electronic reprint

Journal of
Applied Crystallography
ISSN 0021-8898
Editor: **Gernot Kostorz**

Quantum software interfaced with crystal-structure databases: tools, results and perspectives

Yvon Le Page and John R. Rodgers

Journal of
Applied Crystallography
ISSN 0021-8898

Received 2 March 2005
Accepted 1 June 2005

Quantum software interfaced with crystal-structure databases: tools, results and perspectives

Yvon Le Page[a]* and John R. Rodgers[b]

[a]ICPET, National Research Council of Canada, Canada, and [b]Toth Information Systems Inc., Ottawa, Canada. Correspondence e-mail: yvon.le_page@nrc.ca

Version 2.0 of Toth's *Materials Toolkit* runs under Windows and prepares ASCII input files for popular *ab initio* packages such as *ABINIT*, *VASP etc*. Those packages, obtainable from their respective developers, may run in desktop or supercomputer setups with Linux or Windows operating systems. The *Toolkit* input is taken at will from a direct plug into CRYSTMET, with 93000 crystal-structure entries for metals and inorganic compounds, from CIF files of public-domain crystal-structure databases, or cut-and-paste from electronic journals followed by minimal free-format editing. The collection of fully general and highly graphical tools grouped on two command screens operates on the structure description stored in an editable ASCII screen. After the model has been searched, modified and evaluated in a few keystrokes with the above tools, its ASCII input files for a selection of *ab initio* packages are produced by selecting the meaningful flags and run options on a dialog. The tedious structure manipulation or decomposition into multiple simulations is performed in the background. Execution is followed by production of a plain-English job report. Four examples among the numerous possible applications of the *Toolkit* illustrate the fact that daunting topics, like the symmetry of chlorapatite, the voids and channels in the hydrogen-storage material $EuNi_5$, the energy per unit area of the contact plane for spinel twin in diamond, and the hardness of lonsdaleite *versus* diamond, are amenable to processing by materials scientists more versed in experiment than theory. The manual with tutorials and availability information can be found at http://www.tothcanada.com/toolkit/.

1. Introduction

This paper expands without duplication the scope of Le Page *et al.* (2002). In the three-year time span since that paper was printed, the computing speedup due to increased clock speed, improved pipelines and greater in-chip parallelization of the basic operations has been a factor of four with no change in price expected from Moore's law. This progress has therefore divided all computing times printed in that paper by a factor of about eight. What we reported three years ago as a four-hour-long job is now a coffee-break job and a week-long job is a day's work. The software side has not remained idle either, with new *ab initio* packages showing up, introducing new functionalities, *e.g.* on the phonon front, while packages existing at the time are maturing into extremely reliable software with increased speed and refined functionalities.

The above good news has done little to alleviate what we were describing in 2002 as an expectable bottleneck in the preparation and interpretation of quantum jobs. Far from it, this bottleneck in the access to crystallographic data, the production of correct and useful input data, as well as the interpretation of the simulation results has clearly become worse by the above factor of eight. This large factor leaves many easily tractable problems with few people to look after them because the data preparation and interpretation methods generally used have not followed the speedup of the computing, with the consequence that even inexpensive materials-modeling machines are often underexploited today.

Similar to our 2002 paper, the present paper attempts to close the increasingly open gap between what a small bank of affordable computers can compute and what modelers can achieve in a workday in terms of data access, data preparation and interpretation. In this perspective, we have first inserted in Toth's *Materials Toolkit* (White *et al.*, 2002) all the functionalities described by Le Page *et al.* (2002) and then added new ones, producing in this way what is now Toth's *Materials Toolkit* version 2.0, on which we report below. This *Toolkit* is accordingly capable of performing all applications detailed in our 2002 paper, plus new functionalities. We detail some of their applications in §4 below.

2. Purpose and basic design

The purpose of the *Materials Toolkit* is to put at the fingertips of materials modelers an integrated collection of crystallographic tools that assist them in the creation of models of materials in view of quantum computations, or in the interpretation of the results of such computations.

By 'tool' we mean a fully general and modular software function operating on a crystal structure description. Most tools leave that

description unchanged and produce from it derived data like a plot, distances-and-angles *etc.* A few tools actually modify that description permanently like edit, supercells, surfaces *etc.* The actual inner workings of a tool can be quite complex, but its functionality can be easily grasped without special training, *e.g. Plot*, *Symmetry* or *Channels*. The input is very flexible, but each input item has a carefully chosen or computed default value. In most cases, a useful output is produced from just defaults by depressing the button executing the dialog. Tools are independent. They interact only through modification of the crystal structure description.

The collective scope of the tools is extensive and dynamic, but there is no claim that their scope is exhaustive. We have deliberately not included substantial tools that are widely available elsewhere, like Rietveld refinement of powder profiles, or refinement of single-crystal diffraction data. We have instead focused on convenient import/export tools that allow easy interchange of structure or model information with excellent packages capable of performing such tasks. In contrast, we have included a cluster-scattering tool that computes their 'powder pattern' from the Cartesian atom coordinates of the atoms constituting the cluster because we do not feel that such a function is widely available. The Debye sum (Debye, 1915), which is of general application, is computed rather than the convolution of a discrete pattern with a peak-broadening function, which is an approximation for crystalline materials, and, furthermore, cannot be applied to glassy materials, molecules or gases. We have also focused on crucial interactive tools involving visualization like surface building, cluster building, assembling epitaxic multilayer systems, modeling of the contact plane in twins, deposition of molecules or nanotubes on such surfaces or clusters *etc.*

The easy data exchange with other systems or data sources gives some tools a much wider scope than their apparent scope. For example, the molecule builder can quickly assemble straight or branched alkanes and alcohols which are the current focus of fuel-cell research, but it is not currently geared toward creation of molecules of *e.g.* aromatic compounds. It might in the future, but this lack is already fully obviated by the fact that users can paste ASCII Cartesian coordinates for any molecule, cluster or nanotube imported from a foreign molecule builder into the molecule edit screen. In the same spirit of easy data exchange, users can import CIF files from the electronic version of journals, from crystal-structure databases like COD (http://sdpd.univ-lemans.fr/cod) or ICSD (Behrens, 1999). They can also export CIF files of models assembled by the *Toolkit* for further processing through other computing systems.

