SiO$_2$)	9.55	9.56	9.56	93.9	93.9	93.8	9.40	9.40	9.40	94.3	94.3	94.3	1.38 (1.46)
Coesite (SiO$_2$)	7.43	7.49	7.49	90.0	60.8	90.0	7.17	12.37	7.14	90.0	59.7	90.0	2.56 (2.92)
Claudetite (As$_2$O$_3$)	14.00	5.43	5.06	89.8	85.8	81.8	12.87	4.54	5.25	86.2	90.0	90.0	3.46 (4.29)
Corundum (Al$_2$O$_3$) ($a=b=c,\alpha=\beta=\gamma$)	5.13	5.12	5.12	56.2	56.2	56.2	5.13	5.13	5.13	55.3	55.3	55.3	3.91 (3.99)
Corundum (Al$_2$O$_3$) ($a\neq c,\alpha=\beta\neq\gamma$)	4.83	4.83	12.91	90.0	90.0	120.0	4.76	4.76	12.99	90.0	90.0	120.0	3.91 (3.99)
Ice-I (Ice-Ih)	7.45	7.09	7.09	90.5	90.3	89.3	7.83	4.52	7.38	90.0	105.5	90.0	1.04 (0.92)
Ice-II	7.22	12.26	5.91	89.9	106.5	91.0	7.78	12.98	6.24	90.0	90.0	90.0	1.43 (1.18)
Ice-III	6.69	6.50	6.46	89.9	89.5	90.7	6.73	6.73	6.73	90.0	90.0	89.1	1.28 (1.18)
Ice-V	7.02	9.44	8.80	70.1	90.6	92.6	7.53	10.35	9.20	70.7	90.0	89.1	1.53 (1.24)
Ice-VI	5.38	5.94	5.90	90.9	89.5	89.0	5.70	6.18	6.18	90.0	90.0	90.0	1.59 (1.37)
Ice-VIII	6.44	4.36	4.36	90.0	90.5	90.1	6.41	4.45	4.45	90.0	90.0	90.0	1.96 (1.89)
Ice-X (0Gpa)	2.90	2.90	2.90	90.0	90.0	90.0	2.73	2.73	2.73	90.0	90.0	90.0	2.46 (2.96)
Ice-X (62Gpa)	2.72	2.72	2.72	90.0	90.0	90.0	2.73	2.73	2.73	90.0	90.0	90.0	2.98 (2.96)
Ice-XI	7.62	4.26	7.15	90.1	90.2	89.9	7.83	4.52	7.38	90.0	90.0	90.0	1.03 (0.92)
Ice-XIII	9.77	7.05	8.46	90.5	69.5	90.6	10.29	7.47	9.24	70.3	90.0	90.0	1.54 (1.25)
Ice-XIV	3.83	8.07	7.84	90.3	90.0	90.3	4.08	8.35	8.14	90.0	90.0	90.0	1.48 (1.29)
Keggin Dodecatungstophosphoric acid hexahydrate H$_3$(PO$_4$@W$_{12}$O$_{36}$).6(H$_2$O)	13.01	13.43	13.00	90.4	92.2	89.7	12.51	12.51	12.51	90.0	90.0	90.0	4.37 (5.07)
Lime (CaO)	4.95	4.95	4.95	90.0	90.1	90.0	4.82	4.82	4.81	90.0	90.0	90.0	3.07 (3.34)
Mordenite (SiO2)	18.68	21.00	7.75	90.2	90.0	90.0	18.13	20.49	7.52	90.0	90.0	90.0	1.58 (1.71)
Lead oxide (PbO)	3.83	4.74	3.83	89.9	90.0	89.9	3.98	5.02	3.98	90.0	90.0	90.0	10.68 (9.34)
Periclase (MgO)	4.29	4.29	4.29	90.0	90.0	90.0	4.22	4.22	4.22	90.0	90.0	90.0	3.40 (3.57)
Perovskite (CaTiO$_3$)	5.38	5.54	7.84	89.7	90.1	90.1	5.38	5.44	7.64	90.0	90.0	90.0	3.87 (4.04)
Plattnerite (PbO$_2$) (Rutile structure)	3.50	4.66	4.73	89.8	90.0	90.0	3.39	4.96	4.96	90.0	90.0	90.0	10.31 (9.55)
Rutile (TiO$_2$)	4.84	4.84	3.12	90.0	90.0	90.0	4.59	4.59	2.96	90.0	90.0	90.0	3.71 (4.25)
Stishovite (SiO$_2$)	2.82	4.38	4.38	90.0	90.0	90.0	2.67	4.18	4.18	90.0	90.0	90.0	3.70 (4.28)
Zincite (ZnO)	3.33	3.33	5.26	90.0	90.0	60.3	3.25	3.25	5.20	90.0	90.0	60.0	5.33 (5.68)
Spinels													
Chrysoberyl ((BeO)(Al$_2$O$_3$))	4.58	9.58	5.72	90.3	90.0	90.0	4.40	9.33	5.44	90.0	90.0	90.0	3.36 (3.78)
Spinel (MgAl$_2$O$_4$)	8.31	16.63	16.63	90.2	90.0	90.0	8.21	16.41	16.41	90.0	90.0	90.0	3.29 (3.42)
Borates													
Boracite (MgO)$_5$(MgCl$_2$)(B$_2$O$_3$)$_7$	8.54	8.59	12.11	90.0	90.0	90.0	8.55	8.55	12.09	90.0	90.0	90.0	2.93 (2.95)
Borax (Na$_2$B$_4$O$_7$·10(H$_2$O))	12.33	11.27	16.20	105.1	93.7	83.0	12.19	10.74	11.89	73.4	100.4	84.0	1.18 (1.70)
Parahilgardite (Ca$_2$(B$_5$O$_9$)Cl(H$_2$O))	17.19	6.39	6.18	119.6	99.8	84.6	17.50	6.49	6.31	119.2	100.4	84.0	2.84 (2.69)
Ulexite (NaCaB$_5$O$_9$·8(H$_2$O))	6.77	13.62	8.88	97.4	108.2	103.3	6.68	12.87	8.82	110.0	109.1	90.4	1.82 (1.96)
Carbonates													
Aragonite (CaCO$_3$)	4.55	7.80	6.08	90.0	90.0	90.1	4.96	7.97	5.74	90.0	90.0	90.0	3.08 (2.93)
Azurite ((CuO)$_3$(CO$_2$)$_2$(H$_2$O))	5.73	5.12	9.50	91.0	90.1	90.0	5.85	5.01	10.34	92.4	90.0	90.0	4.10 (3.78)
Calcite (CaCO$_3$)	6.32	6.32	6.32	45.2	45.2	45.2	6.36	6.36	6.36	46.1	46.1	46.1	2.87 (2.73)

Table 5 (continued)

	PM6						X-ray						Density (g/cm³)ᵈ
	a	b	c	α	β	γ	a	b	c	α	β	γ	
Cerussite (PbCO₃)	5.84	6.43	7.09	89.3	91.4	89.3	5.18	6.14	8.49	90.0	90.0	90.0	6.66 (6.57)
Dawsonite ((Na₂O)(Al₂O₃)(CO₂)₂(H₂O₂))	6.73	5.98	5.75	90.0	90.2	90.0	6.71	5.21	5.58	90.0	90.0	90.0	2.07 (2.45)
Dolomite (CaMg(CO₃)₂)	4.74	4.74	15.32	90.0	89.9	120.0	4.81	4.81	16.02	90.0	90.0	120.0	3.08 (2.86)
Huntite (CaMg₃ (CO₃)₄)	9.40	9.35	7.67	90.5	90.2	59.8	9.50	9.50	7.82	90.0	90.0	60.0	3.02 (2.88)
Ikaite ((CaO)(CO₂)(H₂O)₆)	8.72	7.85	10.81	91.3	71.3	90.3	8.87	8.23	11.02	90.0	69.8	90.0	1.97 (1.83)
Kalicinite(KHCO₃)	19.63	4.16	4.16	90.9	77.8	90.2	15.19	5.63	3.71	90.0	75.5	90.0	2.00 (2.17)
Magnesite (MgCO₃)	14.91	4.57	7.95	90.0	90.2	90.0	15.02	4.63	8.03	90.0	90.0	90.0	3.11 (3.01)
Natrite (Na₂CO₃)	9.93	6.81	5.66	90.2	90.0	102.0	8.91	6.04	5.24	90.0	90.0	101.3	1.88 (2.55)
Thermonatrite (Na₂CO₃(H₂O))	5.85	11.81	7.81	90.0	89.3	91.6	5.26	10.72	6.47	90.0	90.0	90.0	1.53 (2.26)
Otavite (CdCO₃)	5.56	4.71	15.59	90.0	88.9	90.9	4.93	4.27	16.27	90.0	90.0	90.0	4.21 (5.02)
Smithsonite (ZnCO₃)	4.83	4.83	14.46	89.5	90.0	120.0	4.65	4.65	15.03	90.0	90.0	120.0	4.28 (4.44)
Strontianite (SrCO₃)	6.59	8.65	5.02	90.1	90.0	90.0	6.00	8.36	5.09	90.0	90.0	90.0	3.42 (3.84)
Witherite (BaCO₃)	4.99	7.26	8.49	89.9	89.9	89.9	5.31	6.43	8.90	90.0	90.0	90.0	4.26 (4.32)
Zabuyelite (Li₂CO₃)	5.26	8.80	6.23	67.5	90.0	90.0	4.97	8.36	6.20	65.2	90.0	90.0	1.84 (2.10)
Group IV													
Moissanite (SiC)	5.00	5.36	3.09	90.0	90.0	90.0	5.05	5.33	3.08	90.0	90.0	90.0	3.22 (3.22)
Silicon carbide (SiC)	4.39	4.39	4.39	90.0	90.0	90.0	4.35	4.35	4.35	90.0	90.0	90.0	3.16 (3.24)
Groups III-V													
Aluminum antimonide (AlSb)	5.94	5.94	5.94	90.0	90.0	90.0	6.14	6.14	6.14	90.0	90.0	90.0	4.71 (4.28)
Aluminum arsenide (AlAs)	5.91	5.91	5.91	90.0	90.0	90.0	5.66	5.66	5.66	90.0	90.0	90.0	3.27 (3.73)
Aluminum nitride (AlN)	4.40	4.40	4.40	90.0	90.0	90.0	4.37	4.37	4.37	90.0	90.0	90.0	3.20 (3.25)
Aluminum phosphide (AlP)	5.49	5.49	5.49	90.0	90.0	90.0	5.42	5.42	5.42	90.1	90.1	90.1	2.33 (2.42)
Boron Nitride (BN)	3.64	3.64	3.64	90.0	90.0	90.0	3.61	3.61	3.61	90.0	90.0	90.0	3.42 (3.49)
Gallium arsenide (GaAs)	5.66	5.66	5.66	90.0	90.0	90.0	5.65	5.65	5.65	90.0	90.0	90.0	5.31 (5.32)
Indium arsenide (InAs)	5.97	5.97	5.97	90.0	90.0	90.0	6.06	6.06	6.06	90.0	90.0	90.0	5.93 (5.67)
Group VI													
Cadmium telluride (CdTe)	5.58	5.58	5.58	90.0	90.0	90.0	6.36	6.36	6.36	90.0	90.0	90.0	9.17 (6.20)
Coloradoite (HgTe)	6.61	6.61	6.61	90.0	90.0	90.0	6.32	6.32	6.32	90.0	90.0	90.0	7.56 (8.63)
Lead selenide (PbSe)	4.34	4.34	4.35	90.0	90.0	90.0	6.17	6.17	6.17	90.0	90.0	90.0	23.18 (8.10)
Lead telluride (PbTe)	7.87	7.87	7.87	90.0	90.0	90.0	7.10	7.10	7.10	90.0	90.0	90.0	4.56 (6.22)
Stibnite (Sb₂S₃)	10.67	4.05	11.40	90.0	87.9	90.0	11.29	3.83	11.21	90.0	90.0	90.0	4.58 (4.66)
Stilleite (ZnSe)	5.95	5.95	5.95	90.0	90.0	90.0	5.54	5.54	5.54	90.0	90.0	90.0	4.56 (5.65)
Hydrides													
BH₃NH₃	4.82	4.24	5.25	94.4	90.1	90.1	4.99	4.89	5.39	90.0	90.0	90.0	0.96 (0.78)
Hydroxides													
Brucite (Mg(OH)₂)	3.20	3.21	4.30	89.8	90.0	120.1	3.14	3.14	4.77	90.0	90.0	120.0	2.53 (2.38)
Diaspore (AlO(OH))	4.17	9.75	3.00	90.0	89.9	89.8	4.40	9.43	2.85	90.0	90.0	90.0	3.26 (3.38)
Gibbsite (Al(OH)₃)	9.04	10.38	4.45	90.1	90.0	90.0	9.74	10.16	4.34	90.0	85.5	90.0	2.48 (2.42)
MnO(OH)	7.00	6.23	13.45	90.2	88.4	89.8	5.29	5.80	13.31	89.7	89.6	90.0	2.99 (4.29)
Orthoboric acid (B(OH)₃)	10.97	7.26	7.26	120.1	83.4	100.4	6.35	7.04	7.02	119.8	78.5	87.5	0.84 (1.56)
Nitrates													
Ammonium nitrate (NH₄NO₃)	8.82	13.85	10.76	95.8	89.4	72.1	9.49	12.23	10.76	90.0	89.9	91.3	1.71 (1.70)
Niter (KNO₃)	8.96	5.38	6.59	90.6	89.1	89.5	9.16	5.41	6.43	90.0	90.0	90.0	2.11 (2.11)
Nitratine (NaNO₃)	18.66	5.43	4.08	90.1	96.2	90.5	16.83	5.07	4.39	90.0	90.0	90.0	2.06 (2.26)

