Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

CRYSTMET: a database of the structures and powder patterns of metals and intermetallics

Peter S. White,[a]* John R. Rodgers[b] and Yvon Le Page[c]

[a]Department of Chemistry, CB 3290 Venable Hall, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3290, USA, [b]Toth Information Systems Inc., 2045 Quincy Avenue, Ottawa, Ontario, Canada K1J 6B2, and [c]ICPET, National Research Council of Canada, Ottawa, Ontario, Canada K1A 0R6

Correspondence e-mail:
pwhite@pyrite.chem.unc.edu

Received 18 December 2001
Accepted 11 February 2002

CRYSTMET is a database of critically evaluated crystallographic data for metals (including alloys, intermetallics and minerals) and associated bibliographic, chemical and physical information. Also included are simulated powder diffraction patterns for all of the entries. The database currently contains almost 70000 entries and covers the literature exhaustively from 1922 to the present. The database is available on CD-ROM with search/analysis software for use on personal computers. This software can be used with any database in the appropriate format; currently CRYSTMET and the ICSD databases are available. This paper describes the database content, the procedures used in its construction, the software made available to the user and a number of potential uses for the data.

1. Introduction

Structural scientists benefit greatly from the collection of published crystal structures contained in a series of databases covering the full spectrum of chemical compounds, from proteins (PDB; Berman *et al.*, 2002), through organic and organometallic compounds (CSD; Allen, 2002; Bruno *et al.*, 2002), inorganic compounds (ICSD; Belsky *et al.*, 2002), to intermetallics and metals (CRYSTMET). This paper will address CRYSTMET, a database of alloys, intermetallics, metals and minerals, and will provide a description of the database contents as well as an overview of the data checking procedures and the software available for searching the database.

The database was started in 1960 by D. T. Cromer and A. C. Larson at the Los Alamos Laboratory. Larson brought the compilation to the National Research Council of Canada in 1974, where it was extended by L. D. Calvert. It eventually became part of the Scientific Numeric Databases system (Wood *et al.*, 1996), made available on-line to researchers by the Canada Institute for Scientific and Technical Information (CISTI). In 1996, this service was discontinued, and the production and dissemination of CRYSTMET was transferred to Toth Information Systems, Inc. Since that time, the database has been transferred from the archival flat text file to a relational database, which can run on a PC. A Windows-based search program has been developed, which permits a user to access the database in an intuitive manner and provides a number of data analysis routines. Internally, a suite of programs has been developed to assist editors in the task of inputing and checking the data being added to the database.

Table 1
Some statistics for the fall 2001 release of CRYSTMET.

Number of entries	69 054
Number of entries with coordinates	24 321
Number of distinct journal CODENs	579
Number of distinct structure types	4030
Number of single-crystal studies	26 010
Number of powder studies	42 781

2. The database

CRYSTMET is a database of metals, alloys and intermetallic compounds. As of the fall 2001 release, it contained 69054 entries, of which almost half contain atomic coordinates (Table 1). The earliest entry dates from 1922 and the latest 2001. Each of the crystallographic databases addresses different types of structures that are best described and accessed in different ways. For proteins, the sequence of amino acids is important; in the organic and organometallic database, molecules are important and structures are best classified by connectivity. The inorganic and metals databases both contain non-molecular structures that are better classified by atomic composition. There is also a far higher incidence of isostructural compounds in these databases. In CRYSTMET, the entries can all be assigned to one of about 4000 unique 'structural types', and these can be used for classification and searching.

For inclusion in the database, a compound must be broadly defined as metallic. This is taken to include metallic elements, alloys, intermetallics or minerals composed of elements to the left of the Zintl line in the Periodic Table. Compounds formed with elements to the immediate right of the Zintl line are included. Also included are compounds that define structure types that occur in intermetallics. In addition, a compound must have a clearly defined composition and the space group and unit cell must have been published. It is not required that atomic coordinates be available as there are a large number of compounds in this area where unit-cell dimensions are measured from powder diffraction patterns and a structure type has, or can be, assigned by inspection of the data and comparison with similar compounds. The atomic coordinates are not determined but the probable Wyckoff positions are known.