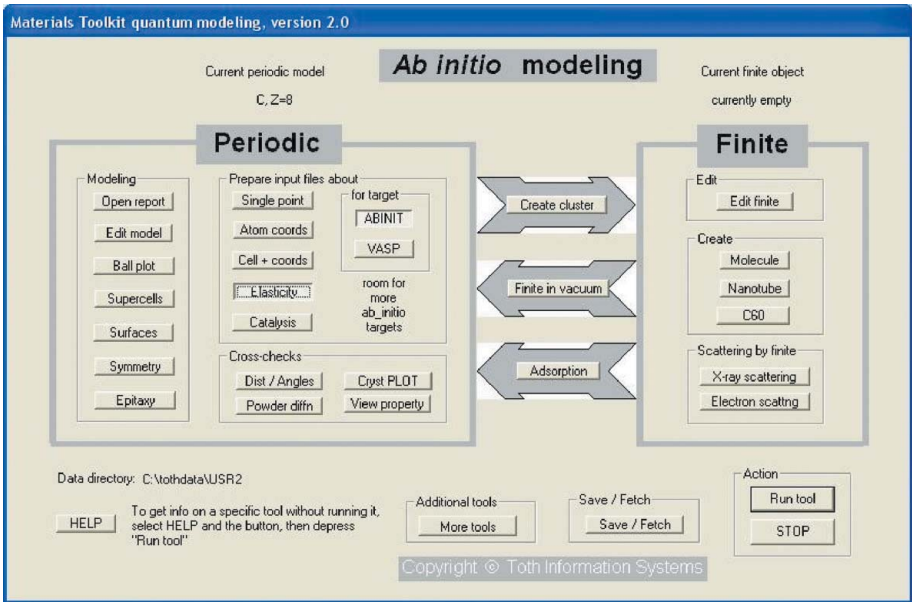

Figure 1
The major control panel.

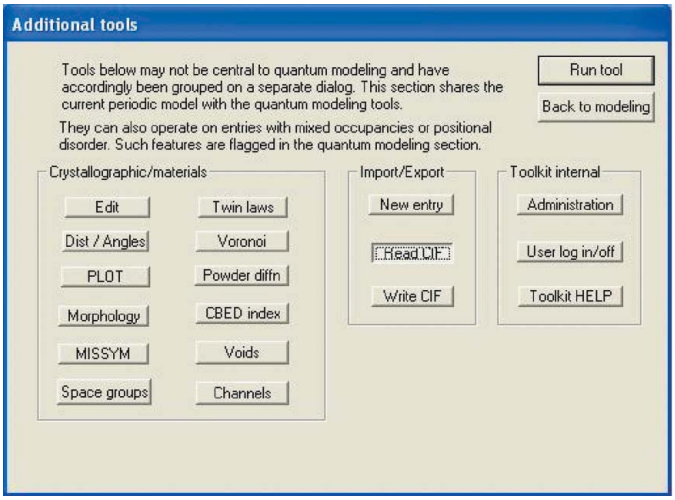

Figure 2
The minor control panel

3. Architecture and tools

There are two major categories of objects frequently manipulated for quantum modeling: the triperiodic objects familiar to crystallographers and the finite objects like clusters or nanotubes. DFT quantum software for materials purposes is based on Car and Parrinello's algorithm (Car & Parrinello, 1985), and thus only operates on triperiodic (3-p) objects. It can nevertheless be made to operate very effectively on surfaces (2-p), nanowires (1-p) or finite objects (0-p) by developing triperiodic arrays of them with vacuum gaps between the planes, the wires or the clusters. Many systems of interest to the modeler today are composite systems assembled from *e.g.* a catalyst surface and a molecule, or C_{60} and nanotubes, with the fullerene either on the nanotubes or in the nanotubes (peapods).

The above considerations naturally lead to a simple design where a system of tools operates on two editable ASCII buffers, one containing a crystal structure (space group, cell, atom types and fractional coordinates) and the other containing atom types and Cartesian coordinates for a finite object.

In order to give modelers the opportunity to draw full benefit from the competing qualities of simplicity and completeness, after considering a number of options, we opted for a pair of control panels, a major one grouping the tools essential for quantum modeling (Fig. 1) and a secondary one from which less frequently used tools or peripheral tools can be accessed (Fig. 2). Full details

including a Help and a Tutorial for each tool are available at http://www.tothcanada.com/toolkit/.

4. Selected applications

The range of applicability of the *Toolkit* is very large. We will only quote four enlightening applications in the three sections below, avoiding overlap with all capabilities already described in Le Page *et al.* (2002). Those are all implemented in *Materials Toolkit* version 2.0.

4.1. Symmetry of chlorapatite

Mackie *et al.* (1972) (MEY) show that the crystal structure of chlorapatite is a monoclinic superstructure of the previously accepted hexagonal structure in space group $P6_3/m$.

Depressing the *Symmetry* button on MEY's structure with the default tolerance of 0.05 Å proposes what amounts to a setting ***ac*-b** of the same structure. In other words, the alternate setting $P112_1/b$ is just reset to its standard orientation $P2_1/c$. If the option 'implement new description' is selected on the dialog, the ASCII buffer visible by depressing *Edit model* now contains the crystal structure description of apatite in the standard setting. MEY's selection of axes is fully justified by the need to preserve the hexagonal orientation of the reference axes in order to describe the transformation as a doubling of the *b* repeat. Users who need to transform a structure from an alternate setting into its standard setting then only need to run the *Symmetry* tool with a small distance tolerance. It is easy to check with the *Distances-and-Angles* tool that the transformation has been successful.

If a 0.5 Å tolerance is used instead of 0.05 Å, the *Symmetry* tool indicates the $P6_3/m$ space-group symmetry with **b** halved (Fig. 3). The tool has identified a pseudo-translation within this tolerance, derived the metric symmetry elements of the new cell with **b** halved as in Le Page (1982) and extracted the corresponding symmetry elements of the structure as in Le Page (1987). The oriented point group of the space group is known from the structural elements and the short lattice translations. This produces a short list of possible space groups. A one-to-one match of the structural elements extracted above with the elements of each standard space group in the short list then produces a single match, which is the correct space group. It also produces the origin shift and axial transformation that transform the original cell and its atomic content into the cell and with the origin that are conventional for the space group. This is in essence an automation of the method proposed in Le Page *et al.* (1996). Trial-and-error shows that the minimum tolerance for which a hexagonal structure is detected is 0.43 Å.

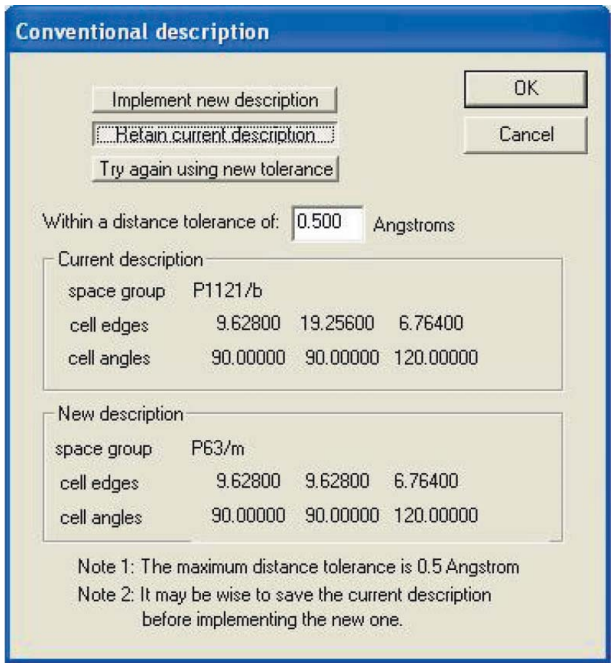

Figure 3
The dialog for the *Symmetry* tool, which extracts the conventional description within a distance tolerance from any valid structure description.

