

## International Union of Crystallography

### Commission on Crystallographic Data

#### The Deposition of Crystallographic Results: Current Problems and Their Causes

BY GÜNTER BERGERHOFF\*

*Anorganisch-Chemisches Institut der Universität Bonn, Gerhard-Domagk-Strasse 1, D-5300 Bonn 1, Federal Republic of Germany*

AND FRANK H. ALLEN,† SHARON A. BELLARD AND CLAIRE V. LUCAS

*Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England*

(Received 3 February 1986; accepted 8 July 1986)

#### Abstract

Deposited or supplementary publications are now an important source of primary crystallographic results. Of 1117 complete publications recently processed by the Cambridge Structural Database, 65.1% had associated deposited data. In 28.6% of the cases, the deposited material was the sole source of atomic coordinates. A large number of depositions (25.0%) presented considerable problems of interpretation. These may be categorized as problems of illegibility (77.5%), poor layout (24.1%) and incompleteness (3.8%). These problems will be experienced by any user of deposited data, but create particular difficulties and delays for database producers who must handle several thousand depositions each year. Specific problems are described in each category, their causes are identified and simple remedies are suggested.

#### Introduction

The Cambridge Structural Database (CSD: Allen *et al.*, 1979; 50 499 entries) and the Inorganic Crystal Structure Database (ICSD: Bergerhoff, Hundt, Sievers & Brown, 1983; 24 000 entries) (1 January 1986) contain details of 93% of all published crystal structures. More than 45% of this material has been published since 1980 and nearly 75% has appeared in the last decade. This massive increase in crystallographic output has prompted more and more journals to use some form of deposited or supplementary publication for the associated numerical results. Structure factor listings have been deposited for many years and this policy is now applied routinely in various

journals to some or all of: (a) crystal data, (b) atomic coordinates (all atoms or H atoms only), (c) atomic vibrational parameters ( $U$  or  $U_{ij}$ ), (d) bond lengths and inter-ionic distances, (e) valence and torsion angles, and (f) miscellaneous geometrical results and illustrations. It should be noted, however, that deposition schemes are not designed solely to conserve valuable space in the primary (printed) journals. They also provide a route by which full details of crystallographic analyses, reported as 'Short Communications' in a variety of journals, may be placed in the public domain. Many hundreds of coordinate sets which might otherwise have been lost (and *were* lost before the late 1970's) are thereby now available both to individuals and to the databases.

Supplementary publications are, therefore, an increasingly important source of primary crystallographic data. Table 1 shows an analysis of the last 1391 entries incorporated in CSD. For those entries where atomic coordinates were reported (in either the primary or supplementary publication) no less than 65% had deposited material in categories (a)–(d) above. For nearly 30% of these entries, the supplementary publication represented the sole source of atomic coordinates. Of greater concern, however, is the high proportion (25%) of deposited material for which severe problems of interpretation were encountered by CSD staff. Similar problems are encountered by ICSD and will also affect any individual who attempts to make use of such documents. This is particularly unfortunate since many major journals now enforce the deposition of items (b), (c) above, either wholly or in part. Crystallographic results are now proving increasingly important for systematic studies in chemistry (see *e.g.* Wilson & Huffman, 1980; Allen, Kennard

\* Chairman and † Member IUCr Commission on Crystallographic Data.

& Taylor, 1983; Buergi & Dunitz, 1983; Brown & Altermatt, 1985), in pharmacology (Vinter, 1985), and in other branches of science. It is therefore important that depositions should be readily interpretable by all users, whether they are crystallographers or not.

Table 1 shows that problems arise principally from the poor quality and presentation of much deposited material. The IUCr Commission on Crystallographic Data is concerned with crystallographic information wherever and however it is published and has raised these points with the editorial offices of several journals. The editorial viewpoint may be summarized as follows: the importance and ready availability of crystallographic deposition is recognized, as is awareness that some material is of poor quality, but this is often only noted after the primary paper has been sent for printing. Requests to authors for better copy would, at this stage, cause delays in the journal production schedule. They prefer that problems of quality and presentation be eliminated before the material reaches the production stage.