Mineral	a	b	c	α	β	γ	a	b	c	α	β	γ	ρ
Phosphates													
Bobierrite (Mg$_3$(PO$_4$)$_2$(H$_2$O)$_8$)	10.46	26.99	4.58	89.4	104.3	93.0	10.07	27.93	4.67	90.0	105.0	90.0	2.16 (2.13)
Calcium phosphate (Ca$_3$(PO$_4$)$_2$)	5.53	19.97	5.34	88.7	56.9	90.3	5.25	18.67	5.25	90.0	60.0	90.0	3.13 (3.47)
Fluorapatite (Ca$_5$(PO$_4$)$_3$F)	9.28	9.29	6.87	89.8	90.0	59.6	9.37	9.37	6.88	90.0	90.0	60.0	3.28 (3.20)
Dorfmanite (Na$_2$HPO$_4$2(H$_2$O))	8.49	11.77	15.90	89.9	90.2	90.1	6.60	10.36	16.87	90.0	90.0	90.0	1.48 (2.05)
Lithiophosphate (Li$_3$PO$_4$)	5.49	6.75	11.62	90.0	90.1	90.1	4.93	6.12	10.53	90.0	90.0	90.0	1.79 (2.42)
Newberyite (MgHPO$_4$·3(H$_2$O))	9.87	10.06	10.32	90.0	90.0	90.4	10.02	10.68	10.20	90.0	90.0	90.0	2.26 (2.12)
Phosphammite ((NH$_4$)$_2$HPO$_4$)	7.75	6.74	10.91	89.9	66.4	90.1	8.03	6.70	11.04	90.0	66.6	90.0	1.68 (1.61)
Potassium dihydrogen phosphate (KH$_2$PO$_4$)	7.69	7.57	7.59	88.0	91.0	88.7	6.97	7.45	7.45	90.0	90.0	90.0	2.05 (2.33)
Potassium oxovanadium(IV) phosphate(V) (KVPO$_5$)	7.66	7.71	8.81	90.2	89.8	90.0	7.61	7.31	8.41	90.0	90.0	90.0	2.56 (2.85)
Sodalite (AlPO$_4$)	8.32	8.33	8.60	89.4	90.5	90.4	9.04	9.04	9.04	90.0	90.0	90.0	2.04 (1.64)
Wagnerite (Mg$_2$PO4F)	12.11	12.42	9.65	90.0	74.6	90.0	11.96	12.68	9.64	90.0	71.7	90.0	3.09 (3.11)
Sulfates													
Alum (KAl(SO$_4$)$_2$·12(H$_2$O))	12.31	12.35	12.27	89.3	89.8	90.5	12.18	12.18	12.19	90.0	90.0	90.0	1.69 (1.74)
Aluminite (Al$_2$SO$_6$ 9(H$_2$O))	6.90	15.33	11.29	90.2	108.0	90.0	7.44	15.58	11.70	90.0	110.2	90.0	2.01 (1.80)
Alunite (KAl$_3$(SO$_4$)$_2$(OH)$_6$)	7.01	7.00	17.38	89.9	89.7	60.2	6.96	6.96	17.35	90.0	90.0	60.0	2.79 (2.84)
Ammonium sulfate (2(NH$_4$)SO$_4$)	10.99	5.75	7.29	91.0	90.1	90.2	10.64	5.99	7.78	90.0	90.0	90.0	1.90 (1.77)
Anglesite (PbSO$_4$)	7.40	6.33	7.00	92.4	91.4	95.1	6.96	8.48	5.40	90.0	90.0	90.0	6.18 (6.32)
Anhydrite (CaSO$_4$)	6.79	8.39	7.09	87.1	90.0	90.2	6.99	6.25	7.00	90.0	90.0	90.0	2.24 (2.96)
Arcanite (K$_2$SO$_4$)	7.55	9.62	6.38	89.9	90.0	90.2	7.48	10.07	5.76	90.0	90.0	90.0	2.50 (2.67)
Barite (BaSO$_4$)	7.44	5.37	9.34	90.1	89.9	90.0	7.15	5.46	8.88	90.0	90.0	90.0	4.15 (4.47)
Celestine (SrSO$_4$)	8.43	5.67	6.93	90.1	89.8	91.1	8.39	5.36	6.89	90.0	90.0	90.0	3.68 (3.94)
Epsomite (MgSO$_4$ (H$_2$O)$_7$)	11.82	11.81	6.19	89.9	90.0	90.1	11.89	12.01	6.86	90.0	90.0	90.0	1.89 (1.67)
Glauberite (CaNaSO$_4$)	8.41	10.59	10.18	50.5	90.1	90.1	8.31	8.53	10.13	76.7	90.0	90.0	2.64 (2.78)
Gypsum (CaSO$_4$·2H$_2$O)	5.58	14.42	6.71	90.2	106.1	90.2	5.67	15.10	6.49	90.0	118.5	90.0	2.21 (2.34)
Hanksite (KCl)$_2$(Na$_2$O)$_{22}$(SO$_3$)$_{18}$(CO$_2$)$_4$	11.65	11.76	21.87	88.0	90.5	119.2	10.47	10.46	21.19	90.0	90.0	120.0	1.99 (2.59)
Hexahydrite (MgSO$_4$·6(H$_2$O))	24.30	6.92	9.65	89.8	78.2	90.0	24.41	7.21	10.11	90.0	81.7	90.0	1.91 (1.72)
Langbeinite (K$_2$Ca$_2$(SO$_4$)$_3$)	10.70	10.38	10.58	89.3	89.0	90.2	10.43	10.43	10.43	90.0	90.0	90.0	2.52 (2.62)
Leonite (K$_2$Mg(SO$_4$)$_2$·4(H$_2$O))	10.73	8.99	13.04	89.9	95.6	90.0	9.85	9.47	11.78	90.0	84.7	90.0	1.95 (2.23)
Mercallite (KHSO$_4$)	9.85	17.85	8.715	90.2	90.0	90.6	9.80	18.96	8.41	90.0	90.0	90.0	2.36 (2.32)
Thenardite (Na$_2$SO$_4$)	6.34	13.28	11.31	90.2	91.1	90.2	5.87	12.30	9.83	90.0	90.0	90.0	1.98 (2.66)
Zinkosite (ZnSO$_4$)	4.93	6.36	8.85	90.1	86.6	90.1	4.77	6.75	8.60	90.0	90.0	90.0	3.87 (3.87)
Sulfides													
Cinnabar (HgS)	4.31	4.31	9.87	90.0	88.9	60.0	4.15	4.15	9.51	90.0	90.0	60.0	7.30 (8.17)
Galena (PbS)	5.87	5.87	5.87	90.0	95.4	90.0	5.94	5.94	5.94	90.0	90.0	90.0	7.86 (7.60)
Greenockite (CdS)	6.70	7.15	4.13	90.0	91.8	89.9	6.75	7.16	4.13	90.0	90.0	89.9	4.85 (4.80)
Hawleyite (CdS)	3.84	3.84	3.84	90.0	87.8	90.0	5.83	5.83	5.83	90.0	90.0	90.0	5.02 (4.84)
Metacinnabar (HgS)	6.29	6.29	6.29	90.0	90.0	90.0	5.85	5.85	5.85	90.0	90.0	90.0	6.21 (7.71)
Niningerite (MgS)	5.20	5.20	5.20	90.0	90.0	90.0	5.20	5.20	5.20	90.0	90.0	90.0	2.66 (2.66)
Sphalerite (ZnS)	5.41	5.41	5.41	90.0	90.0	90.0	5.43	5.43	5.43	90.0	90.0	90.0	4.08 (4.04)
Wurtzite (ZnS)	3.84	3.84	6.30	90.0	90.0	90.0	3.81	3.81	6.23	90.0	90.0	60.0	4.03 (4.12)
Vanadates, tungstates, chromates, molybdates													
Descloizite (PbZn(VO$_4$)(OH))	8.36	6.30	9.35	94.7	88.9	90.8	7.60	6.07	9.44	90.0	90.0	90.0	5.48 (6.17)
Ferberite (FeWO$_4$)	5.00	5.35	4.64	89.9	95.4	89.9	4.95	5.70	4.73	90.0	90.0	90.0	8.17 (7.55)
Phoenicochroite (Pb$_2$CrO$_5$)	5.67	7.74	15.67	42.1	91.8	88.2	5.67	7.14	14.00	64.8	90.0	90.0	7.90 (7.07)
Scheelite (CaWO$_4$)	5.45	5.73	14.43	91.0	87.8	92.4	5.24	5.24	11.38	90.0	90.0	90.0	4.25 (6.12)

Table 5 (continued)

	PM6						X-ray						Density (g/cm^3)[d]
	a	b	c	α	β	γ	a	b	c	α	β	γ	
Tarapacaite (K$_2$CrO$_4$)	10.57	6.05	7.70	89.9	89.8	90.3	10.39	5.92	7.66	90.0	90.0	90.0	2.62 (2.74)
Vanadinite (Pb$_5$(VO$_4$)$_3$Cl)	7.78	11.97	16.74	87.4	92.3	91.6	7.34	10.33	17.89	90.0	90.0	90.0	6.04 (6.93)
Wulfenite (PbMoO$_4$)	5.62	12.49	5.59	90.3	90.0	89.2	5.47	12.18	5.47	90.0	90.0	90.0	6.22 (6.69)
Zincochromite (ZnCr$_2$O$_4$)	8.40	8.40	8.40	90.0	90.0	90.0	8.33	8.33	8.33	90.0	90.0	90.0	5.24 (5.36)
Zirconium tungstate (ZrW$_2$O$_8$)	9.43	9.49	9.50	90.2	89.7	89.8	9.18	9.18	9.18	90.0	90.0	90.0	4.59 (5.04)

[d] X-ray densities in parenthesis
[e] See [31]

broken as a result of geometry optimization. Unit cell parameters for those solids where the calculated PM6 structure inside the unit cell was substantially similar to the reference are presented in Table 5 and Table 6, and heats of formation are given in Table 7.

Elements

Three allotropes of carbon were modeled: diamond, consisting of pure covalent sp^3 bonds, graphite, with sp^2 bonds in-plane and π-stacking or VDW forces between planes, and "glitter", a hypothetical structure composed of 1,4 cyclohexadiene fragments [30]. The simple C–C bond in diamond is reproduced with good accuracy, 1.57 Å versus 1.54 Å observed, as is the in-plane bond in graphite, 1.43 Å versus 1.42 Å. This last result was unexpected in light of the zero band gap of graphite, in that severe difficulties in achieving a SCF are often encountered when systems with very small band-gaps are modeled; graphite appears to be a unique exception to this. Graphite layers are held together by weak VDW forces, which are poorly represented in all NDDO methods. One consequence of this is that the predicted interlayer distance, 3.56 Å, is significantly larger than the observed 3.35 Å. Intermediate between graphite and diamond is the hypothetical glitter, a tetragonal allotrope of carbon composed of single and double bonds. A PM6 calculation predicted that the structure would be essentially the same as that given by DFT methods [31].

The standard state of carbon is graphite, and therefore by definition the value of its ΔH_f is 0.0 kcal/mole-atom. PM6 predicts the ΔH_f of graphite to be 1.29 kcal/mole-atom, in good agreement with the reference value. The higher energy allotrope, diamond, has a reported heat of formation of 0.45 kcal/mole-atom [32]; for diamond, PM6 predicts the ΔH_f to be 1.34 kcal/mole-atom, i.e., 0.05 kcal/mole-atom above graphite. Glitter is a hypothetical allotrope, but is predicted by PM6 to be unlikely to be formed under equilibrium conditions: both PM6 and DFT predict the ΔH_f to be large and positive +9.1 kcal/mole-atom and +11.8 kcal/mole-atom [33], respectively.