If the Cl atom is omitted from the monoclinic description, the switch from monoclinic to hexagonal occurs for a tolerance of 0.16 Å. This shows that the monoclinic distortion is chlorine driven, because it is the only atom in the structure that deviates from higher symmetry by more than the root mean square amplitude of its thermal motion. It is then no surprise that moderate heating of chlorapatite transforms it to hexagonal symmetry with chlorine occupying the 2*b* Wyckoff position with $\bar{3}$.. site symmetry.

The *Symmetry* tool is a major feature of the *Toolkit* and it is central to its smooth operation, *e.g.* for superstructure analysis in model building or for elasticity calculations. It has been exhaustively tested database-wide and is accordingly quite robust. We built a separate application which scanned automatically the whole database with 0.001 Å distance tolerance and was made to flag the entries for which the *Symmetry* analysis resulted in a different space group number or in a different volume for the conventional cell. Careful examination of the flagged entries showed that the tool was right for each entry we checked.

In rare cases, the *Symmetry* tool will produce an internal inconsistency message. This happens when the analysis leads to more than one solution. For example, a model that is hexagonal within a distance threshold d*h* can have several monoclinic descriptions within a distance tolerance d$h - \delta$. If there are two possible monoclinic descriptions, the tool accepts the best one and proposes it. If there are three monoclinic solutions within the threshold, the tool tries to assemble first a rhombohedral solution from them and then an orthorhombic solution. If this fails, the tool acknowledges that it is unable to develop a solution for the requested tolerance. Increasing the distance tolerance will produce the hexagonal description. Decreasing it will produce the best monoclinic solution, but the user would be well advised to also consider other subgroups of the hexagonal solution as possible solutions. The *MISSYM* output which lists the obliquities of twofold axes and the symmetry elements within a given distance tolerance can be of great help in this case.

4.2. Channels in $EuNi_5$

The channels are the locus of points that are at a distance greater than a specified threshold from all atoms in the structure. As the locus of points that are at a given distance from a point is the surface of a sphere centered at the point, it follows that the surface of channels is obtained from the intersections of such spheres. The *Channels* tool produces a three-dimensional plot of the channels in the model within a given plot range combined with a plot of the structure over another given range. This allows examination of the voids and channels both separate from the structure and fitting within the structure.

The $CaCu_5$ structure type includes many structures for earth or rare-earth compounds, a number of which have been investigated as potential hydrogen-storage materials, for example $EuNi_5$ (Gavra *et al.*, 1985). The covalent radius of Ni is 1.15 Å. Fig. 4 plots the channels

that are larger than 1.63 Å in $EuNi_5$. At coarse resolution, this plot takes a few seconds to compute and display. At the ultrafine resolution that was used for the figure, the plot generation takes a minute or so. Two features are quite clear: large cavities midway between Eu atoms and tiny pathways opened up between the Ni atoms in columns parallel to *z*. The pathways are opened up by the stress caused by the combination of the large Eu atom and the small Ni atom within the $CaCu_5$ structure type. At face value, hydrogen molecules or atoms might be able to penetrate such channels, but that is not granted. The high thermal-motion amplitude of the Ni atom due to the gaps created between them by Eu-atom stresses indicates that the diameter at the constriction can fluctuate, allowing H atoms or H_2 to squeeze into a temporarily enlarged channel and hop in this way from cavity to cavity.

Based on the above intuitive considerations, one can then easily insert H_2 at various strategic places in models of $EuNi_5$ and relax the model to see whether the molecule dissociates or not, both in the channels and in the cavities, and use the *Catalysis* tool to produce the series of constrained models that are required to evaluate the energy barrier to be overcome in order to hop from cavity to cavity.

4.3. Model for a material potentially harder than diamond

4.3.1. Atomic model for a (111) slab of the spinel twin of diamond. Depressing the button for the *Twin laws* tool produces the list of the binary twin laws that are possible according to Mallard's law (see Le Page, 2002) for the entry in the *Periodic* ASCII buffer. When operating on a diamond entry, the interactively adjustable defaults of maximum obliquity 3° and maximum twin index of 3 produce two possible twin laws (Fig. 5). Both are twins by reticular merohedry (zero obliquity imposed by lattice symmetry combined with a twin index greater than 1). Both twin laws are acceptable descriptions for the spinel twin that is commonly observed on natural samples of diamond (see Friedel, 1964, pp. 447–448, especially Fig. 485). The dual definition of this twin is due to the fact that $[1\bar{2}1]$ is perpendicular to [111] because the lattice is cubic. As inversion is a symmetry operation of diamond, a 180° rotation about $[1\bar{2}1]$ exchanges the two directions along [111] and amounts then to mirror reflection in (111). This situation of exact equivalence between apparently different descriptions can only happen with twins by merohedry or by reticular merohedry.

Requesting from the *Plot* tool a 5 Å thick (111) disc of diamond with 25 Å diameter viewed down $[\bar{1}01]$ with [111] in the vertical plane (all those requests can be specified interactively from the plot dialog) produced the three-dimensional plot in Fig. 6. The slab surface displays C—C bonds pointing along [111]. The direction of those bonds would therefore be preserved by mirror reflection in (111) or 180° rotation around [111], which are the possible binary twin operations. Diamond being centrosymmetric, mirror reflection about a plane and twofold rotation about its perpendicular are equivalent operations. We therefore consider only the rotation about [111].

We accordingly ask the *Surface* tool to produce an 8 Å thick (111) slab of diamond with 15 Å vacuum gap. Depressing the *Edit model* tool then shows that the corresponding P1 model has 8 atoms in a 2.5, 2.5, 23 Å cell with 90, 90, 60° angles. Depressing the *Symmetry* button tells us that this model has symmetry $P\bar{3}m1$ with the same cell but 120° γ angle. As the *Epitaxy* tool operates only on *P*1 models, we accept the default option of retaining the current description. Fig. 7, which plots the model, is produced by depressing the *Plot* tool. We store it, *e.g.* under a name like '8A diamond (111) slab', using the *Save/Fetch* tool that allows up to ten models to be kept in memory or written to disk together for future use. We then produce a similar but

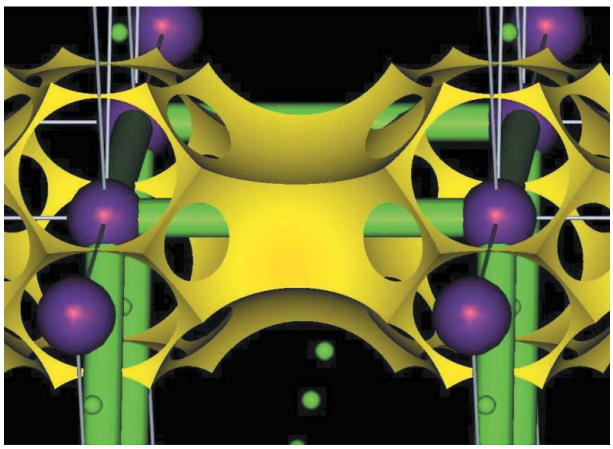

Figure 4
The channels larger than 1.63 Å in $EuNi_5$.