The Commission feels that secondary material is sometimes regarded as of secondary importance throughout the publication process. The purpose of this Commission Report is to highlight the problems caused by such an attitude, in the hope that the situation can be remedied. Despite the fact that *Acta Crystallographica* is essentially blameless in this respect, the Commission feels that it is the only journal where the present comments will reach a significant proportion of the relevant audience, identified as the authors themselves, crystallographers acting as referees, and crystallographers who are members of the editorial boards of non-IUCr journals.

### Current deposition practices and problems

#### *Types of depository and methods of acquisition*

We begin by identifying the five major methods used in storing and distributing deposited material: (1) library or national archive (as hard copy or microfilm); (2) microfiche supplement to a journal; (3) direct deposition with a crystallographic database; (4) 'on application to the authors'; (5) editorial office of a journal. The relative frequency of each method is given in Table 1.

In terms of acquisition by an end-user, methods (1) and (2) are efficient but usually involve a small financial transaction for each document. This is simple if both user and depository are in the same country, otherwise currency exchanges for such amounts can be irritating and expensive. Large-scale users such as CSD and ICSD prefer to subscribe to the supplementary edition and to have accounts with major depositories to ensure rapid receipt. Many individuals prefer to wait until the material appears in the relevant database. Methods (4) and (5) are of declining popularity. Method (4) places

Table 1. *Analysis of deposited material incorporated into the Cambridge Structural Database during the period June–August 1985*

A. Entries processed	1391
B. With coordinates reported	1117 (80.3% of A)
C. With deposited material	727 (65.1% of B)
C1 H coordinates	518 (46.4% of B)
C2 non-H coordinates	319 (28.6% of B)
D. Problems of interpretation	182 (25.0% of C)
E. Breakdown of C and D by source of deposited data and by category of problem.	
	No. of problem entries
Source	No. (%C) Legibility Layout Incomplete Total
1. Library/archive	342 (47.0) 55 17 1 73
2. Microfiche suppl.	204 (28.1) 72 3 2 77
3. With database	126 (17.3) 12 13 1 26
4. From authors	50 (6.9) 2 1 1 4
5. Editorial office	5 (0.7) 0 0 2 2
Totals	727 141 34 7 182
%C	100.0 19.4 4.7 1.0 25.0
%D	— 77.5 24.1 3.8 100.0

an onus on the authors to respond to requests and is open to the criticism that it lacks the permanence and continuity of a large organization and implies that the secondary material is not available to referees and editors. In practice method (4) can work very well (Table 1E), but these results do not include cases of zero response.

Method (3) implies an active working relationship between the journal and the database and has some advantages. The journal forwards relevant material after the refereeing process is complete. It can then be rapidly associated with the printed paper, checked for possible typographic errors and included in the database. The database producer agrees to supply the material freely to any applicant, irrespective of whether the applicant is a database subscriber or not. This is a photocopy of the material for requests received within a few weeks of publication, but is normally a copy of the relevant database entry with any errors detected and (hopefully) rectified. Individuals requesting material deposited with *Acta Crystallographica* generally use method (5), a free and efficient service of the IUCr.

#### *Problems of quality and presentation*

Experience at CSD and ICSD has identified three major sources of difficulty faced by many users of deposited material, arising from inadequate legibility, layout or completeness of the secondary document. Relevant figures under these headings are given in Table 1E for each type of depository identified in the previous section. Databases report that problem material can take up to ten times longer to process, such delays being compounded where clarification and missing or replacement material must be obtained from the authors. *Notes for Authors* (1983) provides explicit instructions for the preparation of hard copy for deposition.

### Legibility

This is by far the most prevalent difficulty. Consider Fig. 1 and note that no special effort was required to locate parts (b) or (c) as examples. Fig. 1(a) presents no problems: it can be readily interpreted and photocopied with ease without significant loss of quality. Fig. 1(c) is the opposite extreme: it is largely illegible and results in maximum delay since replacement material must be requested from the authors. Fig. 1(b) is representative of the vast majority of legibility problems. Although there is reduction in quality it initially appears to be decodable but eventually guesses must be made concerning the presence or absence of negative signs or the identity of certain digits, particularly 0's and 8's, 5's and 6's, 3's and 8's.

