(14) Wenger, C. B.; Childress, J. "Journal Evaluation in a Large Research Library", *J. Am. Soc. Inf. Sci.* **1977**, *28*, 293-299.
(15) Wender, R. W. "Counting Journal Title Usage in the Health Sciences", *Spec. Libr.* **1979**, *70*, 219-226.
(16) Garfield, E. "Citation Analysis As a Tool in Journal Evaluation", *Science (Washington, D.C.)* **1972**, *178*, 471-479.
(17) Cawkell, A. E. "Evaluating Scientific Journals with Journal Citation Reports—A Case Study in Acoustics", *J. Am. Soc. Inf. Sci.* **1978**, *29*, 41-46.
(18) Hirst, G. "Discipline Impact Factors: A Method for Determining Core Journal Lists", *J. Am. Soc. Inf. Sci.* **1978**, *29*, 171-172.
(19) Cezairliyan, A. O.; Lykoudis, P. S.; Touloukian, Y. S. "A New Method for the Search of Scientific Literature through Abstracting Journals", *J. Chem. Doc.* **1962**, *2*, 86-92.
(20) Windsor, D. A. "Rational Selection of Primary Journals for a Biomedical Research Library: The Use of Secondary Journal Citations", *Spec. Libr.* **1973**, *64*, 446-451.
(21) Pope, A. "Bradford's Law and the Periodical Literature of Information Science", *J. Am. Soc. Inf. Sci.* **1975**, *26*, 207-213.
(22) Danilowicz, C.; Szarski, H. "Metoda oceny wartości merytorycznej czasopism naukowych na podstawie danych systemu SDI", *Zagadnienia Nauk.* **1979**, *15*, 55-63.
(23) Stefaniak, B.; Kacprowicz, H.; Dabkowski, H.; Grzecznowska, A. "Selektywna Dystrybucja Informacji. Organizacja i Funkcjonowanie SDI w Polsce", *Prace IINTE* **1979**, *20*, 1-119.
(24) Stefaniak, B. "Investigation of Needs of Users of Scientific and Technical Information System in Chemistry for Information Sources Included in *CACon* Data Base", 1st International Seminar—Foreign Machine-Readable Data bases for Chemistry and Chemical Industry—Processing and Experience. Warsaw 1979, 65-77 (in Russian).
(25) Brookes, B. C. "Numerical Methods of Bibliographic Analysis", *Libr. Trends* **1973**, *22*, 18-43.
(26) The lists come from the 1977 and 1978 "Chemical Abstracts Service Source Index" (*CASSI*) cumulative supplements.

Registration-Identification of Crystalline Materials Based on Lattice and Empirical Formula

JOHN R. RODGERS†
University Chemical Laboratory, Cambridge CB2 1EW, England

ALAN D. MIGHELL*
Center for Materials Science, National Bureau of Standards, Washington, DC 20234

Received August 20, 1980

Data files containing information on solid-state materials are expanding rapidly. Each year several thousand new materials are characterized by X-ray diffraction techniques. Consequently, computer procedures have been developed to register materials entering large data files. We have found registration based on lattice parameters and empirical formula especially effective. In our present registration procedure, the lattice is uniquely represented by the reduced cell and the elements in the formula are uniquely specified by prime numbers. This method has been applied for several years to register new materials for the Cambridge Crystallographic Data File which contains data on more than 25 000 compounds containing organic carbon. The method is now being adapted to register materials for the NBS Crystal Data File. Our experience suggests the desirability of routinely characterizing organic materials by cell parameters in addition to the traditional chemical analysis. A solid-state registry number that would allow one to identify the same compound in different data bases could be derived from the lattice constants and the empirical formula. Such a number would make it possible to distinguish polymorphs and different phases of the same composition.

INTRODUCTION

The recent expansion in the number of crystal structures characterized by X-ray diffraction methods[1] has forced data-base builders to improve techniques for the registration of new materials. Registration is the process by which new entries are added (if acceptable) to a data base and includes the identification of existing entries in the master file that are the same as or related to a given new entry. During registration, the new entry is critically evaluated with respect to the existing entries, and it is assigned a registry number and a reference code. In some cases existing entries may be changed, updated, or deleted. The method of registration (lattice formula) described here has been used at the Cambridge Crystallographic Data Centre since 1977 and will soon be adapted as a general method for registration of new crystalline materials entering the NBS Crystal Data File.[2]

Registration based on lattice formula is an effective procedure to identify previous entries on the same material, isometrically and isostructurally related substances, and materials with related lattices (e.g., sub- and superlattices). For organic materials which crystallize mainly in the lower symmetry crystal systems,[3] our experience has shown that the lattice alone is highly characteristic for a given compound. In fact, it could be used as the basis of a solid-state registration number.

In lattice-formula registration discussed below, the lattice is specified by the reduced cell and the elements of the empirical formula are uniquely defined by prime numbers. The reduced cell is used because it can be uniquely defined[4-6] and convenient algorithms to calculate it[5,6] have been devised. For registration, this cell allows one to locate identical and metrically similar lattices independently of crystal system and lattice centering. Further reasons for using the reduced cell for identification of crystalline materials have been published.[7] Recently, more general matrix methods to establish interlattice relationships have been developed.[8] For registration enhancement and the critical evaluation of data, newer versions of the registration program will be expanded to include these matrix methods.