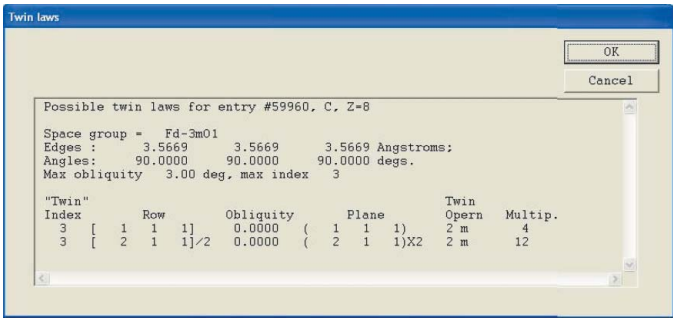

Figure 5
The possible twin laws with maximum obliquity 3° and maximum twin index 3.

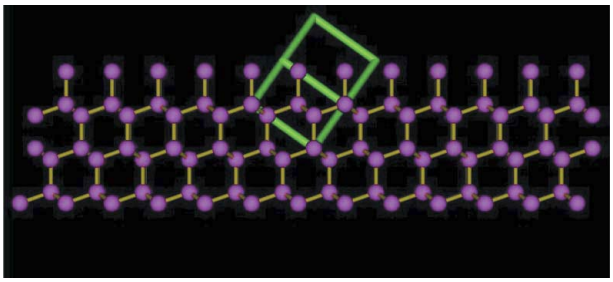

Figure 6
View down $[\bar{1}01]$ with [111] vertical of a 5 Å thick (111) disc of diamond with 25 Å diameter. Rotation of adjacent such slabs by 180° preserves the orientation of the C—C bonds and provides for a straightforward atomic arrangement at the contact plane.

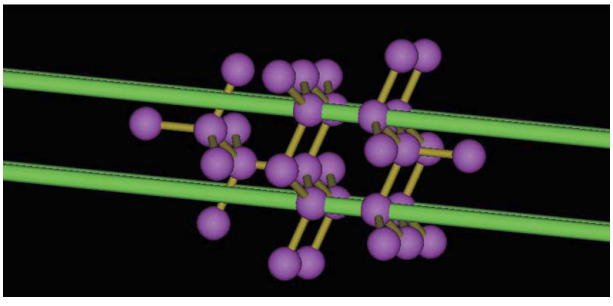

Figure 7
Triperiodical model for 8 Å thick slabs of diamond with 15 Å vacuum gap between them.

rotated diamond slab by using the *Supercells* tool to refer the same model to axes $-a$, $-b$, c and store it in the same way under the name 'Rotated diamond slab' and write the ten-model file to disk under the name 'Diamond slabs', just in case a wrong move spoils or overwrites a model.

Fig. 8 shows the *Epitaxy* dialog. The principle is that two *xy* slabs with **z** perpendicular to **x** and **y** are deposited, one in a local buffer called *substrate* and the other in *epilayer*. They can each be viewed or edited, as well as the *combined* and *multilayer* systems that we will create from them, with the ball viewing tool as frequently as needed. This operation places the last viewed model in the *Periodic* structure buffer. One can then return to the main tools, operate further on this slab, come back to *Epitaxy*, return the adjusted slab to its buffer and proceed with the epitaxic deposition, again as many times as needed.

The epitaxy is performed in two steps.

(i) Positioning of one atom of the epilayer over the substrate is performed by inputting an atom number, read off the plot, from the bottom of the epilayer and from 1 to 3 atom numbers from the top of the substrate layer with corresponding interatomic distances. These numbers are understood to designate those atoms with $0 < x, y, z < 1$, which is always possible by using the *origin change* option of the *Edit* model tool. If one atom is specified, the epilayer atom is understood to be apical at the specified distance over the substrate atom. This distance can be zero, meaning that the two atoms superpose. If two atoms are specified, the epilayer atom bridges them vertically at the specified distances. If three atoms are specified, the epilayer one is positioned above their plane at the specified distances. In case no solution exists, the software says so and expects revised input.

(ii) The lattice match is specified by inputting the two pairs of meshes that match in the epitaxy, (Sh1, Sk1; Sh2, Sk2) from the substrate and (Eh1, Ek1; Eh2, Ek2) from the epilayer. The lattice match is output in the form of the two edge lengths and their angle for each layer. The substrate mesh is then enforced on the epilayer and the combined model is produced. It can be viewed with the ball plot tool accessible from within this dialog.

In order to export the *Combined* model from the *Epitaxy* dialog to the *Periodic* buffer of the *Modeling* dialog, it is sufficient to view it and then depress *Back to Tools*. Its distances-and-angles can then be verified or it can be plotted with the *crystal-chemical plot* utility.

The *Combined* model is itself an *xy* slab. It can be deposited in the *Substrate* buffer. A new slab can then be assembled and deposited in the *Epitaxic* buffer. A three-layer epitaxic system could thus be produced in a few additional key strokes and mouse clicks.

4.3.2. Creation of a periodic multilayer system of spinel twins in diamond. This task uses the third subdialog called *Multilayer* in the *Epitaxy* dialog (Fig. 8). This dialog requests one atom number from the bottom of the *Combined* model and one, two or three atoms from the top, together with one, two or three interatomic distances, with exactly the same meaning and purpose as in the first step of epitaxy. The vector from the first atom to the point characterized by the one, two or three distances becomes the ***c*** repeat of a periodic multilayer system, *i.e.* a genuine triperiodic model for a material.

We applied this to the case of the diamond twin, which produced a *P*1 model for the system. Depressing the *Symmetry* button, we get the model with space-group symmetry $P6_3/mmc$ printed in Fig. 9 and plotted in Fig. 10. The six-atom-thick (111) diamond slabs extend from $z = 1/4$ to $z = 3/4$. At $z = 1/4$ and $z = 3/4$, the slab mirrors into adjacent slabs, giving an infinite stacking of diamond spinel-twinned every six carbon layers.

We prepared in the same graphical way models for diamond twinned every four layers and every two layers. The two-layer model is the structure of lonsdaleite, a mineral occurring where diamond has been subjected to extreme uniaxial compression like meteorite impacts. This unusual mineral is rumoured to be harder than diamond (see *e.g.* Wang & Ye, 2003), but we do not think that an experimental proof of this has been published yet.