Original authors are probably unaware of the processes that may be applied to their material before it reaches a user. These include photocopying (sometimes with reduction), microfilming, enlargement from microfilm or fiche and photocopying of the resultant enlargement. Routine computer printout is seldom suitable for deposition, even when a clean new ribbon is used, since inks used in this process often contain an appreciable blue element. This is especially true for single-pass dot-matrix printers which produce a large proportion of the problems exemplified by Fig. 1(b). If such printers are used then trial photocopies should be made to ensure that clarity can be preserved. A good-quality printer with a daisy wheel (or similar) print head and a new ribbon is infinitely preferable. If such facilities are not available then typed tables are superior to poor computer printout. It is quicker for database producers to locate typing errors in a few coordinates than it is to decode all numerals from partially legible copy.

The results of Table 1E reflect many of the comments above. Where CSD receives the top copy of deposited material [sources (3), (4), (5)] only 14 out of 181 documents (7.7%) had legibility problems. This figure rises to 16.1% for source (1) where photocopies are normally supplied and to 35.3% for material recovered from microfiche [source (2)]: the latter material has been subject to the greatest degree of reprocessing and requires the greatest care by authors in preparing high-quality hard copy.

### Layout

Any or all of the information items (a)–(f) of the *Introduction* may be of interest to the reader, who is probably expecting the information to be displayed as in a printed paper, with a concise tabulation for each category. All too often this is not the case, primarily owing to the use of final refinement output for deposition, which is sometimes as much as the last four least-squares cycles. Very few crystallographic software packages appear to produce concise tables of (particularly) non-redundant geometry tables suitable for

deposition; often the output is voluminous. This is not to say that such outputs are useless or ill-conceived, simply that they are being used for a purpose for which they were not intended. The cooperation of major developers of crystallographic program systems is being sought in this area, but it may be some time before the resulting improvements will filter through the research community. Meanwhile we draw attention to a number of particular problem areas.

ATOM	X	Y	Z
1S	3552( 1)	435( 1)	4167( 1)
10	4348( 2)	1097( 4)	4157( 3)
20	3138( 2)	3368( 3)	3665( 2)
30	2304( 2)	3186( 3)	2782( 2)
40	2046( 3)	-397( 3)	4142( 3)
1C	2689( 3)	1542( 4)	4231( 3)
2C	2381( 3)	1038( 4)	3274( 3)
3C	2947( 3)	2940( 5)	4416( 3)
4C	2272( 3)	1893( 4)	2468( 3)
5C	2068( 4)	816( 5)	4564( 4)
6C	1631( 4)	62( 5)	3183( 4)
7C	1082( 3)	1242( 5)	3988( 4)

(a)

ATOM	X	Y	Z
C1	0.2225(35)	0.9527(36)	0.2724(37)
C2	0.4059(33)	0.9539(32)	0.2335(33)
C3	0.5158(35)	1.0006(34)	0.1491(33)
C4	0.6475(36)	0.9039(37)	0.1112(36)
C5	0.6475(34)	0.8000(32)	0.2494(35)
C6	0.4671(34)	0.8040(31)	0.2930(33)
C7	0.3770(39)	0.0341(35)	0.3752(37)
C8	0.2462(43)	0.7656(39)	0.3510(46)
C9	0.1551(39)	0.8463(43)	0.2614(44)
C10	0.2177(37)	0.9271(38)	0.3005(38)
C11	0.5053(42)	0.0170(42)	0.1070(41)
C12	0.6005(47)	0.7091(50)	0.0245(49)
C13	0.0905(49)	0.0066(61)	-0.0105(48)
C14	0.0178(40)	0.9257(45)	0.0712(42)

(b)

01 1	-751( 2)	-4005( 1)	1110( 1)
01 2	3499( 1)	3135( 1)	1694( 1)
01 3	2221( 1)	5802( 1)	4895( 1)
01 4	2574( 1)	-1701( 1)	3129( 1)
01 5	8911( 2)	5104( 1)	940( 1)
01 6	9518( 1)	7161( 1)	6932( 1)
01 7	8964( 1)	-790( 1)	6686( 1)
01 8	3414( 1)	-2605( 1)	2524( 1)
01 9	1630( 3)	-291( 3)	2845( 2)
01 10	1733( 3)	500( 2)	1917( 2)
01 11	2359( 3)	2005( 3)	2820( 2)
01 12	2100( 2)	1200( 2)	3542( 2)
01 13	6136( 3)	1931( 2)	1762( 2)
01 14	7417( 3)	3274( 2)	2612( 2)
01 15	4648( 3)	2520( 2)	3181( 2)

(c)

Fig. 1. Examples of deposited data of diminishing legibility (a)–(c).