Silicon also crystallizes in the diamond lattice, but the calculated silicon–silicon distance, 2.22 Å, is significantly less than the observed value of 2.35 Å, and while the experimental ΔH_f is, by definition, zero, the calculated heat of formation is −16.6 kcal/mole-atom.

Sulfur forms eight-membered rings, with 16 rings per unit cell. Because the unit cell is so large, and because there is a distinct insulating gap between each ring, the approximation that Γ represents the entire Brillouin zone is valid even when only a single unit cell is used. Within each ring, the sulfur–sulfur distance is 2.04 Å, in perfect agreement with the 2.04 Å observed, but the inter-ring distance is badly

Table 6 Calculated and X-ray structural parameters for silicates. Values of a, b, c are in Ångstroms; α, β, γ in degrees

	PM6						X-ray						Density[d]
	a	b	c	α	β	γ	a	b	c	α	β	γ	
Cyclosilicate													
Benitoite ($BaTiSi_3O_9$)	6.76	11.60	9.82	90.0	90.3	90.0	6.60	11.43	9.71	90.0	90.0	90.0	3.57 (3.75)
Beryl ($Be_3Al_2Si_6O_{18}$)	19.09	19.06	9.02	90.0	90.0	120.0	18.43	18.43	9.24	90.0	90.0	90.0	2.51 (2.63)
Dravite ($NaMgTi_2Al_6(BO_3)(Si_6O_{18})O_4$)	16.14	16.13	7.39	89.9	89.9	60.2	15.95	15.95	7.21	90.0	90.0	60.0	2.99 (3.14)
Inosilicate													
Cummingtonite ($Mg_7Si_8O_{23}(H_2O)$)	9.06	18.38	5.67	90.0	100.0	90.0	9.51	18.19	5.33	90.0	101.9	90.0	2.79 (2.87)
Diopside ($MgCaSi_2O_6$)	5.36	9.11	9.22	90.0	76.3	90.0	5.25	8.90	9.75	90.0	74.4	90.0	3.29 (3.28)
Enstatite ($MgSiO_3$)	5.38	8.96	17.69	90.1	89.9	90.0	5.18	8.82	18.23	90.0	90.0	90.0	3.13 (3.20)
Jadeite ($NaAlSi_2O_6$)	8.68	9.08	5.72	90.1	106.2	90.1	9.47	8.61	5.24	90.0	107.6	90.0	3.10 (3.30)
Johannsenite ($CaMnSi_2O_6$)	9.79	5.26	9.79	75.6	90.2	90.1	9.16	5.29	9.98	90.0	90.0	90.0	3.36 (3.52)
Kosmochlor (Si_2CrNaO_6)	5.69	9.39	10.07	90.0	76.2	90.2	5.27	8.72	9.58	90.0	72.6	90.0	2.89 (3.59)
Spodumene ($LiAlSi_2O_6$)	9.52	6.11	8.04	64.0	86.2	87.4	8.39	5.22	9.46	69.9	90.0	90.0	2.95 (3.18)
Tremolite ($Ca_2Mg_5Si_8O_{22}(OH)_2$)	9.30	18.83	5.58	90.1	77.4	90.0	9.84	18.05	5.27	90.0	75.3	90.0	2.83 (2.98)
Wollastonite ($CaSiO_3$)	9.16	11.12	7.63	100.0	102.1	82.7	10.12	11.07	7.31	99.5	100.5	83.4	3.10 (2.93)
Nesosilicate													
Åkermanite ($Ca_2MgSi_2O_7$)	7.91	7.73	4.77	90.0	87.2	90.1	7.84	7.84	5.01	90.0	90.0	90.0	3.11 (2.94)
Andalusite (Al_2SiO_5)	6.13	7.96	5.90	97.1	85.9	87.9	5.56	7.90	7.80	90.0	90.0	90.0	3.78 (3.14)
Chromium silicate (Cr_2SiO_4)	12.52	11.38	5.77	89.8	90.2	89.0	9.59	11.17	5.70	90.0	90.0	90.0	3.17 (4.26)
Euclase ($(BeO)_2(Al_2O_3)(SiO_2)_2(H_2O)$)	4.83	15.12	4.41	90.0	99.9	90.0	4.78	14.32	4.63	90.0	100.3	90.0	3.03 (3.09)
Forsterite (Mg_2SiO_4)	4.77	6.10	10.23	90.1	89.9	90.1	4.76	6.00	10.22	90.0	90.0	90.0	3.14 (3.20)
Garnet ($Al2O_3(CaO)_3(SiO_2)_3$)	11.75	11.75	11.75	90.0	90.0	90.0	11.86	11.85	11.85	90.0	90.0	90.0	3.69 (3.59)
Kyanite (Al_2SiO_5)	7.96	5.77	5.65	82.1	84.9	78.3	7.85	7.13	5.57	78.9	90.0	74.0	4.29 (3.67)
Pyrope ($(Al_2O_3)(MgO)_3(SiO_2)_3$)	11.36	11.37	11.36	89.8	89.7	89.8	11.55	11.55	11.55	90.0	90.0	90.0	3.65 (3.48)
Sillimanite (Al_2SiO_5)	12.49	7.36	5.73	89.8	90.0	90.0	11.55	7.68	7.49	90.0	90.0	90.0	4.09 (3.24)
Titanite ($CaTiSiO_5$)	7.44	7.00	8.43	89.9	89.6	113.9	7.07	6.57	8.72	90.0	90.0	113.9	3.24 (3.52)
Topaz ($Al_2SiO_4F_2$)	7.11	4.18	8.95	89.7	89.1	89.6	9.33	4.42	8.39	90.0	90.0	90.0	4.60 (3.54)
Uvarovite ($Ca_3Cr_2Si_3O_{12}$)	12.07	12.06	12.06	89.9	90.1	90.0	12.00	12.00	12.00	90.0	90.0	90.0	3.79 (3.85)
Willemite (Zn_2SiO_4)	8.94	8.93	8.95	108.8	108.7	108.7	8.62	8.62	8.63	107.9	107.9	107.9	3.93 (4.26)
Zircon ($ZrSiO_4$)	6.09	6.08	8.53	90.6	90.1	90.3	6.61	6.61	5.98	90.0	90.0	90.0	3.85 (4.66)
Phyllosilicate													
Apophyllite ($(Na_2O)(CaO)_8(SiO_2)_{16}(H_2O)_{15}$)	9.55	9.81	14.91	92.1	85.9	89.0	8.97	8.97	15.77	90.0	90.0	90.0	2.08 (2.28)
Chlorite ($(MgO)_{10}(Al_2O_3)_3(SiO_2)_4(H_2O)_9$)	5.45	9.39	14.16	89.3	97.2	89.2	5.24	9.07	14.29	90.0	97.0	90.0	2.57 (2.64)
Kanemite ($Na_2O(SiO_2)_4(H_2O)_7$)	5.14	20.70	7.60	90.4	90.0	91.1	4.95	20.50	7.28	90.0	90.0	90.0	1.76 (1.93)
Kaolinite ($Al_2Si_2O_5(OH)_4$)	6.59	9.65	5.34	87.5	77.8	88.6	7.40	8.94	5.16	89.9	75.1	88.3	2.59 (2.60)
Lithium disilicate ($Li_2Si_2O_5$)	5.37	14.78	6.46	90.0	90.0	90.0	4.78	14.65	5.68	90.0	90.0	90.0	1.95 (2.50)
Prehnite ($Ca_2Al_2Si_3O_{10}(OH)_2$)	4.36	5.88	18.43	90.4	88.8	89.8	4.65	5.48	18.49	90.0	90.0	90.0	2.90 (2.91)
Pyrophyllite ($AlSi_2O_5(OH)$)	9.43	5.63	20.70	99.6	90.1	90.0	8.90	5.16	18.64	100.8	90.0	90.0	2.21 (2.85)
Talc ($Mg_3Si_4O_{10}(OH)_2$)	5.57	9.61	9.79	91.1	81.2	90.2	5.29	9.17	9.46	90.4	81.3	89.9	2.43 (2.78)
Sorosilicate													
Bertrandite ($(BeO)_4(SiO_2)_2(H_2O)$)	16.26	4.92	8.75	90.1	90.0	90.0	15.27	4.57	8.71	90.0	90.0	90.0	2.26 (2.60)
Danburite ($CaB_2Si_2O_8$)	8.70	8.19	8.24	90.0	90.0	90.0	8.04	8.75	7.73	90.0	90.0	90.0	2.78 (3.00)

Table 6 (continued)

	PM6						X-ray						Density[d]
	a	b	c	α	β	γ	a	b	c	α	β	γ	
Hardystonite ($Ca_2ZnSi_2O_7$)	8.04	8.04	4.82	86.4	86.9	89.8	7.83	7.83	5.01	90.0	90.0	90.0	3.36 (3.39)
Hemimorphite ($Zn_4Si_2O_7$)(OH)$_2$·(H$_2$O))	5.40	10.61	8.32	90.1	89.9	90.0	5.12	10.73	8.37	90.0	90.0	90.0	3.36 (3.48)
Lawsonite ($CaAl_2Si_2O_7(OH)_2$·(H$_2$O))	7.56	6.33	13.34	90.1	90.0	90.0	8.80	5.85	13.14	90.0	90.0	90.0	3.27 (3.09)
Rosenhahnite ($H_2Ca_3Si_3O_{10}$)	9.37	6.70	6.82	88.4	79.2	105.0	9.48	6.81	6.96	82.9	84.1	108.6	3.01 (2.90)
Tanzanite ($Ca_2Al_3(SiO_4)(Si_2O_7)O(OH)$)	15.61	5.52	9.95	90.0	90.0	90.1	16.22	5.54	10.03	90.0	90.0	90.0	3.52 (3.35)
Thortveitite ($Sc_2Si_2O_7$)	7.44	4.46	8.98	89.9	89.7	73.4	6.65	4.69	8.62	90.0	90.0	77.8	3.00 (3.27)
Tectosilicate													
Albite ($NaAlSi_3O_8$)	8.76	12.69	7.04	89.6	65.0	90.2	8.15	12.87	7.11	93.1	63.5	89.7	2.46 (2.62)
Mordenite ($NaAlSi_{11}O_{24}$)	18.67	21.00	7.75	90.2	90.0	90.0	18.13	20.49	7.52	90.0	90.0	90.0	1.58 (1.71)
Nepheline ($KNa_3Al_4Si_4O_{16}$)	10.15	10.29	8.26	90.0	90.0	61.1	10.05	10.05	8.38	90.0	90.0	60.0	2.57 (2.64)
Paracelsian ($SrGa_2Si_2O_8$)	7.05	9.37	9.38	90.1	90.2	90.1	8.40	9.48	9.00	90.0	89.3	90.0	4.40 (3.81)
Sodalite ($Na_4Al_3Si_3O_{12}Cl$)	8.91	8.90	8.90	90.1	90.1	90.0	8.88	8.88	8.88	90.0	90.0	90.0	2.28 (2.30)
Thomsonite ($NaCa_2Al_5Si_5O_{20}6H_2O$)	13.69	13.57	12.74	80.7	75.8	84.2	13.25	13.06	13.10	90.0	90.0	90.0	2.37 (2.36)

[d] X-ray densities in parenthesis

predicted, resulting in a calculated density of 1.54 g/cc, considerably less than the observed 2.06 g/cc. This lack of inter-ring interaction is the likely cause of the calculated heat of formation being 1.45 kcal/mole-atom, rather than being nearer to the reference 0.0 kcal/mole-atom.

Halides

One of the simplest sets of inorganic solids are the alkali metal halides. Most of these crystallize with the rock-salt lattice, the exceptions being cesium chloride, bromide, and iodide, which crystallize in the body-centered cubic or cesium chloride structure. Because of their high symmetry there is only one geometric variable, so for this group of solids, symmetry was used to accelerate the geometry optimization. Geometries were optimized for both the rock-salt and cesium chloride structures for each salt. With the exception of rubidium fluoride, the calculated heats of formation (Table 8) of the isolated molecule and of the crystal are reproduced with good accuracy, although PM6 did a very poor job of predicting which of the two crystal forms was the more stable.