4.3.3. Energy per unit area of the spinel twin of diamond. We create two cell-and-coordinates relaxation jobs by choosing the *Cell* +

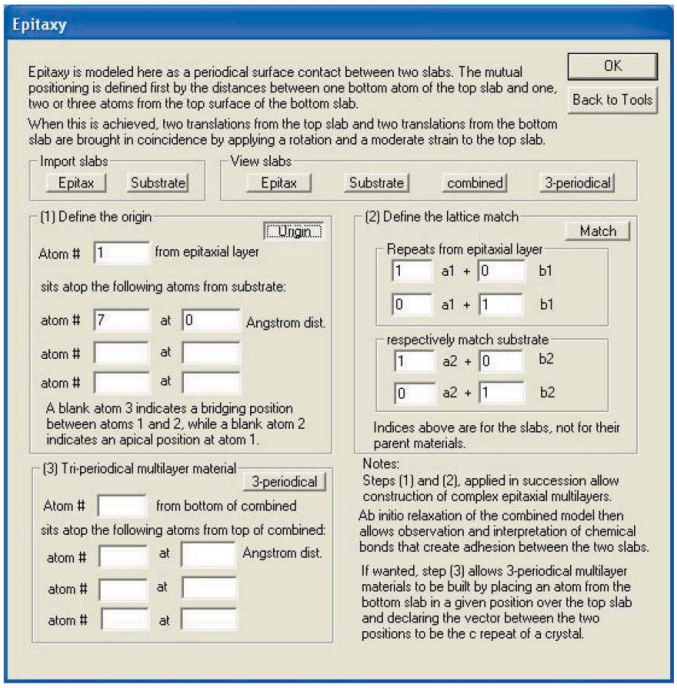

Figure 8
The *Epitaxy* dialog. Explanations on the dialog guide users through the two or three steps required to produce an epitaxic (or twin) contact, or a multilayer model.

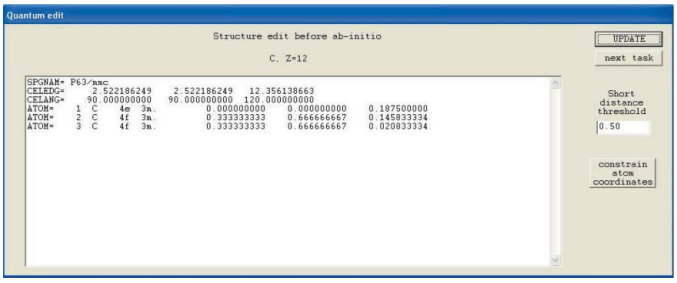

Figure 9
The *Edit model* dialog after graphically assembling the d6 multilayer system and then depressing the *Symmetry* button.

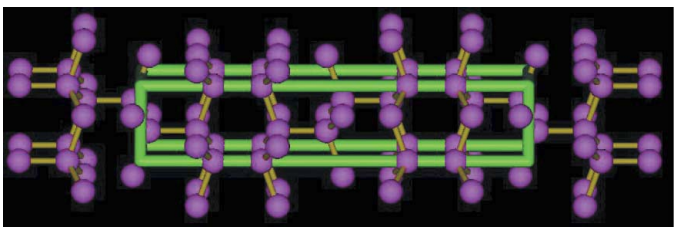

Figure 10
Default structure plot for the d6 multilayer system with the *c* axis horizontal. Diamond layers, six C-atoms thick, extending from $z = 1/4$ to $z = 3/4$ reflect in the spinel twin planes to give the periodical multilayer structure with space group $P6_3/mmc$.

Table 1
Comparison of experimental and *ab initio* results for diamond and for several multilayer systems based on the spinel twin of diamond.

The *ABINIT* diamond cell is not as close to the experimental value because *VASP* was used with GGA PAW potentials while *ABINIT* was used with LDA potentials. *ABINIT* with GGA potentials would presumably have been just as close. The two sets of results agree on the energy difference between diamond and the multilayer systems, as well as trends in the elastic tensor coefficients and their equivalent isotropic coefficients.

Diamond (= d∞)

	Experimental	*VASP*	*ABINIT*
k-mesh		9 × 9 × 9	9 × 9 × 9
a (Å)	3.56691 (Haruna *et al.*, 1992)	3.56306	3.54128
C_{11}	1076 (Kittel, 1996)	1070 (3)	1093 (17)
C_{12}	125	142 (2)	146 (12)
C_{44}	577	566 (4)	589 (24)
LS residual		1.1%	6.56%
Bulk modulus (Pa)	442	452	462
Shear modulus (Pa)	478	523	540
Young modulus (Pa)	1051	1131	1166

Diamond twin multilayer d6

	Model	*VASP*	*ABINIT*
		9 × 9 × 2	9 × 9 × 2
a (Å)	2.52218	2.51457	2.49906
c (Å)	12.35614	12.39919	12.32579
z(C1: 0,0,*z*)	0.18750	0.18686	0.18687
z(C2: 1/3,2/3,*z*)	0.14583	0.14507	0.14507
z(C3: 1/3,2/3,*z*)	0.02083	0.02076	0.02077
$U_{diamond} - U_{multilayer}$ (meV per atom)		−6.43	−5.90
C_{11}		1185 (2)	1218 (7)
C_{12}		109 (10)	112 (7)
C_{13}		50 (6)	52 (4)
C_{33}		1249 (10)	1287 (7)
C_{44}		487 (14)	496 (10)
LS residual		3.2%	2.35%
Bulk modulus (Pa)		449	461
Shear modulus (Pa)		528	541
Young modulus (Pa)		1137	1167

Diamond twin multilayer d4

	Model	*VASP*	*ABINIT*
		9 × 9 × 3	9 × 9 × 3
a (Å)	2.52218	2.51457	2.49663
c (Å)	8.23743	8.28373	8.23631
z(C1: 0,0,*z*)	0.09375	0.09296	0.09294
z(C2: 1/3,2/3,*z*)	0.15625	0.15553	0.15555
$U_{diamond} - U_{multilayer}$ (meV per atom)		−9.46	−9.17
C_{11}		1195 (9)	1225 (8)
C_{12}		111 (9)	112 (8)
C_{13}		44 (5)	44 (4)
C_{33}		1272 (9)	1308 (8)
C_{44}		483 (13)	490 (11)
LS residual		3.1%	2.5%
Bulk modulus (Pa)		451	462
Shear modulus (Pa)		531	542
Young modulus (Pa)		1143	1169

Lonsdaleite = diamond twin multilayer d2

	Experimental (Bundy & Kasper, 1967)	*VASP*	*ABINIT*
		9 × 9 × 4	9 × 9 × 4
a (Å)	2.52	2.50598	2.49007
c (Å)	4.12	4.16906	4.14705
z(C1: 1/3,2/3,*z*)	0.0625	0.06275	0.06284
$U_{diamond} - U_{lonsdaleite}$ (meV per atom)		−25.54	−25.02
C_{11}		1212 (8)	1243 (8)
C_{12}		110 (8)	112 (8)
C_{13}		21 (5)	20 (5)
C_{33}		1325 (8)	1365 (8)
C_{44}		465 (12)	471 (12)
LS residual		2.74%	2.61%
Bulk modulus (Pa)		450	462
Shear modulus (Pa)		531	544
Young modulus (Pa)		1145	1171

coordinates tool on the main dialog (Fig. 2) combined successively with *ABINIT* (Gonze *et al.*, 2002) and *VASP* (Kresse, 1993; Kresse & Hafner, 1993, 1994), both with a 9 × 9 × 2 k-mesh and a Monkhorst–Pack scheme (Monkhorst & Pack, 1976). This creates subdirectories in the *ABINIT* and *VASP* directories with corresponding input files and command file to run the quantum package and interpretation software. In the case of our installation, double-clicking on the command files executes the jobs. Other installations might require light ASCII editing of the input files for the corresponding name and location of potential files to be used by the quantum package.