Atomic coordinates and anisotropic vibrational parameter (a.v.p.) tables are generally interpretable. They occasionally lack column headings, which can cause problems if a.v.p.'s are in a non-standard order, or are  $b$ 's or  $\beta$ 's rather than  $U_{ij}$ . Sometimes integer-form parameters ( $p \times 10^n$ ) are supplied in which the multiplicative power of 10 is not stated and is difficult to deduce. A more common problem is a change in atomic labels in going from printed publication to deposited material. This can make it particularly difficult to relate deposited coordinates with printed geometry or *vice versa*.

Geometrical output from standard program packages can produce severe problems for routine interpretation, such as: tables with redundant information e.g. C(1)—C(2) and C(2)—C(1); cryptic identification of symmetry-related atoms; bonded and non-bonded distances for covalent compounds mixed in the same table, where geometry within a given coordination sphere is generated to locate short intermolecular contacts; and distances, valence angles and, sometimes, torsion angles being generated as parts of one continuous listing.

Apart from problems of interpretation, the sheer bulk of some depositions makes them difficult to store, whilst the output may be impossible to photocopy onto standard A4 sheets. Both CSD and ICSD have received photocopies from depositories where key information is invariably missed by the copier. Here the only solution is to ask the authors to supply a complete version.

#### Completeness

The primary (printed) paper and the secondary deposited material should together form a complete document; all information items which are advertised in the printed paper as having 'been deposited' should actually be presented in the secondary publication. It is the responsibility of authors to ensure that this is so, and to include (at least) items (a), (b), (c) of the *Introduction*. Too often small data items or pieces of vital information fall into the void between primary and secondary publications. Typical examples are: some or all of the crystal data (e.g. cell parameters,  $Z$  value, space group, details of non-standard origin choice); the  $R$  factor; details, even the briefest indication, of the presence of disorder, solvent molecules, or counter ions. In the latter case the deposited coordinate table may contain extra atoms, additional to those required to define the chemical entity described in print. It is often impossible to identify a solvent molecule from a selection of atomic sites which may be disordered and/or be located about a symmetry element. In all of these examples it is necessary for the user to contact the authors to request missing information. Of greatest help is the inclusion, in the printed paper or in the deposition, of a simple chemical diagram which illustrates clearly the connectivity of the system(s) and the atomic nomenclature used in the study.

#### Discussion

This article demonstrates the importance of high-quality deposition documents as sources of *primary* crystallographic results. Both CSD and ICSD are committed to including these results in their files and are in a unique position to comment on current problems. It should be noted that the results of Table 1 are based upon about three months' input to CSD; hence in a full year CSD handles some 3000 deposition documents. The adverse effects of poor-quality material on database producers are emphasized, since it is the databases that provide many crystallographers and other scientists with the deposited data. Database users quite rightly expect as short as possible a time lag between publication of a paper and its appearance on file. Any scheme that divides a report between two different documents and two different sources is bound to introduce delays. This problem is compounded when the material itself suffers from the problems identified in this paper, especially when the sheer volume of information from this source is taken into account. In cases where difficulties with deposited data are encountered by a user, the authors should respond rapidly to requests for clarification. This is not always the present case. It is hoped that authors will take note of the *causes* of the problems described above, and that editors and referees will bear these comments in mind when assessing the quality of supplementary material, which should always accompany the primary manuscript.