Where available, the following data are included in an entry.

(i) ID – a unique identifier for the entry.

(ii) The empirical formula.

(iii) A formula qualifier, *e.g.* α or low-temperature *etc.*

(iv) Structure type, named for the first compound of a type.

(v) Pearson symbol.

(vi) A qualifier for the structure type, *e.g. anti*.

(vii) Mineral name.

(viii) Refinement qualifier, *e.g.* Rietveld, multipole, least-squares is assumed.

(ix) Literature reference, including authors, journal (CODEN), volume, pages, part and year.

(x) Remarks indicating the minimum information required to use the atomic parameters correctly.

(xi) Unit cell.

(xii) Space group symbol, number and crystal system.

(xiii) Z.

(xiv) Density, calculated and measured (including temperature).

(xv) Atomic coordinates, including atom type, Wyckoff symbol, site symmetry, fractional coordinates and occupancy.

(xvi) Atomic displacement parameters.

(xvii) Error record to describe any potential errors in the original publication.

(xviii) Disorder record to indicate any disorder present; this record also describes any twinning problems if present.

(xix) For powder diffraction studies, radiation (and wavelength in the case of synchrotron and neutron studies), apparatus, method of intensity measurement, type of standard used, if pattern is just recorded, indexed or calculated, quoted R factor, R factor for intensities, profile R factor, number of observed lines, temperature, pressure.

(xx) Remarks relating to powder work.

(xxi) Elemental analysis.

(xxii) An indication that a solid solution range was studied

(xxiii) For single-crystal studies, refined or qualitative, number of reflections, counter or photographic data, radiation (and wavelength if synchrotron, neutron, gamma or electron), R factor, intensity R factor, temperature, pressure.

(xxiv) Remarks relating to single-crystal work.

In the original database, the information for each entry was kept in a flat text file format. The on-line version of the database used an indexed sequential file and included a series of indices designed to speed up the searching process. When the decision was made to move the database to the desktop environment and write customized search/analysis software with a graphical user interface (GUI), it was also decided to move the data to a relational database.

In order to accomplish this, the data were divided into several tables, related by the entry ID. The main table (tData) contains the information that exists for the majority of entries and has a single value per entry. This includes formula, structure type, cell dimensions, space group, bibliographic information and a flag indicating if coordinates are part of the entry. In addition, there are two more tables containing atomic coordinates (tAtoms) and atomic displacement (thermal) parameters (tThermal) as well as a table of text fields (tRemarks) that contains all the records that contain free text. These three tables have been separated from the main table for two reasons: first they all have the possibility of several records per database entry, and second they contain data that is not present for each entry, so we do not have to store a large number of blank fields. This set of tables contains all the original information input by the abstractors and in most cases this is stored in character format so that the original data can be reproduced exactly.

The relational database contains two other types of tables. The first contains information that aids the search program in interpreting the original entries; for example, there is a table of the journal CODENs and associated journal titles. Other tables contain atomic properties, such as radii *etc.* A major advantage of putting this information into the database rather than hard coding it into the software is that it can easily be updated without having to redistribute executables. The second additional type of table contains derived data from the entries that enable more sophisticated searches of the database. These tables often contain the data in numeric format, permitting searches on ranges of values. Examples of these tables are: (*a*) reduced cell dimensions for searching using measured unit cells that may not be in the same setting, (*b*) tables of element types, so one can easily search for compounds containing particular elements or groups of elements, and (*c*) lists of author names and initials with all letters lower case for easy bibliographic searches. In addition, during database generation, powder patterns are computed for every entry and the most intense lines are stored in a table for use with a pattern-matching query.

3. Software

With the move of the database to the desktop/relational database platform, a completely new set of software has been developed. These programs fall into three categories: data abstraction and checking, database building and maintenance, and search and analysis. All of these programs rely heavily on a central library of subroutines that perform most of the major crystallographic calculations.

One of the major problems with crystallographic databases is the need to handle space groups in all the settings and origin choices that have been used over the years. This crystallographic library is centred on space-group handling routines that accomplish this task and are capable of computing derived information, such as site symmetries and Wyckoff symbols, which is consistent with the *International Tables of Crystallography*, Vol. A (1983).