GENERAL METHODOLOGY OF LATTICE-FORMULA REGISTRATION

An overall view of the general techniques used in lattice-formula registration is given in Figure 1. The new data (EDIT FILE) are registered against the old data (MASTER FILE). This general procedure has been used by the Cambridge Crystallographic Data Centre to register a new batch of compounds before they are added to the MASTER FILE. A

†University of Adelaide, Adelaide, South Australia, 5001.

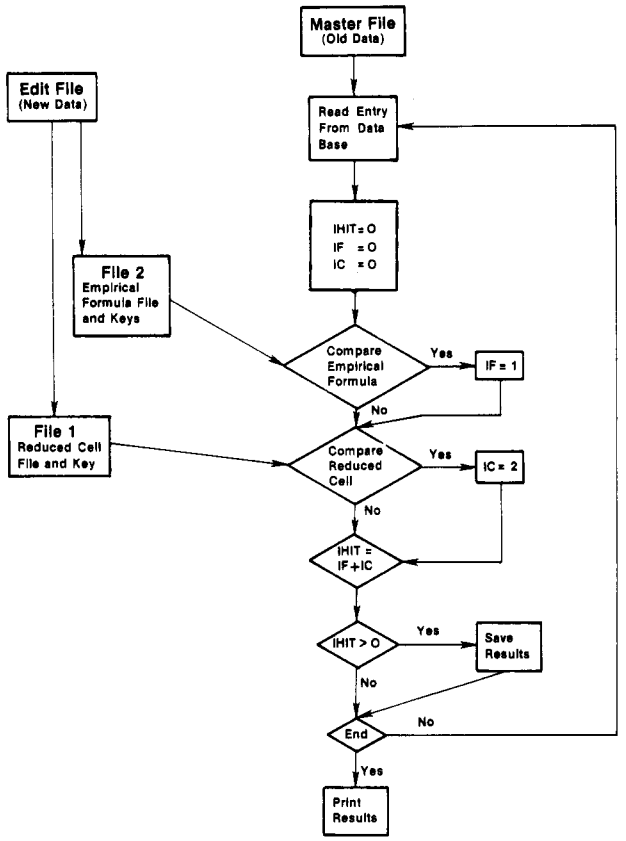

Figure 1. Flow diagram for lattice-formula registration: MASTER FILE (old data) compared against the EDIT FILE (new data).

new batch of about 400 entries is registered against the MASTER FILE (~25 000 entries) once every 5 or 6 weeks. To register a new batch requires about 1 min of computer time on an IBM 370/165.

There are three basic steps in registration. First, two files are generated from the EDIT FILE and kept in core. FILE 1 contains the reduced cell and cell key. FILE 2 contains the empirical formula and formula keys. The keys are used to speed the entire registration process. Second, the cell and formula data for each entry in the MASTER FILE are checked against FILE 1 and FILE 2 for a possible match. Third, a log is kept for all entries of the MASTER FILE that match (cell, formula, or both) one or more entries in the EDIT FILE. The results are printed out at the end of the registration run.

Once the EDIT FILE has been registered against the MASTER FILE, the hits (matches) are printed. Those new batch entries which do not have any hits are new entries. The hits are carefully scrutinized and fall into three types: (a) formula, (b) cell, and (c) cell and formula. Hits of type a usually represent isomeric materials or the same material where the lattice is incorrect for either the new or the old entry. Hits of type b may represent isostructural materials or the same material in which the chemistry is incorrect. Hits of type c almost always represent the same material.

Type c hits can occur for a number of reasons. First, such a hit may result because both the new and the old entry correspond to the identical compound from the same publication. Sometimes the data are inadvertently abstracted more than once from the same publication. The Centre has many sources for abstracting the primary literature. These are *Chemical Abstracts*, *Bulletin Signaletique*, or direct journal scans by members of the Centre. Even though a rigorous bookkeeping system is maintained, some entries are sometimes duplicated. Second, authors may publish the data from the same experiment in several different journals. For example, a brief summary of the crystallographic results of a structural study may be presented in a chemical journal and the complete results published in a crystallographic journal. Third, it is not uncommon for different authors to publish crystallographic details on the same compound. This may be done on purpose because more accurate experimental parameters are desired, or it may be done by mistake because of the inability of the authors to access recent journals or bibliographies on crystallography. Fourth, certain crystallographic details on the same experiment are reported more than once at national and international conferences.

SPECIFICS OF LATTICE-FORMULA REGISTRATION

To show the details of lattice-formula registration, we will now describe (1) the structure of the MASTER FILE (old data), (2) the structure of the EDIT FILE (new data), (3) the formats of FILE 1 and FILE 2, (4) the nature of the linkage between FILE 1 and FILE 2, and (5) an example of how a MASTER FILE entry is checked against FILE 1 and FILE 2 of the EDIT FILE.

(1) Structure of the MASTER FILE. The old material is held in the MASTER FILE and consists of three records per entry (unformatted).