Many of the problems detailed herein can be overcome in the short term by the establishment of closer links between journals and database producers. Links between *Acta Crystallographica*, CSD and ICSD have always been good, and are currently being improved. In the longer term the transmission of crystallographic results in machine-readable form will become more widespread. It should be noted that this method of deposition has been accepted by the Protein Data Bank (Bernstein *et al.*, 1977) since its inception in 1973, and that *Acta Crystallographica* now requires depositions in this form for data relating to biological macromolecules. A common interchange format (e.g. SCFS: Brown, 1983, 1985) for small-molecule results, coupled with international electronic mail and file-transfer facilities, will forge closer links with the original authors. These developments can only improve the accuracy of crystallographic information, both in the printed journals and in the computerized databases.

#### References

- 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. (1979). *Acta Cryst.* B35, 2331–2339.

ALLEN, F. H., KENNARD, O. & TAYLOR, R. (1983). *Acc. Chem. Res.* **16**, 146–153.  
 BERGERHOFF, G., HUNDT, R., SIEVERS, R. & BROWN, I. D. (1983). *J. Chem. Inf. Comput. Sci.* **23**, 66–69.  
 BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F., BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMANOUCI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.  
 BROWN, I. D. (1983). *Acta Cryst.* **A39**, 216–244.

BROWN, I. D. (1985). *Acta Cryst.* **A41**, 399.  
 BROWN, I. D. & ALTERMATT, D. (1985). *Acta Cryst.* **B41**, 240–247.  
 BUERGI, H.-B. & DUNITZ, J. D. (1983). *Acc. Chem. Res.* **16**, 154–161.  
*Notes for Authors* (1983). *Acta Cryst.* **A39**, 174–186, §11.3.  
 VINTER, J. G. (1985). *Chem. Ber.* **21**, 32–38.  
 WILSON, S. R. & HUFFMAN, J. C. (1980). *J. Org. Chem.* **45**, 560–567.

*Acta Cryst.* (1986). **C42**, 1675–1678

## Die Kristallstrukturen von Rubidiumtriiodid und Thalliumtriiodid\*

VON K.-F. TEBBE UND U. GEORGY

Institut für Anorganische Chemie der Universität zu Köln, Greinstraße 6, D-5000 Köln 41, Bundesrepublik Deutschland

(Eingegangen am 6. Dezember 1985; angenommen am 11. Juni 1986)

**Abstract.** Orthorhombic,  $Pnma$ .  $Z = 4$ ,  $\lambda(\text{Mo } K\alpha) = 0.71069 \text{ \AA}$ ,  $T = 300 \text{ K}$ .  $\text{RbI}_3$ :  $M_r = 466.18$ ,  $a = 10.908 (1)$ ,  $b = 6.655 (1)$ ,  $c = 9.711 (1) \text{ \AA}$ ,  $V = 704.96 \text{ \AA}^3$ ,  $D_x = 4.392 \text{ Mg m}^{-3}$ ,  $\mu(\text{Mo } K\alpha) = 19.71 \text{ mm}^{-1}$ ,  $F(000) = 784$ .  $\text{TlI}_3$ :  $M_r = 585.10$ ,  $a = 10.599 (2)$ ,  $b = 6.419 (2)$ ,  $c = 9.436 (2) \text{ \AA}$ ,  $V = 641.98 \text{ \AA}^3$ ,  $D_x = 6.053 \text{ Mg m}^{-3}$ ,  $\mu(\text{Mo } K\alpha) = 39.55 \text{ mm}^{-1}$ ,  $F(000) = 960$ . The crystal structures are isotypic with  $\text{CsI}_3$  and have been refined by full-matrix least squares to  $R_F = 0.027$  for 727 reflections of  $\text{RbI}_3$  and to  $R_F = 0.042$  for 410 reflections of  $\text{TlI}_3$  [ $I \geq 3\sigma(I)$ ]. The triiodide ions  $\text{I}_3^-$  are asymmetrical and slightly bent with  $d = 2.833 (1)$ ,  $3.051 (1) \text{ \AA}$ ,  $\varphi = 178.11 (3)^\circ$  for  $\text{RbI}_3$  and  $d = 2.826 (3)$ ,  $3.063 (3) \text{ \AA}$ ,  $\varphi = 177.89 (9)^\circ$  for  $\text{TlI}_3$ . The asymmetry of the triiodide ion as a function of the size of the cations is discussed for the four isotypic compounds  $\text{MI}_3$ ,  $M = \text{Tl, Rb, Cs, NH}_4$ .