Elements of Group II form two main groups of halides: those with 8 coordinate metal atoms, such as fluorite, CaF_2, and frankdicksonite, BaF_2, and the six-coordinate layer structures of the chlormagnesite, $MgCl_2$, type. Other structures include the rutile lattice, as in sellaite, MgF_2, and the distorted rutile lattice, hydrophilite, $CaCl_2$. For all alkaline earth halides investigated, the calculated structure had the same lattice type as that in the starting geometry.

A more complex halide is carnallite, $KMgCl_3$·6(H$_2$O), composed of isolated potassium and chloride ions, and magnesium ions that are octahedrally coordinated by water molecules. For the X-ray structure, PM6 predicted the charge on the coordination complex ion $[Mg(H_2O)_6]^{2+}$ to be +1.89, on the potassium ion, +0.74, and on the chloride, −0.87. Within the complex ion, the charge on magnesium was +0.79, and, on average, each water molecule had a charge of +0.183. The complex ion thus behaves like an extremely ionic Group II element; the net charge being much greater than that on any Group II metal ion in any halide.

Trisodium hexachlorotitanate crystallizes in a cryolite-like structure [34]. Each titanium ion has one unpaired electron and although the Ti–Cl bond is highly covalent, the smallest distance between the $[TiCl_6]^{3-}$ complex ions is more than 3.6 Å; therefore, from an electronic perspective, the unpaired electrons could be regarded as isolated. To verify the validity of this assumption, the geometry of solid Na_3TiCl_6 was optimized using the UHF Hamiltonian. A cluster of 16 formula units was used, this representing 8 unit cells. In one calculation, the magnetic component of spin, Ms, was set to zero, and in the other it was set to the

Table 7 Comparison of calculated and experimental heats of formation of inorganic compounds (kcal/mol)

	PM6	Reference[a]	Difference
α Quartz (SiO_2)	−189.4	−217.7	28.3
α Crystobalite (SiO_2)	−193.2		
β Tridymite (SiO_2)	−193.2		
β Quartz (SiO_2)	−189.4		
Chabazite (SiO_2)	−190.7		
Coesite (SiO_2)	−187.2		
Mordenite (SiO_2)	−191.2		
Stishovite (SiO_2)	−160.7		
Aluminum nitride (AlN)	−42.7	−76.0	33.3
Aluminum phosphide (AlP)	−26.5	−39.8	13.3
Ammonium chloride (NH_4Cl)	−70.0	−75.1	5.1
Ammonium nitrate (NH_4NO_3)	−88.0	−87.4	−0.6
Ammonium sulfate (($NH_4)_2SO_4$)	−251.5	−282.2	30.7
Anatase (TiO_2)	−229.3	−224.4	−4.9
Andalusite (Al_2SiO_5)	−665.2	−619.5	−45.7
Anglesite ($PbSO_4$)	−221.9	−219.9	−2.0
Anhydrite ($CaSO_4$)	−275.3	−342.9	67.6
Aragonite ($CaCO_3$)	−269.0	−288.6	19.6
Arcanite (K_2SO_4)	−377.5	−343.6	−33.9
Arkelite (ZrO_2)	−286.4	−263.0	−23.4
Barium oxide (BaO)	−182.2	−131.0	−51.2
Barite ($BaSO_4$)	−383.8	−352.1	−31.7
Boron (B)	−8.2	0.0	−8.2
Boron Nitride (BN)	−75.5	−60.8	−14.7
Brookite (TiO_2)	−229.8		
Brucite ($Mg(OH)_2$)	−176.6	−221.0	44.4
Calcite ($CaCO_3$)	−271.7	−288.6	16.9
Calcium Iodide (CaI_2)	−108.8	−127.5	18.7
Cassiterite (SnO_2)	−59.8	−138.1	78.3
Celestine ($SrSO_4$)	−297.4	−291.6	−5.8
Chloromagnesite ($MgCl_2$)	−153.9	−153.3	−0.6
Chrysoberyl (($BeO)(Al_2O_3$))	−490.6	−549.9	59.3
Cinnabar (HgS)	−53.3	−13.9	−39.4
Coloradoite (HgTe)	−20.4	−10.0	−10.4
Corundum (Al_2O_3)	−370.9	−400.5	29.6
Cryolite (Na_3AlF_6)	−871.0	−792.8	−78.2
Enstatite ($MgSiO_3$)	−324.2	−370.2	46.0
Fluorite (CaF_2)	−207.1	−293.0	85.9
Forsterite (Mg_2SiO_4)	−452.9	−520.3	67.4
Frankdicksonite (BaF_2)	−305.3	−288.5	−16.8
Galena (PbS)	−24.8	−24.0	−0.8
Gallium arsenide (GaAs)	−35.5	−17.0	−18.5
Graphite (C)	1.3	0.0	1.3
Greenockite (CdS)	−85.1	−38.7	−46.4
Hawleyite (CdS)	−85.0	−38.7	−46.3
Hydrophilite($CaCl_2$)	−151.4	−190.1	38.7
Indium arsenide (InAs)	−17.6	−14.0	−3.6
Lead selenide (PbSe)	−124.7	−24.6	−100.1
Lead telluride (PbTe)	6.2	−16.9	23.1
Lime (CaO)	−116.2	−151.8	35.7
Magnesite ($MgCO_3$)	−224.8	−265.7	40.9
Magnesium dibromide ($MgBr_2$)	−116.9	−125.3	8.4
Magnesium diiodide (MgI_2)	−67.2	−87.0	19.8

Table 7 (continued)

	PM6	Reference[a]	Difference
Sodium sulfate (Na_2SO_4)	−372.5	−331.5	−41.0
Orthoboric acid ($B(OH)_3$)	−238.3	−261.7	23.4
Lead oxide (PbO)	−85.9	−52.3	−33.6
Lead dioxide (PbO_2) (Rutile structure)	−21.6	−66.3	44.7
Periclase (MgO)	−96.9	−143.7	46.8
Rutile (TiO_2)	−230.1	−225.6	−4.5
Sellaite (MgF_2)	−234.9	−268.7	33.8
Silicon (Si)	−13.5	0.0	−13.5
Silicon carbide (SiC)	−37.6	−15.6	−22.0
Smithsonite ($ZnCO_3$)	−175.7	−194.3	18.6
Soda Niter ($NaNO_3$)	−159.6	−111.8	−47.8
Sphalerite (ZnS)	−40.4	−49.2	8.8
Spinel ($MgAl_2O_4$)	−493.7	−549.5	55.8
Sulfur (S)	1.5	0.0	1.5
Willemite (Zn_2SiO_4)	−375.9	−391.2	15.3
Wurtzite (ZnS)	−39.4	−46.0	6.6
Zincite (ZnO)	−84.6	−83.8	−0.8
Zinkosite ($ZnSO_4$)	−211.0	−234.9	24.0
Zircon ($ZrSiO_4$)	−461.0	−529.9	68.9

[a] Reference values taken from the CRC Handbook [32]

maximum value for 16 unpaired electrons, i.e., Ms=8. Both optimized geometries were essentially the same, geometric differences were negligible, and the two heats of formation were −459.2 kcal/mol and −462.0 kcal/mol, respectively. Therefore. the assumption that the unpaired electrons are electronically isolated, at least when the focus of interest is energetics and structure, is justified.

Oxides

SiO_2 PM6 predicts the Si–O distance in silicon dioxide to be too large by 3–5%, with the result that the predicted densities of the polymorphs of silica were all underestimated. This is best seen in the very dense stishovite, where the predicted density, 3.70, is about 14% less than the observed 4.28. Stishovite has a rutile structure, so the positions of the atoms in the unit cell can be defined using two bond lengths and one angle. PM6 predicts the Si–O–Si angle correctly, 97.8° compared to the observed 98.8°, but overestimates the Si–O bond-lengths, 1.88 Å and 1.87 Å compared to the 1.76 Å and 1.81 Å observed. In α-quartz the predicted Si–O distance, 1.65 Å, is also larger than that observed, 1.61 Å, but, in addition, the Si–O–Si angle opens from the observed 143.7° to 157.2°. This additional decrease in density results in the predicted density of α-quartz being too low by 15.5%. The error in the Si–O–Si angle reaches its maximum in α-crystobalite where PM6 predicts it to be 180.0° instead of the observed 146.8°. This

Table 8 Comparison of calculated and experimental heats of formation of alkali metal halides (kcal/mol)

Salt	Molecule		Crystal		
	PM6	Reference [1]	PM6 (NaCl)	PM6 (CsCl)	Experimental [32]
Lithium fluoride	−81.5	−86.0	−138.2	−139.7	−147.4
Lithium chloride	−46.8	−53.9	−107.4	−108.2	−97.6
Lithium bromide	−36.8	−38.1	−89.6	−87.3	−83.9
Lithium iodide	−19.4	−16.1	−72.0	−73.2	−64.6
Sodium fluoride	−69.6	−64.0	−123.2	−125.7	−137.5
Sodium chloride	−43.4	−49.4	−87.7	−86.8	−98.3
Sodium bromide	−34.2	−37.6	−99.4	−101.7	−86.4
Sodium iodide	−19.0	−19.8	−75.2	−75.4	−68.8
Potassium fluoride	−78.1	−74.9	−185.7	−183.5	−135.9
Potassium chloride	−51.2	−53.4	−112.0	−118.3	−104.4
Potassium bromide	−43.0	−44.3	−99.8	−113.2	−94.1
Potassium iodide	−30.0	−30.0	−82.5	−93.9	−78.4
Rubidium fluoride	−79.2	−101.3	−152.8	−111.3	−133.3
Rubidium chloride	−54.7	−61.3	−117.9	−108.2	−104.1
Rubidium bromide	−43.7	−50.8	−107.7	−103.5	−94.3
Rubidium iodide	−32.1	−24.0	−66.4	−69.6	−79.8
Cesium fluoride	−85.2	−81.9	−155.9	−115.3	−132.6
Cesium chloride	−57.4	−62.9	−128.4	−130.2	−105.8
Cesium bromide	−50.0	−55.0	−106.0	−117.6	−97.0
Cesium iodide	−36.3	−43.5	−86.0	−98.4	−82.8

results in a large error in density, of −20%. Linear Si–O–Si systems are not unknown in nature: in the high temperature form of tridymite, hexagonal β-tridymite [35] and in thortveitite, $(Sc,Y)_2Si_2O_7$ [36], these atoms form a straight line, and, as expected, this angle is also precisely reproduced by PM6.

In addition to the dense phases of silica there are open lattice structures which, of their nature, necessarily have very low densities. Two examples are mordenite and chabazite. Calculated and X-ray structures in these systems are shown in Figs. 8 and 9. Chabazite has the lowest density of any silica polymorph, 1.46. PM6 reproduces these systems but, as with the higher density polymorphs, it underestimates the density, but in this case by about 5%.

H_2O Water forms several polymorphs. In Ice-I or Ice-Ih, the common polymorph of ice, oxygen atoms form a hexagonal lattice and each atom has two hydrogen atoms strongly covalently bound to it and two hydrogen atoms at a larger distance. The locations of the hydrogen atoms can be specified by application of the "ice rules". However, even when that is done correctly, the positions of the hydrogen atoms are not unambiguously defined. That is, although the X-ray structure shows that the oxygen atoms are located precisely on the vertices of the hexagonal lattice, there is statistical disorder in the positions of the hydrogen atoms—between each pair of adjacent oxygen atoms there are two potential positions that a hydrogen

atom can occupy, and in the observed structure each position has a 50% occupancy. This type of disorder must be resolved before a solid state calculation can be carried out, so in Ice-Ih, an arbitrary distribution of hydrogen atoms consistent with the ice rules was used.

Geometry optimization of the Ice-Ih structure gave a recognizable hexagonal lattice, as expected, but as a result of the asymmetric distribution of hydrogen atoms each oxygen atom was slightly displaced relative to the ideal lattice.

All the ices, except Ice-IV and Ice-VII, were modeled successfully. The simple unit cell of Ice-IV contains 16 water molecules but, in order to satisfy the ice rules and have the hydrogen atoms disordered, a large cluster would be needed. Assigning hydrogen atoms in such a cluster in order to achieve both randomness and compliance with the ice rules would be a non-trivial task. However, Ice-VII is the high-temperature (above 5°C) modification of Ice-VIII and differs from Ice-VIII only in that the positions of the hydrogen atoms are disordered. As such, Ice-VII is trivially different from Ice-VIII.