Record 1: Reduced Cell Key (RKEY) and Formula Key (FKEY). The reduced cell key is given by

$$\text{RKEY} = 10^5\text{INT}(10^4(a/b) + 0.5) + \text{INT}(10^4(b/c) + 0.5)$$

where a, b, and c are the reduced cell dimensions. The formula key (FKEY) is given by

$$\text{FKEY} = 10^6\text{INT}(x + 0.5) + 10^3\text{INT}(y + 0.5) + \text{INT}(z + 0.5)$$

where x, y, and z are the element counts for the empirical formula when ordered as C_xH_y . . . and alphabetic on the remaining elements. The choice of element counts rather than element name was dictated because more than 90% of organic compounds contain C, H, and N or O. This method is very selective and can be illustrated by the example C12 H22 O11. There are many studies containing C, H, and O but very few having counts 12, 22, 11. Only the first three elements are chosen because this is only a key and because of the restriction of the word size, [$(2^{31} - 1)$, maximum integer size for a 32-bit word].

Record 2: Reference Code, Directory to Record 3, Element Key, Element Count, Reduced Cell, Author's Cell, Space Group. The reference code is used to identify an entry within the MASTER FILE. The directory to record 3 contains the starting addresses for different fields of the bibliographic material held in the final record for each entry. The element key (EKEY) is a single prime number indicating all the element types for the compound. Each element type is assigned a prime number (p_i) (except for carbon which is contained in all organic compounds) starting with the prime number 3. Then

$$\text{EKEY} = p_1p_2p_3 \ldots p_n$$

where p_1, p_2, etc., are the prime numbers assigned to the elements.

By use of prime numbers, each combination of element types produces a unique number. EKEY can serve a dual purpose. First it permits the matching of element types or formulas in the MASTER FILE with that of the new entries. Second, EKEY permits a rapid search for any given element. If an entry has the element types C, H, N, O, P, Ni which are assigned prime numbers –, 3, 5, 7, 43, 47, then EKEY = 212 205. A single match of EKEY would indicate the presence of all five elements (C is always present). This is much more

Table I. Definitions of Keys, Arrays, and Variables Used in Registration

edit file	master file	name	definition
R	RKEY	reduced cell key	$[10^9(a/b) + 10^4(b/c)]^{a,b}$
F	FKEY	formula key	$(10^6x + 10^3y + z)^{a,c}$
E	EKEY	prime number key	$p_1p_2 \cdots p_n{}^d$
EC	ECNT	element count	an array containing the number of each element
RC	RCELL	reduced cell	an array containing the reduced cell parameters
ISEQ		sequence number	position in EDIT FILE
	NCAR	no. of carbon atoms	FKEY/10^6 (integer only)
	DET	1st digit of a/b ratio	RKEY/10^8 (integer only)
LINK			the value in LINK(I) can be decoded to link data on a given entry in FILE 1 and FILE 2
RMAP			RMAP(DET) gives the position in the R array with a value of 0.DET or greater
FMAP			FMAP(NCAR) gives the position in the F array with NCAR or more carbon atoms

[a] See text for representation as an integer in the computer. [b] a, b, and c are the edges of the reduced cell. [c] x = C, y = H, z = next element in alphabetical order. [d] The p's represent the prime numbers assigned to the elements.

economical of computer time than five separate comparisons. The presence of Ni alone is indicated by MOD(EKEY,47) = 0 and the presence of O and P is given by MOD(EKEY, 7 × 43) = 0.

Element count (ECNT) is an array containing the exact value for each element, e.g., for C10 H12 values in the array would be 10, 12 and for C8 H7 N1, 0.45 (H2 O1), the values in the array would be 8, 7.9, 1, 0.45.

Reduced cell (RCELL) is an array containing the six reduced cell parameters: a, b, c, α, β, γ; the author's cell is stored in an array containing the six parameters of the author's cell; the space group is included if present.

Record 3: Bibliographic and Chemical Information. This record is a continuous character string for which the starting positions of each field are contained in the directory key of record 2. This record contains the compound name, synonym, formula, authors, journal references, and chemical classifications.

(2) Structure of the EDIT FILE. The data for the new materials (EDIT FILE) to be registered against the MASTER FILE are organized in nearly the same way. The definitions for the keys and arrays that are set up for the new entries are the same as those for the MASTER FILE and are given in Table I. This file contains all necessary data for input into the Cambridge Crystallographic Data Centre files. The EDIT FILE goes through a series of programs which checks the syntax and correctness of each element of the entry. When this EDIT FILE is error free, it is ready for registration.

(3) Registration Files: FILE 1 and FILE 2. From the EDIT FILE, two temporary files are created that contain the keys and data required for registration. FILE 1 contains the data and keys for cell registration (R, RC, ISEQ). FILE 2 contains data and keys for formula registration (F, E, EC, ISEQ). ISEQ is the sequence number of a given entry in the EDIT FILE.

After processing new entries, FILE 1 and FILE 2 are sorted on increasing values of R and F, respectively. An example for a batch of 10 compounds for FILE 1 and FILE 2 is given in Tables II and III. Corresponding entries in the two tables have the same sequence number (ISEQ). The registration program reads these two files into core and the data are stored in arrays. Data on the same compound in FILE 1 and FILE 2 are linked through an array called LINK.

(4) Linkage of FILE 1 and FILE 2 and Array Pointers. An example of the linking of FILE 1 and FILE 2 is illustrated in Table IV. The value of LINK(I) in the array LINK can be decoded so one can find the positions in the R and F arrays that contain data on the same compound. Thus the integer numbers LINK(I)/10^3 and MOD(LINK(I), 1000) give the positions in the R and F arrays that contain data on the same material. LINK(1) gives the positions of the reduced cell, empirical formula, and associated keys for entry 1 in the EDIT FILE. For example in Table IV, LINK(1) = 5001. Consequently, the position of the reduced cell key in array R = LINK(1)/10^3 = 5; and the position of the formula key in array F = MOD(LINK(1), 1000) = 1.