We did the same thing for diamond as well as the two-carbon layer and the four-carbon layer models. Results in the IRE reference system (Brainerd, 1949) are grouped in Table 1 where we designate the models d∞, d6, d4 and d2, in view of the number of diamond layers between two twin planes. The two quantum packages agree remarkably well on fractional coordinates and on the small difference in total energy per C atom between diamond and the twinned models. Considering that there are two twin planes per cell with an *ab* mesh area of about 5.47 $Å^2$, one concludes that the energy associated with a unit area of twin plane is 7.05, 6.91 and 9.33 meV $Å^{-2}$ for the models containing respectively 12, 8 and 4 C atoms per cell and the numbers from *VASP*. The somewhat higher value for the lonsdaleite model is probably due to a repulsion between the twin planes caused by additional second neighbors for all C atoms and not just some of them. This moderate energy explains both that lonsdaleite is a rare mineral and that diamond twins are not exceptional in nature.

We attribute the difference in optimized cell parameters between *ABINIT* and *VASP* to the use of LDA potentials with *ABINIT versus* GGA PAW with *VASP*. In view of the diamond results, it is clear that *VASP*'s cell data are closer to the experimental results. This is a systematic effect between LDA and GGA that we discussed previously in Le Page & Saxe (2002). We trust that this difference would have vanished if we had used GGA potentials with *ABINIT*. However, the LDA numbers are just as consistent as the GGA numbers, giving the same fractional coordinates for atoms and the same energy per unit area of twin plane.

4.3.4. Elastic tensor of diamond and of the multilayer models. The relaxed model is imported from the job report of the optimization into the *Edit model* dialog *via* cut-and-paste. Depressing *Elasticity* combined successively with *ABINIT* and *VASP* submitted with the same parameters as the optimization jobs creates again directories with ASCII input files and command files to run them. Executing the command files creates job reports listing their independent elastic tensor coefficients calculated according to Le Page & Saxe (2001, 2002) and reported here in Table 1. The user has been completely buffered from the complexities of task creation for distortions beside their strain magnitudes and from the complexities of the extraction of stress results from quantum output files, their least-squares processing in terms of the independent tensor coefficients and the extraction of derived data.

4.3.5. Is lonsdaleite really harder than diamond? Among the values listed in the job report like the bulk modulus, Young's modulus, the average speed of sound and the Debye temperature calculated according to Anderson (1963), the report lists the shear modulus for the isotropic material. This modulus has been linked to the hardness of materials, *i.e.* to their measured resistance to various indentation tests and to their ranking in the streaking test. Both *ABINIT* and *VASP* results indicate that lonsdaleite has an isotropic shear modulus that is marginally higher than that of diamond. It is commonly accepted that the hardness of materials is related to their equivalent isotropic shear coefficient. At face value, the results of quantum simulations indicate that lonsdaleite is then marginally

harder than diamond, meaning that it should read higher than diamond in the Mohs scale and in indentation tests.

We have doubts about the above for the following reason. Only isotropic materials like glass or cryptocrystalline materials like silica have a single shear coefficient at the scale of a grain of fine abrasive. In contrast, all crystalline materials have two principal shear coefficients in each direction except along the directions of three-, four- and sixfold symmetry axes where the two coefficients are necessarily equal due to symmetry, and sometimes along other directions where they are accidentally equal. It is also well known that sometimes material A streaks material B while material B also streaks material A. This is likely to be due to the anisotropy in the large and in the small shear moduli, and it explains why gem diamonds are polished with diamond powder. Figs. 11 and 12 are radial plots for the big and the small shear coefficients of respectively lonsdaleite and diamond produced with the *Property* tool. They look very different, hinting of possibly different minimum shear coefficients. The cylindrical symmetry about Z of the shear coefficients of lonsdaleite is a well known consequence of the hexagonal symmetry of the material. It is clear from the three-dimensional plot that the minima of its small shear coefficient are then along [100] and [001], where both are 465 GPa while its maximum is along directions making a 68.4° angle with the Z axis, with a value of 539.35 GPa. For diamond, the minimum is along [101] with a value of 464.5 GPa and the maximum is along [100] with a value of 566 GPa, with a secondary local maximum along [111] where its value is 498.3 GPa. These numerical values are produced by the *Property* tool.

In the above perspective of the anisotropy of the small shear of the materials, diamond seems to be a harder material overall because the minimum of its small shear coefficient is identical to that of lonsdaleite, while the maximum of its small shear is significantly larger. Each of them should streak the other, but the analysis of the anisotropy of the small shear coefficient seems to slant slightly in favor of diamond, *i.e.* opposite to the conclusion derived from the isotropic shear coefficient. However, cryptocrystalline diamond (or lonsdaleite) would truly claim an isotropic shear around 540 GPa and be harder than both single-crystal diamond and single-crystal lonsdaleite. This might explain the materials 'harder than diamond' obtained by compression of fullerene (*e.g.* in Blank *et al.* 1994).

5. Discussion

The applications in §4 are only four among very many possible combinations of the tools from §3 towards the solution of materials problems. Beside the design of ultrahard materials or of hydrogen-storage materials, the *Toolkit* clearly also has applications in property mining within structure types using database entries, in the design of materials for electronics and photonics applications, in the understanding of the properties of nanomaterials and for exploring catalysis problems.