With the exception of Ice-X, each polymorph optimized to the correct structure. That is, the structure of each calculated polymorph was qualitatively similar to that of the observed polymorph. Ice-X is unique in that the oxygen atoms form a body-centered cubic lattice with hydrogen atoms equidistant between half of the adjacent pairs of oxygen atoms: all oxygen atoms are symmetrically tetrahe-

Fig. 8 Part of the unit cell of Mordenite, SiO$_2$. *Left* X-ray structure, *right* optimized PM6 mirror image structure
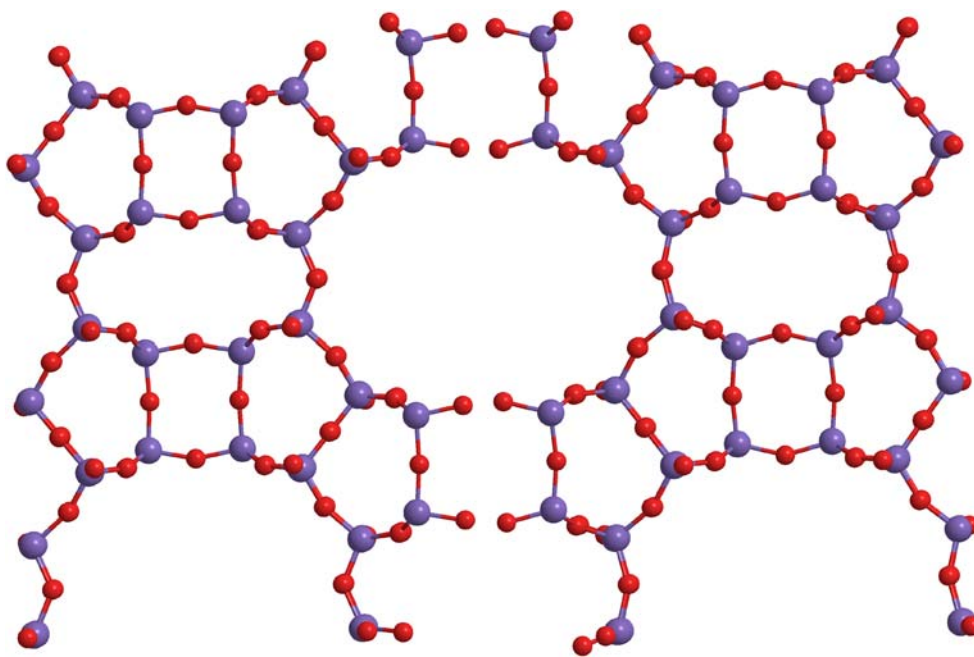

drally coordinated by hydrogen atoms. Unconstrained optimization of this structure resulted in a disordered, essentially amorphous, solid. When symmetry constraints were imposed, the geometry of Ice-X optimized to give an O-H distance of 1.214 Å. Ice-X is stable at pressures above 62 GPa and, at that pressure, PM6 predicts the O–H distance to be 1.179 Å, in good agreement with the X-ray value [37] of 1.181 Å. At a much higher pressure, about 480 GPa, PM6 predicts that the antifluorite structure would become the preferred polymorph; in this structure each hydrogen bonds to four oxygen atoms and each oxygen bonds to eight hydrogen atoms.

Al$_2$O$_3$ An attempt [12] by Gale to use the MNDO, AM1, and PM3 methods to predict the structure of corundum met with only limited success. The failure to reproduce the observed structure was attributed to deficiencies, mainly the lack of *d*-orbitals, in the basic semiempirical methods. To a large degree these deficiencies have been corrected in PM6, as can be seen by comparing the structures predicted by the various methods with the known X-ray structure (Table 9).

TiO$_2$ Within each of the polymorphs of titanium dioxide, all the titanium atoms are in the same approximately octahedral environment, surrounded by six oxygen atoms.

Fig. 9 Chabazite, SiO$_2$ *Left* X-ray structure, *right* optimized PM6 structure
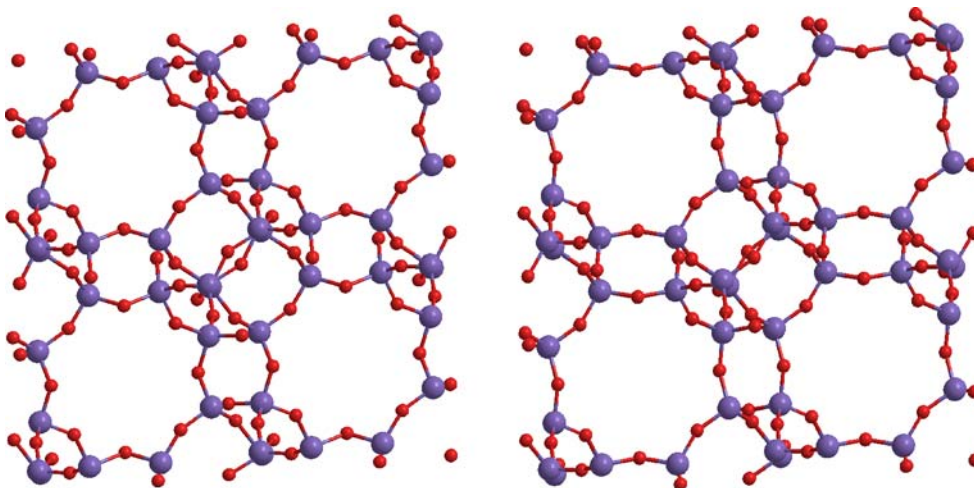
This is illustrated in Fig. 10 for rutile, the commonest polymorph. The structures of all three polymorphs were qualitatively reproduced by PM6, but the Ti–O bond-length was over-estimated by 5%, as shown in Table 10 for rutile, resulting in the predicted density being too low by about 16%.

B(OH)₃ Orthoboric acid, B(OH)₃, forms a layer structure in which individual molecules within each layer are held together by hydrogen bonds, and the layers are held together by VDW forces and very weak electrostatic forces. Although the layer structure is reproduced with good accuracy, the predicted inter-layer distance, 11.0 Å, is much greater than the observed 6.35 Å. The lack of any significant inter-layer interaction can be attributed to the absence of a VDW core–core term in boron. It is likely that if such a term were added, the correct interlayer interaction would be reproduced.

During the survey of organic solids, the structure of the very complicated organometallic pentakis(tetraethylammonium) bis(meso-tetraphenyl-porphyrinato-zinc) tetraconta-oxo-silicon-dodeca-molybdenum bromide (CSD entry PIJFUJ) was accurately reproduced. In contrast, problems had been encountered with various other complicated solids, particularly those involving protonated species. To determine the ability of PM6 to model complex ionic systems, a solid containing an archetypal proton ion, Zundel's cation [38], $[H_5O_2]^+$, was used. This ion had been identified [39] in Keggin's [40] pentahydrate, dodecatungstophosphoric acid hexahydrate. X-ray analysis indicated that the $[H_5O_2]^+$ ion has D_{2h} symmetry and that the O–O separation is 2.38 Å. However, a PM6 calculation on the isolated ion predicted its structure to be C_{2h} with an O–O separation of 2.51 Å. This was not unexpected: ab-initio calculations [41, 42] indicate that the ion should have C_2 symmetry with the C_s structure being only about 0.4 kcal/mol higher in energy, and that the D_{2h} structure should be significantly higher in energy. When D_{2h} symmetry was used, PM6 predicted that the separation would decrease to 2.45 Å, and the energy would increase by 7.8 kcal/mol. Disorder had been reported in Keggin's pentahydrate, with Zundel's ion assuming two different orientations, each having a 50% occupancy. After resolving the disorder, an unconstrained geometry optimization was performed. Examination of the optimized structure revealed that the Zundel ion had been destroyed and that the proton had migrated to the polyoxometalate. This phenomenon, the neutralization of a polyanion, had also been observed in tetrakis(2-carboxypyridinium) octacyano-molybdenum(iv), 4 $[C_6H_6NO_2]^+$ $[Mo^{IV}(CN)_8]^{4-}$, CSD entry PYCMOA, where PM6 predicted that protons on the pyridinium ions would migrate to the metal complex.

Other AB-type solids

A number of solids of the type AB are formed from elements of Group III and V, while others involve elements of Group IV, and still more involve elements of Group VI and heavy elements. Some of these occur naturally, such as wurtzite and sphalerite (zinc sulfide), and coloradoite (mercury telluride), while others are formed synthetically, often by chemical vapor deposition methods. In most of these materials, each atom of type A is tetrahedrally coordinated to four atoms of type B. This results in two types of packing, best exemplified by the two polymorphs of zinc sulfide. For convenience all compounds of this type will be grouped together.

Many of these compounds are semiconductors, and have small band-gaps between the occupied and virtual orbitals. A consequence of this is that the cluster used must be very large in order to minimize errors arising from the cluster approximation. In addition, it was anticipated that the narrow band-gap would give rise to difficulties in solving the self-consistent field equations—such difficulties had frequently occurred when molecules that had small HOMO–LUMO gaps were being studied. Surprisingly, the SCF equations were solved using default options, albeit more iterations than normal were needed.

Symmetry was used to accelerate the geometry optimization of those solids that had the sphalerite structure; such solids have only one adjustable parameter. In most cases, the calculated density for the optimized structure was close to that expected, the exceptions being cadmium telluride, which PM6 predicts to be too dense, and coloradoite, where the density was predicted to be too low.

Table 9 Calculated and observed structure and ΔH_f of α-corundum

	Experimental	MNDO[a]	AM1[a]	PM3[a]	PM6
a (Å)	4.76	4.85	4.69	5.28	4.83
c (Å)	12.99	13.12	12.33	15.47	12.91
ΔH_f (kcal/mol)	−400.5	−266.5	−241.9	−236.4	−370.9

[a] Results for MNDO, AM1, and PM3 taken from Gale [12]

Fig. 10 Unit cell of rutile (TiO_2). Crossed-eyes stereo view
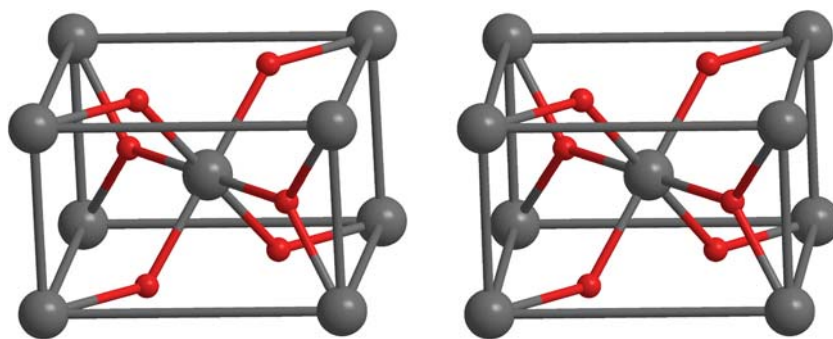
Carbonates, nitrates, and borates

Of the simple carbonate minerals that crystallize in the calcite lattice (Fig. 11), PM6 correctly reproduces the structures of magnesite, $MgCO_3$, smithsonite, $ZnCO_3$, otavite, $CdCO_3$, and calcite, $CaCO_3$ itself (Table 11). PM6 was unable to reproduce the structures of two other carbonates: rhodochrosite, $MnCO_3$, and siderite, $FeCO_3$. The UHF method was used in modeling both of these minerals because they contain transition metal ions that have unpaired electrons. In the case of rhodochrosite, each metal ion has five d-electrons in an approximately octahedral environment. Because the crystal field of the carbonate ions is too weak to cause a large separation of the t_{2g} and e_g levels, Mn^{II} would be in the high-spin configuration. This implies that the local electronic structure would be approximately $^6A_{1g}$, and that there would be no angular terms arising from the d-electrons. That is, a Mn^{II} ion would be expected to behave like a Group II ion. The failure of PM6 to reproduce the calcite structure for rhodochrosite indicates a fault in the parameterization of manganese, most likely attributable to errors in the core-core terms.

The other common polymorph of calcite is aragonite. Other minerals that have the aragonite structure are cerussite, $PbCO_3$, strontianite, $SrCO_3$, and witherite, $BaCO_3$. Of these, PM6 predicts correctly the structures of aragonite, strontianite, and witherite, albeit some carbonate groups were rotated slightly about their C_3 axis. In cerussite, the carbonate groups were tilted out of plane, giving rise to a herringbone-like structure.

The densities of all Group I carbonates are predicted by PM6 to be too low. This includes the simple carbonates natrite, Na_2CO_3, and zabuyelite, Li_2CO_3, potassium hydro-

gen carbonate, kalicinite, and sodium carbonate monohydrate, thermonatrite.

Nitrates

PM6 correctly reproduces the structures of ammonium nitrate and potassium nitrate (niter), but the predicted structure of sodium nitrate, nitratine, is completely incorrect; as with other sodium compounds, the predicted Na–Na distances are unrealistically small.