The arrays RMAP and FMAP are filled in next (see bottom of Table IV). These arrays give the ranges for search so that one does not need to search the entire R or F arrays to make a match. In Table IV, FMAP(3) gives the first position in the F array for which there are three or more carbon atoms in the formula. Likewise RMAP(5) gives the position in the R array for which the magnitude of the a/b ratio of the reduced cell is 0.5 or greater. RMAP has a dimension of 10 because the determinative ratio a/b can vary from 0.0001 to 1.000. FMAP has a dimension of 20 because an analysis of the empirical formula file (1977) for the MASTER FILE showed that most of the entries had carbon counts less than 20. Since 1977 there has been an increase in the size of the molecules studied by X-ray diffraction techniques and the dimension of FMAP will be increased in the next version of the program.

(5) Registration of an Entry of the MASTER FILE Against FILE 1 and FILE 2. For a given MASTER FILE entry, one calculates the number of carbon atoms (NCAR) by FKEY/10^6. One then looks in FMAP(NCAR) to find the location of where to start searching in the F array in FILE 2. Likewise one calculates the first digit of the a/b ratio (DET) by RKEY/10^8. One then looks in RMAP(DET) to find the location of where to start searching in the R array in FILE 1. To overcome problems in rounding and precision in reported cell and formula values, ranges are used: RMAP(DET − 1)–RMAP(DET + 1) and FMAP(NCAR −

Table II. Reduced Cell Data for FILE 1: Cell Data Extracted from a Typical EDIT FILE Consisting of Ten Entries

ISEQ[a]	a	b	c	α	β	γ	a/b	b/c	R	ISEQF1[b]
1	3.65	10.01	10.60	90.0	90.0	90.0	0.3646	0.9443	364609943	5
2	3.83	16.48	18.13	114.8	90.0	90.0	0.2324	0.9089	232409089	2
3	4.02	6.14	21.49	90.9	90.0	91.3	0.6547	0.2857	654702857	10
4	4.83	14.31	14.38	98.1	90.0	90.0	0.3375	0.9951	337509951	4
5	5.43	10.22	13.13	90.0	90.0	103.2	0.5313	0.7783	531307783	7
6	5.90	15.10	29.95	99.9	90.0	90.0	0.3907	0.5041	390705041	6
7	5.99	18.88	19.72	90.0	93.0	90.0	0.3172	0.9574	317209574	3
8	6.54	11.78	11.78	90.0	90.0	90.0	0.5552	1.0000	555210000	8
9	6.68	41.18	41.18	60.0	90.0	90.0	0.1574	1.0000	157410000	1
10	7.04	12.18	18.20	91.0	90.0	90.0	0.5780	0.6692	578006692	9

[a] ISEQ is the sequence position in the EDIT FILE. [b] ISEQF1 is the sequence position in FILE 1.

Table III. Formula Data for FILE 2: Formula Data Extracted from a Typical EDIT FILE Consisting of Ten Entries

ISEQ[a]	formula	E[b]	F	ISEQF2[c]
1	$C_3 H_4 O_3$	21	3004003	1
2	$C_{14} H_6 F_2 O_3$	231	14006002	5
3	$C_4 H_4 N_2 O S$	1785	4004002	2
4	$C_{10} H_6 Br N O_{29} H_2 O$	1995	10008001	4
5	$C_4 H_8 N_2 O_4 Pt$	3885	4008002	3
6	$C_{15} H_{11} Br O$	399	15011001	7
7	$C_{24} H_{20} Cu N_2 O_{49} H_2 O$	3045	24022001	9
8	$C_{16} H_{12} S_4 Sn$	2091	16012004	8
9	$C_{37} H_{51} I O_{10} S$	4641	37051001	10
10	$C_{15} H_{10} Fe_2 O_6$	483	15010002	6

[a] ISEQ is the sequence position in the EDIT FILE. [b] Element prime number. The prime numbers assigned to the elements are -, C; 3, H; 5, N; 7, O; 11, F; 13, I; 17, S; 19, Br; 23, Fe; 29, Cu; 31, Co; 37, Pt; 41, Sn. [c] ISEQF2 is the sequence position in FILE 2.

1)–FMAP(NCAR + 1). For example, in the case of the formulas, the location in the F array for the bottom of the range is FMAP(NCAR – 1) and the top of the range is given by FMAP(NCAR + 1). The range is searched by the binary look-up method.[9]

For a given MASTER FILE entry, the values of RKEY and FKEY are then matched (only the most significant digits) against the arrays R and F (in FILE 1 and FILE 2) in the appropriate ranges.