3.1. Data abstraction and checking

The process of abstracting data and checking that it is correct is very time consuming and at times tedious. Most of the data in CRYSTMET are abstracted from the original journal articles by editors and transcribed into text files using the original CRYSTMET flat format. The preliminary checking routines ensure that these files are in a valid format and all the required fields are present. The entries are then subject to rigorous checking by a second editor. There is quite a large redundancy of information in a published crystal structure and use is made of this fact in data checking, using a procedure that is similar to that performed by journals such as *Acta Crystallographica*. Examples of a few of these checks are are as follows:

(i) The space group symbol, space group number and crystal system must be consistent.

(ii) The unit-cell parameters must be consistent with the space group (*e.g.* angles for orthorhombic groups must be 90°) and the computed volume should equal the reported volume.

(iii) The formula and Z can be used to compute the density and compare it with published values, as well as making sure that it is a reasonable value. If atomic coordinates are published, then these can also be used to confirm the formula and density.

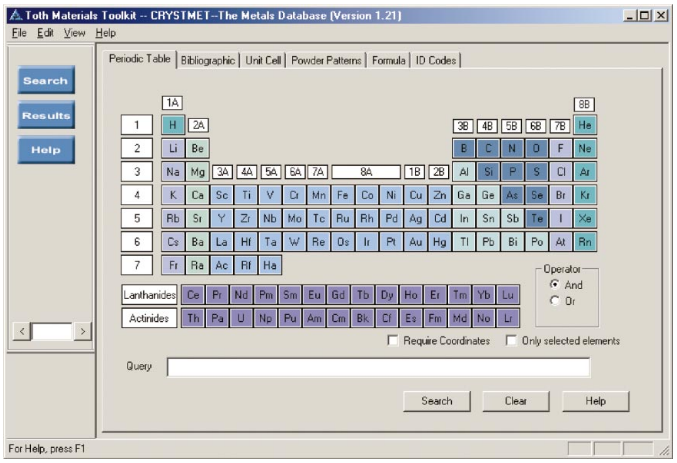
(iv) If the author gives site symmetries and/or Wyckoff symbols in the table of coordinates, then these are compared with computed values.

(v) A list of bond distances and angles is generated and compared with expected as well as published values.

(vi) If a structure type is given, then the atoms should be in the same Wyckoff positions as the definitive structure for that type.

(vii) If a powder diffraction pattern is published, a simulated pattern can be generated from the data and compared with the original.

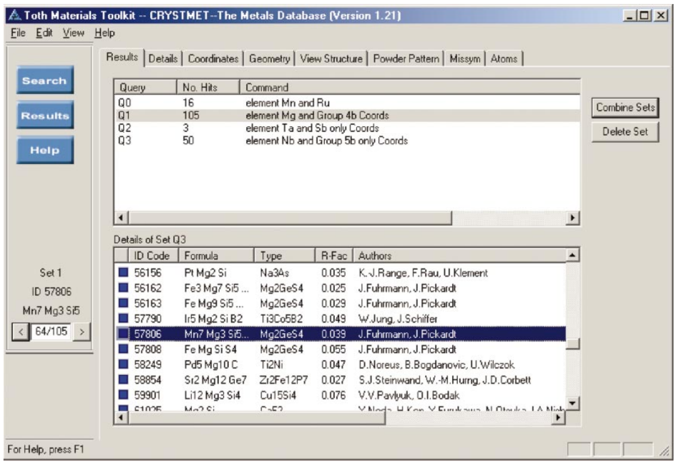
Any abnormalities in these check results cause the original source to be re-examined and if necessary the author is contacted. If no structure type had been assigned earlier, then a search is made for possible types. If no match can be made then this structure is flagged as a new structure type. Much of