Borates

Several borates were examined, including the simple sodium salt, borax. Like other solids that contain a large amount of sodium, its optimized PM6 structure was severely in error. With that single exception, the structures of the borates were reproduced with good accuracy. Almost all minerals have some symmetry in their unit cell, the most well known exception being the naturally-occurring calcium chloroborate, parahilgardite, $Ca_2(B_5O_9)Cl(H_2O)$. PM6 was able to reproduce the unit cell of this almost unique

Table 10 Interatomic distances in rutile

Distance	PM6	X-ray
Ti-O	2.022	1.981
Ti-O'	2.080	1.948
Ti-Ti	3.116	2.959
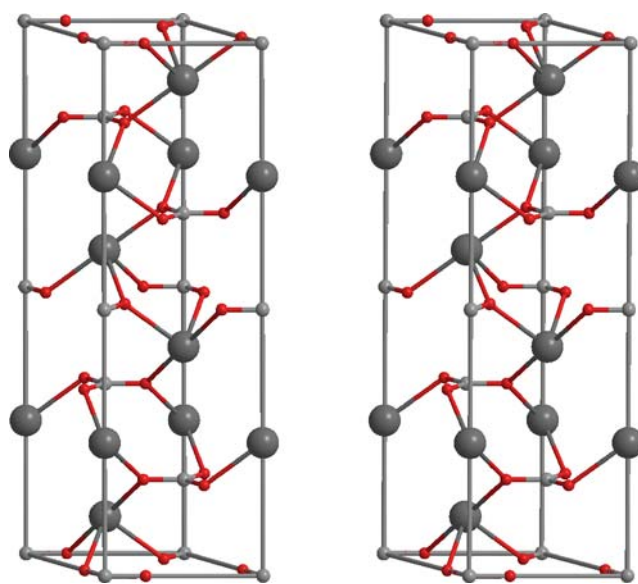
Fig. 11 Unit cell of calcite. Crossed-eyes stereo view

Table 11 Interatomic distances in calcite

Distance	PM6	X-ray
Ca-O	2.297	2.389
C-O	1.289	1.210
Ca-Ca	3.989	4.036

mineral with good accuracy, despite the fact that the unit cell was completely devoid of any symmetry (Fig. 12). This system was unique among the solids studied in that, unlike all the organic compounds, there were no readily identifiable discrete molecular fragments, and, unlike the other inorganic solids, it is devoid of any elements of symmetry.

Molybdates, tungstates chromates vanadates, sulfates, and phosphates

This group of solids is characterized by the presence of tetrahedral oxyanions, e.g., $[SO_4]^{2-}$, $[PO_4]^{3-}$, $[MoO_4]^{2-}$, $[WO_4]^{2-}$, and $[CrO_4]^{2-}$, with counterions ranging from the highly ionic Group I and Group I-like ions, e.g., $[NH_4]^+$, to softer cations such as Pb^{2+} and Zn^{2+}. With the exception of descloizite, $PbZn(VO_4)(OH)$, the internal structures of all systems modeled were quantitatively reproduced. In descloizite, the distance between lead and zinc ions was predicted to be unrealistically small, which, in turn, resulted in severe distortions to the positions of the other entities in the unit cell. There are diatomic core–core repulsion parameters in PM6 for Pb and Zn, but clearly the magnitude of the term was too small. As with other errors

of this type, this fault could readily be corrected by re-parameterizing the core–core terms only.

Charges on the ions are significantly reduced from the formal values. Thus, in fluorapatite (Fig. 13, Table 12), calcium ions have charges of +1.35 and 1.29, depending on environment, fluoride is −0.77, and each phosphate group has a net charge of −1.95, all indicating a high degree of covalence in the bonds.

Strong hydrogen bonding can occur in some sulfates and phosphates, which can be illustrated by salts in which the acid is partially neutralized, such as compounds of the type $MHSO_4$, MH_2PO_4, and M_2HPO_4, where M is a Group I element. A common structure in systems of this type consists of a hydrogen atom positioned between two oxygen atoms on two different acid groups. The positions of the protons as determined by X-ray analysis are of limited reliability, so for the purposes of comparison the oxygen–oxygen distances were used. In mercallite, potassium hydrogen sulfate (Fig. 14), PM6 predicts the oxygen–oxygen distance (Table 13) in perfect agreement with the X-ray structure. PM6 also predicts the hydrogen atoms to be asymmetrically positioned between the two oxygen atoms, this also being in accord with the published X-ray structure. Potassium dihydrogen phosphate is more complicated in that each phosphate forms four hydrogen bonds to four different phosphate groups (Fig. 15). Again, PM6 predicts (Table 14) the oxygen–oxygen distance with good accuracy, and predicts the hydrogen atoms to be asymmetrically positioned; however, in both the hydrogen sulfate and hydrogen phosphate PM6 underestimates the degree of asymmetry.

Fig. 12 Unit cell of parahilgardite. Crossed-eyes stereo view

○ H
● Ca
● B
● O
● Cl
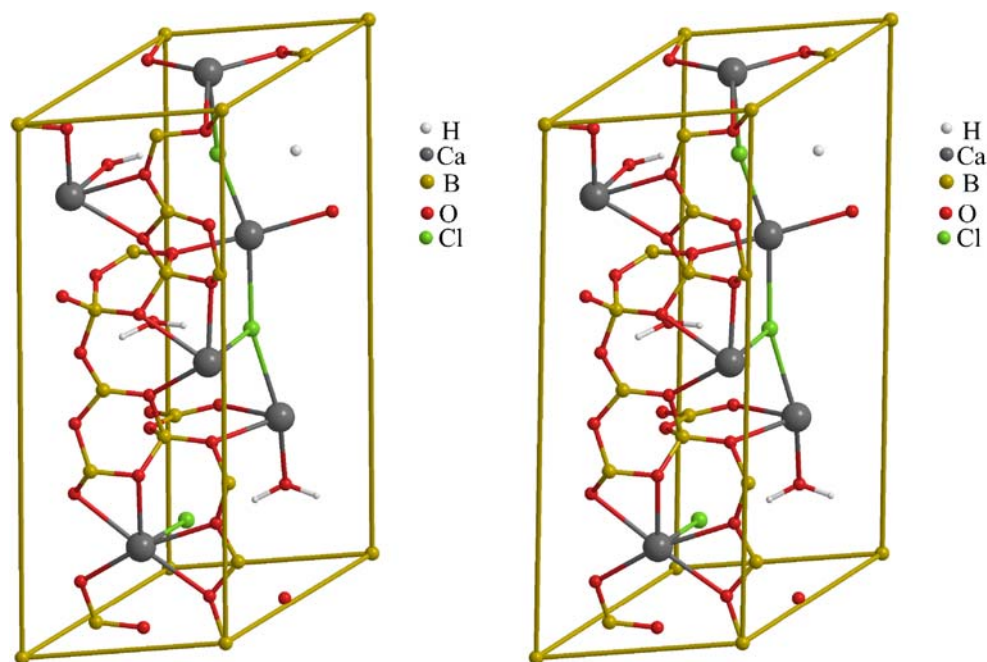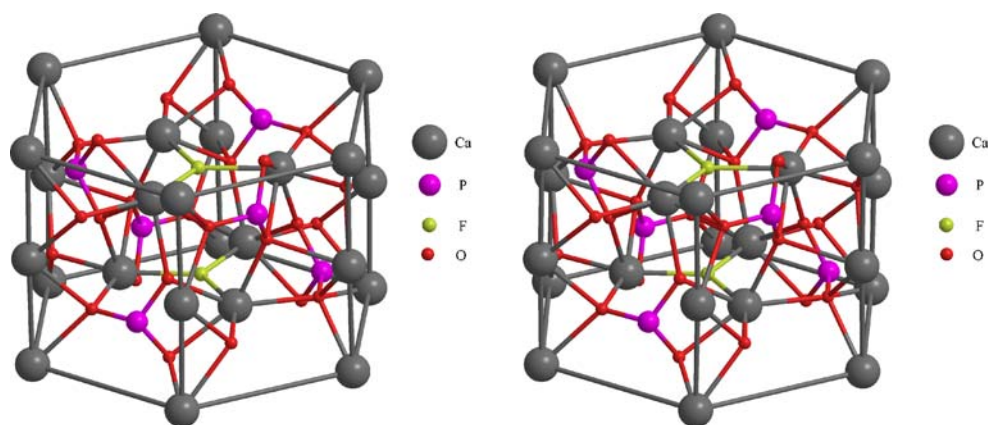
The potassium ion is relatively small in comparison to a phosphate ion, so in KH_2PO_4 the phosphate ions can form a three-dimensional lattice. If a larger cation, e.g., tetramethylammonium, is used, the three-dimensional phosphate lattice becomes energetically unfavorable relative to a one-dimensional chain of phosphates. This structure was also reproduced accurately.

PM6 reproduces the structures of most of the phosphates, but, as expected, in the case of sodium phosphate the predicted structure was qualitatively incorrect.

Silicates

Silicates form the largest and most complicated group of minerals and exhibit a very wide range of properties. From a chemical perspective, however, they are relatively simple: most silicates involve SiO_4 tetrahedra in various degrees of polymerization, and in turn these tetrahedra interact with fully oxidized metal ions. In addition, for any given empirical formula, naturally occurring minerals represent the most stable, or almost the most stable, structure. Some structures might be quite complicated; nevertheless, because they formed under equilibrium conditions, they are near or at the energy minimum, and, for any given empirical formula, an accurate computational model should not be able to generate a polymorph that is significantly more stable than any that occurs naturally. This is in

contrast to organic chemistry, where high energy forms of the same empirical formula occur frequently: for example, dimethyl ether is the high energy isomer of C_2H_6O, with ethanol being the low energy form. For the purposes of this study, the silicates can be divided into families based on the degree of aggregation of the silicate units.

Isolated SiO_4 tetrahedra: nesosilicates The simplest of the silicates are the nesosilicates, which contain isolated SiO_4 groups. With the exception of the three polymorphs of Al_2SiO_5, kyanite, andalusite, and sillimanite, the structures of the nesosilicates are reproduced with good accuracy. In the aluminum silicates there are two distinct anions, an $[SiO_4]^{4-}$ ion and an isolated oxygen ion, $[O]^{2-}$. PM6 incorrectly predicts these ions to coalesce to form a SiO_5 moiety.

Table 12 Interatomic distances in fluorapatite. Ca_1 Ca on a vertex, Ca_2 Ca on an edge, Ca_3 Ca attached to F

Distance	PM6	X-ray
Ca_1-Ca_1	5.415	5.408
Ca_1-Ca_2	3.450	3.456
Ca_1-O	2.305	2.338
Ca_3-O	2.386	2.375
Ca_3-F	2.330	2.361
P-O	1.580	1.585
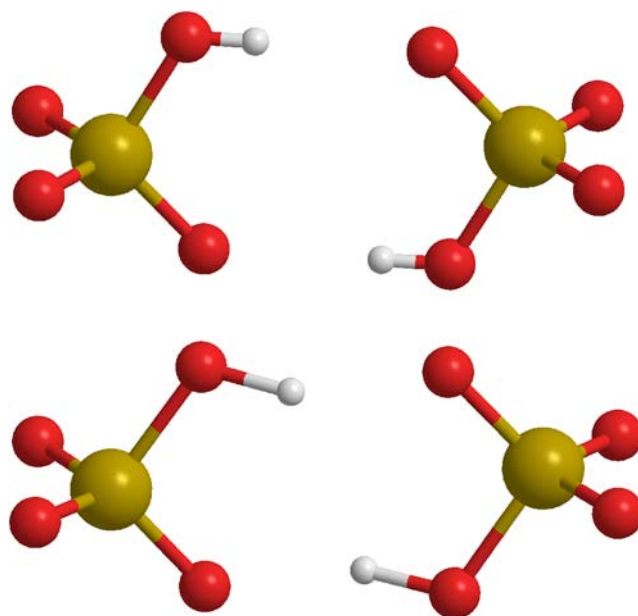
Fig. 14 Structure of hydrogen sulfate in $KHSO_4$. *Top* X-ray structure, *bottom* PM6 structure

Table 13 Interatomic distances in mercallite ($KHSO_4$)

Distance	X-ray	PM6
$O(H) \cdots O$	2.63	2.63
O–H	0.73	1.04
S–S	4.40	4.60

Double and triple tetrahedra: sorosilicates All the sorosilicates examined were reproduced with good accuracy. In most of the minerals that contained the $[Si_2O_7]^{6-}$ moiety the Si-O-Si system is bent. As expected, PM6 reproduced the bent structure. The only exception was thortveitite, an important source of scandium, which exists in nature as $(Sc,Y)_2Si_2O_7$, but which, for the purposes of this work, was simplified to $Sc_2Si_2O_7$. In the crystal structure of this mineral [36], the $[Si_2O_7]^{6-}$ moiety has C_{2h} symmetry; consequently the Si–O–Si angle is precisely 180°. When the structure of thortveitite was optimized using PM6 the symmetry did not change: the predicted Si–O–Si angle was also 180°.