For a match of RKEY against R(ISEQF1) in the R array, the corresponding value F(ISEQF2) in the F array is found via LINK. If no match is found for RKEY in the range specified in R, then a formula match is done of FKEY against the specified range in the F array. The above method for mapping the arrays R and F by the array LINK allows the search for RKEY and FKEY to be carried out in any order. If a match is not found for either the cell or formula, records 2 and 3 of the MASTER FILE are skipped by a dummy read. If a match is found, then record 2 and 3 are read. Record 2 holds the reduced cell values RCELL(a, b, c, α, β, γ) which are matched with those in the array RC. A full comparison of these cells is made. The conditions for equality are given by $\Delta \leq 0.05$ for the cell edges and $\Delta \leq 1$ degree for the angles. These values of tolerances were arrived at by experience. If there is a formula match, then the Δ's for the cell edges are increased (see also section on Improvements for Lattice Matching).

If a match for FKEY vs. F(ISEQF2) exists, the corresponding match for the MASTER FILE element key (EKEY) and FILE 2 element key (E) is performed. If there is a match, a complete comparison of element counts for all element types is performed.

The matching of the MASTER FILE entry against FILE 1 and FILE 2 is continued throughout the specified range. A machine generated log is kept for the types of hits for each EDIT FILE entry vs. the MASTER FILE. The types of hits are identified as follows: *formula match, **cell match, ***cell and formula match.

This registration procedure is also carried out within the batch of new entries to check if the same material is in the batch more than once.

DISCUSSION

Lattice-formula registration has been incorporated in the routine operating procedures of the Cambridge Crystallographic Data Centre[1] since Jan 1977. It has proved to be a very practical method to register organic materials. The number of new materials registered against the MASTER FILE for this period was about 14000. In addition the cross-referencing of Crystal Data Volume 3[10] (~12000 entries) and the back-referencing of this volume to Crystal Data Volume 1 (~8000 entries) were done with this method.

We have noted from our experience that the following types of errors in the literature occur: (1) incorrect chemistry assigned to the material especially in preliminary publications, (2) an incorrect number of solvent molecules reported, and (3) lattice centering not given. Registration, however, is still possible in spite of these errors. For example, the errors in chemistry do not present a serious problem as the lattice is usually sufficient to register the material. Registration in spite of lattice centering errors is possible by an extension of lattice-matching techniques (discussed later).

Although the present registration scheme works well for organic materials, a few minor modifications are necessary for inorganic substances which are commonly ionic and crystallize in the higher symmetry space groups. For example, the RKEY method is not selective in choosing reasonable ranges for searching the reduced cells for inorganic substances. In the preparation of a comprehensive registration scheme for all materials, it is envisaged that the cell volume would be used instead of RKEY.

Examples of Registration. Examples of registration are given in Figures 2–6. In Figure 2, the incoming entry is ?DNINAP01. It has hit with three entries in the MASTER FILE: DNNAPH, DNNAPH01, and DNTNAP10. Two of these entries are isomeric with ?DNINAP01 having hit only on the formula (*type). The third entry DNTNAP10 is in fact a true match having hit on formula and cell (***type).

Figure 3 illustrates isomorphous entries. The two hits shown (ESULYB and HOESUL) have matched on reduced cell (**type) only. On examination of the formula, it can be seen

Table IV. FILE 1 (Cell Data) and FILE 2 (Formula Data) Sorted on R and F, Respectively[a]

	FILE 1										FILE 2								
ISEQ[b]	R		RC					ISEQF1[c]	LINK	ISEQF2[c]	F	E	EC					ISEQ[b]
9	157410000	6.68	41.18	41.18	60.0	90.0	90.0	1	5001	1	3004003	21	3.0	4.0	3.0			1
2	232409089	3.83	16.48	18.13	114.8	90.0	90.0	2	2005	2	4004002	1785	4.0	4.0	2.0	1.0	1.0	3
7	317209574	5.99	18.88	19.72	90.0	93.0	90.0	3	10002	3	4008002	3885	4.0	8.0	2.0	4.0	1.0	5
4	337509951	4.83	14.31	14.38	98.1	90.0	90.0	4	4004	4	10008001	1995	10.0	8.0	1.0	1.0	30.0	4
1	364609943	3.65	10.01	10.60	90.0	90.0	90.0	5	7003	5	14006002	231	14.0	6.0	2.0	3.0		2
6	390705041	5.90	15.10	29.95	99.9	90.0	90.0	6	6007	6	15010002	483	15.0	10.0	2.0	6.0		10
5	531307783	5.43	10.22	13.13	90.0	90.0	103.2	7	3009	7	15011001	399	15.0	11.0	1.0	1.0		6
8	555210000	6.54	11.78	11.78	90.0	90.0	90.0	8	8008	8	16012004	2091	16.0	12.0	4.0	1.0		8
10	578006692	7.04	12.18	18.20	91.0	90.0	90.0	9	1010	9	24022001	3045	24.0	22.0	1.0	2.0	50.0	7
3	654702875	4.02	6.14	21.49	90.9	90.0	91.3	10	9006	10	37051001	4641	37.0	51.0	1.0	10.0	1.0	9
RMAP(1) . . . RMAP(10): 1, 2, 3, 7, 7, 10, 10, 10, 10, 10																			
FMAP(1) . . . FMAP(20): 1, 1, 1, 2, 4, 4, 4, 4, 4, 4, 5, 5, 5, 5, 6, 8, 9, 9, 9, 10																			