Toth Materials Toolkit -- CRYSTMET--The Metals Database (Version 1.21)
File Edit View Help
Periodic Table | Bibliographic | Unit Cell | Powder Patterns | Formula | ID Codes
Search
Results
Help
1A 2A 3A 4A 5A 6A 7A 8A 1B 2B 3B 4B 5B 6B 7B 8B
1 2 3 4 5 6 7
H He
Li Be B C N O F Ne
Na Mg Al Si P S Cl Ar
K Ca Sc Ti V Cr Mn Fe Co Ni Cu Zn Ga Ge As Se Br Kr
Rb Sr Y Zr Nb Mo Tc Ru Rh Pd Ag Cd In Sn Sb Te I Xe
Cs Ba La Hf Ta W Re Os Ir Pt Au Hg Tl Pb Bi Po At Rn
Fr Ra Ac Rf Ha
Lanthanides Ce Pr Nd Pm Sm Eu Gd Tb Dy Ho Er Tm Yb Lu
Actinides Th Pa U Np Pu Am Cm Bk Cf Es Fm Md No Lr
Query
Search
Clear
Help
For Help, press F1

Operator: ● And ○ Or

□ Require Coordinates □ Only selected elements

Figure 1
A view of the main search screen in *Materials Toolkit* showing the main search page. Additional searches can be specified using the tabs across the top.


Toth Materials Toolkit -- CRYSTMET--The Metals Database (Version 1.21)
File Edit View Help
Results | Details | Coordinates | Geometry | View Structure | Powder Pattern | Missym | Atoms
Search
Results
Help
Combine Sets
Delete Set
Set 1
ID 57806
Mn7 Mg3 Si5
64/105
For Help, press F1

Query	No. Hits	Command
Q0	16	element Mn and Ru
Q1	105	element Mg and Group 4b Coords
Q2	3	element Ta and Sb only Coords
Q3	50	element Nb and Group 5b only Coords

Details of Set Q3

ID Code	Formula	Type	R-Fac	Authors
56156	Pt Mg2 Si	Na3As	0.035	K.-J.Range, F.Rau, U.Klement
56162	Fe3 Mg7 Si5 ...	Mg2GeS4	0.025	J.Fuhrmann, J.Pickardt
56163	Fe Mg9 Si5 ...	Mg2GeS4	0.029	J.Fuhrmann, J.Pickardt
57790	Ir5 Mg2 Si B2	Ti3Co5B2	0.049	W.Jung, J.Schiffer
57806	Mn7 Mg3 Si5...	Mg2GeS4	0.039	J.Fuhrmann, J.Pickardt
57808	Fe Mg Si S4	Mg2GeS4	0.055	J.Fuhrmann, J.Pickardt
58249	Pd5 Mg10 C	Ti2Ni	0.047	D.Noreus, B.Bogdanovic, U.Wilczok
58854	Sr2 Mg12 Ge7	Zr2Fe12P7	0.027	S.J.Steinwand, W.-M.Hurng, J.D.Corbett
59901	Li12 Mg3 Si4	Cu15Si4	0.076	V.V.Pavlyuk, O.I.Bodak

Figure 2
A view of the ‘Results’ page. The top pane shows summaries of all the searches that have been performed and the bottom one shows the current set of results in more detail.

this checking has been automated, using software based on the crystallographic library as well as a number of routines from the *NRCVAX* suite of programs (Gabe *et al.*, 1989).

3.2. Building the database

New entries are collected into batches for insertion into the database. The program for inserting new data reads the entries one at a time and performs a number of consistency checks on the data. As part of this process, it adds computable information, which may be missing from the entry, such as calculated density, Pearson symbol, site symmetries and Wyckoff symbols. It then computes a number of quantities for the derived tables, such as the reduced cell, a list of elements with the number present in the formula as well as their row and column in the Periodic Table, and a list of individual authors split into family name and initials. It finally computes a simulated powder diffraction pattern and generates a table of the strongest peaks. In the case of entries with no atomic coordinate information, a list of all possible peak positions is computed. When an entry is ready, a check is made that the ID code for this entry does not already exist, and it is inserted into the database, modifying all the tables necessary.

Once the new database has been built, it is possible to perform a number of checks for potential problems that may have escaped the checking routines. These problems tend to be associated with the consistency of the database as a whole, for example, minor variations in an author's name, for which attempts are made to ensure that only one spelling is used across the database.