Chains: inosilicates With the exception of spodumene, where PM6 predicted the aluminum-silicon distances to be too short, the structures of all the chain silicates were reproduced with good accuracy. As with the aluminum silicates, the fault in spodumene can be attributed to errors in the Al–Si core–core parameters. In the case of cummingtonite, $Mg_7Si_8O_{23} \cdot (H_2O)$, hydrogen atoms were added before the structure was optimized.

Cyclosilicates Three cyclosilicates were examined: benitoite, $BaTiSi_3O_9$, beryl, $Be_3Al_2Si_6O_{18}$, and dravite, $NaMgTi_2Al_6(BO_3)_3(Si_6O_{18})O_4$. All were reproduced with good accuracy, as illustrated by the bond lengths and angles (Table 15) in the simplest of these, beryl (Fig. 16). Dravite has a particularly complicated structure, involving seven different elements, in which the positions of the cations Na^+, Mg^{2+}, Al^{3+}, and Ti^{4+} and the anions $[O]^{2-}$, $[BO_3]^{3-}$ and $[Si_6O_{18}]^{12-}$ are determined by roughly equal contributions of covalent and ionic terms.

Sheets: phyllosilicates The structures of the phyllosilicates are reproduced with modest accuracy. In talc (Fig. 17) for example, in-plane distances are reproduced well, but the inter-plane separation (Table 16) is only poorly modeled. This fault can be attributed to the absence of strong inter-plane interactions: any errors in the weak long-range interatomic terms would give rise to disproportionately large errors in inter-plane separations.

Fig. 15 Detail of structure of dihydrogen phosphate in KH_2PO_4 (upper pair) and in $(CH_3)_4NH_2PO_4$. *Left* X-ray structures on left, *right* PM6 structure
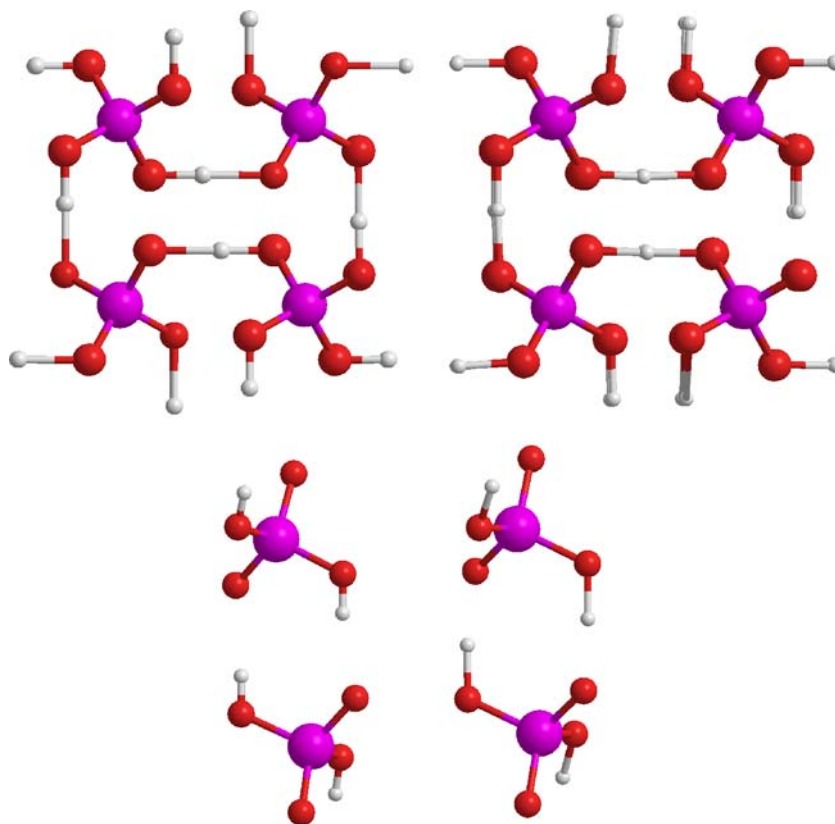
Table 14 Interatomic distances in hydrogen phosphate in KH_2PO_4

Distance	X-ray	PM6
O(H)···O	2.49	2.53
O–H	1.07	1.20
O–H	1.43	1.33
P–P	4.11	4.26

The mineral chlorite has a variable stoichiometry, (Fe, Mg, Al)$_6$(Si, Al)$_4$O$_{10}$(OH)$_8$, which does not allow a simple resolution into a fixed stoichiometry. In it, the presence of main-group metals of different oxidation states (Mg^{II} and Al^{III}) in the cationic part of the formula requires that, in order to maintain electrical neutrality, for any given ratio of magnesium to aluminum, there should be a corresponding composition in the anionic part. In nature, this condition would be satisfied by the random distribution of magnesium, aluminum, and silicon atoms in the lattice. Computationally, however, such a distribution is currently not possible, and fixed stoichiometries are required. In the extremes, these would be Mg$_6$Si$_4$O$_{10}$(OH)$_8$ and Mg$_2$Al$_8$O$_{10}$(OH)$_8$. However, the lattice structures of these two extremes would likely be so different to that of chlorite that comparison between any computed result for one of these extremes and that of chlorite would be meaningless. In order to obtain a meaningful result, an unusually large cluster was used, in which the ratios of the elements was as near to that of chlorite as possible.

Frameworks: tectosilicates With the exception of thomsonite, NaCa$_2$Al$_5$Si$_5$O$_{20}$·6H$_2$O, PM6 predicted the internal structures of the tectosilicates accurately. As with the other aluminosilicates, in thomsonite the predicted aluminum-silicon distance was unrealistically short.

Discussion

With a few exceptions, the geometries of individual organic molecules and ions and their packing arrangement in the crystal lattice were reproduced with good accuracy. In the original PM6 article it was shown that the average error in predicted bond-lengths in organic compounds was about 2–3%. The effect of intermolecular forces on the geometries of the component molecules is likely to be small, so by implication the accuracy of prediction of the geometries of those molecules would be similar to that reported in the earlier article. By far the largest structural error in organic solids involves intermolecular separations. The values of these are determined by several forces, from the weak VDW and π-stacking attractions, in cyclohexane and

coronene, through simple hydrogen bonding of the type found in sucrose, to very strong intermolecular interactions, usually ionic, best typified by the Zwitterions, such as the simple amino acids. Errors in intermolecular separations could be determined by a direct comparison of calculated and reference structures. However, because errors in intermolecular separation have a direct impact on the unit cell dimensions, and consequently on the density, a convenient and reliable estimate of the accuracy of prediction of intermolecular separations of molecules in the unit cell can be obtained from a comparison of predicted and X-ray densities.

The internal structures of three organic solids were incorrectly predicted using PM6; for all other solids the internal packing was qualitatively correct. Two of the faulty solids involved the oxalate and water moieties. From analysis of other systems known to contain neutral species, but which PM6 predicts to be ionic, it appears that PM6 unrealistically favors the formation of the hydronium ion. The fault in sodium acetate trihydrate appears to arise from faults in sodium core–core parameters, in that similar faults were found in several other inorganic sodium-containing compounds. Interestingly, the fault was not present in the closely related species sodium hydrogen acetate, a system chosen because of the presence of the hydrogen-bonded structure, [Ac$_2$H]$^-$. When the three solids that were badly predicted were removed from consideration, the average error in density decreased to 6.1%, with most of this attributable to the 2% average error in intermolecular separations. This error was not evenly distributed within the organic solids in that when interaction energies were large, as in the ionic solids, in particular Zwitterions, PM6 reproduced the intermolecular separations with good accuracy, but when the interaction energies were small, as in species where only VDW forces are involved, such as methionine, errors in intermolecular separations were often quite large.

The structures of most of the inorganic solids were predicted with useful accuracy, with the highest accuracy being exhibited by minerals in which metal atoms interact with oxygen, and the oxygen then interacts with a main-

Table 15 Interatomic distances and angles in beryl

Geometric quantity	X-ray	PM6
Al-Be	2.754	2.656
Al-O	1.928	1.906
Si-O(Al)	1.655	1.610
Si-O(Si)	1.611	1.607
Si-O-Si	168.1	165.3
Al-O-Si	128.6	137.1

Fig. 16 Unit cell of beryl.
Crossed-eyes stereo view
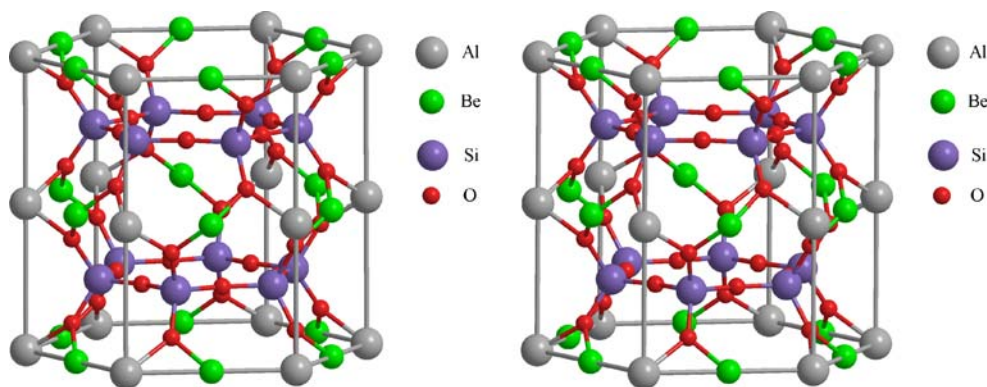
group element. Most of the silicates, phosphates, and sulfates fall into this group. One reason for this can be attributed to the procedure used in developing PM6, in that a large quantity of reference data for systems that had metal-oxygen or main-group oxygen bonds was used during the parameterization. This naturally resulted in increased emphasis being placed on the structure and thermochemistry of oxygen-containing systems.

On the other hand, inorganic systems exhibit a much wider range of types of interaction than those found in organic chemistry, and this makes the general application of PM6 to inorganic solids more difficult. An implication of the fact that PM6 uses the Voityuk diatomic core-core term is that parameters must be present for each pair of elements in a system, unless the pair of elements is separated by a distance sufficiently large that there would be no significant core-core interaction. Only a limited number of types of interactions were surveyed in this work; even within that number, several instances were found in which the PM6 values of the Voityuk parameters were severely in error and gave rise to results that were nonsense. Because of this, great care should be exercised in determining the suitability of PM6 for modeling solids that include diatomic interactions of types not found in any of the species reported here.

In general, PM6 reproduced the X-ray densities of both organic and inorganic crystals with useful accuracy, Fig. 18, the R^2 value being 0.94 and the least squares fit was $\rho_{PM6}=0.936\rho_{X\text{-}ray}+0.121$. The average unsigned error in the predicted densities of inorganic solids, 9.3%, is higher than that of organic solids, 6.9%, although the average signed error was -2.5% compared to the equivalent $+3.9\%$ for organic solids.