[a] Cell and formula data on the same entry are tied together through LINK. FMAP(NCAR) gives the position in the F array to start searching on formulas with NCAR or more carbon atoms. RMAP(DET) gives the position in the R array to start searching on a/b values of 0.DET or greater. [b] ISEQ is the sequence position in the EDIT FILE. [c] ISEQF1 and ISEQF2 are the sequence positions in FILE 1 and FILE 2, respectively.

```
?DNINAP01
#COMPND 1,8-Dinitronaphthalene
#QUAL orthorhombic form, at 97 deg. C
#FORMUL C10 H6 N2 O4                                    ...NEW ENTRY
#AUTHOR M. Ciechanowicz-Rutkowska
#REF 208,22,185,1977
#CLASS 1/24
#CELL a = 11.475  b = 15.002  c = 5.425  SG P212121
#SYSCAT sys O cat 3


1,5-Dinitronaphthalene
C10 H6 N2 O4                                    ...MASTER FILE ENTRY
isomer  J. Trotter
Acta Cryst. 00,13,95,1960
a = 7.760  b = 16.320  c = 3.70  SG P21/A
α = 90.  β = 101.80  γ = 90.
*DNNAPH  1/24        *POSSIBLE CHANGE FOR DNINAP01...DNNAPH01


1,5-Dinitronaphthalene
C10 H6 N2 O4                                    ...MASTER FILE ENTRY...
isomer  N. G. Sevastyanov, G. S. Zhdanov, M. M. Umanskij
Zh. Fiz. Kim. 115,22,1153,1948
No cell parameters given
*DNNAPH01  1/24      *POSSIBLE CHANGE FOR DNINAP01...DNNAPH02


1,8-Dinitronaphthalene
C10 H6 N2 O4                                    ...MASTER FILE ENTRY
true match  A. A. Akopyan, A. I. Kitaigorodskij, T. T. Struhkov
Zh. Strukt. Kim. 082,6,729,1965
a = 11.352  b = 14.934  c = 5.376  SG P212121
***DNTNAP10  1/24      *POSSIBLE CHANGE FOR DNINAP01...DNTNAP01
```

Figure 2. Registration of the new entry shows that the MASTER FILE contains two isomers and one identical material.

```
?004724
#COMPND nona-aquo-Dysprosium(iii) ethylsulfate
#FORMUL 3(C2 H5 O4 S1 1-), H18 O9 Dy1 3+          ...NEW ENTRY
#AUTHOR J. Albertsson, I. Elding
#REF 107,33,1460,1977
#CLASS  1/11
#CELL a = 13.9287  c = 7.0537  SG P63/M
#SYSCAT sys H cat 1


nona-aquo-Ytterbium(iii) ethylsulfate
isomorphous  3(C2 H5 O4 S1 1-), H18 O9 Yb1 3+          ...MASTER FILE ENTRY
J. Albertsson, I. Elding
Acta Cryst. B107,33,1460,1977
a = 13.899  b = 13.899  c = 7.025  SG P63/M
**ESULYB  1/11        *POSSIBLE CHANGE FOR 004724...ESULYB01


nona-aquo-Holmium(iii) ethylsulfate (neutron study)
isomorphous  3(C2 H5 O4 S1 1-), H18 O9 Ho1 3+          ...MASTER FILE ENTRY
C. R. Hubbard, C. O. Quicksall, R. A. Jacobson
Acta Cryst. B107,30,2613,1974
a = 13.920  b = 13.920  c = 7.03  SG P63/M
**HOESUL  1/11        *POSSIBLE CHANGE FOR 004724...HOESULN1
```

Figure 3. Registration shows that the MASTER FILE contains two materials that are isomorphous with the new material.

that the new entry contains Dy, while the MASTER FILE entries contain Yb and Ho.

Figure 4 illustrates a case in which the new entry ?PBPCPD and its associated MASTER FILE entries have a history. The new entry has hit with two entries in the MASTER FILE, one on cell and formula (***type) and the other on cell only (**type). Preliminary structural information on PMBZPD was presented at the Yugoslavian Crystallographic Meeting in 1976. Additional structural information [the presence of a solvent molecule (PDCPPB)] was presented at the European Meeting at Oxford in 1977. The new entry has the same crystal structure as PDCPPB but has more structural information such as atomic coordinates, bond lengths, etc.

Figure 5a gives an example that demonstrates the power of lattice matching in registration. The new entry, ?HY-

```
?PBPCPD
#COMPND Dichloro(2,11-bis(diphenylphosphinomethyl)-
        benzo(c)phenanthrene) palladium benzonitrile
        solvate
#Formul C44 H34 C12 P2 Pd1, C7 H5 N1          ...NEW ENTRY...
#Author F. Bachechi, L. Zambonelli, K. M. Venanzi
#REF 010,60,2815,1977
#CLASS  1/86
#CELL a = 14.361  b = 13.044  c = 11.897  SG P-1
     α = 105.97  β = 100.27  γ = 94.76
#SYSCAT sys A cat 3


Dichloro-(2,11-bis(diphenylphosphinomethyl)benzo(c)phenanthrene) palladium
C44 H34 C12 P2 Pd1                              ...MASTER FILE ENTRY...
F. Bachechi, L. Zambonelli, F. J. Reed, L. M. Venanzi
Izv. Jugosl. Cent. Kristalografiju, Ser. A379,11,55,1976
a = 14.361  b = 13.044  c = 11.897  SG P-1
α = 105.97  β = 100.27  γ = 94.76
**PMBZPD  1/86        *POSSIBLE CHANGE FOR PBPCPD...PMBZPD01


Dichloro-(2,11-bis(diphenylphosphinomethyl)benzo(c)phenanthrene)
palladium benzonitrile solvate
C44 H34 C12 P2 Pd1, C7 H5 N1                    ...MASTER FILE ENTRY
F. Bachechi, L. Zambonelli, D. K. Johnson, K. M. Venanzi
Eur. Cryst. Meeting 245,170,1977
a = 14.361  b = 13.044  c = 11.897  SG P-1
α = 105.97  β = 100.27  γ = 94.76
***PDCPPB  1/86       *POSSIBLE CHANGE FOR PBPCPD...PDCPPB01
```