3.3. Search and analysis

Of most interest to the end users of the database is the software environment for searching the database and displaying the results. For CRYSTMET this is accomplished *via* a program called *Materials Toolkit*. The philosophy is that we have a single extendable framework that provides capabilities for searching one or more databases and analysing the results of these searches. The data has to be in the correct format; currently CRYSTMET and the ICSD inorganic structures database are available. The *Materials Toolkit* consists of a user interface, which presents the user with a series of tabbed pages. Each page can provide some function, *e.g.* input a bibliographic search or display the details of an entry. In the background are several general libraries of functions that are available to all pages. Along with the library of crystallographic functions mentioned, the general libraries provide a connection to the database, which will perform a query and return a list of 'hits' along with some of the basic information about them (formula, authors, *R* factor, coordinate availability *etc*), and a routine that will return all the details about a single entry. The framework maintains a list of the queries that the user has performed, and employs the concept of a 'current query' and 'current entry' available to all the analysis pages. This architecture makes it relatively easy to add functionality to the program, as if one simply adds a tab page and appropriate code, the framework will then pass it pointers to the current query and entry.

The current version of *Materials Toolkit* supports two sets of tabbed pages, selected by the buttons on the left (Fig. 1): one for formulating searches and the other for viewing their results.

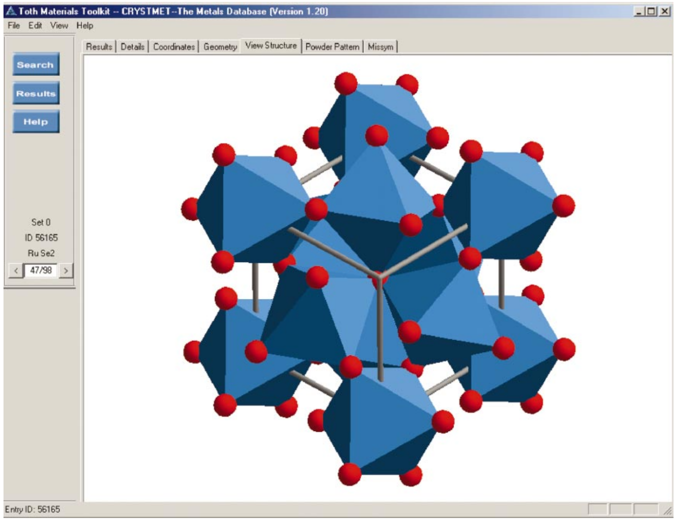


Figure 3
A view of *Materials Toolkit* showing the structure of $RuSe_2$, in inorganic mode.

4. *Materials Toolkit*

The databases of inorganic and metal structures, unlike the protein, organic and organometallic databases, contain structures composed of a wide variety of elements. Therefore, one of the easiest methods of performing an initial search is simply to select the elements one wants to be present. Consequently, the first search page that the user is presented with is a Periodic Table and most queries can be input by clicking on the required elements. It is also possible to select rows or columns of the table to search for groups of atoms.

The user can choose other searches as necessary: 'Bibliographic', searches by author, journal, year *etc.*; 'Unit Cell', permitting input of a unit cell, which is then reduced and compared to the table of reduced cells for all entries; 'Powder Patterns', for which the user inputs the most intense lines from a powder pattern, and optionally any known elements in the sample; 'Formula', permitting searches based on formula, structure type, Pearson symbol, space group, crystal system or on a combination of these fields; 'ID Codes', allowing the rapid retrieval of known entries.

Once the search has been specified, the user is presented with the opportunity of restricting the results based on the presence of coordinates, the *R* factor, the crystal system and the number of distinct elements present.

Several queries may be performed and the results viewed by selecting the 'Results' pages. The first page (Fig. 2) lists the results of all the queries with details of the selected ('Current') query shown in the lower panel. The user can create new queries by performing logical operations on existing queries, so that complex searches can be constructed.