**Einleitung.** Die einander isotypen Kristallstrukturen von  $\text{CsI}_3$  (Runsink, Swen-Walstra & Michelsen, 1972) und  $\text{NH}_4\text{I}_3$  (Tebbe *et al.*, 1985) sind inzwischen sehr genau bekannt. Strukturmerkmal ist ein schichtartiger Aufbau aus kationisch und anionisch gemischten Netzen mit parkettartiger Teilstruktur der auffallend unsymmetrischen Triiodidionen. Während die extreme Verzerrung des  $\text{I}_3^-$  beim  $\text{NH}_4\text{I}_3$  zusätzlich durch Wasserstoffbrückenbindungen verursacht wird, ist diese beim  $\text{CsI}_3$  allein in der asymmetrischen elektrostatischen Umgebung begründet. Um den Einfluß der Größe des Kations auf die Geometrie des Triiodidions und -verbands abschließend diskutieren zu können

(Tebbe, 1977), wurden die restlichen bekannten isotypen Verbindungen  $\text{RbI}_3$  und  $\text{TlI}_3$  ebenfalls röntgenstrukturanalytisch mit Einkristallmethoden bearbeitet.

Die als einziges unsolvatisiertes Rubidiumpolyiodid darstellbare Verbindung  $\text{RbI}_3$  (Briggs & Patterson, 1932) ist bisher offensichtlich röntgenkristallographisch nicht charakterisiert worden. Für das im System  $\text{TlI}/\text{I}_2$  neben  $\text{Tl}_3\text{I}_4$  (Tebbe, 1973) gesicherte Polyiodid  $\text{TlI}_3$  liegt bereits der Isotypie-Nachweis mit  $\text{CsI}_3$  vor (Hazell, 1963).

**Experimentelles.**  $\text{RbI}_3$ . Zunächst unbeabsichtigte Darstellung der Verbindung bei Versuchen zur Gewinnung von Bleihalogeniden ungewöhnlicher Zusammensetzung und Struktur (Georgy & Tebbe, 1984), Identifizierung und Charakterisierung der in ihrem kristallinen Bau meistens stark gestörten Substanz über röntgenographische Pulver- und Einkristallaufnahmen; gezielte Darstellung durch Lösen von  $\text{RbI}$  in ethanolischer Iodlösung (molarer Ansatz, Wells & Wheeler, 1892), Bildung weniger großer, stark verwachsener Kristalle; Züchtung geeigneter Einkristalle durch Eindunsten einer mit dem gleichen Volumen  $n$ -Hexan überschichteten  $\text{RbI}_3$ -Lösung; würfelförmlicher schwarzer Kristall; Berechnung der Zellparameter aus den Winkelpositionen von 18 zonalen und axialen Reflexen mit  $7,5 \leq \theta \leq 14,0^\circ$ ; CAD4, Enraf-Nonius,  $\text{Mo } K\alpha$ -Strahlung,  $\omega/\theta$ -Abtastung, Basis-Scanbreite  $\pm 0,8^\circ$ , Apertur 1,00 mm, variable Scangeschwindigkeit  $1,55 \leq v \leq 6,71^\circ \text{ min}^{-1}$ ;  $2\theta \leq 56^\circ$ ,  $0 \leq h \leq 14$ ,  $0 \leq k \leq 8$ ,  $0 \leq l \leq 12$ , 1021 Reflexe, Meßzeit 20,0 h; 3 Standardreflexe, keine zeitliche Änderung, Polarisations- und Lorentzkorrektur, numerische Absorptionskorrektur (Kristallbeschreibung 100: 0,078,  $0, \pm 1,0$ : 0,063, 011: 0,071,  $1,0, \pm 1$ : 0,068,  $\bar{1},0, \pm 1$ : 0,051 mm);

\* Untersuchungen an Polyhalogeniden. 8. Teil 7: Tebbe, Freckmann, Hörner, Hiller & Strähle (1985).