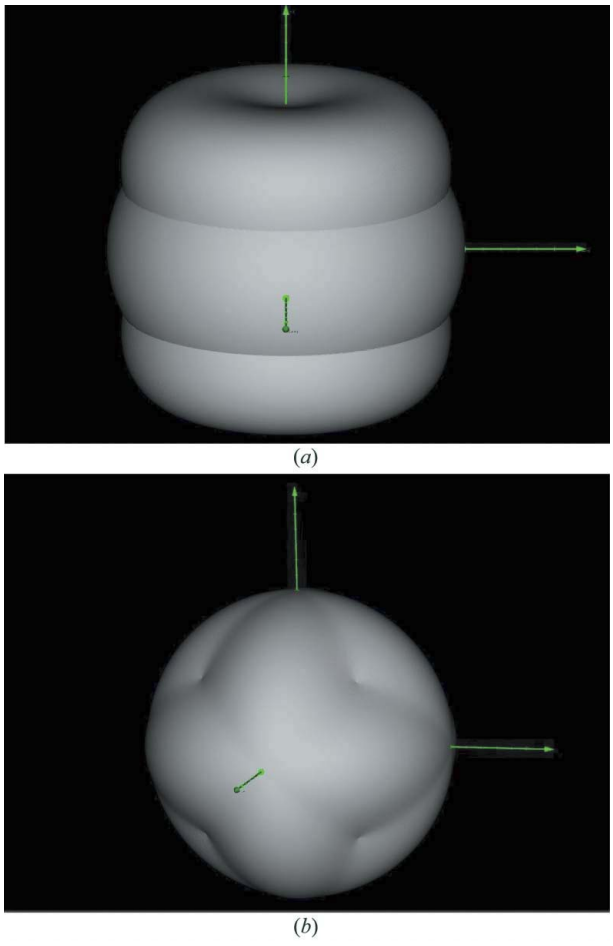

Figure 11
Radial plots for the large shear coefficient of lonsdaleite (top) and diamond (bottom).

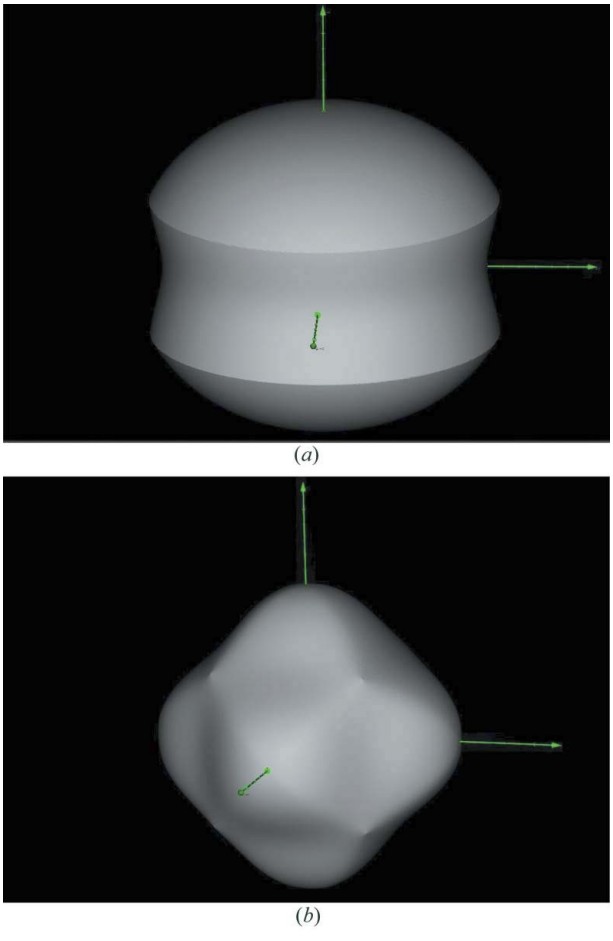

Figure 12
Radial plots for the small shear coefficient of lonsdaleite (top) and diamond (bottom).

The *Toolkit* tools are designed to be as general as possible. The direct plug into a database makes this point relatively easy to ensure for the possible symmetries, settings and origin choices. In its current form, the *Toolkit* is geared toward quantum computations on materials. Such computations become extremely slow for models with more than about 70 atoms. All atom-related arrays in the *Toolkit* are dynamic arrays, meaning that the *Toolkit* could theoretically handle structures with 10,000 or more atoms. However, a first limitation appears over around 250 independent atoms. It is not due to the *Toolkit* itself, but it seems to be a limitation in the system implementation of edit dialogs. Although the editing itself appears to work on the dialog, the string returned to the program is not modified. This limitation could be handled either by editing the atom list 250 atoms at a time, or by calling *Notepad* to perform the editing.

A second limitation that we are aware of appears around 2000 atoms with the *Edit model* tool for quantum editing. Upon each refresh of this dialog, this tool checks in the background that no interatomic distance is closer than a threshold. The default value of that threshold is 0.5 Å, but the algorithm used for checking those distances is a very general algorithm geared toward ensuring completeness in even odd cases rather than speed. This algorithm is well adapted to *e.g.* spelling out all the neighbors of an atom within a 10 or 15 Å radius. A slight delay in the processing of the *Edit model* dialog becomes perceptible when the model exceeds 200 or 300 independent atoms, and prohibitive over about 2000 atoms. As we intend to interface with quantum packages capable of handling thousands of atoms, like *SIESTA* (Soler *et al.*, 2002), we will then assemble a dedicated distance calculation that excels at speed for short interatomic distances.

We do not feel that these known limitations constitute a genuine problem for the only current purpose of the *Toolkit*, which is the preparation of models for quantum computations, because these computations become prohibitively slow for as little as 100 atoms.

As the *Toolkit* was initially derived from modules for handling crystal-structure database entries, it is very robust for genuine crystallographic data, but it is not necessarily foolproof at all stages. It usually, but not always, spots typing errors and returns gracefully to the main dialog. It handles with no problems 500 Å cell edges or 178° cell angles, but we have no clues about how it would behave if one entered negative cell edges or cell angles greater than 180°. One of the things that the *Toolkit* will obstinately refuse to do is change the hand of a system with the *Supercells* tool. It insists on right-handed sets of new axes because of the amazing confusion that this could create, both crystallographically and on the derived physical properties (Donnay & Le Page, 1978). The only way to end up with the enantiomorph is to input it into the *Toolkit*. This may sound inflexible to the expert, but the *Toolkit* is not designed for the expert but rather for the non-specialist. On the positive side of this dilemma, the *Symmetry* tool has no problems with handedness and is comfortable with enantiomorphic pairs of space groups or with origin selection in non-centrosymmetric space groups with both right-handed screws and left-handed screws.

In our views, the enhanced input flexibility of the *Toolkit* combined with a limited selection of meaningful tools is preferable to a proliferation of tiny utilities not achieving much each. We may nevertheless create at a later date a dialog dedicated to such utilities because, in spite of their individual simplicity, they retain some global usefulness and convenience combined with the guarantee that no typo will spoil the model.

The perspectives opened by the *Materials Toolkit* environment are considerable. First, the access to data and literature is straightforward in the sense that the *Toolkit* is integrated with the CRYSTMET crystal structure database with ~93000 entries. CRYSTMET is up-to-date for metals and intermetallic structures published since the discovery of X-ray diffraction. It is also up-to-date for inorganic entries published since 2000. It is quite easy to import data from other crystal-structure databases that output CIF files like ICSD (Behrens, 1999) with about 70000 entries, mostly inorganic, or Crystallography Open Database (COD) (http://sdpd.univ-lemans.fr/cod) with about 12000 entries, many of them inorganics or minerals, because the *Toolkit* has a tool to extract the structure model from such CIF files. Many electronic journals also make their structure results available in the form of CIF files. For most journals, this represents only the past five to ten years, which is nevertheless a lot of structure data. *Acta Crystallographica* now supplies its results in CIF form from the very beginning in 1948. Together, those database and Internet sources probably encompass well over 250000 materials that are either at users' fingertips or within easy reach.