Subsequent pages in the 'Results' section allow the user to examine the current entry in more detail. A 'Details' page presents the bibliographic and unit-cell information along with any comments or remarks contained in the database. 'Coordinates' is simply a list of atomic coordinates including Wyckoff symbols, site symmetries and occupancies.

The 'Geometry' page lists interatomic distances and (optionally) angles about each atom. Equivalent distances and angles are grouped together and the correct multiplicity factor is listed. The user can set the minimum and maximum distances to be tabulated. There is also a graphical display showing the environment of each atom out to a distance of 10 Å.

The 'View Structure' page (Fig. 3) allows the user to visualize the structure of an entry. This page, by default, attempts to generate a chemically meaningful plot of an entry. For intermetallics this is generally a ball and stick plot of the unit cell. If the program decides that the structure is better described by a mixture of spheres and polyhedra, then this representation is used. The user is then free to modify the diagram extensively. The orientation of the plot can be changed using the mouse as a virtual trackball, and the plot can be resized and moved around the display. The user can also spin the plot about either or both of the vertical and horizontal axes. There are also a number of options available that can change the appearance of the plot. 'Atom Details' changes the colour and diameter of an atom as well as its bonding radius, representation (polyhedron, bonds, isolated or omitted), and type (cation, anion, metal, non-bonding). 'Bond Details' changes the colour and radius of bonds. 'Packing Range' enables the user to select the volume of the plot. This may be defined: (*a*) using the unit cell and specifying ranges along each of the three axes, (*b*) as a slab or rod, with a cylinder axis either along a reciprocal axis or perpendicular to a plane, at a specified centre position and with a specified length and diameter, or (*c*) as a sphere with a given position and radius. 'Cell and Axis Details' governs how the cell edges and Cartesian axes are represented. 'Set Viewpoint' rotates the plot so one is viewing either down an axis or perpendicular to a plane. 'Background' sets the background colour. 'Print Background' determines if the background should be included when a plot is sent to the printer. By default, this is set to off to conserve ink.

Once the plot has been modified, the user has a number of choices. The standard Windows print functions are available, so the plot can be sent to a printer at whatever quality level the printer will support. It is also possible to generate a JPEG file at the screen resolution.

As many intermetallic structures are examined using powder techniques, the 'Powder Pattern' page displays a simulated diffraction pattern if adequate information is available in the entry. The user can change the radiation, the maximum 2θ and the overall $B_{\rm iso}$ used to compute the pattern (an overall parameter is used because a large number of entries do not include atomic displacement parameters).

Finally, there is a page which runs *MISSYM* (Le Page, 1988) on the entry. This has a number of uses, including checking the descriptions of structures for missed or non-crystallographic symmetry.

5. Applications

The availability of a comprehensive compilation of metals structural data is proving useful in a number of fields.

5.1. Identification of unknown materials from diffraction measurements

If a unit cell can be determined, as is often the case when a single crystal is available, then the database can be searched for compounds with similar cell dimensions. If for some reason, *e.g.* the sample is a mineral with a varying composition or the cell was measured at a different temperature than the database entry, then the tolerance of the match can be adjusted to widen the search. In such cases, if some information about chemical composition is available, this can be used to restrict the results.

When the unit-cell dimensions are not known but a powder diffraction pattern is available, a search can be made for the strongest diffraction peaks. Again, the availability of some chemical information is extremely useful in narrowing the search. For example, searching for an intense peak with a *d* spacing of 1 Å resulted in 1761 hits; adding the information that the sample contained Fe reduced this to 83 possibilities; and if the sample contained S the number of matches dropped to two. The database not only contains the strongest peaks for entries with coordinates, it also contains peak positions for those entries with only symmetry and cell dimensions. Using this expanded data with some chemical information increases the number of entries that can be searched.

5.2. Structure analysis

The assignment of structure types to the entries in CRYSTMET can be a very powerful approach to structure analysis. If it is possible to find an entry in the database that has the same structure type as the compound under investigation, then it is possible to use that entry as a starting point for assigning atomic coordinates. Even if the choice can be limited to one of a small number of structure types, computing simulated diffraction patterns and comparing them with observed patterns can limit the number of choices.