Figure 4. A new entry with a history.

```
(a)
?HYCAMS
#COMPND Hycanthone methanesulfonate
#FORMUL 0.88 (C20 H25 N2 O2 S1 1+),                ...NEW ENTRY
        0.12 (C20 H23 N2 O2 S1 1+),
        C1 H3 O3 S1 1-
#AUTHOR C. H. Wei, J. R. Einstein
#REF 107,34,205,1978
#CLASS 1/39  2/39
#CELL a = 26.1935  b = 8.7997  c = 10.6373  β = 116.42  SG P21/A
#SYSCAT sys M cat 3


1-(2-(Diethylaminoethyl)amino)-4-(hydroxymethyl)-thioxanthene-9-
one methanesulfonate
C20 H25 N2 O2 S1 1+, C1 H3 O3 S1 1-          ...MASTER FILE ENTRY
C. H. Wei, J. R. Einstein
Am. Cryst. Assoc., Ser.2395,2,34,1977
a = 26.193  b = 8.800  c = 10.637  β = 116.42  SG P21/C
**HYCANM  1/39        *POSSIBLE CHANGE FOR HYCAMS...HYCANM01


(b)
?00495802
#COMPND Methylammonium tetrachloromanganese(ii)
#FORMUL 2(C1 H6 N1 1+), Cl4 Mn1 2-                ...NEW ENTRY
#AUTHOR A. Daoud, A. Thrierr-Sorel, R. Perret, B. Chaillot
#REF 231,22,857,1977
#CLASS 1/3
#CELL a = 7.19  b = 19.40  c = 7.28  SG CMCA
#SYSCAT sys O cat 1


Methylammonium tetrachloromanganese(ii)  (neutron study)
2(C1 H6 N1 1+), Cl4 Mn1 2-                       ...MASTER FILE ENTRY
G. Hegger, D. Mullen, K. Knorr
Phys. Status Solidi 199,31,455,1975
a = 7.276  b = 7.215  c = 19.410  SG ABMA
***MATCHN01  1/3       *POSSIBLE CHANGE FOR 00495802...MATCHN02
```

Figure 5. Two examples of the matching of identical compounds. (a) shows that the match is made in spite of apparent chemical differences.

CAMS, has hit on cell (**type) with one entry in the MASTER FILE. Inspection of these two entries shows that the formulas of ?HYCAMS and HYCANM are different. Also ?HYCAMS has a different space group for the same labeled set of axes. Conventional methods for identifying a match of these two entries would prove difficult. The chemical names

```
?TBFCOP
#COMPND cis-bis(1-(2-Thienyl)-4,4,4-trifluoro-1,3-
        butanedionato)-bis(4-methylpyridine)-cobalt(ii)
#FORMUL  C28 H22 F6 N2 O4 S2 Co1                    ...NEW ENTRY
#AUTHOR J. A. Pretorius, J. C. A. Boeyens
#REF 042,40,407,1978
#CLASS 1/77  1/83
#CELL a = 9.393  b = 17.756  c = 18.042  β = 94.8  SG C2/C
#SYSCAT sys M cat 3
     Reduced Cell  a = 9.39  b = 10.04  c = 18.04
                   α = 87.76 β = 85.20  γ = 62.12



#TFBZNP
#COMPND cis-bis(1-(2-Thienyl)-4,4,4-trifluoro-1,3-
        butanedionato)-bis(4-methylpyridine)-zinc(ii)
#FORMUL C28 H22 F6 N2 O4 S2 Zn1                     ...NEW ENTRY
#AUTHOR J. A. Pretorius, J. C. A. Boeyens
#REF 042,40,407,1978
#CLASS 1/77  1/83
#CELL a = 9.411  b = 17.788  c = 18.131  β = 94.794  SG C2/C
#SYSCAT sys M cat 3
     Reduced Cell  a = 9.41  b = 10.06  c = 18.04
                   α = 87.76 β = 85.20  γ = 62.12
```

Figure 6. Lattice-formula matching of entries within a batch of new entries.

would pose the first problem and a formula lookup would not give a match. Likewise a search by Chemical Abstract registry number would prove unsuccessful. The lattice check, however, provided the match. Figure 5b gives an example of a straightforward cell and formula match. In this example, the new entry has a different set of axes, hence a different space group setting.

Figure 6 gives an example of lattice-formula checking within a batch of new entries. It shows an isostructural match, the different elements being Co and Zn.

Improvements for Lattice Matching. Lattice-matching techniques can be further enhanced by employing matrix procedures for elucidating cell and lattice relationships.[8] Using these procedures one can register materials in spite of certain experimental errors. Two areas in which lattice matching could be improved by matrix techniques are outlined below.