Systems that are badly predicted

Of the systems whose properties were badly predicted, three types could be identified. In the first group, illustrated by lead selenide and 2, 4, 6-tribromoaniline (TBA), the origin of the error could be traced to the values of individual diatomic parameters used in PM6. These were either incorrect, as in lead selenide, where the PM6 value for Voityuk's lead–selenium core–core repulsion was much too small, and in TBA where the bromine–nitrogen parameters were either physically unrealistic or absent. Errors of this type could be easily corrected by carrying out a small parameterization operation involving only the faulty parameters and using a training set consisting of examples of the two atoms in close proximity.
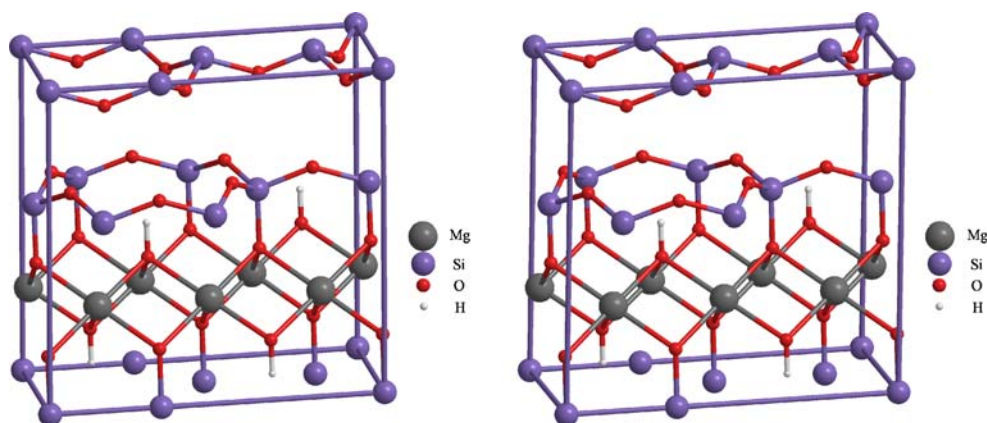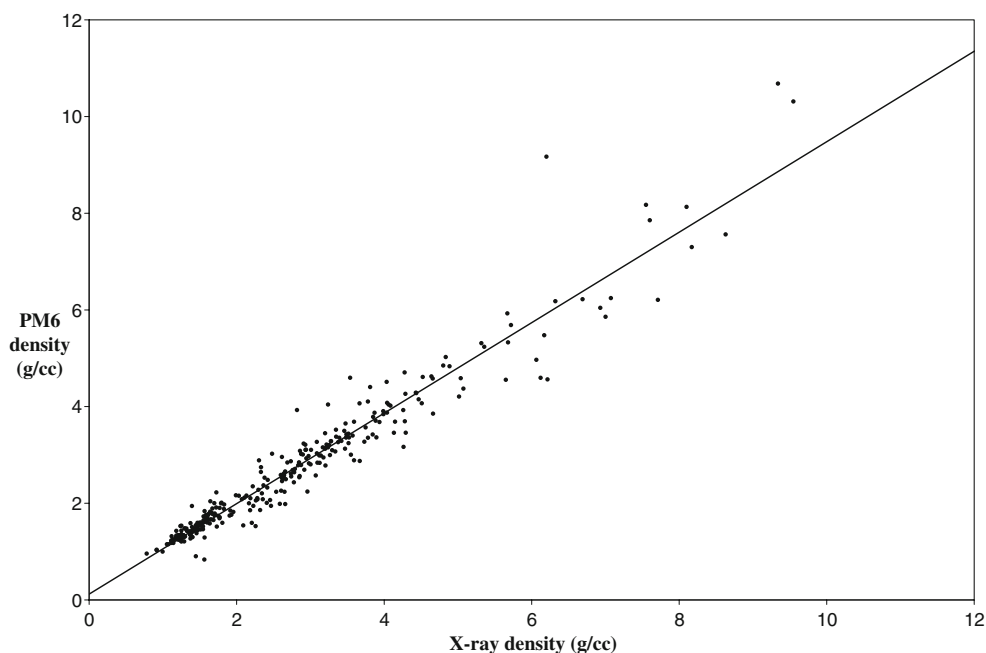
Fig. 17 Unit cell of talc.
Crossed-eyes stereo view

Table 16 Interatomic distances in talc

Distance	PM6	X-ray
Mg-O(H)	2.060	2.062
Mg-O(Si)	2.084	2.080
Si-O(Mg)	1.629	1.624
Si-O(Si)	1.674	1.622
O–O[a]	3.762	3.095

[a] Distance between layers

The second type of error was found in some solids for which no equivalent error was found in the isolated molecule. This is best illustrated by the Group I halides, where PM6 predicted the lowest energy structure to be either rock salt or cesium chloride, almost at random. During the development of PM6, only small representatives of I-VII species were used. Apparently, these systems were too small to allow the lattice properties to be accurately characterized. Errors of this type were therefore not immediately obvious, and only became apparent when full solid-state calculations were done. The implication is that future parameterizations should include solids in the training set. The best solids for use in parameterization would be those that were badly predicted by PM6. Addition of solids to the parameterization would be unlikely to cause a significant increase in error in the prediction of molecular properties. Although the use of solids in parameterization is impractical at present, it is likely that increases in computer power will make such calculations possible in the not-too-distant future.

The third type of error is specific to organic compounds, where PM6 predicts some ions to be too stable. Thus the aminophenols are predicted to exist in the solid state as the Zwitterions rather than as the neutral species. The hydrate of oxalic acid was predicted to exist as oxalate and hydronium ions, a prediction completely in variance with the X-ray structure. No instances were found where PM6 under-estimated the stability of ions relative to their neutral counterparts. That is, the error was completely systematic and unequivocally indicated a fault in the parameterization. The origin of this error most likely lies in the set of atomic electronic parameters, rather than in the diatomic core–core parameters, and as such could only be corrected by a re-parameterization. There was no evidence that any faults were due to an underlying defect in the set of approximations.

Accuracy of geometry vs hardness

In general, the structures of solids that are mechanically extremely hard, seven or more on Moh's scale, are predicted with good accuracy, whereas the geometries of many softer solids, such as most organic species, the transition metal carbonates, and various layer silicates, particularly the micas, are predicted with significantly less accuracy. This inverse relationship of mechanical hardness and computational accuracy can be rationalized by consideration of the interatomic forces involved. In hard solids, all atoms are connected to adjacent atoms by strong covalent bonds, and, of their nature, those bonds have large force constants. In general, semiempirical methods predict such geometries with good accuracy. Any tendency to deviation

Fig. 18 Comparison of calculated and X-ray densities

from the expected geometry would result in a large energy penalty, therefore distortions in predicted geometries are likely to be small.

Conversely, in mechanically weak solids some atoms interact only weakly with adjacent atoms. Thus the layers in the micas and in boric acid are held together by low energy electrostatic and VDW forces. Much effort was expended during the development of PM6 in attempting to accurately model hydrogen bonds and other weak interactions because of their importance in biochemical systems. This work has apparently been of limited success, and, as a result, in solids that are mechanically weak very small changes in energy are often associated with large geometric distortions.

The nature of this inverse relationship can be used in future method development, where, currently, the use of solids in parameterization is precluded because of the heavy computational effort required. At the present time, however, solids can be used in surveys. There, the hardness-accuracy relationship can be used in the choice of solids to be used, in that potential errors in any new method are more likely to be detected in a survey of mechanically weak rather than very hard solids.

Crystal packing automatically considered

In order to accurately reproduce molecular structures determined from X-ray analysis, theoretical methods must not only be able to accurately reproduce gas-phase geometries, but must also be able to include crystal packing forces. When these forces are small, the gas-phase structure is a good approximation to that of the solid, but the converse is also true, in that, in systems where there are large inter-molecular interactions, particularly strong ionic forces, a solid state calculation is essential in order to accurately reproduce such structures. This was most evident in the case of HMPN-H(+), where the gas-phase B3LYP geometry is a significantly worse match to the observed crystal structure than that generated by the PM6 solid state calculation, despite the intrinsically higher accuracy of B3LYP. In this system, it is obvious that the packing forces arising from the electrostatic interaction of the $[PF_6]^-$ ion with the HMPN-H(+) have significant geometric consequences.

Conclusions

The newly developed PM6 method has been demonstrated to reproduce the geometries and heats of formation of many solids, both organic and inorganic, with useful accuracy, although in a few cases the geometries predicted by PM6 were severely in error. Most of these errors can be attributed to incorrect values of the Voityuk core–core parameters. Three examples of this type of error were

identified: the Pb–Sb, Br–N, and Pb–Zn interactions all had serious errors. Where diatomic parameters were either missing or improperly optimized, as a result of a fault in the training set used, a simple correction to the method can be made. This would involve only a re-optimization of the faulty diatomic parameters, and would not alter the performance of PM6 when applied to other systems. One general error was found in the treatment of organic compounds: many compounds that exist in the solid state as neutral species were incorrectly predicted by PM6 to be ionized. Examples include the aminophenols and compounds involving oxalic acid and water, such as oxalic acid dihydrate and barium oxalate oxalic acid dihydrate. In contrast to the errors in the values of the Voityuk parameters, errors of this type involve electronic properties and can only be corrected by a complete re-parameterization. The error was not detected during the development of PM6 because solids were not used. Many solids, particularly those that are mechanically weak, can be used as sensitive detectors of potential faults in new methods, and as such they should be used for quality control in future method development work.

References

1. Stewart JJP (2007) J Mol Modeling 13:1173–1213
2. Dewar MJS, Thiel W (1977) J Am Chem Soc 99:4907–4917
3. Dewar MJS, Thiel W (1977) J Am Chem Soc 99:4899–4907
4. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) J Am Chem Soc 107:3902–3909
5. Stewart JJP (1989) J Comp Chem 10:209–220
6. Stewart JJP (1989) J Comp Chem 10:221–264
7. Born M, von Kármán T (1912) Z Physik 13:297–309
8. Ewald P (1921) Ann Phys 64:253–287
9. Stewart JJP (1985) Quant Chem Prog Exch 5:62–63
10. Dobrotvorskii AM, Evarestov RA (1974) Phys Status Solidi B 66:83–91
11. Perkins PG, Stewart JJP (1980) J Chem Soc, Faraday (II) 76:520–533
12. Gale JD (1997) Faraday Discuss 106:219–232
13. Dewar MJS, Sabelli NL (1961) Proc R Soc London, Ser A 264:431–444
14. Dewar MJS, Sabelli NL (1962) J Chem Phys 66:2310–2316

15. Dewar MJS, Sabelli NL (1962) J Phys Chem 66:2310–2316
16. Klopman G (1964) J Am Chem Soc 86:4550
17. Ohno K (1964) Theo Chim Acta 2:219–224
18. Pople JA, Santry DP, Segal GA (1965) J Chem Phys 43:S129–S135
19. Pople JA, Beveridge DL, Dobosh PA (1967) J Chem Phys 47:2026–2033
20. Stewart JJP; MOPAC2007, Version 7.290W (2007) Stewart COMPUTATIONAL CHemistry; Colorado Springs, CO
21. Allen FH (2007) In: Acta Cryst, vol B58. Cambridge Crystallographic Data Centre, Cambridge, UK, pp 380–388
22. Chickos JS, Acree WE Jr (2002) Journal of Physical and Chemical Reference Data 31:537–698
23. http://www.pdb.org/; 98 (2007) Research Collaboratory for Structural Bioinformatics, The San Diego Supercomputer Center, San Diego, CA
24. Wood SP, Tickle IJ, Treharne AM, Pitts JE, Mascarenhas Y, Li JY, Husain J, Cooper S, Blundell TL, Hruby VJ (1986) Science 232:633–636
25. Raab V, Gauchenova E, Merkoulov A, Harms K, Sundermeyer J, Kovacević B, Maksić ZB (2005) J Am Chem Soc 127:15738–15743
26. Allwood BL, Moysak PI, Rzepa HS, Williams DJ (1985) Chem Soc Chem Commun:1127–1129
27. Camilleri P, Marby CA, Odell B, Rzepa HS, Sheppard RN, Stewart JJP, Williams DJ (1989) JChem Soc Chem Commun:1722–1723
28. Chen S, Guzei IA, Yu L (2005) J Am Chem Soc 127:9881–9885
29. http://rruff.geo.arizona.edu/AMS/amcsd.php (2007) Mineralogical Society of America and the Mineralogical Association of Canada
30. Bucknum MJ, Hoffmann R (1994) J Am Chem Soc 116:11456–11464
31. Bucknum MJ, Pickard CJ, Stamatin I, Castro EA (2006) J Theor Comp Chem 5:175–185
32. Lide DR (ed) (2004) CRC Handbook of Chemistry and Physics; CRC, Boca Raton
33. Bucknum MJ, Castro EA (2007) J Chem Theory Comput 2:775–781
34. Hinz D, Gloger T, Meyer G (2000) Z Anorg Allg Chem 626:822–828
35. http://cst-www.nrl.navy.mil/lattice/ (2007) Center for Computational Materials Science of the United States Naval Research Laboratory
36. Bianchi R, Pilati T, Diella V, Gamaccioli CM, Mannucci G (1988) Am Mineralogist 73:601–607
37. Hemley RJ, Jephcoat AP, Mao HK, Zha CS, Finger LW, Cox DE (1987) Nature:737–740
38. Zundel G, Metzger H (1968) Z Physik Chem (Frankfurt) 58:225–245
39. Brown GM, Noe-Spirlet M-R, Busing WR, Levy HA (1977) Acta Cryst B33:1038–1046
40. Keggin JF (1933) Nature 131:908–909
41. Valeev EF, Schaefer HF (1998) J Chem Phys 108:7197–7208
42. Xie Y, Remmington RB, Schaefer HF (1994) J Chem Phys 101:4878–4884