A similar approach has been used in interpreting extended X-ray absorption fine structure (EXAFS) results, and a custom version of the *Materials Toolkit* that creates files for input to an EXAFS simulation program is available from Rigaku Corp. (Taguchi & White, 2002).

5.3. Property prediction

One of the major challenges in materials science is the design of materials with a desired set of properties. One way to achieve this goal is by exploitation of the information for the wide ranges of compounds that are contained in CRYSTMET.

One approach to utilizing this information is to use structure/property maps in which a variety of coordinate systems are used to arrange the data. Some possible indices are atomic radii, electronegativity, atomic number, heats of reaction, transition temperature or any property of interest. For example, in structure analysis, as described above (§5.2), it often proves useful to plot atomic number on one axis for each atom (*i.e.* two axes for binary compounds or three for ternaries) against structure type. Structures of the same type are observed to form clusters and it is reasonable to predict that undetermined compounds inside a cluster will be of the same type. A full survey of this approach is given by Burdett & Rodgers (1994).

5.4. Structure and property calculation

Once a structure type is known it is now possible, even on a desktop PC, to perform *ab initio* computations on solid-state intermetallic compounds. These computations run in hours, rather than days or months, on such a computer and the results have been found to be consistent with experiment. There are several uses for such computations.

Where a user can identify a range of compounds with a consistent structure type, such a computation can be performed on formulations for which there is no information, or where only cell dimensions, space group and structure type are reported. The result will be a set of atomic coordinates, which can be used in lieu of experimentally determining the structure. Such coordinates have recently been computed for a number of compounds with entries that contained only space group and unit-cell information, and have been added to CRYSTMET (with appropriate comments that this is calculated data). The addition of these coordinates means that powder pattern simulations are now possible and this information can be used to help identify these compounds.

A benefit of these computations is that once complete, it is possible to derive a large number of physical properties from the results. These computed properties can be used in conjunction with structure/property maps to fill in the gaps, or structure/property maps can be used to predict interesting formulations, which can then be investigated computationally. The result is a method of investigating the structure and properties of a wide range of materials that can then be used to guide experimental investigations (Le Page *et al.*, 2002).

6. Availability

The *Materials Toolkit* is available with the CRYSTMET and/or the ICSD databases from Toth Information Systems, Inc. A demonstration version of the software with a limited database can be found at http://www.TothCanada.com. Custom versions of the software and the CRYSTMET database are also available from the Rigaku Corp. (2002) for use with their EXAFS simulation software and Rietveld refinement package.

References

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.

Belsky, A., Hellenbrandt, M., Karen, V. L. & Luksch, P. (2002). *Acta Cryst.* B**58**, 364–369.

Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L., Jain, S., Fagan, P., Marvin, J., Ravichanran, V., Schneider, B., Thanki, N., Padilla, D., Weissig, H., Westbrook, J. D. & Zardecki, C. (2002). *Acta Cryst.* B**58**, 899–907.

Bruno, I. J., Cole, J. C., Edgington, P. R., Kessler, M., Macrae, C. F., McCabe, P., Pearson, J. & Taylor, R. (2002). *Acta Cryst.* B**58**, 389–397.

Burdett, J. K. & Rodgers, J. R. (1994). *Encyclopedia of Inorganic Chemistry*, Vol. 7, pp. 3934–3952. New York: Wiley.

Gabe, E. J., Le Page, Y., Charland, J.-P., Lee, F. L. & White, P. S. (1989). *J. Appl. Cryst.* **21**, 384–387.

Le Page, Y. (1988). *J. Appl. Cryst.* **21**, 983–984.

Le Page, Y., Saxe, P & Rodgers, J. R. (2002). *Acta Cryst.* B**58**, 349–357.

Rigaku Corp. (2002). Rigaku Corporation, 3–9–12 Matsubara-cho, Akishima-shi, Tokyo 196, Japan.

Taguchi, T. & White, P. S. (2002). *Acta Cryst.* B**58**, 343–348.

Wood, G. H., Rodgers, J. R., Gough, S. R. & Villars, P. (1996). *J. Res. Natl Inst. Stand. Technol.* **101**, 205–215.