The first area in which cell matching could be improved concerns the matching of "different" reduced cells from the same lattice. It is possible for two reduced cells (e.g., derived from cell parameters in different publications on the same material) to be based on the same cell edges but different cell angles. Although infrequent, this situation can arise because of the interaction of experimental errors with the testing of the inequalities inherent in the special conditions for reduction. This problem can conveniently be remedied by matrix techniques. Thus in the cell match part of registration, if the edges and volumes match but one or more of the angles do not, one could shift to the matrix method. A computer program has been written on the basis of the published algorithm[8] that finds (if it exists) a matrix (or matrices) that relates the two cells. If an appropriate transformation matrix can be found (i.e., one with integral elements and with a determinant of one), then the two cells define the same lattice.

The second area in which lattice matching could be improved concerns the matching of derivative lattices. This is desirable because in some cases cell centering is not published (e.g., a face-centered cubic cell reported as a primitive cell) or a sub- or supercell may be determined experimentally.[7] Consequently, one sometimes needs to match cells with volume ratios of $^1/_2$, 2, $^1/_3$, 3, $^1/_4$, 4, etc. In other cases (e.g., in the cells determined by indexing procedures on powder data), one may obtain a cell that defines a lattice that bears a composite relationship to the correct lattice.[8,11] In all these cases, lattice registration is still possible by using the matrix approach[8] with the nature of the lattice relationship deduced from the properties of the relevant transformation matrix (matrices). Alternatively, the matching of a lattice against a sublattice or superlattice (but not a composite lattice) can be accomplished by systematically calculating and applying the appropriate upper triangular matrices.[12]

CONCLUSIONS

Our experience indicates that unknown materials can conveniently be identified by using lattice-matching techniques similar to those used in registration. This can be accomplished by using, for example, the NBS Crystal Data File for the MASTER FILE with the unknown corresponding to the EDIT FILE (i.e., the EDIT FILE has only one compound). A segment of this file (soon the entire file) is now available to the scientific community via the NIH-EPA Chemical Information System.[13] Also it is planned to produce a tape version of the file with identification search software. An unknown can be identified once a primitive cell of the lattice is known. This can be accomplished in an average time of about 6 h with an automated single-crystal diffractometer if suitable crystals are available. It is expected that simpler and cheaper methods will be devised to obtain the required primitive cell.

Because of their uniqueness, especially for organic and biochemical materials, lattice parameters should routinely be determined for new materials. Such parameters plus the traditional chemical analysis and other data would accurately characterize the material. For example, a drug sometimes exists in polymorphic forms that can be distinguished by their lattice parameters. In fact, a highly specific solid-state registry number could be based on lattice parameters (e.g., the reduced cell) and the empirical formula. Such a number can be used instead of, in conjunction with, or appended to the CAS registry number.

ACKNOWLEDGMENT

We thank Antonio Santoro and Judith Stalick for many helpful discussions in connection with this work.

REFERENCES AND NOTES

(1) Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B. G.; Kennard, O.; Motherwell, W. D. S.; Rodgers, J. R.; Watson, D. G. "The Cambridge Crystallographic Data Centre: Computer-Based Search, Retrieval, Analysis and Display of Information", *Acta Crystallogr. Sect. B* **1979**, *B35*, 2331–2339.
(2) A computer file of crystallographic data compiled by the NBS Crystal Data Center, National Bureau of Standards, U.S. Department of Commerce.
(3) Mighell, A. D.; Rodgers, J. R. "Lattice Symmetry Determination", *Acta Crystallogr. Sect. A* **1980**, *A36*, 321–326.
(4) Niggli, P. "Kristallographische und Strukturtheoretische Grundegriffe"; Akademische Verlagsgesellschaft: Leipzig, 1928; Vol. 7, part 1.
(5) Santoro, A.; Mighell, A. D. "Determination of Reduced Cells", *Acta Crystallogr., Sect. A* **1970**, *A26*, 124–127.
(6) "International Tables for X-ray Crystallography", 3rd ed.; Kynoch Press: Birmingham, 1969; Vol. 1, pp 530–535.
(7) Mighell, A. D. "The Reduced Cell: Its Use in the Identification of Crystalline Materials", *J. Appl. Crystallogr.* **1976**, *9*, 491–497.
(8) Santoro, A., Mighell, A. D.; Rodgers, J. R. "Determination of the Relationship Between Derivative Lattices", *Acta Crystallogr., Sect. A*, 1980, *A36*, 796–800.
(9) Knuth, D. E. "Algorithms", *Sci. Am.* **1977**, *236*, 63–80.
(10) "Crystal Data Determinative Tables", 3rd ed.; U.S. Department of Commerce, National Bureau of Standards, and the JCPDS-International Centre for Diffraction Data: Swarthmore, PA, 1972, 1973, 1978; Vol. 1–4.
(11) Mighell, A. D.; Stalick, J. K. "The Reliability of Powder Indexing Procedures", *NBS. Spec. Publ.* **1979**, *No. 567*, 393–403.
(12) Santoro, A.; Mighell, A. D. "Coincidence-Site Lattices", *Acta Crystallogr., Sect. A.* **1973**, *A29*, 169–175 (See Appendix for matrix generation algorithm).
(13) NIH-EPA Chemical Information System. Available to the public via Information Sciences Corporation, Suite 500, 918 16th St., NW, Washington, DC 20006