Second, its modular tools automate the traditional stumbling blocks that make materials modeling such an arid and frustrating topic accessible only to theoretically inclined elites. Creation of general supercells, derivation of the symmetry of a given structure model, preparation of constrained input data for structures and for distortions of structures for processing by *ab initio* software, extraction of the independent elastic tensor coefficients from results gleaned in the output files of quantum packages are just a few of those stumbling blocks that are now automated. Creation of epitaxic models and of the constrained tasks for catalysis studies could easily be added to the list. All those functions operate seamlessly on the periodic model.

Third, by seamlessly chaining those tools in a creative manner, users can perform amazingly complex modeling tasks with the 64-bit accuracy of all *Toolkit* calculations. If a gap were to exist in these tools for a given task, the model is easy to export, process externally and re-import in CIF form or in cut-and-paste form. What used to be daunting and time-consuming tasks accessible to a select few can now be achieved in a few mouse clicks by materials scientists more versed in the experimental aspects of the problem than with the arcane details of file preparation for execution using popular *ab initio* packages.

Materials Toolkit therefore brings the modeling part of many problems from such cutting-edge research domains as clusters of metals and alloys, catalysis on surfaces, fuel cell and lithium battery research, adhesion of ultrahard coating materials, multilayer epitaxic systems, nanotubes, fullerenes and peapods *etc.* within the realm of experimentally rather than theoretically inclined materials scientists. It is our hope that this package will help to expand the volume if not the spectrum of attempts to mine crystal-structure databases for exceptional properties or for combinations of properties never measured before.

In the perspective of the above paragraph, the *Materials Toolkit* is also likely to have an impact on the teaching of computational materials science by enabling undergraduate students to perform cutting-edge computer experiments of current post-graduate level within the short time allotted for undergraduate lab assignments.

Materials Toolkit is an open-ended project on the quantum interface front, on the tool front and on the target front. We intend to interface with additional quantum engines in the near future, especially phonon codes like *PWscf* (http://www.pwscf.org/), or recent codes like *SIESTA* (Soler *et al.*, 2002) that have different limitations than the relatively mature workhorse packages that we have interfaced so far. We also intend to expand some tools and add new ones as the need arises and time permits. About targets, we have assembled a different selection of the same buttons grouped on an

appropriate dialog, and use it internally at Toth Information Systems conveniently to pre-process abstracted crystal-structure database entries for inclusion into CRYSTMET. Different groupings of tools could be useful for non-modeling materials scientists or for specific teaching purposes, like familiarization with the topology of structure types or the crystal chemistry of mineral families.

We thank Dr Peter White from the University of North-Carolina for passing CRYSTMET data to the *Toolkit*, and Dr Dennis Klug from the Steacie Institute for Molecular Science, National Research Council of Canada, for comments and suggestions following tests of a preliminary version of the *Toolkit*.

References

Anderson, O. L. (1963). *J. Phys. Chem. Solids*, **24**, 909–917.
Behrens, H. (1999). *Acta Cryst.* A**55**(Suppl.), Abstract C07.ID.001.
Blank, V., Popov, M., Buga, S., Davydov, V., Denisov, V. N., Ivlev, A. N., Mavrin, B. N., Agafonov, V., Ceolin, R., Szwarc, H. & Rassat, A. (1994). *Phys. Lett. A*, **188**, 281–286.
Brainerd, J. G. (1949). *Proc. IRE*, **37**, 1378.
Bundy, F. P. & Kasper, J. S. (1967). *J. Chem. Phys.* **46**, 3437–3446.
Car, R. & Parrinello, M. (1985). *Phys. Rev. Lett.* **55**, 2471–2474.
Debye, P. (1915). *Ann. Phys. (Leipzig)*, **46**, 809–823.
Donnay, J. D. H. & Le Page, Y. (1978). *Acta Cryst.* A**34**, 584–594.
Friedel, G. (1964). *Leçons de Cristallographie*, 2nd ed. Paris: Blanchard.
Gavra, Z., Murray, J. J., Calvert, L. D. & Taylor, J. B. (1985). *J. Less-Common Met.* **105**, 291–301.
Gonze, X., Beuken, J.-M., Caracas, R., Detraux, F., Fuchs, M., Rignanese, G.-M., Sindic, L., Verstraete, M., Zerah, G., Jollet, F., Torrent, M., Roy, A., Mikami, M., Ghosez, Ph., Raty, J.-Y. & Allan, D. C. (2002) *Comput. Mater. Sci.* **25**, 478–492.
Haruna, K., Maeta, H., Ohashi, K. & Koike, T. (1992). *Jpn J. Appl. Phys.* **31**, 2527–2529.
Kittel, C. (1996). *Introduction to Solid State Physics*. New York: Wiley.
Kresse, G. (1993). PhD thesis, Technische Universität Wien, Austria.
Kresse, G. & Hafner, J. (1993). *Phys. Rev. B*, **48**, 13115–13118.
Kresse, G. & Hafner, J. (1994). *Phys. Rev. B*, **49**, 14251–14269.
Le Page, Y. (1982). *J. Appl. Cryst.* **15**, 255–259.
Le Page, Y. (1987). *J. Appl. Cryst.* **20**, 264–269.
Le Page, Y. (2002). *J. Appl. Cryst.* **35**, 175–181.
Le Page, Y., Klug, D. D. & Tse, J. S. (1996). *J. Appl. Cryst.* **29**, 503–508.
Le Page, Y. & Saxe, P. W. (2001). *Phys. Rev. B*, **63**, 174103.
Le Page, Y. & Saxe, P. W. (2002). *Phys. Rev. B*, **65**, 104104.
Le Page, Y., Saxe, P. W. & Rodgers, J. R. (2002). *Acta Cryst.* B**58**, 349–357.
Mackie, P. E., Elliott, J. C. & Young, R. A. (1972). *Acta Cryst.* B**28**, 1840–1848.
Monkhorst, H. J. & Pack, J. D. (1976). *Phys. Rev. B*, **13**, 5188–5192.
Soler, J. M., Artacho, E., Gale, J. D., García, A., Junquera, J., Ordejón, P. & Sánchez-Portal, D. (2002). *J. Phys. Condens. Matter*, **14**, 2745–2779.
Wang, S. Q. & Ye, H. Q. (2003). *J. Phys. Condens. Matter*, **15**, L197–L202.
White, P. S., Rodgers, J. R. & Le Page, Y. (2002). *Acta Cryst.* B**58**, 343–348.